Efficient long fragment editing technique enables rapid construction of genetically stable bacterial strains

Chaoyong Huang[1,2,#], Liwei Guo[1,#], Jingge Wang[1], Ning Wang[1], Yi-Xin Huo[1,2,*]

[1]Key Laboratory of Molecular Medicine and Biotherapy, School of Life Sciences, Beijing Institute of Technology, No. 5 South Zhongguancun Street, Beijing, China, 100081

[2]SIP-UCLA Institute for Technology Advancement, 10 Yueliangwan Road, Suzhou Industrial Park, Suzhou, China, 215123

[#]These authors contributed equally to this work: Chaoyong Huang, Liwei Guo

*To whom correspondence should be addressed. Email: huoyixin@bit.edu.cn. Phone number: +86 18910254522

Abstract

Background: Bacteria are versatile living systems that enhance our understanding of nature and enable biosynthesis of valuable molecules. Long fragment editing techniques are of great importance for accelerating bacterial chromosome engineering to obtain desirable and genetically stable strains. However, the existing genomic editing methods cannot meet the needs of researchers.

Results: We herein report an efficient long fragment editing technique for complex chromosomal engineering in *Escherichia coli*. The technique enabled us to integrate DNA fragments up to 12 kb into the chromosome, and to knock out DNA fragments up to 187 kb from the chromosome, with over 95% positive rates. We applied this technique for *E. coli* chromosomal simplification, resulting in twelve individual deletion mutants and four cumulative deletion mutants. The simplest chromosome lost a 370.6 kb DNA sequence containing 364 open reading frames. In addition, we applied the technique to metabolic engineering and constructed a genetically stable plasmid-independent isobutanol production strain that produced 1.3 g/L isobutanol via shake-flask micro-aerobic fermentation.

Conclusions: These results suggested that the technique is a powerful chromosomal engineering tool, highlighting its potential to be applied in different fields of synthetic biology.

Keywords: chromosomal engineering, long fragment editing, genetic stability, genome simplification, metabolic engineering

Background

As a class of versatile living systems, bacteria are useful in many fields of synthetic biology. In bacteria, genetic information contained on the single-copy chromosome determines the characteristics of a specific strain. To understand bacterial characteristics and utilize them to explore the world and serve human life, researchers frequently conduct chromosomal engineering to reprogram the genetic information of bacteria. Through DNA editing, researchers can add desired exogenous genetic information to or delete unwanted endogenous genetic information from the bacterial chromosome. The long fragment editing technique is of great importance in accelerating bacterial chromosome engineering to obtain genetically stable strains. For example, the long fragment deletion technique can help to simplify the bacterial chromosome to explore the minimal genome of a specific strain [1, 2], and the long fragment insertion technique can help to expand the bacterial chromosome to archive the increasing information of the human world [3]. In metabolic engineering, plasmid maintenance requires continuous antibiotic use, which has led to biosafety issues and elevated industrial cost [4]. The long fragment editing technique is an ideal tool for constructing plasmid-independent and high-production strains.

To accelerate the process of chromosomal engineering, researchers have reported many methods for generating insertions and deletions in bacterial chromosome. Homologous recombination with polymerase chain reaction (PCR) fragments forms the basis of these methods [5, 6]. However, since RecA-mediated homologous

recombination with linear DNA is of low efficiency, researchers created the desired mutagenesis on a suitable plasmid before recombining it into the genome [7-9]. To enhance the efficiency of homologous recombination, the bacteriophage-derived λ-Red system was introduced into bacteria on either the genome or plasmids. Genomic editing based on λ-Red recombinases is referred to as recombineering [10-12]. In recombineering, an antibiotic resistance gene is required as a selectable marker. To remove the selectable marker after genomic editing, researchers introduced counter-selection systems or site-specific recombination systems, including FLP/*FRT* and Cre/*loxP* [13, 14]. Though recombineering can handle the insertion and deletion of short DNA fragments [15-17], the editing efficiency decreases dramatically for long fragments [11]. Moreover, eliminating selectable markers and plasmids is complicated and time-consuming, and the residual *FRT* or *loxP* site may influence a new round of genomic editing [13]. Generating a double-strand break (DSB) in the target DNA is an effective strategy for improving the efficiency of long-fragment manipulations. Though the homing endonuclease I-SceI is efficient for cleaving double-stranded DNA (dsDNA), researchers had to integrate an 18-bp recognition site into the target DNA before inducing DNA cleavage [18-20]. Engineered endonucleases, such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), can be programmed to recognize and cleave the genome at a specific locus. However, these approaches require engineering new enzymes for each target sequence [21-24]. Recently, clustered regularly interspaced short palindromic repeats

(CRISPR)/CRISPR-associated protein 9 (Cas9) technology was developed based on research into the adaptive immune system of *Streptococcus pneumoniae* [25]. Cas9 endonuclease complexed with a designed single-guide RNA (sgRNA) can generate DSB in a specific protospacer sequence where a proper protospacer-adjacent motif (PAM) exists [25-27]. The technique relies on sgRNA-directed cleavage at the target site to kill wild-type cells, thus circumventing the need for selectable markers or counter-selection systems. Changing the 20-bp spacer sequence can reprogram the specificity of the Cas9-sgRNA complex, making CRISPR/Cas9 technology much more convenient than ZFNs and TALENs. Many methods based on the CRISPR/Cas9 technology are efficient for short-sequence editing in *Escherichia coli*. However, for long-fragment editing, the fragment length and editing efficiency still have much room for improvement.

We herein report an efficient long fragment editing technique for complex chromosome engineering in *E. coli*. The technique made full use of CRISPR/Cas9 and recombination technologies, enabling us to integrate DNA fragments up to 12 kb into the chromosome. We were also able to knock out DNA fragments up to 187 kb from the chromosome. Notably, the high performance of the technique was independent of high transformation efficiency, making the technique applicable to researchers of limited experience. Furthermore, the technique has been successfully applied in chromosomal simplification and metabolic engineering, demonstrating its potential as a genetic engineering tool for constructing genetically stable bacterial strains.

Results

Development of CRISPR/Cas9-assisted recombination system (CARS)

The CARS constructed in this study is a two-plasmid system that consists of five elements: a Cas9-expressing cassette induced by L-arabinose; an sgRNA-expressing cassette induced by L-arabinose; a λ-Red recombination system induced by isopropyl-β-D-thiogalactopyranoside (IPTG); a donor DNA-generation system; and a plasmid curing system for eliminating the two plasmids independently or together from cells (Fig. 1).

Specifically, Cas9 protein and λ-Red recombinases (Gam, Beta, and Exo) were expressed by the p15A-P_{araB}-Cas9-P_{T5}-Redγβα plasmid (plasmid#1), which contained a p15A replication origin and a kanamycin-resistant (Kan^R) gene. Targeting sgRNA was expressed by the pSC101-P_{araB}-sgRNA-Donor plasmid (plasmid#2) containing a pSC101 replication origin and an ampicillin-resistant (Amp^R) gene (Fig. 1). There are two types of plasmid#2, the first of which contains two sgRNA-expressing cassettes, and the other containing one sgRNA-expression cassette. The variant of plasmid#2 depends on the type of genomic editing. The *araB* promoter, which is strict and induced by L-arabinose, controlled Cas9 and sgRNA expression so that DNA cleavage was only initiated when the inducer was present. The T5 promoter, which is strong and induced by IPTG, controlled λ-Red recombinase expression to ensure homologous recombination took place in time after DNA cleavage. Donor DNA, which served as a template to introduce sequence deletions, insertions, or replacements, was constructed and integrated into the plasmid#2 (Fig. 1). The

plasmid-borne donor DNA could avoid nuclease attack and copy itself along with the replication of plasmid#2. The target site (N20 + PAM) on the genome was added to plasmid#2 in the flanks of donor DNA, thus the donor DNA could be cut off from plasmid#2 during genomic editing. This generated linear donor DNA that participated in homologous recombination with the cleaved genomic DNA (Fig. 1). At an appropriate concentration of L-arabinose, the expression levels of Cas9 and sgRNA were enough for cleaving the single-copy genome, but insufficient for cleaving all copies of the plasmid#2 (about five copies). Therefore, cells still possessed resistance to Amp. To construct the plasmid curing system, we used the temperature-sensitive pSC101 replication origin for plasmid#2 and added the sucrose-sensitive *sacB* gene to plasmid#1 as a counter-selection marker (Fig. 1).

Each cycle of editing started with the transfection of plasmid#2 into cells containing plasmid#1 (Fig. 1 and Figure S1). Then, we cultivated the correct transformants containing the two plasmids for cell reproduction before adding inducers to trigger DNA cleavage and DSB repair. Theoretically, sgRNA guides Cas9 to recognize and cleave the target DNA, generating DSB in the genome and plasmid#2. Then, the λ-Red recombinases mediate homologous recombination between the broken genome and linear donor DNA. This transfers the desired mutation from the donor DNA to the genome, destroying the target site (Fig. 1 and Figure S1). The cells acquiring the desired mutation survive, and the cells with an unrepaired genome undergo cell death. Thus, plating liquid cultures on agar medium containing Kan and Amp allowed the selection of desired clones. Colonies growing

on the plates were further verified through PCR and sequencing. Then, correct mutants were cultivated at 40 ℃ in medium containing only Kan to eliminate plasmid#2 (Figure S2a). The cultures were inoculated into fresh medium to prepare competent cells for a new round of editing (Figure S1). Each cycle of editing required only three days. After the final round of editing, plasmid#1 and plasmid#2 were eliminated by incubating the correct clones at 40 °C in antibiotic-free medium and plating the cultures on agar medium containing sucrose (Figure S1 and Figure S2).

CARS-mediated long fragment integration

To evaluate the ability of CARS to mediate long fragment integration, we tried to insert fragments of different lengths (3 kb, 6 kb, 9 kb, and 12 kb) into the *lacZ* gene of *E. coli* strain MG1655 (Fig. 2a). We constructed four different versions of plasmid#2 harboring the corresponding donor DNA and expressing the same sgRNA targeting the *lacZ* gene. The four inserted fragments came from the F plasmid of *E. coli* strain XL1-Blue, and they had no homology with the MG1655 genome. The insertion of these fragments would inactivate the *lacZ* gene encoding β-galactosidase. Thus, we could differentiate edited and unedited colonies via blue-white selection. The edited colonies were white in a Luria-Bertani (LB) plate containing IPTG and X-gal, while the unedited colonies were blue. We also identified edited clones though PCR. One pair of primers (F1/R1) was designed for the verification of 3-kb insertion (Fig. 2a), and correct clones obtained much larger PCR products than the control (Figure S3a). Two pairs of primers were designed for the verification of 6-kb, 9-kb, and 12-kb insertions (Fig. 2a). The correct clones obtained the desired PCR products using both

F1/R2 and F2-X/R1 (X=1, 2, 3), while the control did not (Figure S3b–d). The PCR products were further verified by sequencing. Based on the results of blue-white selection, PCR, and sequencing, we determined the editing efficiencies and positive rates. The editing efficiencies in these four insertion experiments were 1.2×10^{-3}, 1.2×10^{-3}, 9.6×10^{-4}, and 7.2×10^{-4}, respectively (Fig. 2b). The positive rates in the four insertion experiments were 97.3%, 98.3%, 96.7%, and 98.3%, respectively (Fig. 2b). These results indicated that both Cas9-mediated DNA cleavage and λ-Red-mediated DSB repair were efficient in our experiments. We found that the small-proportion negative colonies ($<5\%$), commonly called “escapers” [27, 28], came from two sources. More than half of the “escapers” did not undergo cleavage by Cas9, probably because of the limited induction time and intensity of L-arabinose. The remaining “escapers” acquired deletions of unknown length in the target site, which was likely due to the presence of A-EJ repair [29, 30]. We tried to insert a 15-kb fragment into the *lacZ* gene, but failed, because the corresponding plasmid#2, which was over 20 kb in size, was difficult to construct. The 12-kb insertion is sufficient for application in metabolic engineering. To highlight the advantages of our method, we compared CARS to three representative methods that performed relatively well in long fragment insertion. These data came from published articles [28, 31, 32]. Our method performed much better than the others when comparing both largest insertion length and positive rate (Fig. 2c).

CARS-mediated long fragment knockout

Firstly, we successfully deleted a 99.9-kb fragment, starting at 565,156 and ending at

665,088, in the MG1655 genome (Fig. 3a). To determine the relationship between editing performance and the length of the deleted fragment, we selected seven fragments of different lengths within the 99.9-kb fragment for individual deletion. The lengths of these fragments were 9.1 kb, 21.5 kb, 30.6 kb, 39.4 kb, 59.8 kb, 79.8 kb, and 99.9 kb (Fig. 3a). To delete these fragments, we constructed seven different versions of plasmid#2 harboring two sgRNA-expressing cassettes. One sgRNA targets the same site (TS1) in the genome, and the other targets different sites (TS2-1–TS2-7) (Fig. 3a). Based on the results of PCR and sequencing, we determined their editing efficiencies and positive rates (Fig. 3b). As demonstrated, all positive rates were over 95%, similar to the results in long fragment insertion experiments. The deletion of 9.1-kb, 21.5-kb, 30.6-kb, 39.4-kb, 59.8-kb, and 79.8-kb fragments resulted in similar editing efficiencies, and the deletion of the 99.9-kb fragment resulted in lower editing efficiencies (Fig. 3b). We found that the 99.9-kb fragment knockout strain grew much more slowly than MG1655, while the 79.8-kb fragment knockout strain had a similar growth rate to MG1655 (Figure S4a and S4d). This phenomenon implied that the terminal region of the 99.9-kb fragment contained some genetic information that was important, but not essential, for cell survival. The decrease in editing efficiency of the 99.9-kb deletion experiment was probably due to the lower viability of edited cells. In this study, we also successfully deleted other long fragments in the genome (Fig. 4d). To highlight the advantages of our method, we compared CARS with four representative methods that performed relatively well in long fragment deletion. The data came from published articles [28, 33-35]. In comparison to these data, our

method performed much better in terms of both largest deletion length and positive rate (Fig. 3c).

Identification of nonessential sequence and chromosomal simplification

According to previous reports, the MG1655 chromosome harbors 4497 genes, including 4296 protein-encoding genes and 201 RNA-encoding genes [36, 37]. Researchers at Keio University identified the essentiality of all protein-encoding genes in *E. coli* K-12 by single gene deletion, generating the Keio collection [38, 39]. This provided important information for us to identify potential nonessential long fragments in the MG1655 genome. To delete a long fragment, we needed to construct a plasmid#2 that expressed a pair of sgRNA targeting two flanks of the fragment and harboring the corresponding donor DNA (Fig. 4a). To delete a long fragment harboring a limited number of essential genes, we added these genes to the corresponding plasmid#2 between the two homologous arms. Therefore, the essential genes remained in the chromosome after genomic editing, and the edited cells survived (Fig. 4b and 4c). For each long fragment deletion, we designed two pairs of primers for PCR verification. The first primer pair targets DNA sequences within the long fragment, and the second primer pair targets the adjacent sequences outward the two homologous arms (Fig. 4d and Figure S5). The correct clones did not obtain PCR product using the first primer pair, but obtained the corresponding PCR products using the second. On the contrary, the unedited control clone obtained the corresponding PCR products using the first primer pair, but did not obtain PCR products using the second (Fig. 4e and Figure S6).

234 Altogether, we successfully deleted twelve long nonessential fragments in the
235 MG1655 genome (Table 1), including the 99.9-kb fragment (No. 3) mentioned in the
236 previous section. These fragments are located in different regions of the genome, and
237 their lengths range from 52.0 to 186.7 kb. Among the twelve fragments, No. 3, No. 8,
238 and No. 11 harbor one essential gene; No. 1 and No. 4 harbor two essential genes; and
239 No. 9 harbors three essential genes (Table 1). Based on the results of PCR and
240 sequencing, we determined the editing efficiencies and positive rates (Fig. 4f). All
241 positive rates were over 95%, and the editing efficiencies ranged from 2.3×10^{-4} to
242 1.3×10^{-3}. The deletion of fragments No. 3, No. 4, and No. 7 led to much lower
243 editing efficiencies than that from deletion of the other fragments. By measuring
244 growth curves of the twelve knockout strains, we found that the No. 3, No. 4, and No.
245 7 knockout strains grew much slower than other knockout strains, and the No. 4
246 knockout strain grew slowest (Figure S4). This may have led to the lower editing
247 efficiencies in the deletion experiments of fragments No. 3, No. 4, and No. 7. The
248 results indicated that these fragments were important, but not essential, for cell
249 growth.

250 After deleting twelve long fragments individually, we tried to construct cumulative
251 deletion mutants. Here, we used MG1655-ΔNo. X to represent the MG1655 mutant
252 that loses fragment No. X (X=1, 2, 3, …, 12). As No. 1 was the longest fragment
253 deleted in this study (Table 1), we chose to construct cumulative deletion mutants on
254 the basis of strain MG1655-ΔNo. 1. Though iterative editing, we successfully deleted
255 fragment No. 9 from MG1655-ΔNo. 1, generating strain MG1655-ΔNo. 1/ΔNo. 9 that

lost a total of 270.7 kb of the DNA sequence, containing 268 open reading frames (ORFs) (Fig. 4g). We then tried to delete a third fragment on the basis of MG1655-ΔNo. 1/ΔNo. 9. According to the growth curves of single deletion mutants, the knockout of fragment No. 2, No. 5, No. 6, No. 8, No. 10, or No. 12 had no apparent influence on cell growth (Figure S4). Therefore, we attempted to delete these fragments individually in MG1655-ΔNo. 1/ΔNo. 9. As a result, we successfully obtained strains MG1655-ΔNo. 1/ΔNo. 9/ΔNo. 2, MG1655-ΔNo. 1/ΔNo. 9/ΔNo. 5, and MG1655-ΔNo. 1/ΔNo. 9/ΔNo. 6. The three knockout strains lost a total of 324.1 kb, 370.6 kb, and 368.7 kb of the DNA sequence containing 315, 364, and 368 ORFs, respectively (Fig. 4g). We failed to knock out fragments No. 8, No. 10, and No. 12 in MG1655-ΔNo. 1/ΔNo. 9 despite repeating the experiments several times, implying that these fragments were all essential for the survival of MG1655-ΔNo. 1/ΔNo. 9.

Metabolic engineering of *E. coli* for producing isobutanol

Higher alcohols such as isobutanol and n-butanol show promise in becoming the next generation of biofuels, due to their higher energy density, higher vapor pressure, and relatively low hydroscopicity [40, 41]. To illustrate the potential application of CARS in metabolic engineering, we used the system to modify the *E. coli* chromosome for producing isobutanol. Firstly, we constructed a chassis strain named JW74 based on MG1655 with six rounds of genomic editing (Fig. 5a). The competency of JW74 was 170-fold that of MG1655, making it much easier to transfect exogenous DNA. We then built a 7.9-kb operon and integrated it into the JW74 chromosome, thus displacing fragment No. 5 (Fig. 5a) and generating strain SH258. Fragment No. 5 was

278 99.9 kb in length, and the corresponding knockout strain grew slightly faster than its
279 parental strain (Figure S4f). The operon consists of five structural genes and 5′ and 3′
280 untranslated regions (UTRs). The 5′ UTR contains a strong bacterial
281 ribosome-binding site [42] and a T7 promoter, which naturally controls the expression
282 bacteriophage T7 RNA polymerase [43]; the 3′ UTR contains a T7 terminator. The
283 five structural genes are *alsS*, *ilvC*, *ilvD*, *kivD*, and *adhA* (Fig. 5a). Among the five
284 genes, *ilvC* and *ilvD* came from *E. coli*, *alsS* came from *Bacillus subtilis* [44], and
285 *kivD* and *adhA* came from *Lactococuus lactis* [45] (Fig. 5b). In order to initiate
286 transcription of the operon, we introduced the T7 RNA polymerase-encoding gene
287 controlled by the T5 promoter [46] to the SH258 genome, generating the SH274 strain
288 (Fig. 5a). Though the T5 promoter is a strong inducible promoter repressed by LacI, it
289 served here as a strong constitutive promoter. This is because SH274 is a
290 *lacI*-defective strain. In traditional metabolic engineering, introducing a
291 high-copy-number fermentation plasmid is a commonly used strategy to overexpress
292 enzymes related to the target products. Therefore, we constructed the
293 pColE1-P_{T5}-*alsS*-*ilvC*-*ilvD*-*kivD*-*adhA* plasmid and transfected it into JW74,
294 generating the SH279 strain.

295 We used the strains SH274 and SH279 to conduct micro-aerobic fermentation in
296 shake flasks containing 20 mL M9 medium. Briefly, the acetolactate synthase (AlsS)
297 converts pyruvate, the intermediate product of glycolysis, into 2-acetolactate. This is
298 then transformed into 2,3-dihydroxy-isovalerate by ketol-acid reductoisomerase
299 (IlvC). As the substrate of dihydroxyacid dehydratase (IlvD),

2,3-dihydroxy-isovalerate is converted into 2-ketoisovalerate, which is transformed into isobutyraldhyde by 2-ketoisovalerate decarboxylase (KivD). Finally, isobutyraldhyde is catalyzed by alcohol dehydrogenase (AdhA), generating isobutanol (Fig. 5b). During fermentation, samples were taken every 12 hours to measure the OD600 value and isobutanol titer (Fig. 5c). As a result, isobutanol reached a maximum titer of 1.3 g/L after 48 hours of SH274 fermentation (Fig. 5c). To our knowledge, this was the first attempt to produce isobutanol without introducing a high-copy-number fermentation plasmid, and isobutanol production was higher than many reports using such a plasmid [47, 48]. For strain SH279, isobutanol reached a maximum titer of 5.5 g/L after 48 hours (Fig. 5d). This is 4.2 fold that of SH274, indicating that the SH274 strain has much room for improvement. In future study, we therefore plan to increase the copy number of the operon P_{T7}-*alsS-ilvC-ilvD-kivD-adhA*-T_{T7} in the SH274 genome to strengthen the expression of related enzymes.

Discussion

In this study, we developed CARS for genomic editing in *E. coli*. As a versatile tool, CARS is efficient for different kinds of genetic modifications, including sequence insertion, sequence deletion, and sequence displacement. The CARS is particularly suitable for long-fragment manipulation with a high editing efficiency and positive rate. With the aid of CARS, we were able to integrate DNA fragments up to 12 kb into the chromosome, and to knock out DNA fragments up to 187 kb from the

chromosome. In the 12-kb insertion experiment, the positive rate was 98.3%, and the editing efficiency was 7.2×10^{-4}. In the 187-kb deletion experiment, the positive rate was 97.7%, and the editing efficiency was 1.2×10^{-3}. Other researchers have conducted a significant amount of work on long-fragment editing in *E. coli* using the CRISPR/Cas9 technique. Li et al. developed a CRISPR/Cas9-assisted recombineering method that enabled the insertion of 8-kb exogenous DNA, yielding a positive rate of 15% [28]. Utilizing a similar strategy, Chung et al. developed a method that enabled insertion of 7-kb exogenous DNA, and the positive rate was 61% in the presence of a selectable marker [31]. Li et al. reported a modified method that enabled consecutive insertion of DNA fragments, and the appropriate fragment size was 3–4 kb [32]. The system developed by Li et al. was also used to delete DNA fragments up to 12 kb, and the positive rate was 90% [28]. Zhao et al. descripted a method that enabled the deletion of DNA fragments up to 100 kb with a 75% positive rate, but a specific target site (N20 + PAM) must be integrated into chromosome through recombineering before inducing the deletion process [34]. Su et al. combined the CRISPR/Cas9 system and *Mycobacterium tuberculosis*-derived non-homologous end-joining (NHEJ) system and completed the deletion of a 17-kb fragment in a homologous recombination-independent manner, with a 17% positive rate [33]. Similarly, Zheng et al. introduced a *Mycobacterium smegmatis*-derived NHEJ system and deleted a 123-kb fragment with a 36% positive rate [35]. NHEJ-mediated methods generate stochastic DNA indels in the target region, which makes chromosome editing inaccurate. In these studies, the editing efficiencies ranged from 10^{-7} to 10^{-4} [30].

Compared with existing methods, CARS had a higher editing efficiency and positive rate. In addition, the high performance of CARS was independent of the high-competency host strain, making the technique applicable to experimenters of limited experience. Using this technique, "3*N* + 1" days are sufficient for "*N*" rounds of editing. To our knowledge, the 12 kb fragment inserted and the 187 kb fragment deleted in this study are the longest fragments manipulated in *E. coli* using CRISPR/Cas9 technology. Theoretically, CARS has the potential to be used for insertion of fragments over 12 kb, as long as one is able to construct a large plasmid#2 that is over 17 kb. Similarly, CARS has the potential to be used for deletion of fragments over 187 kb, as long as one can determine a fragment of this length that is nonessential for the survival of the host strain. At present, CARS can only manipulate one target in a single round of genetic editing. In future studies, we will try to upgrade the system to manipulate more targets at the same time.

As a powerful chromosome engineering tool, CARS has great application potential. In this study, to demonstrate its potential, we have applied CARS in genome simplification and metabolic engineering. *E. coli* has been the prominent prokaryotic organism in research laboratories since the origin of molecular biology, and is arguably the most completely characterized single-cell life form [49]. According to previous studies, different *E. coli* strains possess different genome sizes. For example, MG1655, an *E. coli* K-12 strain, has a 4.6-Mb genome that harbors 4497 genes, including 4296 protein-encoding genes and 201 RNA-encoding genes [36, 37]. Functional analyses have shown that *E. coli* cells grown under given conditions use

366 only a fraction of their genes [50]. As Koob et al. have proposed, deletion of genes
367 that are nonessential under a given set of growth conditions could identify a
368 minimized set of essential *E. coli* genes and DNA sequences [51]. In past decades,
369 researchers have explored nonessential sequences and removed them from the *E. coli*
370 genome individually or cumulatively, trying to construct a minimized genome [1, 2,
371 52, 53]. Though these work is extremely important, the methods utilized to delete
372 nonessential sequences are very complicated and time-consuming. To remove a long
373 fragment from the genome, researchers have tried all classical recombination
374 techniques both alone or in combination, including Flp/*FRT*, Cre/*loxP*, λ-Red, Tn5
375 transposon, and phage P1 transduction [1, 2, 52, 54]. Compared with these methods,
376 the technique we proposed saves time and is simple to conduct. Using this technique,
377 we have constructed twelve individual-deletion and four cumulative-deletion strains
378 on the basis of MG1655, with the simplest genome lacking a 370.6 kb sequence
379 containing 364 ORFs. Although some of the deletions generated could coexist in a
380 single strain, many deletions that were viable individually were not viable when
381 combined with other deletions. These results clearly indicate that some genes are not
382 dispensable simultaneously, despite being dispensable individually. The genes
383 belonging to this group may be those involved in alternative metabolic pathways. This
384 observation also suggests that the number of essential genes is greater than estimated,
385 and further illustrates the utility of our combinatorial-deletion approach for functional
386 study of the *E. coli* genome.

387 Microorganisms are versatile living systems for achieving biosynthesis of valuable

388 molecules contributing to chemical, energy, and pharmaceutical processes [55-59].
389 Plasmids have been commonly used for domesticating microbial materials to obtain
390 desired cellular functions, due to simplicity of genetic manipulation. Inspired by
391 nature, antibiotics have been widely used to minimize phenotype variation of
392 plasmid-containing microbes. However, the use of antibiotics may result in
393 multidrug-resistant species by horizontal gene transfer, and metabolic burden leading
394 to suboptimal production of target compounds [4]. The addition of antibiotics not only
395 increases the cost, but also contaminates final products in industrial settings.
396 Chromosomal integration is a good alternative to plasmids and provides more stability
397 for artificially introduced genetic information. The technique we developed is
398 efficient for chromosomal integration. In this study, we integrated the isobutanol
399 synthetic pathway into a chassis strain derived from MG1655, generating a
400 genetically stable metabolic engineering strain that produced 1.3 g/L isobutanol in a
401 shake flask. As expected, productivity of this engineering strain was lower than the
402 strain containing a high-copy-number fermentation plasmid, mainly due to the low
403 expression of related enzymes. In future studies, we will endeavor to increase
404 isobutanol production by integrating more copies of the isobutanol synthetic pathway
405 into the chromosome.
406

407 **Conclusions**

408 Overall, this study proposed an efficient chromosomal engineering tool for the
409 insertion and deletion of long DNA fragments in *E. coli*, and demonstrated the tool's

potential in synthetic biology by successfully applying it in genome simplification and metabolic engineering.

Methods

Strains and culture conditions

E. coli strain DH5α (American Type Culture Collection – ATCC® 68233™) served as the host strain for molecular cloning and plasmid manipulation. MG1655 (ATCC® 47076™) served as the genetic material in editing experiments unless otherwise stated. Strains involved in this study are listed in Table S1. Verification primers used in genomic editing experiments are listed in Table S2. LB medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) was used for cell growth in all cases unless otherwise noted. Solid medium contained 20 g/L agar. Super optimal broth with catabolite repression (SOC) medium (20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) was used for cell recovery. M9 medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10 mg/L VB_1, 40 g/L glucose, and 4 g/L yeast extract) was used for shake-flask fermentation. The working concentrations of Amp and Kan were 0.1 g/L and 0.025 g/L, respectively. The working concentrations of IPTG, X-gal, glucose, and sucrose in media or cultures were 1 mM, 0.1 g/L, 10 g/L, and 20 g/L, respectively. The working concentration of L-arabinose was 20 mM in liquid media and 5 mM in solid media. Details of reagents and media used in this study are listed in Table S3.

Plasmid construction

Plasmids involved in this study are listed in Table S4. Complete sequences of plasmids p15A-P_{araB}-Cas9-P_{T5}-Redγβα, pSC101-P_{araB}-sgRNA-Donor-T1, pSC101-P_{araB}-sgRNA-Donor-T2, and pSC101-P_{araB}-sgRNA-Donor-T3 are presented in Notes S1–S4. CRISPR target sequences designed in this study are listed in Table S5. The construction of plasmid pSC101-P_{araB}-sgRNA-Donor was the key step in a specific genomic editing experiment. When constructing the pSC101-P_{araB}-sgRNA-Donor plasmid containing one sgRNA expression chimera, pSC101-P_{araB}-sgRNA-Donor-T1 served as the parental plasmid. First, a specifically designed donor DNA was integrated into pSC101-P_{araB}-sgRNA-Donor-T1 to construct an intermediate plasmid. The donor DNA contained two homologous arms of approximately 500 bp. Then, a specific spacer (20 bp) was inserted into the intermediate plasmid between the *araB* promoter and the gRNA scaffold via single PCR and single Gibson Assembly. The spacer introduced by PCR served as the overlap in Gibson Assembly. When constructing the pSC101-P_{araB}-sgRNA-Donor plasmid containing two sgRNA expression chimeras, pSC101-P_{araB}-sgRNA-Donor-T2 and pSC101-P_{araB}-sgRNA-Donor-T3 served as the parental plasmids. First, a specifically designed donor DNA was integrated into pSC101-P_{araB}-sgRNA-Donor-T2 to construct an intermediate plasmid. Then, the intermediate plasmid and pSC101-P_{araB}-sgRNA-Donor-T3 were combined to construct the pSC101-P_{araB}-sgRNA-Donor plasmid through PCR and Gibson Assembly. The two specific spacers introduced by PCR served as overlaps in Gibson Assembly. Detailed

454 construction procedures of the pSC101-P_{araB}-sgRNA-Donor plasmid are illustrated in
455 Figure S7.

456 **Procedures of genomic editing, plasmids curing, and iterative editing**

457 First, the Kan^R plasmid p15A-P_{araB}-Cas9-P_{T5}-Redγβα (plasmid#1) was transfected
458 into the target strain such as MG1655 to obtain the corresponding transformants such
459 as MG1655/plasmid#1. A series of temperature-sensitive Amp^R plasmids were
460 constructed to express specific sgRNA and generate specific donor DNA, and these
461 plasmids were generally named pSC101-P_{araB}-sgRNA-Donor (plasmid#2). Then,
462 specific plasmid#2 was transfected into the MG1655/plasmid#1 strain, and the
463 MG1655/plasmid#1/plasmid#2 strain was screened in a LB plate with Amp, Kan, and
464 glucose at 30 °C. One or several single colonies were inoculated into 2 mL LB
465 medium, and the culture was cultivated at 30 °C for two hours. Then, 2 μL Amp, 2 μL
466 Kan, and 20 μL IPTG were added to the culture. After one hour, 20 μL L-arabinose
467 was added, and the cultures were cultivated for another three hours before plating. A
468 1-μL or 0.1-μL aliquot of the culture was plated onto a LB plate containing Amp, Kan,
469 and L-arabinose, and the plate was incubated overnight at 30 °C. Positive mutants
470 were verified by colony PCR and sequencing. The flowchart of genomic editing is
471 shown in Fig. 1 and Figure S1. The positive mutant was cultivated in LB medium in
472 the presence of only Kan at 40 °C for 12 hours to remove the temperature-sensitive
473 Amp^R plasmid#2 (Figure S3a). Then, the obtained edited strain containing only
474 plasmid#1 was used as the starting strain for the next round of genomic editing. The
475 Kan^R plasmid#1 is not stable in the host strain in the absence of Kan. When the final

round of genomic editing was completed, the edited strain was cultivated in LB medium without Kan at 40 °C for 12 hours to remove both Amp^R plasmid#2 and sucrose-sensitive Kan^R plasmid#1 (Figure S3b). The overnight culture was diluted for plating on a LB plate containing sucrose. Theoretically, colonies grown on the plate are plasmid-free. For further verification, single colonies were inoculated into LB medium with or without corresponding antibiotics. The flowchart of plasmid curing and iterative editing is shown in Figure S1.

Calculation of positive rate and editing efficiency

One hundred colonies in the LB plate containing Amp, Kan, and L-arabinose were tested by colony PCR to screen for positive mutants. Twenty of the positive mutants were further verified via sequencing. The positive rate was calculated as the proportion of positive colonies to the total number of colonies. In blue-white selection experiments, positive colonies were also recognized by their color. White colonies were positive, and blue colonies were negative. One control group was set along with the experimental group to calculate editing efficiency. In the control group, L-arabinose was not added, and thus no Cas9 protein or sgRNA were expressed. All other conditions and processes were the same as for the experimental group. The editing efficiency was calculated as the proportion of positive colonies in the experimental group to the total number of colonies in the control group.

Measurement of growth curve and transformation efficiency

For measuring the growth curve, one single colony was inoculated into 5 mL LB medium, and the culture was cultivated at 37 °C for 12 hours. Then, 1 mL seed liquid

was inoculated into 100 mL fresh LB medium, and the culture was cultivated at 37 °C in a 220-rpm shaker. During the 12-hour cultivation, samples were taken every hour to measure the optical density at a wavelength of 600 nm (OD600) of the culture using an ultraviolet spectrophotometer (V-5100, Shanghai Metash Instruments Co., Ltd) at 600 nm. For measuring transformation efficiency, pure pUC19 was used as supercoiled DNA. First, 1 μL pUC19 (1 ng/μL) was added to one tube of competent cells (100 μL). Next, the mixture was incubated for 30 minutes before conducting heat-shock for one minute in a 42 °C water bath. Then, the tube was placed on ice for two minutes before adding 900 μL 37 °C SOC medium, and the tube was shaken at 200–230 rpm (37 °C) for 40 minutes. Finally, 100 μL of the cultures were plated on a LB plate containing Amp, and the plate was incubated overnight at 37 °C. The transformation efficiency is $N \times 10^4$ CFU/μg pUC19 ("*N*" refers to the number of transformants obtained in the plate).

Shake-flask fermentation and product detection

For testing isobutanol production, single colonies of engineered strains were inoculated into 5 mL LB media containing the appropriate antibiotics, and the cultures were cultivated at 37 °C for 12 hours. Then, 200-μL seed liquid was transferred to airtight shake flasks containing 20 mL antibiotic-free M9 medium for micro-aerobic fermentation. During the 72-hour fermentation, samples were taken every 12 hours to test the biomass and the titer of isobutanol. Biomass was evaluated by measuring the OD600 of fermentation broth with an ultraviolet spectrophotometer (V-5100, Shanghai Metash Instruments Co., Ltd). For measuring isobutanol concentration, the

fermentation broth was centrifuged at 1400 × *g* for 10 minutes. The supernatant was tested via a gas chromatograph (PANNA GCA91, Shanghai Wangxu Electric Co., Ltd), with high-purity isobutanol as the standard and high-purity n-pentanol as an internal reference.

List of abbreviations

PCR: polymerase chain reaction; DSB: double-strand break; dsDNA: double-stranded DNA; ZFN: zinc-finger nuclease; TALEN: transcription activator-like effector nuclease; CRISPR: clustered regularly interspaced short palindromic repeats; Cas9: CRISPR-associated protein 9; sgRNA: single-guide RNA; PAM: protospacer-adjacent motif; CARS: CRISPR/Cas9-assisted recombination system; IPTG: isopropyl-β-D-thiogalactopyranoside; Kan^R: kanamycin-resistant; Amp^R: ampicillin-resistant; LB: Luria-Bertani; ORF: open reading frame; UTR: untranslated region; NHEJ: non-homologous end-joining; ATCC: American Type Culture Collection; SOC: super optimal broth with catabolite repression; OD600: optical density at a wavelength of 600 nm; TS: target site; LHA: left homologous arm; RHA: right homologous arm; F: forward primer; R: reverse primer.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declared no conflicts of interests.

Funding

This work was jointly supported by the National Key R&D Program of China (No. 2019YFA0904100) and the National Natural Science Foundation of China (No. 31961133014)

Authors' contributions

Conceptualization: [Yi-Xin Huo]; Methodology: [Chaoyong Huang], [Liwei Guo]; Formal analysis and investigation: [Chaoyong Huang], [Liwei Guo], [Jingge Wang], [Ning Wang]; Writing - original draft preparation: [Chaoyong Huang], [Liwei Guo]; Writing - review and editing: [Yi-Xin Huo]; Funding acquisition: [Yi-Xin Huo]; Resources: [Yi-Xin Huo]; Supervision: [Yi-Xin Huo]. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable

Additional materials

File name: Additional file 1

File format: .pdf

Title of data: Supplementary Figure 1–7, Supplementary Table 1–5, and

Supplementary Note 1–4.

Data description: Strains involved in this study are listed in Supplementary Table 1. Verification primers used in genomic editing experiments are listed in Supplementary Table 2. Details of reagents and media used in this study are listed in Supplementary Table 3. Plasmids involved in this study are listed in Supplementary Table 4. CRISPR target sequences designed in this study are listed in Supplementary Table 5. Complete sequences of plasmids p15A-ParaB-Cas9-PT5-Redγβα, pSC101-ParaB-sgRNA-Donor-T1, pSC101-ParaB-sgRNA-Donor-T2, and pSC101-ParaB-sgRNA-Donor-T3 are presented in Supplementary Notes 1–4. Other related data mentioned in the results section are presented in the Supplementary Figures.

References

1. Kato J, Hashimoto M. Construction of consecutive deletions of the *Escherichia coli* chromosome. Mol Syst Biol. 2007;3:132.

2. Kato J, Hashimoto M. Construction of long chromosomal deletion mutants of *Escherichia coli* and minimization of the genome. Methods Mol Biol. 2008;416:279–293.

3. Shipman SL, Nivala J, Macklis JD, Chruch GM. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. Nature. 2017;547:345–349.

4. Mignon C, Sodoyer R, Werle B. Antibiotic-free selection in biotherapeutics: now and forever. Pathogens. 2015;4:157–181.

5. Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. Nucleic Acids Res. 1993;21:3329–3330.

6. Wilson RB, Davis D, Mitchell AP. Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. J Bacteriol. 1999;181:1868–1874.

7. Russell CB, Thaler DS, Dahlquist FW. Chromosomal transformation of *Escherichia coli* recD strains with linearized plasmids J Bacteriol. 1989;171:2609–2613.

8. Link AJ, Phillips D, Church GM. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. J Bacteriol. 1997;179:6228–6237.

9. Pósfai G, Kolisnychenko V, Bereczki Z, Blattner FR. Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. Nucleic Acids Res. 1999;27:4409–4415.

10. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc. 2009;4:206–223.

11. Jeong J, Cho N, Jung D, Bang D. Genome-scale genetic engineering in *Escherichia coli*. Biotechnol Adv. 2013;31:804–810.

12. Pines G, Freed EF, Winkler JD, Gill RT. Bacterial recombineering: genome engineering via phage-based homologous recombination. ACS Synth Biol. 2015;4:1176–1185.

13. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in

Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA. 2000;97:6640–6645.

14. Wang H, Bian X, Xia L, Ding X, Müller R, Zhang Y, Fu J, Stewart AF. Improved seamless mutagenesis by recombineering using *ccdB* for counterselection. Nucleic Acids Res. 2014;42:e37.

15. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009;460:894–898.

16. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol. 2010;28:856–862.

17. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science. 2011;333:348–353.

18. Tischer BK, von Einem J, Kaufer B, Osterrieder N. Two-step Red-mediated recombination for versatile high-efficiency markerless DNA manipulation in *Escherichia coli*. Biotechniques. 2006;40:191–197.

19. Yu BJ, Kang KH, Lee JH, Sung BH, Kim MS, Kim SC. Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. Nucleic Acids Res. 2008;36:e84.

20. Yang J, Sun B, Huang H, Jiang Y, Diao L, Chen B, Xu C, Wang X, Liu J, Jiang W, et al. High-efficiency scarless genetic modification in *Escherichia coli* by using lambda red

recombination and I-SceI cleavage. Appl Environ Microbiol. 2014;80:3826–3834.

21. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 2010;11:636–646.

22. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol. 2013;14:49–55.

23. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. Science. 2011;333:1843–1846.

24. Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, et al. Targeted genome editing across species using ZFNs and TALENs. Science. 2011;333:307.

25. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–821.

26. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA. 2012;109:e2579–2586.

27. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol. 2013;31:233–239.

28. Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang Y, Chen T, Zhao H. Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 meditated genome editing. Metab Eng. 2015;31:13–21.

29. Chayot R, Montagne B, Mazel D, Ricchetti M. An end-joining repair mechanism in

Escherichia coli. Proc Natl Acad Sci USA. 2010;107:2141–2146.

30. Huang C, Ding T, Wang J, Wang X, Guo L, Wang J, Zhu L, Bi C, Zhang X, Ma X, et al. CRISPR-Cas9-assisted native end-joining editing offers a simple strategy for efficient genetic engineering in *Escherichia coli*. Appl Microbiol Biotechnol. 2019;103:8497–8509.

31. Chung ME, Yeh IH, Sung LY, Wu MY, Chao YP, Ng IS, Hu YC. Enhanced integration of large DNA into *E. coli* chromosome by CRISPR/Cas9. Biotechnol Bioeng. 2017;114:172–183.

32. Li Y, Yan F, Wu H, Li G, Han Y, Ma Q, Fan X, Zhang C, Xu Q, Xie X, et al. Multiple-step chromosomal integration of divided segments from a large DNA fragment via CRISPR/Cas9 in *Escherichia coli*. J Ind Microbiol Biotechnol. 2019;46:81–90.

33. Su T, Liu F, Gu P, Jin H, Chang Y, Wang Q, Liang Q, Qi Q. A CRISPR-Cas9 assisted non-homologous end-joining strategy for one-step engineering of bacterial genome. Sci Rep. 2016;6:37895.

34. Zhao D, Feng X, Zhu X, Wu T, Zhang X, Bi C. CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. Sci Rep. 2017;7:16624.

35. Zheng X, Li SY, Zhao GP, Wang J. An efficient system for deletion of large DNA fragments in *Escherichia coli* via introduction of both Cas9 and the non-homologous end joining system from *Mycobacterium smegmatis*. Biochem Biophys Res. 2017;485:768–774.

674 36. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S,
675 Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, et al.
676 EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res.
677 2013;41:D605–D612.
678 37. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi
679 R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, et al. The EcoCyc
680 database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res.
681 2017;45:D513–D550.
682 38. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y,
683 Furubayashi A, Kinjyo S, Dose H, Hasegawa M, et al. Update on the Keio collection of
684 *Escherichia coli* single-gene deletion mutants. Mol Syst Biol. 2009;5:335.
685 39. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M,
686 Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene
687 knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:8.
688 40. Saini M, Wang ZW, Chiang CJ, Chao YP. Metabolic engineering of *Escherichia coli* for
689 production of n-butanol from crude glycerol. Biotechnol Biofuels. 2017;10:173.
690 41. Liang S, Chen H, Liu J, Wen J. Rational design of a synthetic Entner-Doudoroff
691 pathway for enhancing glucose transformation to isobutanol in *Escherichia coli*. J Ind
692 Microbiol Biotechnol. 2018;45:187–199.
693 42. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators.
694 Nature. 2000;403:335–338.
695 43. Rong M, He B, McAllister WT, Durbin RK. Promoter specificity determinants of T7

RNA polymerase. Proc Natl Acad Sci USA. 1998;95:515–519.

44. Atsumi S, Li Z, Liao JC. Acetolactate synthase from *Bacillus subtilis* serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in *Escherichia coli*. Appl Environ Microbiol. 2009;75:6306–6311.

45. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature. 2008;451:86–89.

46. Bujard H, Gentz R, Lanzer M, Stueber D, Mueller M, Ibrahimi I, Haeuptle MT, Dobberstein B. A T5 promoter-based transcription-translation system for the analysis of proteins in vitro and in vivo. Methods Enzymol. 1987;155:416–433.

47. Chen CT, Liao JC. Frontiers in microbial 1-butanol and isobutanol production. FEMS Microbiol Lett. 2016;363:fnw020.

48. Lan EI, Liao JC. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. Bioresour Technol. 2013;135:339–349.

49. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997;277:1453–1462.

50. Tao H, Bausch C, Richmond C, Blattner FR, Conway T. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J Bacteriol. 1999;181:6425–6440.

51. Koob MD, Shaw AJ, Cameron DC. Minimizing the Genome of *Escherichia coli*. Ann N Y Acad Sci. 1994;745:1–3.

52. Yu BJ, Sung BH, Koob MD, Lee CH, Lee JH, Lee WS, Kim MS, Kim SC. Minimization

of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat Biotechnol. 2002;20:1018–1023.

53. Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, et al. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. Mol Microbiol. 2005;55:137–149.

54. Kang Y, Durfee T, Glasner JD, Qiu Y, Frisch D, Winterberg KM, Blattner FR. Systematic mutagenesis of the *Escherichia coli* genome. J Bacteriol. 2004;186:4921–4930.

55. Paddon CJ, Keasling JD. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol. 2014;12:355–367.

56. Fang H, Li D, Kang J, Jiang P, Sun J, Zhang D. Metabolic engineering of *Escherichia coli* for de novo biosynthesis of vitamin B_{12}. Nat Commun. 2018;9:4917.

57. Yu Y, Zhu X, Xu H, Zhang X. Construction of an energy-conserving glycerol utilization pathways for improving anaerobic succinate production in *Escherichia coli*. Metab Eng. 2019;56:181–189.

58. Huo YX, Ren H, Yu H, Zhao L, Yu S, Yan Y, Chen Z. CipA-mediating enzyme self-assembly to enhance the biosynthesis of pyrogallol in *Escherichia coli*. Appl Microbiol Biotechnol. 2018;102:10005–10015.

59. Huo YX, Cho KM, Rivera JG, Monte E, Shen CR, Yan Y, Liao JC. Conversion of proteins into biofuels by engineering nitrogen flux. Nat Biotechnol. 2011;29:346–351.

740

741 **Table 1. Long fragments deleted in the MG1655 genome**

Fragment No.	Starting site	End site	Length	Essential gene
No. 1	240,056	426,771	186,715	*yagG*, *hemB*
No. 2	499,529	552,955	53,426	None
No. 3	565,456	665,088	99,932	*entD*
No. 4	990,473	1,127,061	136,588	*fabA*, *serT*
No. 5	1,449,596	1,549,490	99,894	None
No. 6	1,549,491	1,647,484	97,993	None
No. 7	2,349,152	2,430,141	80,989	None
No. 8	2,442,420	2,517,306	74,886	*argW*
No. 9	2,822,534	2,906,555	84,021	*ispF*, *ispD*, *ftsB*
No. 10	3,610,719	3,689,415	78,696	None
No. 11	3,824,765	3,876,879	52,114	*selC*
No. 12	4,198,958	4,251,002	52,044	None

742

743 **Figure legends**

744 **Fig.1** Constitution of CARS and schematic of genomic editing. LHA: left
745 homologous arm. RHA: right homologous arm.

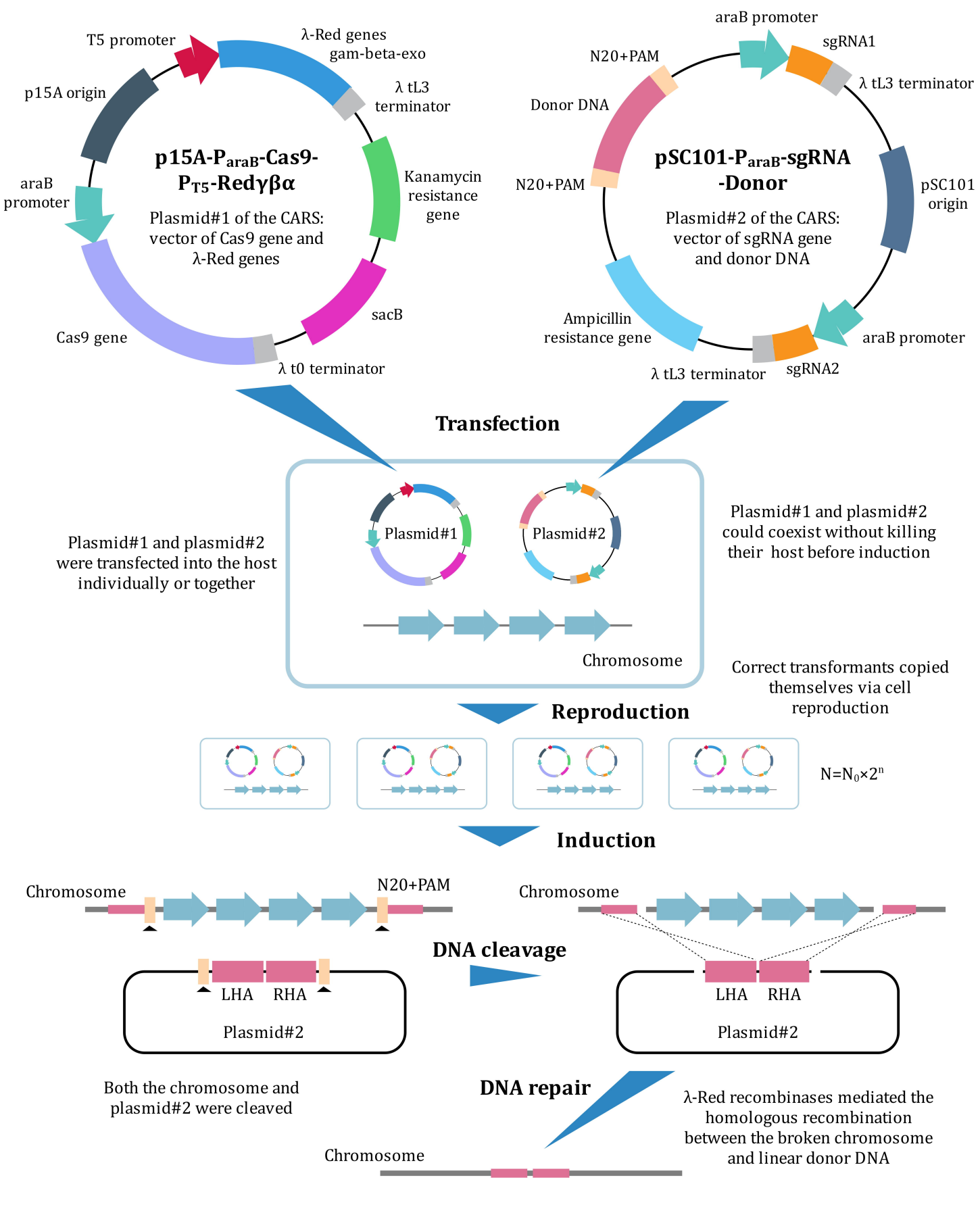
746 **Fig.2** CARS mediated long fragment integration. (**a**) Schematic of fragment
747 integration of different length. TS: target site. LHA: left homologous arm. RHA: right
748 homologous arm. F: forward primer. R: reverse primer. (**b**) Editing efficiencies and
749 positive rates in four editing experiments. (**c**) Comparison of largest insertion length
750 and positive rate between three reported methods and our method. Data are expressed
751 as means ± s.d. from three independent experiments.

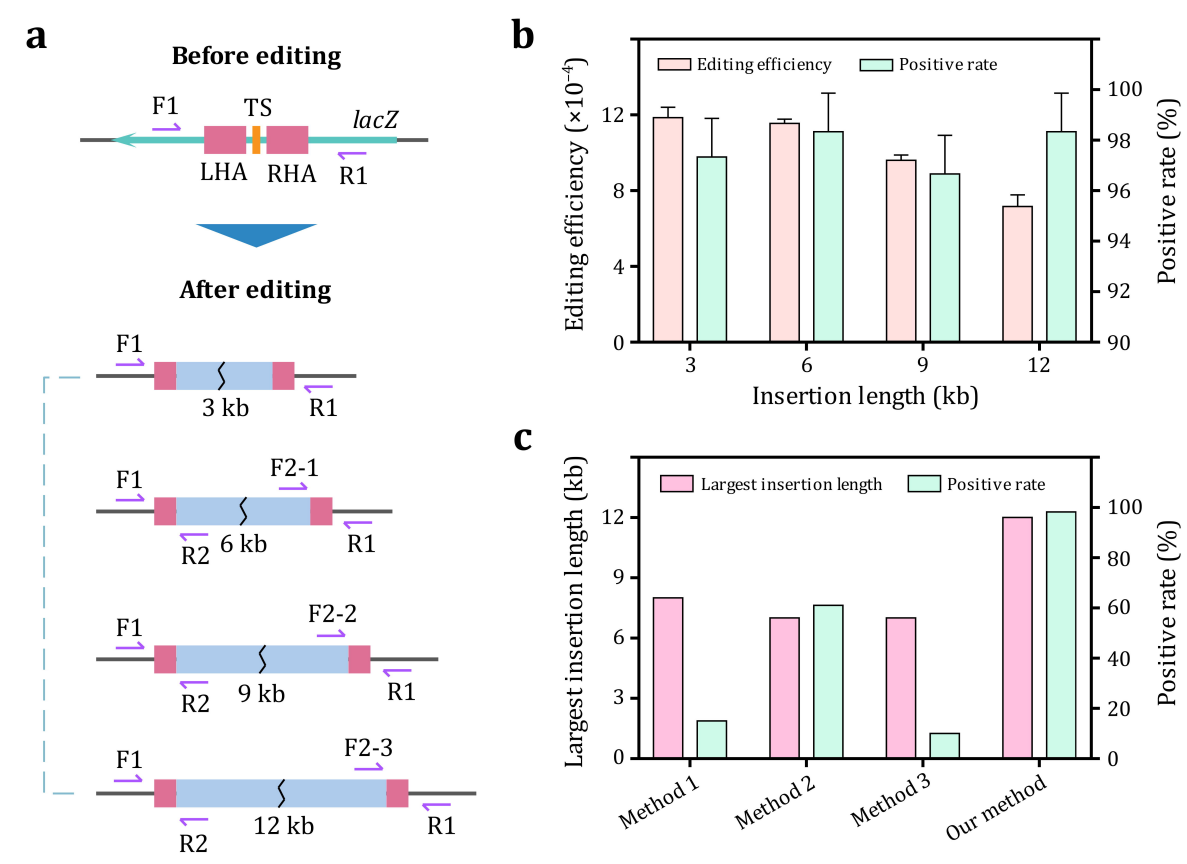
752 **Fig.3** CARS mediated long fragment knockout. (**a**) Schematic of fragment deletion of
753 different length. TS: target site. (**b**) Editing efficiencies and positive rates in seven
754 editing experiments. (**c**) Comparison of largest deletion length and positive rate

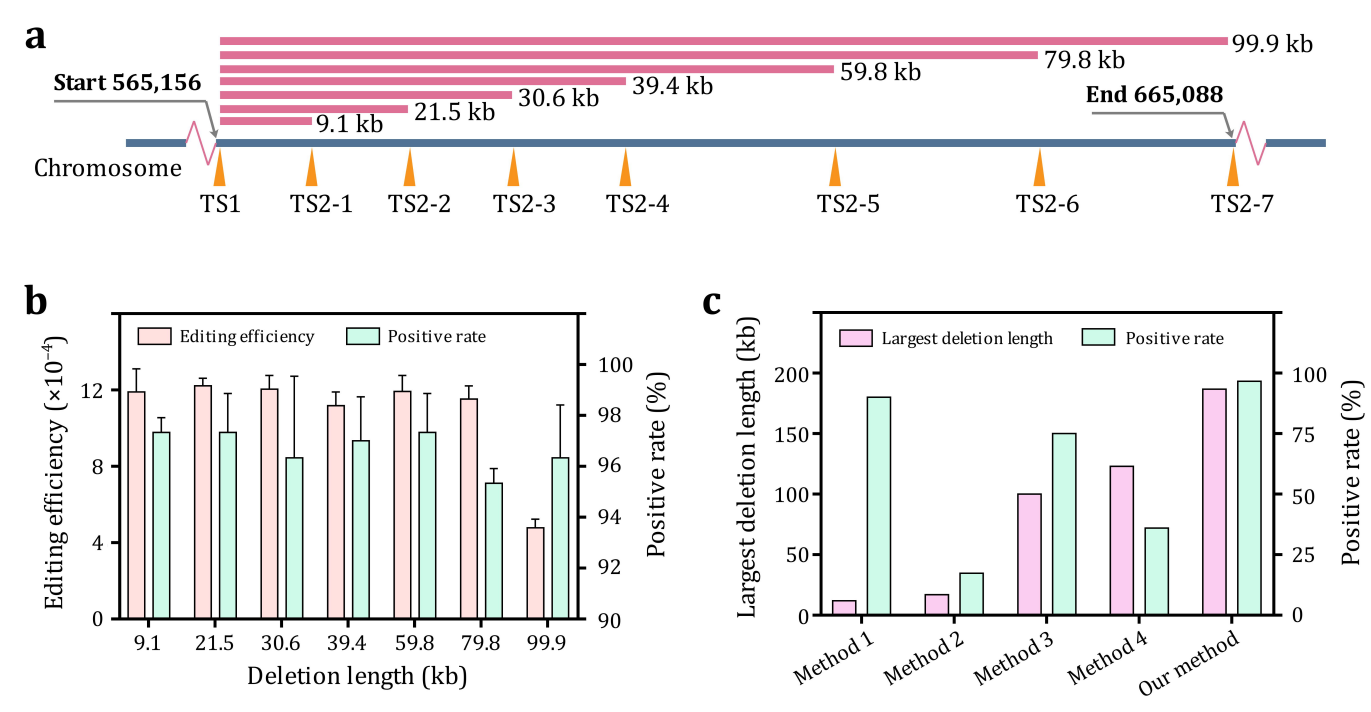
between four reported methods and our method. Data are expressed as means ± s.d. from three independent experiments.

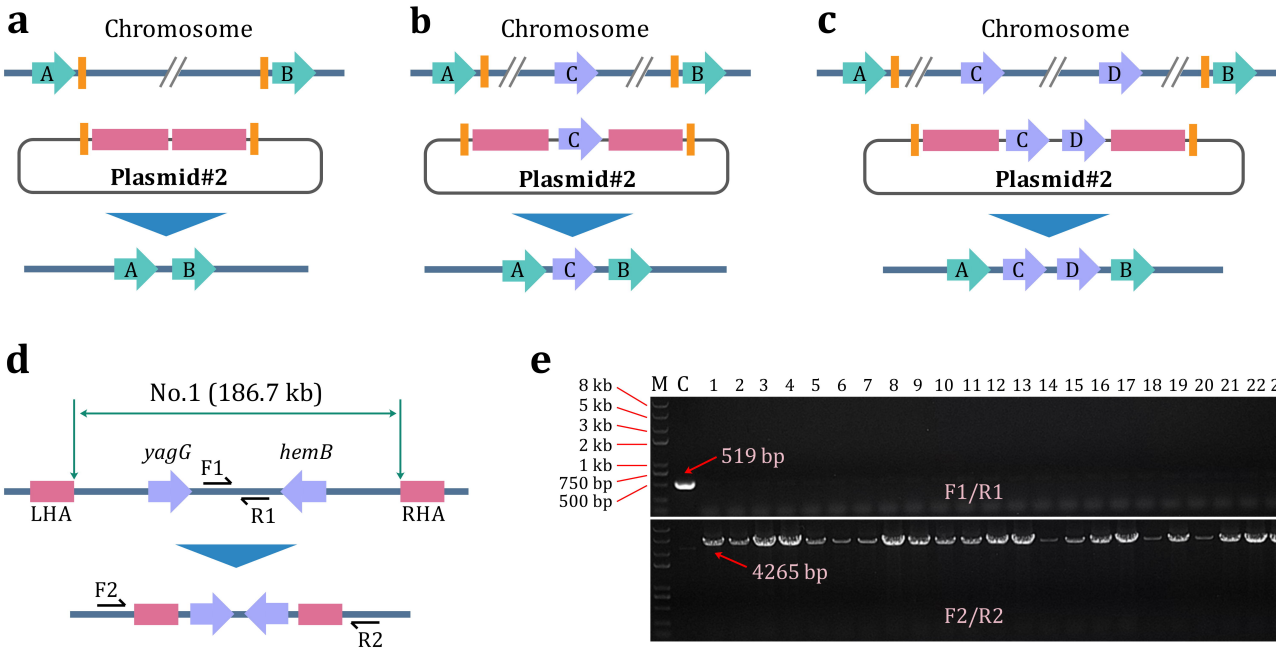
Fig.4 Deletion of nonessential sequence and chromosomal simplification. (**a**) Deletion of long fragment containing no essential gene. (**b**) Deletion of long fragment containing one essential gene. (**c**) Deletion of long fragment containing two essential genes. (**d**) Schematic of the deletion of fragment No.1. LHA: left homologous arm. RHA: right homologous arm. F: forward primer. R: reverse primer. (**e**) Representative results of PCR verification in the deletion experiment of fragment No.1. (**f**) Results in the deletion experiments of twelve nonessential fragments. (**g**) Summary of cumulative deletion. Data are expressed as means ± s.d. from three independent experiments.

Fig.5 Metabolic engineering of *E. coli* for producing isobutanol. (**a**) Construction of strain SH274 in the basis of strain JW74. (**b**) The synthetic pathway of isobutanol from glucose. (**c**) Results of isobutanol fermentation of strain SH274. (**d**) Results of isobutanol fermentation of strain SH279. Data are expressed as means ± s.d. from three independent experiments.


T5 promoter
λ-Red genes
gam-beta-exo
λ tL3
terminator
p15A origin
Kanamycin
resistance
gene
araB
promoter
p15A-ParaB-Cas9-
PT5-Redγβα
Plasmid#1 of the CARS:
vector of Cas9 gene and
λ-Red genes
sacB
Cas9 gene
λ t0 terminator
araB promoter
sgRNA1
λ tL3 terminator
N20+PAM
Donor DNA
pSC101
origin
pSC101-ParaB-sgRNA
-Donor
Plasmid#2 of the CARS:
vector of sgRNA gene
and donor DNA
N20+PAM
Ampicillin
resistance gene
araB promoter
λ tL3 terminator
sgRNA2
Transfection
Plasmid#1
Plasmid#2
Chromosome
Plasmid#1 and plasmid#2
were transfected into the host
individually or together
Plasmid#1 and plasmid#2
could coexist without killing
their host before induction
Reproduction
Correct transformants copied
themselves via cell
reproduction
N=N0×2n
Induction
Chromosome
N20+PAM
LHA
RHA
Plasmid#2
DNA cleavage
Chromosome
LHA
RHA
Plasmid#2
Both the chromosome and
plasmid#2 were cleaved
DNA repair
λ-Red recombinases mediated the
homologous recombination
between the broken chromosome
and linear donor DNA
Chromosome


a
Before editing
F1
TS
lacZ
LHA
RHA
R1
After editing
F1
3 kb
R1
F1
F2-1
R2
6 kb
R1
F1
F2-2
R2
9 kb
R1
F1
F2-3
R2
12 kb
R1
b
Editing efficiency
Positive rate
Editing efficiency (×10⁻⁴)
Positive rate (%)
Insertion length (kb)
c
Largest insertion length
Positive rate
Largest insertion length (kb)
Positive rate (%)
Method 1
Method 2
Method 3
Our method


a
Start 565,156
End 665,088
99.9 kb
79.8 kb
59.8 kb
39.4 kb
30.6 kb
21.5 kb
9.1 kb
Chromosome
TS1
TS2-1
TS2-2
TS2-3
TS2-4
TS2-5
TS2-6
TS2-7
b
Editing efficiency
Positive rate
Editing efficiency (×10⁻⁴)
Positive rate (%)
Deletion length (kb)
9.1
21.5
30.6
39.4
59.8
79.8
99.9
c
Largest deletion length
Positive rate
Largest deletion length (kb)
Positive rate (%)
Method 1
Method 2
Method 3
Method 4
Our method

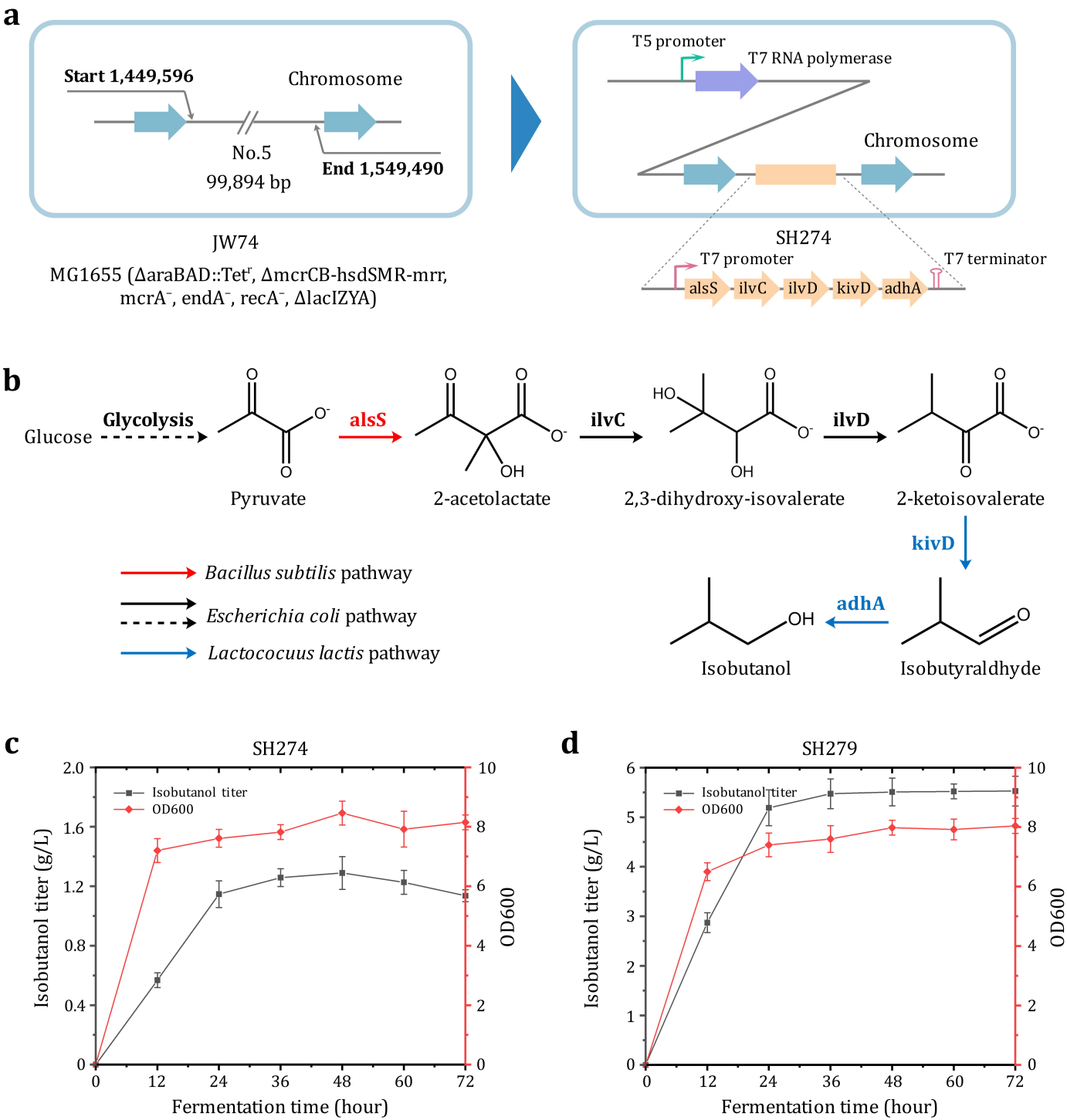


f

Fragment No.	Length	Editing efficiency	Positive rate (%)
No.1	186.7 kb	1.20 (± 0.09) × 10^{-3}	97.7 (± 2.6)
No.2	53.4 kb	1.29 (± 0.02) × 10^{-3}	96.3 (± 1.2)
No.3	99.9 kb	4.79 (± 0.36) × 10^{-4}	96.3 (± 1.7)
No.4	136.6 kb	2.31 (± 0.28) × 10^{-4}	95.0 (± 0.8)
No.5	99.9 kb	1.28 (± 0.02) × 10^{-3}	95.0 (± 0.8)
No.6	98.0 kb	1.19 (± 0.07) × 10^{-3}	96.0 (± 1.4)
No.7	81.0 kb	4.44 (± 0.27) × 10^{-4}	98.0 (± 0.8)
No.8	74.9 kb	1.22 (± 0.05) × 10^{-3}	95.0 (± 0.8)
No.9	84.0 kb	1.14 (± 0.03) × 10^{-3}	97.7 (± 2.6)
No.10	78.7 kb	1.16 (± 0.06) × 10^{-3}	95.7 (± 1.2)
No.11	52.1 kb	1.09 (± 0.02) × 10^{-3}	96.7 (± 1.7)
No.12	52.0 kb	1.22 (± 0.04) × 10^{-3}	97.0 (± 2.2)

g

Cumulative deletion	Deleted length	ORF number
ΔNo.1 + ΔNo.9	270.7 kb	268
ΔNo.1 + ΔNo.9 + ΔNo.2	324.1 kb	315
ΔNo.1 + ΔNo.9 + ΔNo.5	370.6 kb	364
ΔNo.1 + ΔNo.9 + ΔNo.6	368.7 kb	368


a
Start 1,449,596
Chromosome
No.5
99,894 bp
End 1,549,490
JW74
MG1655 (ΔaraBAD::Tet^r, ΔmcrCB-hsdSMR-mrr, mcrA$^-$, endA$^-$, recA$^-$, ΔlacIZYA)
T5 promoter
T7 RNA polymerase
Chromosome
SH274
T7 promoter
T7 terminator
alsS
ilvC
ilvD
kivD
adhA
b
Glycolysis
Glucose
alsS
ilvC
ilvD
Pyruvate
2-acetolactate
2,3-dihydroxy-isovalerate
2-ketoisovalerate
kivD
adhA
Isobutanol
Isobutyraldhyde
Bacillus subtilis pathway
Escherichia coli pathway
Lactococuus lactis pathway
c
SH274
Isobutanol titer
OD600
Isobutanol titer (g/L)
OD600
Fermentation time (hour)
d
SH279
Isobutanol titer
OD600
Isobutanol titer (g/L)
OD600
Fermentation time (hour)