IRI-CCE: A Platform for Evaluating CRISPR-based Genome Editing Tools and Its Components

Rahul Mahadev Shelake[1,†,*], Dibyajyoti Pramanik[1,†], Jae-Yean Kim[1,2,*]

[1] Division of Applied Life Science (BK21 Four Program), Plant Molecular Biology and Biotechnology Research Center, Gyeongsang National University, Jinju 52828, Korea

[2] Division of Life Science, Gyeongsang National University, 501 Jinju-daero, Jinju 52828, Korea

[†]These authors contributed equally to this work.

Correspondence to: Jae-Yean Kim: kimjy@gnu.ac.kr (0000-0002-1180-6232) and Rahul Mahadev Shelake: rahultnau@gmail.com (0000-0003-0691-560X)

Running title: Dual expression platform for genome editing

Abstract

Rapid assessment of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-based tools and validation of cloned CRISPR-plasmid vectors are critical aspects for the successful application of genome editing (GE) in different organisms, including plants. Also, the selection of active guide RNAs (gRNAs) is a determining factor to achieve efficient GE at the targeted locus. Incidentally, plasmid vectors capable of ectopic expression of Cas enzyme and sgRNAs in bacteria will facilitate the quick screening and reliability of cloned plasmids and the functionality of designed gRNAs. We report a platform for *in vivo* rapid investigation of CRISPR components in *Escherichia coli* (IRI-CCE), which is compatible with plant-transforming binary vectors comprising plant promoters/terminators. Besides, IRI-CCE analyses reveal distinct features of cytidine (PmCDA1, evoCDA1, APOBEC3A) and adenine (ABE8e) base editors. Finally, we show the IRI-CCE platform as a reliable and rapid method for screening functional gRNAs to achieve successful GE outcomes. Base editor-based CRISPR components expressed by promoters of different strengths led to the establishment of IRI-CCE platform for three major applications: optimization of novel CRISPR tools, validation of cloned CRISPR plasmids, and to know gRNA functionality.

Keywords: Base editing, CRISPR, genome editing, heterologous expression

Background

The adoption of the bacterial immune system, CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein), for targeted genetic manipulation has transformed the field of genome editing (GE) research [1]. The CRISPR-based tools allow programmable engineering of all three major building blocks of life, i.e., DNA, RNA, and protein [2]. In this regard, various approaches have been adopted to develop novel CRISPR-based tools or validate CRISPR components before their robust application in microbes, animals, plants, and other eukaryotes [3, 4]. Most of the novel CRISPR-based tools are initially developed and tested for mammalian and bacterial GE that is later optimized for plant applications. Additional engineering and optimization are required to use newly devised CRISPR tools in different hosts [5]. The design of an independent plasmid vector system with desired selection-marker genes, specific promoter-terminator combinations for Cas enzyme, or single guide RNA (sgRNA) expression is desirable before commencing the GE experiments.

Moreover, it is well-known that the target selection and features of the gRNA sequence are the significant determinants in successful GE experiments [6]. Hence, the design of a functional gRNA helps efficient editing at target sites in the genome of selected organisms, including plants. Several *in silico* tools can predict the on-target and off-target activities of candidate gRNAs. However, available bioinformatics tools predict the suitable gRNA candidates only in a limited number of bacterial and plant species. The algorithms of plant-specific tools are based on results obtained from gRNA datasets of non-plant hosts [7]. Also, predicted outcomes do not necessarily reflect the actual *in vitro* or *in vivo* sgRNA activity in heterologous hosts owing to several factors. Notably, low or no sgRNA activity could be a consequence of three potential factors: the presence of inhibitory motifs in the sgRNA sequence [8], the failure to form

a functional Cas9-gRNA complex [6], or inefficient *in vivo* binding of active Cas9-sgRNA complex to the target site [9]. Although detrimental intra-gRNA interactions hindering active Cas9-gRNA complex formation can be investigated by *in vitro* assays, it is impossible to validate the efficacy of sgRNA expression or to identify inhibitory motifs due to the use of preassembled ribonucleoprotein (RNP) complexes comprising purified Cas9 and synthetic sgRNA. Therefore, it is highly desirable to quickly assess the *in vivo* sgRNA activity in the targeted host or heterologous organisms.

In several instances, the cloning of new plasmid vectors, genetic delivery of CRISPR plasmids into host organisms, tissue culture, and analysis of GE outcomes are costly and time-consuming processes in eukaryotic experiments [10, 11]. Thus, an analogous platform for simultaneous validation of a cloned CRISPR/Cas9 plasmids, including sgRNA components, may help to save the investment of resources and time. It may also provide first-hand knowledge about the applicability of the designed plasmid vectors for testing new CRISPR technologies in higher organisms. Incidentally, plasmid vectors with the ability of ectopic expression of Cas enzyme and sgRNAs in bacteria will facilitate the quick screening and reliability of cloned plasmids and editing activities of designed gRNAs.

Although virus-derived plant promoters are expected to remain silent in bacteria due to divergent features of transcriptional machinery between prokaryotes and eukaryotes, some virus-derived plant promoters have been reported to drive the transcription of downstream coding genes in *Escherichia coli* [12, 13]. For example, the transcriptional machinery of *E. coli* can recognize Cauliflower Mosaic Virus (CaMV) 35S promoter (p35S) [14, 15], which is also the most often used promoter for expression of *Cas* gene in plant GE. Although the p35S is less active, this aspect is of great significance for rapid validation of cloned plasmids in *E. coli* or

design a heterologous expression platform. Moreover, the toxic effects of BE components fused with nickase (nCas9) or dead Cas9 (dCas9) under the control of strong promoters are earlier reported in bacteria [16, 17]. In this regard, combination of differential-strength promoters for the optimal expression of base editing (BE) components including sgRNAs preferably might avoid the toxicity issue in *E. coli*.

Here, we compared a set of promoters for their transcriptional activity in *E. coli* (**Fig. 1**). Subsequently, a deaminase-based BE system was employed to establish an efficient and rapid evaluation platform to assess the differential-strength promoter-driven base conversion in *E. coli*, termed as *in vivo* rapid investigation of CRISPR components in *E. coli* (IRI-CCE). The nCas9-based BE (nCas9-BE) does not need to make double-stranded breaks (DSBs) in the DNA sequence for nucleotide substitution [18, 19]. In animals and plants, primarily nCas is employed to stimulate the nicking of non-deaminated strands that favor desired BE outcomes by tweaking the cellular DNA damage responses [20]. The sgRNA-nCas9-BE complex binds the protospacer adjacent motif (PAM) in the target locus to form an R-loop, in which the partially open non-target strand (NTS) provides a possible substrate for base conversion. Depending on the type of deaminase used, editing can occur within or near the gRNA region, described as deamination or editing window [21]. We report that the efficient sgRNA expression was driven by the consensus sequence of Arabidopsis U6 promoter (75 bp) in *E. coli* cells. The modular cloning approach described in the present work enables the interchangeable use of designed parts independent of their context in the IRI-CCE platform.

Different versions of cytidine (CBE) and adenine (ABE) deaminases were analyzed for their substitution efficiency from C-to-T and A-to-G, respectively. Comparison between the differential-strength promoter-driven editing activities revealed distinct features of BEs,

including CBEs (Target-AID system based on sea lamprey cytidine deaminase, PmCDA1 [19]; evolved version of PmCDA1, evoCDA1 [22]; human APOBEC3A, A3A [23]) and most active ABE version, ABE8e [24]. Also, we demonstrate the potential of an optimized IRI-CCE platform for screening suitable gRNAs for GE experiments by comparing the sets of active and inactive gRNAs. Moreover, editing outcomes of all the analyzed gRNAs imply that the promoter choice is crucial in achieving improved GE.

Results

Virus-derived plant promoters drive the gene expression in *E. coli*

Because the GE activity differs according to the expression level of CRISPR/Cas machinery, we sought to screen a set of bacterial, virus-derived plant, and native-plant promoters (**Fig. 1a, Supplementary Fig. 1**) for transcriptional activities in *E. coli*. The virus-derived plant and native-plant promoters were chosen due to their frequent use to express the Cas enzyme in plant GE [25–28]. We constructed superfold GFP (sfGFP)-expressing transcriptional units using an appropriate combination of promoters and terminators (**Fig. 1b**). The transformed 10-beta *E. coli* cells were grown and amounts of sfGFP fluorescence were plotted in **Fig. 1c**.

The bacterial promoter pGlpT showed the highest expression levels among all the tested promoters. No activity was observed in the case of two native-plant promoters (pRbc and pRPS5a). All three p35S-based promoters expressed a comparable amount of sfGFP, which consonant with the previous reports [14, 15]. Based on the level of sfGFP fluorescence, the promoter order from highest to lowest strength was determined as pGlpT>pEc1>p35S(L)I>p35S(S)>p35S(L)>pRbc/pRPS5a. Further, relative mRNA transcripts of *sfGFP* were measured using quantitative real-time PCR (qRT-PCR) that showed a consistent pattern to that of protein amount except for pEc1 (**Supplementary Fig. 2**). Interestingly, pEc1

showed maximum transcriptional activity among all the tested promoters. However, the mRNA abundance of *sfGFP* was not correlating with the fluorescence activity of sfGFP protein for pEc1, possibly due to poor mRNA-protein correlations at the analyzed time-point [29]. pGlpT and p35S(L)I showed significantly higher activity than p35S(L), pRbc, and pRPS5a. Overall, pGlpT and p35S(L)I promoters exhibited the highest protein expression levels among the tested bacterial and virus-derived plant promoter groups, respectively.

Different promoters drive the base editing with variable efficiency in the devised platform

Apart from the selectable markers, the CRISPR-based BE construct consists of two major transcriptional units: deaminase-fused Cas protein and sgRNA expression units driven by two separate promoters. Previous bacterial BE studies consisted of a deaminase fusion with partially impaired nCas or inactive dCas enzyme [16, 17, 30, 31]. We chose the nCas9 (D10A)-based CBE system in the present study because of the apparent direct use of constructed plasmids in eukaryotes like plants. To avoid the lethal effect of nCas9-based CBE in *E. coli*, we opted to explore different promoters for expression of CRISPR machinery.

Although the pJ23119 promoter has successfully been used for sgRNA expression in *E. coli* and other bacteria [16, 17, 32], its activity in plants is not yet explored. In plants, the Arabidopsis U6 promoter (pAtU6) is often used to drive the sgRNA expression. We first sought to inspect ten different RNA polymerase III (RNA Pol III)-dependent promoters for conserved -10 (TATAAT) and -35 (TTGACA) elements (**Supplementary Fig. 3**). The set of promoters were chosen because of earlier studies reporting efficient sgRNA expression in different species, including bacteria (pJ23119), plants (pAtU6, pOsU3, pOsU6, pMtU6.6, pZmU3, pTaU3, pTaU6), yeast (pSNR52), and human (phU6). We detected partial or fully conserved -10 and -35 elements in all the promoters. Remarkably, the DNA sequence alignment of pJ23119 and pAtU6

showed 54.28% identity with fully conserved -10 and partially conserved -35 elements (**Supplementary Fig. 4**). Therefore, in view of developing a heterologous promoter for sgRNA expression, pAtU6 was examined for optimal sgRNA expression in *E. coli*.

We employed a PmCDA1-based Target-AID [19] system that induces C:G to T:A base-pair mutations in the editing window located approximately 1 to 8 bases from the distal end of the PAM (counting positions 21-23 for PAM) in 20 bp gRNA [19]. Two gRNAs were tested containing alternate Cs at even (Test gRNA1, **Fig. 2b**) and odd (Test gRNA2, **Fig. 2c**) positions spanning from 1 to 10 in the 20 bp gRNA. The target regions for the Test gRNAs were synthesized and cloned into the universal target-acceptor module (**Supplementary Fig. 5**). Next, we adopted two approaches for BE testing. Firstly, a two-component system consisted of a target plasmid and a CRISPR plasmid including Pro-sgRNA and Pro-nCas9-PmCDA1-Linker-1xUGI (hereafter nCas9-PmCDA1) (**Supplementary Fig. 6a**). Secondly, a single-component system composed of the target, sgRNA, and nCas9-PmCDA1 assembled into a single plasmid (**Supplementary Fig. 6b**). The copy number of individual plasmids in a two-component system significantly varies due to several factors [33–35]. Also, the maintenance of multiple plasmids increases the metabolic burden on host cells and makes it challenging to obtain synchronized copy numbers in a single cell [36]. To avoid these issues, we optimized the testing platform using a single plasmid vector containing all the three bioparts, i.e., target, nCas9-PmCDA1, and sgRNA. Cloning of the required target region into a universal target-acceptor using MoClo kit protocol [27] permits the assessment of any intended gRNA.

After the transformation of individual constructs in 10-beta *E. coli*, single colonies were cultured for plasmid isolation and target-region sequencing. pEc1-driven nCas9-PmCDA1 shown poor transformation efficiency probably attributable to lethal effect on bacterial growth. Base

conversion (C-to-T) frequency of nCas9-PmCDA1 was found to be 100% in the case of pGlpT, p35S(L)I, p35S(L), and p35S(S) with variable efficiency in the editing window ranging from 1 to 6 position (**Fig. 2d-e**). In comparison, pRbc and pRPS5a clones shown no C-to-T editing in the target region. Calculated editing efficiency showed a variable percentage of C-to-T editing (**Fig. 2d-f**), implying that after transforming the plasmid vectors into competent cells, the grown individual clones were indeed mixed populations. Similar outcomes for base editing were previously reported in other bacteria, like *Corynebacterium glutamicum* [37], *Clostridium beijerinckii* [38], *Staphylococcus aureus* [30], and *Agrobacterium* sp. [17]. For pGlpT, four colonies showed an average of 92% editing at targeted C4 in Test gRNA1 and 79.5% editing at targeted C5 in Test gRNA2 (**Fig. 2d-e**). A variable range of editing efficiencies was observed for each C position in nCas9-PmCDA1 driven by different promoters, possibly due to differential expression of CRISPR components. Consequently, the optimized method including differential-strength promoters for expression of CRISPR components was named IRI-CCE, an *In vivo* Rapid Investigation of CRISPR Components in *E. coli*.

Next, we targeted *E. coli* genes to evaluate the BE efficacy and potential toxicity mediated by differential-strength promoters in the genomic context of bacteria. Five different gRNAs were chosen for editing three genes, namely *galK*, *rpoB*, and *rppH* (**Supplementary Fig. 7**). The pGlpT and p35S(L)I-driven Target-AID efficiently edited the available Cs in the editing window of all the tested sites. Because the editing efficiency of BE versions may differ due to distinct genetic backgrounds in different *E. coli* strains, we examined the p35S(L)I-driven base conversion efficiency in three more strains, including DH5α, DB3.1, and BL21(DE3). Irrespective of the genetic background of strain, we found comparable C-to-T mutations in tested

gRNAs in all the three bacterial strains suggesting the broader applicability of the IRI-CCE platform across different *E. coli* strains.

IRI-CCE platform reveal unexplored editing features of broad-range CBEs in *E. coli*

Base editors with broader editing windows are generally applied for mutagenesis of a user-defined region. The editing windows vary according to BE type, target site, sequence context, and experimental conditions [20]. Several latest BE tools are not yet characterized for *E. coli* use. To characterize the editing features due to differential expression of CRISPR components, two CBE types (evoCDA1 and A3A) were tested in the IRI-CCE platform with distinctive features (different range of editing window and no sequence context-preference). In original reports, editing windows for evoCDA1 and A3A positioned from 1 to 14 [22] and 1 to 17 [23]. A core-editing region was observed between 3 to 12 positions for evoCDA1 in human cells and 3 to 9 positions for A3A in plants. To examine C-to-T conversion in *E. coli*, in addition to two Test gRNAs, we designed Test gRNA3 by mimicking the sequence of TaVRN1-gRNA1 (**Fig. 3a**) that comprises the Cs in the range of 1 to 17 bp [23].

Constitutive expression of CRISPR components using different promoters showed an expanded editing window for evoCDA1 and A3A, possibly owing to higher nCas9-BE expression (**Fig. 3**). Among the tested CBEs, evoCDA1 showed the C-to-T conversion with the broadened editing window from -6 to 14 position with pGlpT and from -1 to 11 position with p35S(L)I (**Fig. 3b**). The expression of nCas9-A3A using pGlpT and p35S(L)I promoter also displayed distinct editing windows ranging from -6 to 17 and 2 to 17, respectively (**Fig. 3c**). For pGlpT-evoCDA1, C-to-T conversion at positions -1 to 9 was the highest (18%-100%), whereas, for pGlpT-A3A, C-to-T mutation at position 3 to 10 was the highest (43%-100%).

On the other hand, for p35S(L)I-evoCDA1, C-to-T conversion at positions 2 to 7 was the highest (57%-100%), whereas for p35S(L)I-A3A, C-to-T mutation at position 4 to 10 was the highest (38%-100%). Notably, G:C at position -5 was mutated to A:T in both the CBE types with moderate efficiency across all the three gRNAs (up to 43%), implying the C-to-T conversion in the target (non-deaminated) strand. The upstream (position -3 to -1) and middle (position 12 to 16) regions of target sites were not accessible to A3A deamination in the tested combinations of gRNAs and respective target sites (**Fig. 3d**). Although direct comparison with earlier reports is not possible due to earlier mentioned parameters, these findings demonstrate the robust applicability of the IRI-CCE platform to validate the cloned plasmids and reveals novel aspects of broad-range CBE tools.

ABE8e shows efficient A-to-G editing with differential-strength promoters in *E. coli*

Recently, an evolved ABE variant, TadA-8e (ABE8e), has been characterized and shown improved A-to-G base conversion *in vitro* and human cells [24] but not yet applied in bacteria. We sought to investigate the ABE8e activity in IRI-CCE using two synthetic gRNAs (**Fig. 4a-b**). The editing window for ABE8e-nCas9 was reported as position 4 to 8 in human cells [24]. Editing levels of ABE8e-nCas9 showed significant increases in the case of pGlpT compared to that of p35S(L)I (**Fig. 4c-d**). We observed the editing window of ABE8e from positions 3 to 7 and 5 to 7 for pGlpT and p35S(L)I, respectively (**Fig. 4e**). For pGlpT-ABE8e, A-to-G conversion at positions 5 to 7 was the highest (74%-100%). For p35S(L)I-ABE8e, A-to-G mutation at positions 5 to 7 was the highest but in the lower range (21%-33%), suggesting an increase of editing activity in a promoter strength-dependant manner.

Most of the previous studies have used gRNAs with a single A in the target region for testing ABE systems. The gRNAs with the multiple adenosines (As) in a target region exhibited

different A-to-G editing efficiencies for each position in *E. coli* and *S. aureus* [30]. Moreover, the editing window can vary with the target site, host species, genotype, and experimental conditions [20, 39]. Therefore, it is not reasonable to compare the data about editing activity and window length from different systems. Overall, the editing windows for nCas9-ABE8e in the IRC-CCE platform are slightly narrower than the editing windows for the tested nCas9-CBEs.

IRI-CCE platform allows screening of active and inactive gRNAs

Pre-screening of gRNAs before commencing GE studies is crucial in certain organisms like plants to avoid loss of time for gRNA screening, lower editing activities, and waste of resources. Given that the efficiencies predicted by *in silico* tools and the actual GE activities may differ significantly, *in vitro* or preferentially *in vivo* validation of chosen gRNAs is vital before the real GE test[40]. In that scenario, the IRI-CCE platform may facilitate the simultaneous evaluation of designed plasmid vectors and editing efficacy of the selected gRNAs that enable potential avoidance of incorrect plasmids and bad gRNAs.

In some instances, gRNAs predicted as inactive by bioinformatic tools or design criteria were reported to work efficiently under *in vivo* conditions [7]. On the other hand, even if *in silico* parameters are not promising for a particular set of gRNAs, one needs to choose those gRNAs for editing a specific region of targeted locus due to PAM constraints. We analyzed sets of active and inactive gRNAs reported in the previous studies (**Fig. 5a**). Based on the A or C availability in the canonical editing window, we opted for a particular BE type to edit the target sites of individual gRNAs. The target regions, including the PAM, were cloned using two BsmBI type IIS sites in the universal target-acceptor modules (**Supplementary Fig. 5 and 6**). Firstly, the set of three active gRNAs (AtGL1-gRNA1 [25], SlMlo1-gRNA2, and SlPelo-gRNA1 [26]) targeting

three sequences from the plant genome showed higher base conversion efficiencies (**Fig. 5b-d**) with differential-strength promoters as observed under *in planta* conditions.

Secondly, we analyzed ABE8e activities for two inactive gRNAs (Inactive gRNA1 and Inactive gRNA2) that displayed poorer performance [6]. The lower editing activity was attributed to internal sgRNA interactions that allow binding with Cas9, but it prevents target DNA recognition by the Cas9-sgRNA complex (**Supplementary Fig. 8**). Besides, inactive gRNA competes with active gRNAs for binding to the Cas9 protein, thereby reducing the editing activities. We investigated the base conversion activity at DNA templates for both the gRNAs and discovered that both the sgRNAs led to no BE frequencies in the target DNA region with low-strength p35S(L)I (**Fig. 5e-f**). Moreover, even a high-strength promoter (pGlpT)-driven ABE8e expression led to relatively lower BE activity (36% and 23.5% for Inactive gRNA1 and 2, respectively).

Thirdly, we analyzed two gRNAs predicted as inactive due to internal base-pairing (7 bases of 20 bp) within guide sequence (Os-gRNA22) or lower GC content (20%) with T-rich repeats (Os-gRNA23) [41]. The target DNA template and sgRNA cassettes were cloned along with BE transcription units to examine the editing frequencies. Sanger sequencing of the target region showed pGlpT-driven BE activity in both the gRNAs, implying that the Os-gRNA22 and Os-gRNA23 are functional gRNAs (**Fig. 5g-h**). Notably, the sgRNA expression is typically driven by RNA Pol III-dependent promoters in eukaryotes [42]. T-rich repeats act as a termination signal for RNA Pol III that seems did not affect Os-gRNA23 expression in *E. coli*. While the pGlpT-driven CBE activity was higher for Os-gRNA23 in the target site, p35S(L)I-driven CBE constructs showed negligible editing with no apparent reason. In all the tested target sites, any base at position 18 was mutated to A plausibly during single-stranded break (SSB)

repair induced by Cas9 nickase. Overall, these data indicate that the IRI-CCE platform is helpful to distinguish the active and inactive gRNAs. Taken together, the IRI-CCE platform could provide first-hand information of *in vivo* gRNA activity and help to reduce the number of gRNAs one needs to use in higher eukaryotic systems.

Discussion

The several features of *E. coli* contribute to make it a model microorganism in molecular biology that includes the short life cycle, simple genetic manipulations, low cost, and in-depth knowledge of genome [43]. In the current study, we developed an *E. coli*-based IRI-CCE platform to investigate three key factors: CRISPR components, cloned plasmids, and gRNA functionality with several advantages. First, dual expression promoters for sgRNA and CRISPR components allow heterologous expression in *E. coli* without additional cloning steps. Second, modular parts are exchangeable for cloning and investigation of different CRISPR tools. Third, IRI-CCE enables the rapid validation of designed plasmid vectors (expression of sgRNA, Cas9, and other bioparts). Forth, it is easy to optimize novel CRISPR tools using the IRI-CCE that allow shuffling of desired parts. Fifth, quick validation of gRNA functionality helps to reduce the burden of gRNAs screening in eukaryotes. Finally, this method does not need any protein purification or high-throughput sequencing techniques like targeted deep DNA sequencing, reducing the operation cost. Also, the approach reported here could be readily scaled to a large number of constructs and BEs.

Existing platforms have some pros and cons for the pre-screening of gRNAs and preliminary check of CRISPR outcomes. For instance, the combination of protoplast culture and pre-assembled RNP complexes is one of the handy methods for assessing the editing activities of sgRNAs in plants [44]. Nevertheless, protoplast-based assays require protoplast isolation,

limiting only model plant species [45]. In addition, the protoplast system and whole plant differ significantly in terms of epigenetic marks [46]. Cas9-sgRNA RNP complexes require additional expertise, equipment, and laboratory reagents to express and purify Cas enzyme, further adding up to the cost. Likewise, protoplast regeneration with a high risk of genomic instability remains a bottleneck for this platform [46]. Transient infiltration assay proved to be a comparatively quick *in vivo* method for analyzing sgRNA efficacy, but the lower activity and inconsistency are the significant obstacles [47]. Therefore, it remains a constant demand to develop rapid and scalable strategies; consequently, the IRI-CCE platform serves as an ideal tool to overcome most of the said bottlenecks. The strengths and limitations of each method are summarized in **Supplementary Table 1**.

The attempt to analyze the BE activity of bacterial (pJ23119) promoter-driven gRNA expression was not successful owing to the toxicity caused by BE components. Likewise, the pEc1-mediated CBE test also showed poor transformation efficiency with a mix of incorrect clones. The toxicity of BE components is earlier reported in bacteria [16, 17]. When nCas9-PmCDA1 was expressed together with pJ23119-sgRNA, no correct clones were obtained in *E. coli*, possibly due to toxic effects of constitutively-expressed BE components [16]. Even the cloning of dCas9-PmCDA1 with pJ23119-sgRNA showed the wrong clones in *E. coli* [16] and *Agrobacterium* sp. [17], suggesting cells cannot retain the functional dCas9-PmCDA1-expressing plasmids. A high rate of SSBs induced by nCas9 during pJ23119-mediated sgRNA production seems difficult to be restored by the bacterial system, ultimately losing the plasmid. Although dCas9 does not introduce SSB, earlier dCas9-PmCDA1 data suggested that the surplus amount of UGI compromises the genome integrity [16].

Previous studies reported that the use of protein degradation (LVA) tag fusion with dCas9-Target-AID reduced the toxicity to *E. coli* cells [16]. By producing the optimal amount of BE and sgRNA might reduce the toxicity to a tolerable level for a bacterial cell as observed in nCas9-BE together with pAtU6-driven sgRNA expression. As described in our study, the combination of pGlpT-nCas9-BE or p35S(L)-nCas9-BE with pAtU6-sgRNA expression units proved to be the most appropriate to edit the target site in the same plasmid vector or *E. coli* genome. Besides, bacterial codon optimization was not preferred for several bioparts that might be contributing to optimal expression levels. Therefore, our data show that fine-tuning the expression of BEs and sgRNAs by using the optimal combination of promoters is a crucial factor to minimize the toxic effects in the IRI-CCE platform.

Sanger sequencing followed by analysis with the EditR program is sufficient to know the gRNA functionality in a cost-effective way. EditR detects and quantifies base mutation pattern from fluorescence-based Sanger sequencing file [48]. Indeed, high throughput techniques like next-generation sequencing (NGS) provide accurate and precise information about the distribution of base mutation at a specific position in the editing window. However, to evaluate the functionality of cloned sgRNA or CRISPR components, one does not need to use expensive techniques like NGS. BE interpretation with EditR is mainly limited by the quality of Sanger sequencing data. Perhaps the isolation of target-containing plasmids facilitates an easier way to prepare high-quality DNA templates for Sanger sequencing rather than using a PCR-based approach in NGS-based targeted deep DNA sequencing [49]. Also, NGS data analysis needs in-depth knowledge of bioinformatics and may take longer time. Therefore, method of GE detection can be decided on case-by-case basis. An off-target event is not a significant concern in the present system because off-targeting may kill the cell or may not affect the anticipated results

and can thus be undetermined. Apparently, examining the off-target sites for selected gRNAs in the *E. coli* genome is highly desirable to avoid unforeseen issues.

Sequence preference is a common characteristic of BE tools. Also, deaminase interaction with Cas protein and target DNA region alters the location and size of the editing window [50]. The indispensable role of endogenous DNA repair mechanism is a decisive factor that may produce variable BE results in different organisms [51]. Initially, no sequence preference was reported for ABE8e, PmCDA1, evoCDA1, and A3A. Nevertheless, we observed significant differences between the targeting competencies of three CBEs owing to the differential-strength of promoters used for Cas-BE expression (**Supplementary Fig. 9**). The evoCDA1 is an evolved version of PmCDA1 [22] that showed higher catalytic activity and expanded editing window than its wild-type counterpart. We observed higher C-to-T editing in the expanded window up to position -6 in target and non-target strand that was not detected in the previous work for evoCDA1 or A3A.

The comparison between the editing windows revealed that the accessibility to the target site was totally different for evoCDA1 and A3A. Some Cs in the editing window were not accessible to A3A deamination (**Supplementary Fig. 9**). These observations highlight the importance of considering BE type and promoter strength that best suit the expected outcome. Also, temperature affects the enzymatic activities of CRISPR components, thereby influencing the GE outcome [2]. Mainly, the performance of Cas enzymes is best at 37°C. So, besides the promoter activity and evolved versions, better editing efficacy in IRI-CCE may resulted from the favorable temperatures for enzymatic activities.

Additional factors need to be considered while using a validated gRNA sequence in the plant system. Factors that impact sgRNA production, the Cas9-sgRNA complex formation or

stability, the efficiency of target DNA-binding and cutting would be identified in the IRI-CCE platform. However, more efforts are needed to understand uncovered aspects. For example, the editing efficiency of each Cas9-sgRNA complex differs significantly, depending on the target locus. In higher eukaryotes, editing differences arise from the sequence bias in DNA repair machinery and accessibility of target gene due to complex chromatin structures in the host genome [9, 52]. Also, the bigger genome size of eukaryotic cells influences the dynamics of target search. Therefore, IRI-CCE is missing those features common to plant and other eukaryotic cells. While our data show the promising way to confirm the cloned plasmids for CRISPR work, there are multiple opportunities to further expand the IRI-CCE utility. For example, IRI-CCE allows testing of pegRNA functionality because it is one of the critical aspects for the successful application of prime editing [53, 54].

Conclusions

In summary, nCas9-BE components expressed by promoters of different strengths led to the establishment of IRI-CCE platform for three major applications: optimization of novel CRISPR tools, validation of cloned CRISPR plasmids, and to know gRNA functionality. IRI-CCE analyses showed variable length of the editing window for specific BE types. Each independent BE type has its specific editing pattern for a given target site and promoter type. Because of its easy cloning and heterologous applicability, the IRI-CCE platform may be suited for rapid evaluation of cloned plasmid vectors and applicable to a broader range of *E. coli* strains being used for cloning. This platform facilitates the verification of gRNA-related factors such as cloning, sgRNA production, Cas9-sgRNA complex formation, and DNA recognition. Through the applications described here and through further improvements, IRI-CCE has the potential to

make a widespread impact on the characterization of CRISPR components and rapid assessment of newer CRISPR technologies.

Methods

E. coli strains

The four *E. coli* strains were used: 10-beta, DH5α, DB3.1, and BL21(DE3). The information about strain genotype is provided in **Supplementary Table 2**. All the primers and plasmids used in this study are listed in **Supplementary Table 3 and 4**, respectively. The *E. coli* strains were cultured in Luria-Bertani (LB) broth with appropriate antibiotics.

Plasmid construction and cloning

The required bioparts were amplified by conventional PCR using a high-fidelity Taq DNA polymerase (Phusion Taq, Thermo Fisher Scientific, Waltham, MA). The cloning of different plasmid vectors was designed and performed by following the principle of MoClo [55] and Golden Gate assembly [27] protocols using BpiI/BsaI Type IIS enzyme digestion-ligation. The DNA recognition sites of restriction enzyme BsaI and BpiI were removed from internal sequences of bioparts during PCR amplification to make them suitable for Golden Gate cloning (a procedure also termed domestication). The DNA oligonucleotide pairs for the pEc1 (SJM901+RBSTL2) promoter and TerL3S2P21 terminator were annealed and ligated into BpiI-digested acceptor plasmids. Various promoters including pGlpT, p35S(Long), p35S(Short), pRbc, and pRPS5a were used from the pYTK001 (Addgene #65108) [56], pICH51266 (Addgene #50267) [27], pICH51277 (Addgene #50268) [27], pICH45195 (Addgene #50275) [27], and

pKI1.1R (Addgene #85808) [28], respectively. The promoter DNA sequences provided in **Supplementary Fig. 1** Similarly, DNA sequences of Terminators, Ter35S and TerHSP were adopted from the pICH41414 (Addgene #50337) [27] and pKI1.1R (Addgene #85808), respectively.

Three cytidine deaminases: PmCDA1 (Target-AID) [19], evoCDA1 [22], and APOBEC3A [23], were PCR amplified from PmCDA1-1xUGI (Addgene #79620), evoCDA1 pBT277 (Addgene #122608), and A3A-PBE-ΔUGI (Addgene #119770), respectively. For all three deaminases, the same linker regions were used as reported in the previous studies. The plant-codon optimized ABE8e was synthesized from Bioneer Co. (Daejon, Korea). PmCDA1-1xUGI was fused to the C terminus of nCas9. The evoCDA1, A3A, and ABE8e were fused to the N terminus of nCas9 with XTEN linker. The 2xUGI biopart (template: Addgene #122608) was fused to the C terminus of evoCDA1-nCas9 and A3A-nCas9. The sgRNA expression was driven by bacterial pJ23119 (synthetically cloned) or plant AtU6 promoter (pICSL01009, Addgene #46968). Nickase Cas9 (D10A) was generated by the PCR method using previously optimized Cas9 as a template (Level 1 hCas9 module, Addgene #49771). Desired sgRNA sequence was PCR amplified using plasmid pICH86966::AtU6p::sgRNA_PDS (Addgene #46966) as a template and cloned together with either pJ23119 or pAtU6 for sgRNA expression.

For cloning of target region of the desired sgRNA sequence, a pair of oligonucleotide DNAs that contains the target sequence with PAM was annealed and ligated into BsmBI-digested universal target-acceptor plasmid L1 or L2 having optimized sfGFP [56] at downstream side (**Supplementary Fig. 5 and 6**).

Bacterial transformation, plasmid isolation, and Sanger sequencing

Ligation products of all the steps during the cloning were transformed into competent cells of *E. coli* 10-beta strain by a heat-shock method. The bacterial culture was spread on the LB media containing desired antibiotic and incubated at 37 °C for 18-24 h. The cloned plasmids and BE activities were confirmed by Sanger sequencing at Solgent Ltd. (Daejeon, Korea) or Cosmogentech Ltd. (Seoul, Korea). For mutagenesis assay, the different *E. coli* strains were transformed with the appropriate plasmids using the heat-shock method and were pre-cultured for 1 h with 1 ml of LB medium. After incubation for 1 h at 37 °C, the fraction of cell cultures were spread on LB agar (1.5%) plates containing selection antibiotics with needed concentrations. The next day, individual colonies from the plate were inoculated in 3 ml LB broth with appropriate antibiotics and cultured at 37 °C. The culture time was varied according to the experimental parameters and mentioned in appropriate sections. The plasmid isolation was done using Plasmid Mini-Prep Kit from BioFact Co. Ltd. (Daejeon, Korea) for Sanger sequencing analysis.

Promoter activity analysis

As shown in **Fig. 1b**, all the tested promoters were cloned with the bacterial codon-optimized sfGFP sequence at the downstream side, followed by a termination signal. Bacterial cultures were grown from the transformed single colonies in LB broth with shaking at 37°C for 24 h. The grown cultures were incubated for 24 h at 4°C, and then OD_{600} values were normalized to 1. The cells were harvested at 10000 rpm for 1 min and washed with 1x PBS (Phosphate buffered saline). Images were captured in blue light with a FluoroBox from CELLGENTEK Co., Ltd. (Deajeon, South Korea). The fluorescence intensity level was quantified using the ImageJ software [57], and corrected total cell fluorescence (CTCF) was calculated using the following

formula: CTCF = Integrated Density - (Area of selected cell x Mean fluorescence of background readings).

For qRT-PCR, RNA was extracted using the following method. *E. coli*-carrying plasmids (**Fig. 1b**) were grown in LB medium. Cells were harvested during the exponential growth phase (12 h with OD_{600} value 0.5). Total RNA was extracted using RNeasy Protect Bacteria Mini Kit from Qiagen. For all samples, 600 ng of total RNA was used for complementary DNA (cDNA) synthesis using a QuantiTect Reverse Transcription Kit from Qiagen following the manufacturer's instructions. To estimate the relative *sfGFP* transcript, the qRT-PCR reactions were carried out using the KAPA SYBR FAST qPCR kit from Kapa Biosystems (MA, USA) with *sfGFP*-specific primer sets (**Supplementary Table 3**). Cycling of PCR consisted of pre-denaturation at 95°C for 5 min followed by 40 cycles of a denaturation step at 95°C for 10 min, an annealing step at 60°C for 15 s, and final extension step at 72 °C for 20 s using the CFX384 Real-Time System from Bio-Rad, Hercules (CA, USA). The qRT-PCR reactions were performed with independent biological replicates. Relative *sfGFP* transcript values normalized against internal control 16S ribosomal RNA (*rrsA*) gene. Data analyses were performed by the $2^{-\Delta\Delta Ct}$ method [58].

Evaluation of editing activities

The single colonies were cultured, and the plasmid vectors containing synthetic targets were purified using Plasmid Mini-Prep Kit from BioFact Co. Ltd. (Daejeon, Korea) for further sequencing analysis. For sequencing analysis of genomic loci, the target fragments were PCR amplified using target-specific primers from the randomly picked colonies and then analyzed by Sanger sequencing. Sanger sequencing data analysis was performed using SnapGene software (GSL Biotech; available at snapgene.com). The editing efficiency was determined either by the

ratio of non-edited to edited colonies from the randomly picked cells or by the nucleotide conversion rate using online tool EditR [48] from C-to-T and A-to-G for CBE and ABE, respectively. The data were statistically analyzed and plotted in GraphPad Prism 9.0.0 (www.graphpad.com).

Data Availability

The authors state that all data necessary for confirming the conclusions presented in this article are represented fully within the article or can be provided by the authors upon request. All data generated or analyzed during this study are included in this published article and its supplementary information files.

Contributions

J.Y.-K. and R.M.S. conceived this study and designed the research. R.M.S. designed constructs and performed experiments. R.M.S. and D.P. performed analyses. R.M.S. and D.P. wrote the paper and analyzed the data. J.Y.-K. supervised the research activities. All authors contributed to discussions and manuscript review.

Acknowledgements

This work was supported by the National Research Foundation of Korea (grants NRF 2020M3A9I4038352, 2020R1A6A1A03044344) and the Program for New Plant Breeding Techniques (NBT, grant PJ01478401), Rural Development Administration, Korea.

Competing interests

The authors declare no competing interests.

References

1. Anzalone A V., Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol. 2020;38:824–44. doi:10.1038/s41587-020-0561-9.

2. Pramanik D, Shelake RM, Kim MJ, Kim J-Y. CRISPR-Mediated Engineering across the Central Dogma in Plant Biology for Basic Research and Crop Improvement. Mol Plant. 2021;14:127–50. doi:10.1016/j.molp.2020.11.002.

3. Adli M. The CRISPR tool kit for genome editing and beyond. Nat Commun. 2018;9. doi:10.1038/s41467-018-04252-2.

4. Shelake RM, Pramanik D, Kim JY. Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. Plant Biotechnol Rep. 2019;13:423–45. doi:10.1007/s11816-019-00562-z.

5. Arndell T, Sharma N, Langridge P, Baumann U, Watson-Haigh NS, Whitford R. GRNA validation for wheat genome editing with the CRISPR-Cas9 system. BMC Biotechnol. 2019;19:1–12.

6. Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF. Internal guide RNA interactions interfere with Cas9-mediated cleavage. Nat Commun. 2016;7:1–7. doi:10.1038/ncomms11750.

7. Naim F, Shand K, Hayashi S, O'Brien M, McGree J, Johnson AAT, et al. Are the current gRNA ranking prediction algorithms useful for genome editing in plants? PLoS One. 2020;15:1–12. doi:10.1371/journal.pone.0227994.

8. Graf R, Li X, Chu VT, Rajewsky K. sgRNA Sequence Motifs Blocking Efficient

CRISPR/Cas9-Mediated Gene Editing. Cell Rep. 2019;26:1098-1103.e3. doi:10.1016/j.celrep.2019.01.024.

9. Isaac RS, Jiang F, Doudna JA, Lim WA, Narlikar GJ, Almeida R. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. Elife. 2016;5:1–14.

10. Marshall R, Maxwell CS, Collins SP, Jacobsen T, Luo ML, Begemann MB, et al. Rapid and scalable characterization of CRISPR technologies using an E. coli cell-free transcription-translation system. Mol Cell. 2018;69:146-157.e3. doi:10.1016/j.molcel.2017.12.007.

11. Oh Y, Lee B, Kim H, Kim SG. A multiplex guide RNA expression system and its efficacy for plant genome engineering. Plant Methods. 2020;16:1–11. doi:10.1186/s13007-020-00580-x.

12. Jacob D, Lewin A, Meister B, Appel B. Plant-specific promoter sequences carry elements that are recognised by the eubacterial transcription machinery. Transgenic Res. 2002;11:291–303.

13. Jopcik M, Bauer M, Moravcikova J, Boszoradova E, Matusikova I, Libantova J. Plant tissue-specific promoters can drive gene expression in Escherichia coli. Plant Cell Tissue Organ Cult. 2013;113:387–96.

14. Assaad FF, Signer ER. Cauliflower mosaic virus P35S promoter activity in Escherichia coli. MGG Mol Gen Genet. 1990;223:517–20.

15. Lewin A, Jacob D, Freytag B, Appel B. Gene expression in bacteria directed by plant-specific regulatory sequences. Transgenic Res. 1998;7:403–11.

16. Banno S, Nishida K, Arazoe T, Mitsunobu H, Kondo A. Deaminase-mediated multiplex genome editing in Escherichia coli. Nat Microbiol. 2018;3:423–9. doi:10.1038/s41564-017-

0102-6.

17. Rodrigues SD, Karimi M, Impens L, Van Lerberge E, Coussens G, Aesaert S, et al. Efficient CRISPR-mediated base editing in Agrobacterium spp. Proc Natl Acad Sci U S A. 2021;118:1–8. doi:10.1073/pnas.2013338118.

18. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533:420–4. doi:10.1038/nature17946.

19. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science. 2016;353:aaf8729. doi:10.1126/science.aaf8729.

20. Huang TP, Newby GA, Liu DR. Precision genome editing using cytosine and adenine base editors in mammalian cells. Nat Protoc. 2021.

21. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet. 2018;19:770–88. doi:10.1038/s41576-018-0059-1.

22. Thuronyi BW, Koblan LW, Levy JM, Yeh WH, Zheng C, Newby GA, et al. Continuous evolution of base editors with expanded target compatibility and improved activity. Nat Biotechnol. 2019;37 September:1070–9. doi:10.1038/s41587-019-0193-0.

23. Zong Y, Song Q, Li C, Jin S, Zhang D, Wang Y, et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat Biotechnol. 2018;36:950–3. doi:10.1038/nbt.4261.

24. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, et al. Phage-assisted

evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat Biotechnol. 2020;38:883–91. doi:10.1038/s41587-020-0453-z.

25. Hahn F, Mantegazza O, Greiner A, Hegemann P, Eisenhut M, Weber APM. An efficient visual screen for CRISPR/CAS9 activity in arabidopsis thaliana. Front Plant Sci. 2017;8 January:1–13.

26. Pramanik D, Shelake RM, Park J, Kim MJ, Hwang I, Park Y, et al. CRISPR/Cas9-Mediated Generation of Pathogen-Resistant Tomato against Tomato Yellow Leaf Curl Virus and Powdery Mildew. Int J Mol Sci. 2021;22:1878.

27. Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JDG, et al. A Golden Gate modular cloning toolbox for plants. ACS Synth Biol. 2014;3:839–43.

28. Tsutsui H, Higashiyama T. pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in Arabidopsis thaliana. Plant Cell Physiol. 2017;58:46–56. doi:10.1093/pcp/pcw191.

29. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165:535–50. doi:10.1016/j.cell.2016.03.014.

30. Zhang Y, Zhang H, Wang Z, Wu Z, Wang Y, Tang N, et al. Programmable adenine deamination in bacteria using a Cas9-adenine-deaminase fusion. Chem Sci. 2020;11:1657–64. doi:10.1039/c9sc03784e.

31. Zheng K, Wang Y, Li N, Jiang F-F, Wu C-X, Liu F, et al. Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. Commun Biol. 2018;1:32. doi:10.1038/s42003-018-0035-5.

32. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell. 2013;152:1173–83.

33. Wein T, Hülter NF, Mizrahi I, Dagan T. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. Nat Commun. 2019;10:1–13. doi:10.1038/s41467-019-10600-7.

34. Velappan N, Sblattero D, Chasteen L, Pavlik P, Bradbury ARM. Plasmid incompatibility: More compatible than previously thought? Protein Eng Des Sel. 2007;20:309–13.

35. Goldsmith M, Kiss C, Bradbury ARM, Tawfik DS. Avoiding and controlling double transformation artifacts. Protein Eng Des Sel. 2007;20:315–8.

36. Yang S, Liu Q, Zhang Y, Du G, Chen J, Kang Z. Construction and Characterization of Broad-Spectrum Promoters for Synthetic Biology. ACS Synth Biol. 2018;7:287–91.

37. Wang Y, Liu Y, Liu J, Guo Y, Fan L, Ni X, et al. MACBETH: Multiplex automated Corynebacterium glutamicum base editing method. Metab Eng. 2018;47 March:200–10.

38. Li Q, Seys FM, Minton NP, Yang J, Jiang Y, Jiang W, et al. CRISPR–Cas9 D10A nickase-assisted base editing in the solvent producer Clostridium beijerinckii. Biotechnol Bioeng. 2019;116:1475–83.

39. Katti A, Foronda M, Zimmerman J, Diaz B, Zafra MP, Goswami S, et al. GO: A functional reporter system to identify and enrich base editing activity. Nucleic Acids Res. 2020;48:2841–52.

40. Kim DH, Lee J, Suh Y, Lee K. Necessity for Validation of Effectiveness of Selected Guide

RNA In Silico for Application of CRISPR/Cas9. Mol Biotechnol. 2021;63:140–9.

41. Liang G, Zhang H, Lou D, Yu D. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci Rep. 2016;6:21451. doi:10.1038/srep21451.

42. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science (80-). 2013;339:823–6. doi:10.1126/science.1232033.

43. Jia B, Jeon CO. High-throughput recombinant protein expression in Escherichia coli: Current status and future perspectives. Open Biol. 2016;6.

44. Woo JW, Kim J, Kwon S Il, Corvalán C, Cho SW, Kim H, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol. 2015;33:1162–4. doi:10.1038/nbt.3389.

45. Jansing J, Schiermeyer A, Schillberg S, Fischer R, Bortesi L. Genome Editing in Agriculture: Technical and Practical Considerations. Int J Mol Sci. 2019;20. doi:10.3390/ijms20122888.

46. Fossi M, Amundson K, Kuppu S, Britt A, Comai L. Regeneration of solanum tuberosum plants from protoplasts induces widespread genome instability. Plant Physiol. 2019;180:78–86.

47. Charrier A, Vergne E, Dousset N, Richer A, Petiteau A, Chevreau E. Efficient targeted mutagenesis in apple and first time edition of pear using the CRISPR-Cas9 system. Front Plant Sci. 2019;10 February:1–12.

48. Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahante JE, Webber BR, et al. EditR: A Method to Quantify Base Editing from Sanger Sequencing. Cris J. 2018;1:239–50.

49. Wang Y, Zhou L, Tao R, Liu N, Long J, Qin F, et al. SgBE: A structure-guided design of sgRNA architecture specifies base editing window and enables simultaneous conversion of

cytosine and adenosine. Genome Biol. 2020;21:1–16.

50. Cheng TL, Li S, Yuan B, Wang X, Zhou W, Qiu Z. Expanding C–T base editing toolkit with diversified cytidine deaminases. Nat Commun. 2019;10:1–10. doi:10.1038/s41467-019-11562-6.

51. Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat Biotechnol. 2021;39:41–6. doi:10.1038/s41587-020-0609-x.

52. Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Methods. 2015;12:823–6. doi:10.1038/nmeth.3473.

53. Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576:149–57. doi:10.1038/s41586-019-1711-4.

54. Lin Q, Jin S, Zong Y, Yu H, Zhu Z, Liu G, et al. High-efficiency prime editing with optimized , paired pegRNAs in plants. Nat Biotechnol. doi:10.1038/s41587-021-00868-w.

55. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. PLoS One. 2011;6.

56. Lee ME, DeLoache WC, Cervantes B, Dueber JE. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth Biol. 2015;4:975–86.

57. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5. doi:10.1038/nmeth.2089.

58. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time

quantitative PCR and the 2-ΔΔCT method. Methods. 2001;25:402–8.

Figure Legends

Fig. 1 Screening of different promoters to evaluate the protein expression activities.

a. Information of bacterial, virus-derived plant, and native-plant promoters along with their additional components and size used in the present study is summarized. The details about the source are provided in Supplementary Information. **b.** Schematic representation of superfold GFP (sfGFP)-expressing transcriptional units and sfGFP fluorescence in 10-beta E. coli cells. **c.** Relative expression levels of sfGFP driven by seven different promoters were inferred from the corrected total cell fluorescence (CTCF). Statistical analysis of three independent biological replicates was performed and plotted using GraphPad Prism 9.0 software. The significant differences were calculated by one-way analysis of variance followed by Tukey's multiple comparisons test and are indicated at the top (****P<0.0001, ns-not significant). Bars represent mean, and error bars represent the standard error of the mean (SEM) values.

Fig. 2 Evaluation of C•G to T•A conversion by PmCDA1-mediated cytosine base editor (CBE) driven by different promoters in *E. coli* cells.

a. Schematic representation of IRI-CCE (*in vivo* rapid investigation of CRISPR components in *Escherichia coli*) platform for testing the new genome editing tools, reliability of cloned plasmids and editing activities of designed gRNAs. Dual expression promoters enable the production of CRISPR components (Cas9-fusions and sgRNAs) in heterologous hosts including *E. coli* and plants. **b and c.** Two independent gRNAs were tested containing alternate Cs at even (Test gRNA1, **b**) and odd (Test gRNA2, **c**) positions spanning from 1 to 10 in the 20 bp protospacer length. The numbers assigned by counting the distal base as 1 from the protospacer adjacent motif (PAM), i.e., GGG. Schematic representation of the plasmid vectors showing a synthesized target region for gRNA and two transcriptional units (TUs) composing of nCas9 (D10A) fused with PmCDA1-1xUGI, and AtU6 promoter-sgRNA unit. **d and e.** PmCDA1-based C•G to T•A editing activities. Graph values show mean percentage on y axis and the tested protospacer positions on x axis. The graph bar shows the mean of percentage values and error bars

indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates. **f.** Summary of C•G to T•A editing activity windows of tested gRNAs for PmCDA1-CBE expressed by different promoters. Dark green zone (>40%) indicates the ideal editing window with higher editing efficiencies, and green (20-40%) or light green (<20%) zone regions indicate lesser C•G to T•A conversion.

Fig. 3 evoCDA1 and APOBEC3A expressed under the control of different promoters show expanded editing windows with efficient C-to-T editing.

a. Nucleotide sequences of three gRNAs (underlined in black) along with additional N-terminal regions (underlined in red). The numbers are assigned by counting the distal base as 1 from the protospacer adjacent motif (PAM), i.e., GGG. **b and c.** evoCDA1 (**b**) and APOBEC3A (A3A) (**c**) driven C•G to T•A editing activities for three substrates targeted in three independent experiments. The designed BE constructs are represented in the upper side, and obtained results depicted in bar graphs. G-5 position (yellow shade) data showed G•C to A•T conversion, suggesting editing happened in the target strand. Graph values show the mean percentage on the y-axis and the tested protospacer positions on the x-axis. The graph bar shows the mean of percentage values, and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates. **d.** Summary of C•G to T•A editing activity windows of tested gRNAs for evoCDA1 and A3A expressed by different promoters. Dark green zone (>40%) indicates the ideal editing window with higher editing efficiencies, and green (20-40%) or light green (<20%) zone regions indicate lesser C•G to T•A conversion. The pink zone indicates the inaccessible regions for C•G to T•A editing.

Fig. 4 Adenine base editor ABE8e efficiently converts A•T to G•C in *E. coli* cells.

a. Nucleotide sequence of Test gRNA1 (underlined in black) showing available adenine (A) residues (green) and schematic diagram of designed ABE8e plasmid vectors. **b.** Test gRNA2 sequence (highlighted with underline) with available As and graphical representation of designed ABE8e plasmid vectors. **c and d.** The A•T to G•C conversion efficiencies of ABE8e driven by pGlpT and p35S(L)I

promoters in the target region for Test gRNA1 (**c**) and Test gRNA2 (**d**). The graph bar shows the mean of percentage values, and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates. **e.** Summary of A•T to G•C editing windows of tested gRNAs for ABE8e.

Fig. 5 Screening of designed gRNAs in the IRI-CCE platform provides first-hand knowledge about the sgRNA functionality.

a. Set of gRNAs used for validating the potential use of IRI-CCE in gRNA screening. The base editor (BE) type was chosen depending on the availability of A or C in the editing window. Frequencies of base conversions by utilized BE types in target regions of AtGL1-gRNA1 (**b**), SlMlo1-gRNA2 (**c**), and SlPelo-gRNA1 (**d**), Inactive gRNA1 (**e**), Inactive gRNA2 (**f**), Os-gRNA22 (**g**) and Os-gRNA23 (**h**). The graph bar shows the mean of percentage values, and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates.

Supplementary Information

- **Supplementary Fig. 1**. Promoter sequences from different organisms used in Fig. 1.
- **Supplementary Fig. 2**. Evaluation of promoter activities at the transcriptional level in *E. coli*.
- **Supplementary Fig. 3**. Promoter sequences used for sgRNA expression in different organisms.
- **Supplementary Fig. 4**. Promoter sequence alignment used for sgRNA expression in bacteria (J23119) and plant (Arabidopsis U6).
- **Supplementary Fig. 5**. The scheme followed for cloning the target regions as DNA templates for gRNA binding and subsequently CRISPR editing.
- **Supplementary Fig. 6**. Approaches analyzed for the establishment of IRI-CCE platform.

- **Supplementary Fig. 7**. Differential-strength promoter-driven cytosine base editing (CBE) activities in *E. coli* genome.
- **Supplementary Fig. 8**. Features of inactive gRNAs from Fig. 5.
- **Supplementary Fig. 9.** Editing windows for CBEs (PmCDA1, evoCDA1, APOBEC3A) and ABE reported in the present study expressed under the promoters of different strengths.
- **Supplementary Table 1** Strengths and limitations of methods/platforms to evaluate the gRNA functionality.
- **Supplementary Table 2** *Escherichia coli* strains used in present study.
- **Supplementary Table 3** Primer sequences used for cloning and sequencing are summarized.
- **Supplementary Table 4** Plasmids used in present work.
- **Supplementary Table 5** DNA sequences of CRISPR and BE components.

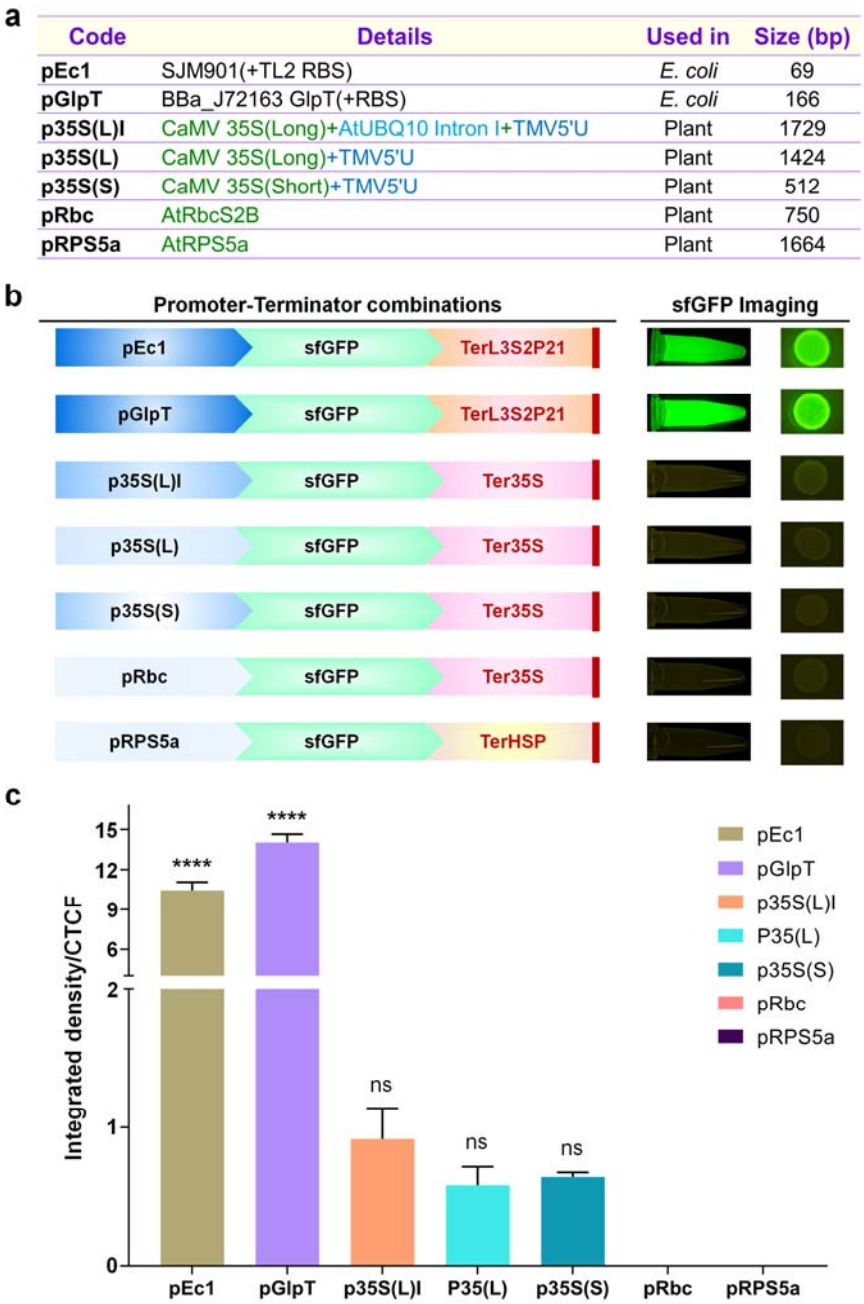

Fig. 1 Screening of different promoters to evaluate the protein expression activities. **a.** Information of bacterial, virus-derived plant, and native-plant promoters along with their additional components and size used in the present study is summarized. The details about the source are provided in Supplementary Information. **b.** Schematic representation of superfold GFP (sfGFP)-expressing transcriptional units and sfGFP fluorescence in 10-beta *E. coli* cells. **c.** Relative expression levels of sfGFP driven by seven different promoters were inferred from the corrected total cell fluorescence (CTCF). Statistical analysis of three independent biological replicates was performed and plotted using GraphPad Prism 9.0 software. The significant differences were calculated by one-way analysis of variance followed by Tukey's multiple comparisons test and are indicated at the top (****P<0.0001, ns-

not significant). Bars represent mean, and error bars represent the standard error of the mean (SEM) values.

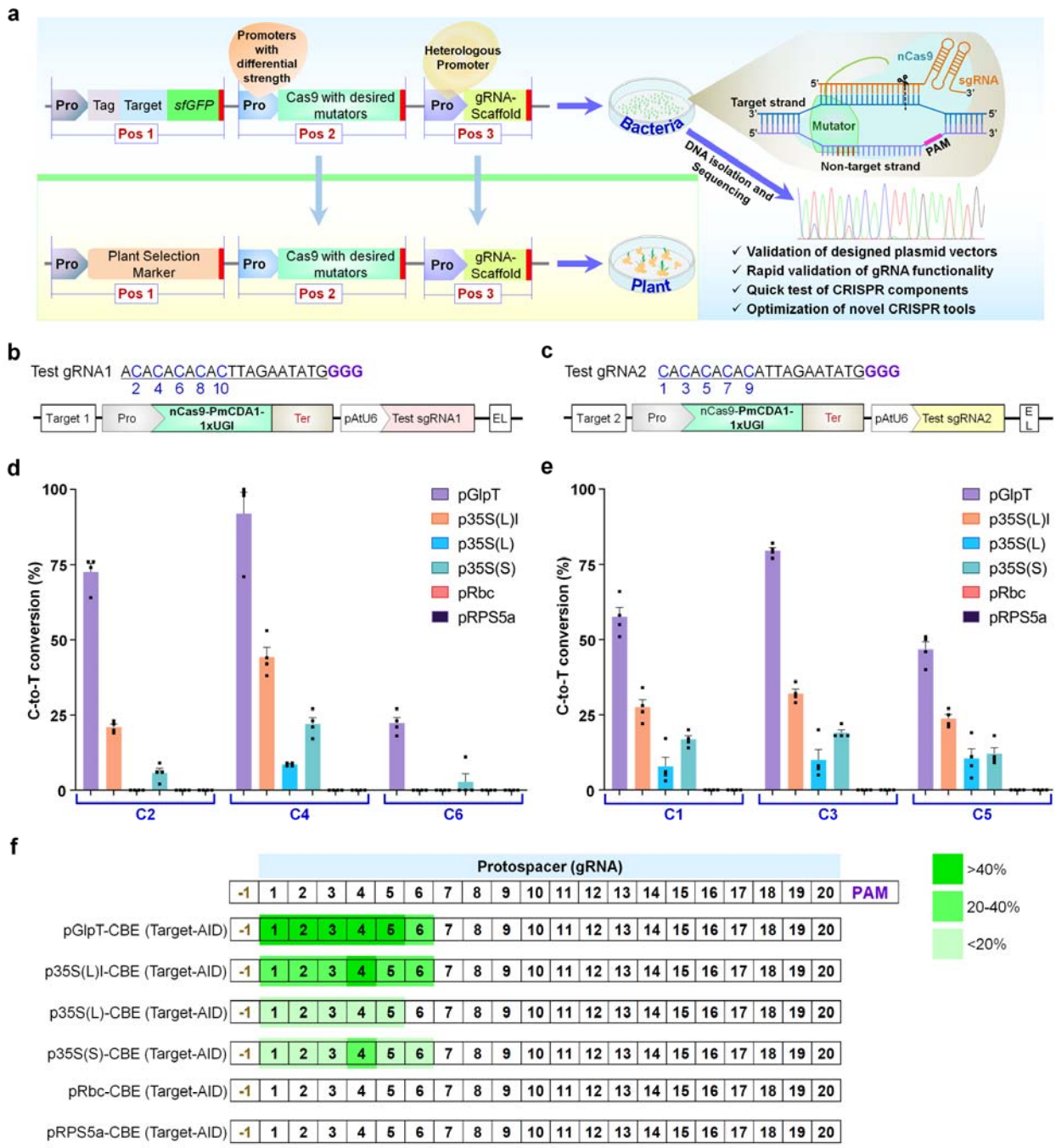

Fig. 2 Evaluation of C•G to T•A conversion by PmCDA1-mediated cytosine base editor (CBE) driven by different promoters in *E. coli* cells.

a. Schematic representation of IRI-CCE (*in vivo* rapid investigation of CRISPR components in *Escherichia coli*) platform for testing the new genome editing tools, reliability of cloned plasmids and editing activities of designed gRNAs. Dual expression promoters enable the production of CRISPR components (Cas9-fusions and sgRNAs) in heterologous hosts including *E. coli* and plants. **b and c.** Two independent gRNAs were tested containing alternate Cs at even (Test gRNA1, **b**) and odd (Test gRNA2, **c**) positions spanning from 1 to 10 in the 20 bp protospacer length. The numbers assigned by counting the distal base as 1 from the protospacer adjacent motif (PAM), i.e., GGG. Schematic representation of the

plasmid vectors showing a synthesized target region for gRNA and two transcriptional units (TUs) composing of nCas9 (D10A) fused with PmCDA1-1xUGI, and AtU6 promoter-sgRNA unit. **d and e.** PmCDA1-based C•G to T•A editing activities. Graph values show mean percentage on y axis and the tested protospacer positions on x axis. The graph bar shows the mean of percentage values and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates. **f.** Summary of C•G to T•A editing activity windows of tested gRNAs for PmCDA1-CBE expressed by different promoters. Dark green zone (>40%) indicates the ideal editing window with higher editing efficiencies, and green (20-40%) or light green (<20%) zone regions indicate lesser C•G to T•A conversion.

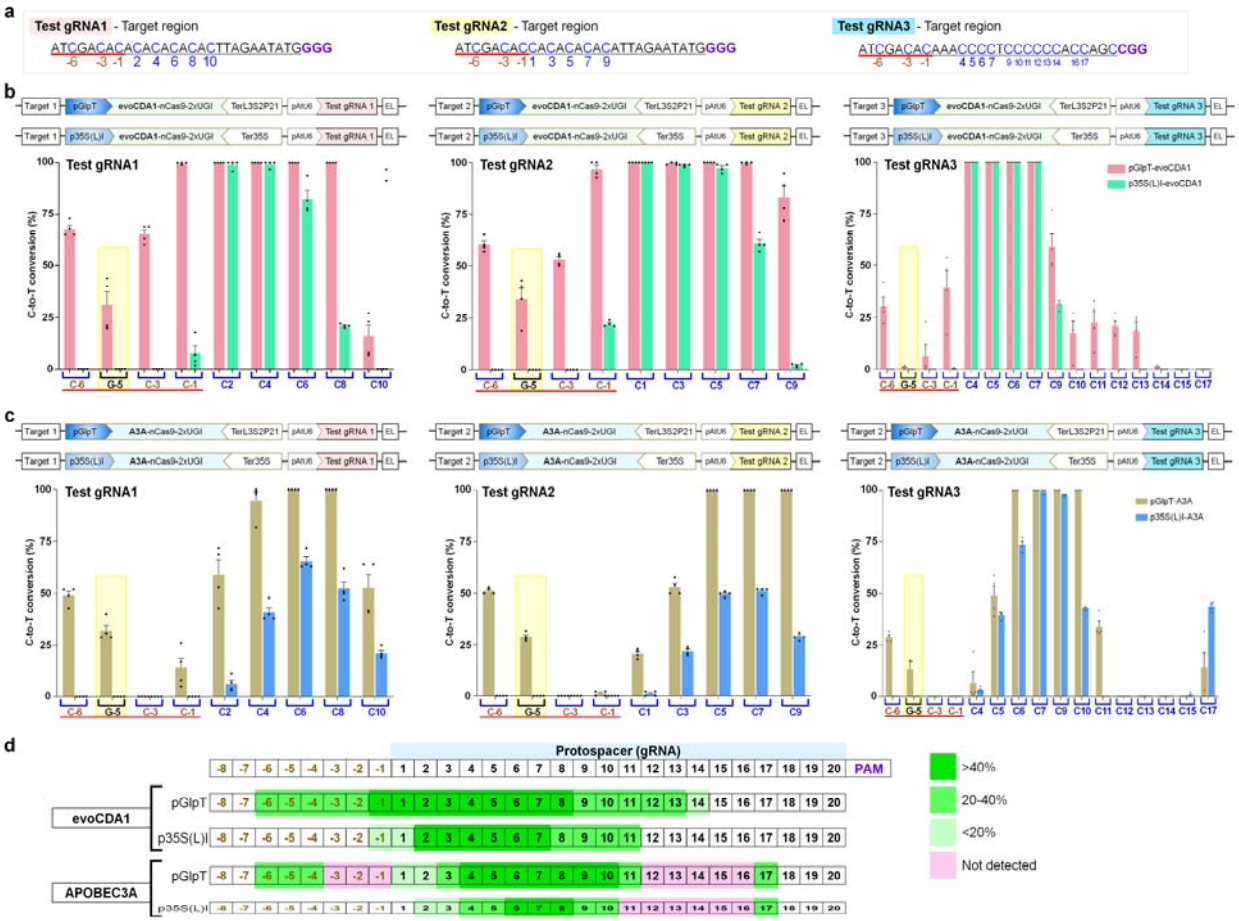

Fig. 3 Characterization of broad-range evoCDA1 and APOBEC3A cytosine base editors (CBEs) driven by differential-strength promoters.

a. Nucleotide sequences of three gRNAs (underlined in black) along with additional N-terminal regions (underlined in red). The numbers are assigned by counting the distal base as 1 from the protospacer adjacent motif (PAM), i.e., GGG. **b and c.** evoCDA1 (**b**) and APOBEC3A (A3A) (**c**) driven C•G to T•A editing activities for three substrates targeted in three independent experiments. The designed BE constructs are represented in the upper side, and obtained results depicted in bar graphs. G-5 position (yellow shade) data showed G•C to A•T conversion, suggesting editing happened in the target strand. Graph values show the mean percentage on the *y*-axis and the tested protospacer positions on the *x*-axis. The graph bar shows the mean of percentage values, and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates. **d.** Summary of C•G to T•A editing activity windows of tested gRNAs for evoCDA1 and A3A expressed by different promoters. Dark green zone (>40%) indicates the ideal editing window with higher editing efficiencies, and green (20-40%) or light green (<20%) zone regions indicate lesser C•G to T•A conversion. The pink zone indicates the inaccessible regions for C•G to T•A editing.

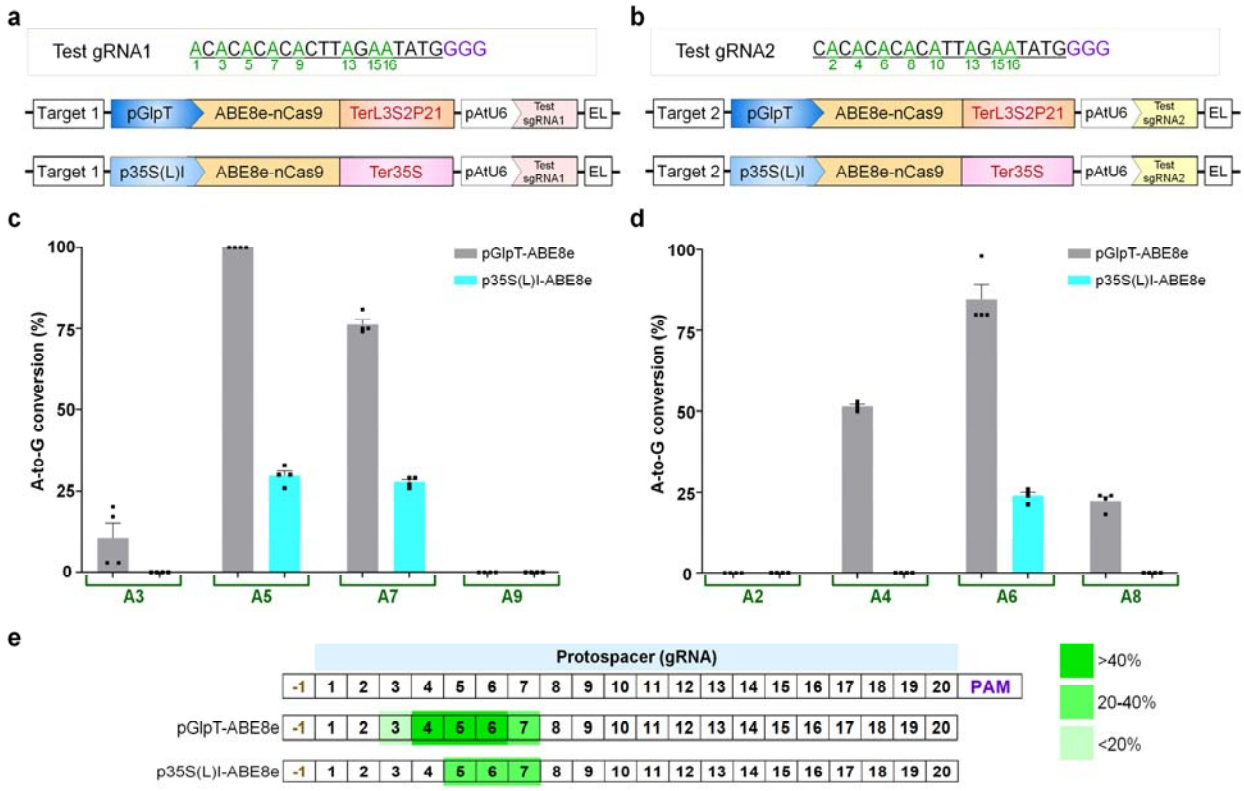

Fig. 4 Adenine base editor ABE8e efficiently converts A•T to G•C in *E. coli* cells.
a. Nucleotide sequence of Test gRNA1 (underlined in black) showing available adenine (A) residues (green) and schematic diagram of designed ABE8e plasmid vectors. **b.** Test gRNA2 sequence (highlighted with underline) with available As and graphical representation of designed ABE8e plasmid vectors. **c and d.** The A•T to G•C conversion efficiencies of ABE8e driven by pGlpT and p35S(L)I promoters in the target region for Test gRNA1 (**c**) and Test gRNA2 (**d**). The graph bar shows the mean of percentage values, and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates. **e.** Summary of A•T to G•C editing windows of tested gRNAs for ABE8e.

a

Code	Guide RNA Sequence	Reported Activity (Reason)	Source	BE used
AtGL1- gRNA1	GGAAAAGTTGTAGACTGAGATGG	Active	Hahn et al., 2017	ABE8e
SlMlo1- gRNA2	GTACAAAGTTAATCAAGAATAGG	Active	Nekrasov et al. 2016	ABE8e
SlPelo- gRNA1	TCCAGCATCATTCAGTTGTGGGG	Active	Pramanik et al. 2021	PmCDA1
Inactive gRNA1	GGTGAAGCAGCGGACAGCAGTGG	Inactive (Internal sgRNA interactions)	Thyme et al. 2016	ABE8e
Inactive gRNA2	GGGATGAGCATCGGGTAGCCTGG	Inactive (Internal sgRNA interactions)	Thyme et al. 2016	ABE8e
Os- gRNA22	CTTAATTTCTGTTAGGTTCTTGG	Inactive (7 internal base pairs)	Liang et al. 2016	ABE8e
Os- gRNA23	AGTACAAGTTTATTTTTTTCTGG	Inactive (20% GC content, T-rich repeats)	Liang et al. 2016	PmCDA1

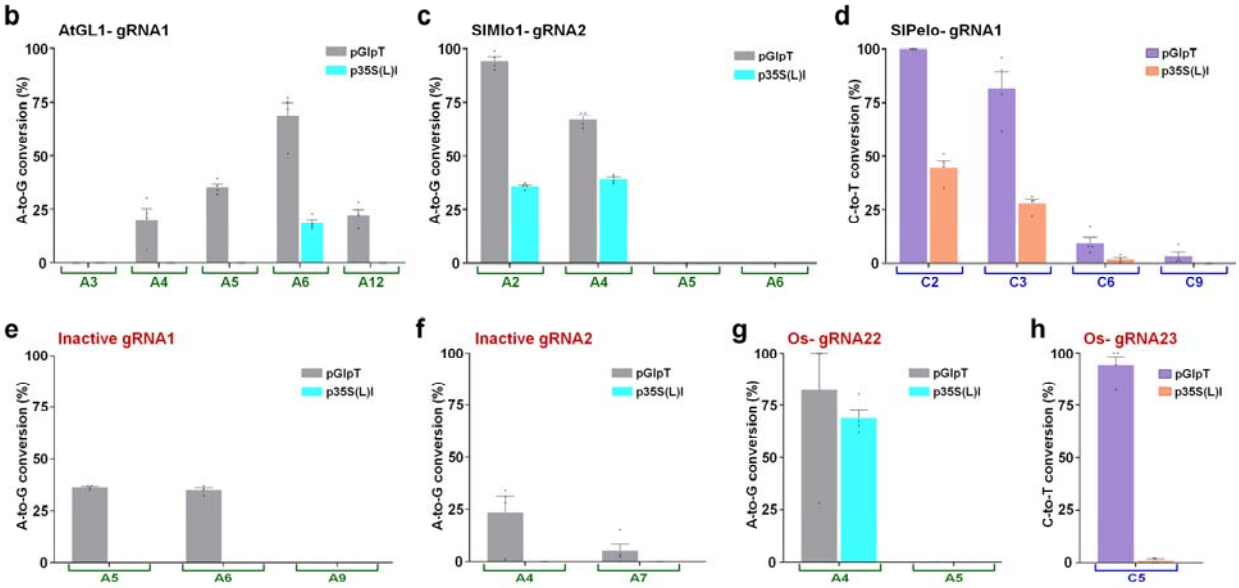

Fig. 5 Screening of designed gRNAs in the IRI-CCE platform provides first-hand knowledge about the sgRNA functionality.

a. Set of gRNAs used for validating the potential use of IRI-CCE in gRNA screening. The base editor (BE) type was chosen depending on the availability of A or C in the editing window. Frequencies of base conversions by utilized BE types in target regions of AtGL1-gRNA1 (**b**), SlMlo1-gRNA2 (**c**), and SlPelo-gRNA1 (**d**), Inactive gRNA1 (**e**), Inactive gRNA2 (**f**), Os-gRNA22 (**g**) and Os-gRNA23 (**h**). The graph bar shows the mean of percentage values, and error bars indicate standard error of the mean (mean ±s.e.m.) of four independent biological replicates. Dots indicate the individual biological replicates.