Exploiting Type I-B CRISPR Genome Editing System in *Thermoanaerobacterium Aotearoense* SCUT27 And Engineering The Strain For High-Level Ethanol Production

Kaiqun Dai
  South China University of Technology

Hongxin Fu
  South China University of Technology

Xiaolong Guo
  South China University of Technology

Chunyun Qu
  South China University of Technology

Jufang Wang (✉ jufwang@scut.edu.cn)
  South China University of Technology   https://orcid.org/0000-0002-9908-1001

Research

Abstract

Background: Thermophilic microbes for biofuels and chemicals have attracted great attention due to their tolerance of high temperature and wide range of substrate utilization. *Thermoanaerobacterium aotearoense* SCUT27 has the ability of glucose and xylose co-utilization in lignocellulosic biomass. Polygene manipulation was a bottleneck since it was hindered by available markers for selection. In this study, the endogenous Type I-B CRISPR/Cas system was developed for multiplex genome editing in SCUT27.

Results: The protospacer-adjacent motif (PAM) was identified by *in silico* and orotidine-5'-phosphate decarboxylase (*pyrF*) and then lactate dehydrogenase (*ldh*) were chosen as the editing target to assess the toxicity of this immune system and gene editing efficiency. The mutants could be repeatedly obtained with an editing efficiency of 58.3-100%. Higher transformation efficiency was observed after optimization of some editing strategies. Furthermore, a new method was performed for screening mutants of plasmid curing (recycling of the editing plasmid) for multiplex genome editing based on the negative selection marker *tdk*, and then *ldh* and arginine repressor (*argR*) were knocked out successively. The mutant SCUT27/Δ*ldh*/Δ*argR* had the prominent advantages over SCUT27 for ethanol production with enhanced ability to metabolize xylose. When cultured under various lignocellulosic hydrolysates, the mutant showed a satisfactory performance with the ethanol titer and yield improved by 147.42−739.40% and 112.67−267.89%, respectively, compared with SCUT27, as well as the enhanced tolerance to inhibitors.

Conclusion: The multi-gene editing by native CRISPR/Cas system is a promising strategy to engineer SCUT27 for higher ethanol production with lignocellulosic hydrolysates.

Background

Lignocellulosic biomass (LCB) is globally available and sustainable feedstocks containing sugar-rich platform, which can be converted to biofuels and specialty products through appropriate processing [1]. *Thermoanaerobacterium aotearoense* SCUT27 (termed SCUT27), a Gram-positive anaerobe, is a natural ethanol producer and a promising cell factory for the conversion of lignocellulosic hydrolysates to biofuels and bioproducts through high-temperature fermentation [2] (Fig. 1). Unlike *Yeasts* and *Pseudomonas*, SCUT27 can efficiently utilize xylose and also co-utilize it with glycose. When glucose or xylose as the substrate, the mutant with deletion of lactate dehydrogenase (*ldh*) exhibited increase of 31.0% and 31.4% in cell yield, and 2.37- and 2.1-fold in the yield of ethanol, especially without lactate. It might be due to increased acetate and ATP levels, and redirection of more NADH flowing to ethanol production [3]. Arginine repressors (ArgRs) were involved in several key metabolic pathways and might be the global regulator in SCUT27 [4]. In previous study, SCUT27/Δ*argR* exhibited improved co-utilization of glucose and xylose, and increased ethanol yield by 82.95%, due to the enhanced xylose utilization and the higher energy level [4, 5]. Based on the results above, multiplex genome editing targeting these key genes in SCUT27 has much greater potential for higher ethanol production.

However, there were several problems hindering the metabolic engineering in SCUT27, such as the limited selective markers and absence of efficient genetic engineering tools which can facilitate engineering the strain continuously, and make a traceless editing without markers (avoid the interference from them). Thus, there is an urgent demand to develop other technologies for strain improvement.

Clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated (Cas) system is a prokaryotic RNA-guided adaptive immune system in bacteria and archaea against invading foreign DNAs such as phages and plasmids [6, 7]. It has been reported that approximately 47% of sequenced bacteria and 87% of sequenced archaea harbor CRISPR/Cas loci, all of which share similar features consisting of identical direct repeats (DR) separated by variable sequences called spacers along with a suite of associated Cas genes. The CRISPR system can be classified into two classes and six types (33 subtypes) based on the Cas protein and the architecture of CRISPR/Cas loci [8]. CRISPR/Cas9-based editing system has been well understood and applied maturely with high efficiency in a board range of organisms [9–13]. However, the utilization in thermophiles is quite rare until now. Although some exogenous thermo-tolerant Cas9 had been found, their severe toxicity to cells led to low efficiency of transformation and editing [9, 10]. Besides, the Type II system with Cas9 has a more complex PAM sequence and shorter single guide RNA (sgRNA) that may limit the selection of editing locations or off-target [10, 14]. To make an editing easily, many researchers utilized the endogenous CRISPR/Cas system for genome editing. Pyne et al. utilized the native Type I-B CRISPR/Cas system in *Clostridium pasteurianum* as an efficient genome editing tool, and successfully deleted the *cpaAIR* gene [15]. Zhang et al. exploited the CRISPR/Cas system in *Clostridium tyrobutyricum* to make multiplex genome editing with deletion of *spo0A* and *pyrF* by a single transformation, and also integrated the alcohol dehydrogenase gene to replace *cat1* for n-butanol production [16]. Zhou et al. successfully deleted *spo0A* and *aldh* by the customized endogenous CRISPR/Cas system in *C. butyricum*, yielding an efficiency high up to 100% [17]. Liu et al. found the endogenous subtype II-A CRISPR/Cas system and used it to deplete native plasmids for enhancing cell growth and integrate of a L-lactate dehydrogenase gene into the chromosome, then led to both enhanced cell growth and lactic acid production [18].

In this study, we firstly reported the native type I-B CRISPR/Cas system in *T. aotearoense* strains in which an in-depth analysis for elements of this immune system was lacking. *In silico* CRISPR arrays analysis and *in vivo* toxicity assay were performed and revealed that the motif 'TTA' at 5'-end of the predicted protospacers was a functional PAM site for targeting of this endogenous system. By using a constitutive strong promoter P*kan* for CRISPR mini-array expression, we successfully deleted *ldh* with a high editing efficiency which could be up to 100%. Some attempts were also made to improve the transformation efficiency in this process. Besides, for improving ethanol production, multi-gene editing system has been developed here based on a novel way for plasmid recycling. With the deletion of both *ldh* and *argR*, the mutant showed a better performance on ethanol production with a yield of 0.35-0.40 g /g, especially feeding with the xylose. Then, we assessed the engineered strain SCUT27/Δ*ldh*/Δ*argR* for its ability to metabolize xylose-rich lignocellulosic hydrolysates.

Results

Identification and activity test of the Type I-B CRISPR/Cas system in SCUT27

Recently, the Type II CRSIPR/Cas9 system tools have been reported for thermophiles based on some thermostable cas9 derived from *Geobacillus stearothermophilus, Acidothermus cellulolyticus*, and *Geobacillus thermodenitrificans* T12 [9, 10]. For genome engineering in SCUT27, we also tried to construct the plasmids pLY2_*ldh*-HA12 and pLY2_*ldh*-HA12 (Table 1) containing the thermostable *cas9* gene from *G. stearothermophilus* (under the control of constitutive promoter P*adhE* and riboswitch-controlled inducible promoter P*kan*-RS$_{pbuE}$, respectively), a sgRNA expressing module targeting *ldh* and a repairing template. However, no transformants were obtained by electro-transformation with several tries. It indicated that the genome editing with CRISPR/Cas9 system was difficult to be used in our strain probably due to the high toxicity of the potent Cas9 RNP (the heterologous nuclease Cas9 was non-lethal as the pLY1 could transformed into SCUT27), and the limited transformation efficiency with pIKM1. Even using the inducible promoter P*kan*-RS$_{pbuE}$, it seemed that the leaky expression of Cas9 with sgRNA led to cell death. Therefore, we turned to exploit the endogenous CRISPR/Cas system for genome editing in SCUT27.

Table 1 Strain and plasmids used in this study.

Strains/plasmids	Description	Sourse
Strains		
SCUT27	Wild type of *T. aotearoense* SCUT27 (GDMCC 60765)	Our lab
SCUT27/Δ*ldh::ermR*	Deletion of *ldh* (v518_0188) in SCUT27	Our lab
SCUT27/Δ*argR::kanR*	Deletion of *argR* (v518_1864) in SCUT27	Our lab
SCUT27/Δ*ldh*	Deletion of *ldh* in SCUT27	This study
SCUT27/Δ*tdk*	Deletion of *tdk* (v518_2650) in SCUT27	This study
SCUT27/Δ*tdk*/ Δ *ldh*	Deletion of *ldh* in SCUT27/Δ*tdk*	This study
SCUT27/Δ*tdk*/Δ*ldh*/Δ*argR*	Deletion of *argR* in SCUT27/Δ*tdk*/ Δ *ldh*	This study
E. coli DH5α	Host cells for plasmids amplification	Tiangen
Plasmids		
pIKM1	Original plasmid for cloning and expression with Amp^R and Kan^R	Mai et al.[30]
pBluescript II SK(+)	Primordial suicide plasmid	Our lab
pLY1	Insert *GeoCas9* gene under promoter P*adhE* in pIKM1	This study
pLY2	Insert *GeoCas9* gene under promoter P*kan*-RS*pbuE* in pIKM1	This study
pLY1_*ldh* -HA12	Plamid used for *ldh* deletion	This study
pLY2_*ldh* -HA12	Plamid used for *ldh* deletion	This study
pKQ1	Inset P*kan* and terminator (298 bp of sequence located at the 3′ end of the endogenous CRISPR array)	This study
pKQ1_*pyrF1*	Plamid used for *pyrF* (v518_1373) inference	This study
pKQ1_*pyrF2*	Plamid used for *pyrF* inference	This study
pKQ1_*pyrF3*	Plamid used as a control for targeting *pyrF* with a PAM of 'GGC'	This study
pKQ1_*ldh* -HA12	Plasmid used for *ldh* deletion	This study
pKQ2_*ldh* -HA12	The P*kan* in pKQ1_*ldh* -HA12 was replaced with P*cat1*	This study
pKQ3_*ldh* -HA12	The P*kan* in pKQ1_*ldh* -HA12 was replaced with P*adhE*	This study
pKQ4_*ldh* -HA12	The P*kan* in pKQ1_*ldh* -HA12 was replaced with P*kan*-RS*pbuE*	This study
pKQ1_*argR* -HA12	Palsmid used for *argR* deletion with sgRNA_*argR*	This study
pKQ1-P*cat1* _*ldh* -HA12	Strong promoter to express gRNA of *ldh*	This study
pKQ1-P*kan*-RS*pbuE*_*ldh* -HA12	Inducible promoter to express gRNA of *ldh*	This study
pKQ1_Δ*ldh*-H12::*tdk*	Overexpress *tdk* based on pKQ1_Δ*ldh*-H12	This study
pKQ1_Δ*argR*-H12::*tdk*	Overexpress *tdk* based on pKQ1_Δ*argR*-H12	This study
pBlu_*tdk*-H12	Plasmid used for *tdk* deletion	This study
pIKM1-P*ldh*-*ldh*	Overexpress *ldh* with P*ldh*	This study
pIKM1-P*adhE*-*ldh*	Overexpress ldh with P*adhE*	This study
pIKM1-P*cat1*-*ldh*	Overexpress ldh with P*cat1*	This study
pIKM1-P*kan*-RS*pbuE*-*ldh*	Overexpress ldh with P*kan*-RS*pbuE*	This study

CRISPRCasFinder, a web tool to identify clustered regularly interspaced short palindromic repeats and the presence of Cas genes [19], was used here to mine the native CRISPR/Cas system on chromosome of SCUT27 (GCF_000512105.1 *Thermoanaerobacterium aotearoense* SCUT27 assembly genomic.fna). Results showed that SCUT27 possessed the native CRISPR/Cas system which was mainly assigned to Class 1 Type I-B. Seven major CRISPR loci were found in SCUT27 genome sequence (Table S2), but intact Cas genes (v518_0414-0422) were only identified at the upstream and the downstream of the largest CRISPR array containing 55 spacers (Fig. 2a; Table S2). The other CRISPR arrays uncoupled with Cas

genes were defined as orphan CRISPR. This may suggest that the longest array created by numerous spacer acquisition events and co-expression with a complete functional set of Cas genes might constitute an active CRISPR functional unit within the bacterial genome [20]. As a result, this longest CRISPR array (Positon: 14334–18020 in NZ_AYSN01000014.1) including 55 spacers along with direct repeats (DR) was used for further research.

To recognize protospacers by CRISPR/Cas system, a functional PAM flanking at the 5′ end of the protospacer is necessary. The primary approach for PAM identification was to identify the putative original sequences of the invading elements that ended up as CRISPR spacers during the 'adaptation' stage in SCUT27. The searching program *CRISPRTarget* [21], was used to analyze the CRISPR spacers of the strain by aligning these sequences against the existing genome sequences in various databases such as phage, plasmid and so on, as described by Pyne [15]. We set out to analyze spacer sequences of which score were greater than 20 points *in silico*. However, there was no obvious pattern of PAM with the only 55 spacers in the largest CRISPR array in SCUT27. Then we searched for the closest-related strain *T. thermosaccharolyticum* M0795, which possessed the similar Cas operon (Thethe_02655-02664) as SCUT27 and up to 266 spacers on its largest CRISPR array (the same DR motif as SCUT27: GTTTTTAGCCTACCTATAAGGAATTGAAAC) to identify the PAM sequence. From the results of spacer-protospacer matching analysis in M0795, we found the high-frequency patterns of PAM motif 'TTA' at the 5′ end of the protospacers (Fig. 2b, 2c). Besides, the PAM sequence 'TNA' or 'TNA' referring to *Thermoanaerobacter sp.* was also predicted in some literatures with various *in silico* tools [22–24]. In some closely related *Clostridium* such as *C. pasteurianum*, *C. difficile*, *C. tyrobutyricum* and *C. thermocellum*, PAM site 'TNA' was defined in their reported subtype I-B system as well [9, 15, 16, 25]. Based on the above analysis, the most confident PAM site 'TTA' in SCUT27 was characterized and chosen in the following experiments.

Then toxicity assay was performed to evaluate the activity of native CRISPR system. Through designing a plasmid containing a CRISPR mini-array cassette (Fig. S1b, c), the expression of sgRNA further forming ribonucleoprotein (RNP) on the target site mediated single strand DNA break. The artificial mini-array composed of 30 bp DR from the largest CRISPR array of SCUT27 and a 37 bp spacer (responding to the target protospacer with a functional PAM) was constructed (Fig. S1a, b). Like most bacteria, SCUT27 does not encode proteins responsible for non-homologous end joining. Thus, if the CRISPR/Cas system was functional, DNA cutting caused by the active RNP complex could not be repaired and no transformants would be observed. Here, the mini-array on the plasmid was designed to target *pyrF* gene (V518_1373) in SCUT27 (Fig. S1). The 921 bp *pyrF* ORF contains a total of 44 potential PAMs (TTA) on double strands of DNA. Two spacers (37 nt downstream of 'TTA') were chosen for targeting *pyrF* locus and one non-target spacer (PAM motif as 'GGC') was designed as a negative control (Fig.S1a). With a strong promoter P*kan* for sgRNA expression, we observed a 21-fold decrease in electro-transformation efficiency of SCUT27 (from ~ 48.3 to 2.3 CFU/μg DNA), indicating that CRISPR-mediated interference was happening and around 95 % cell killing (cutting efficiency) was observed (Table 2, Fig. S1d). The promoters upstream of Cas operon and sgRNA were strong enough to mediate a single-chain cut. Some background colonies (escape mutants) containing the killing plasmids in experimental groups may be

due to spontaneous mutations in Cas operon (specifically Cas3), which would be expected to inactivate the encoded functional protein [9, 26].

Table 2
Effect of spacers of different genes on transformation efficiency.

Plasmids	5' PAM	Spacer sequence in sgRNA (protospacer)	Transformation efficiency (CFU/µ DNA)
pKQ1_pyrF1	TTA	TACGTAGCTCCAACAACTGCACCTATAGAACTATATC	2.5 ± 0.9
pKQ1_pyrF2	TTA	CGTCATGTATATATGGTGTCATGGTATCTTCTCCCAT	2.1 ± 0.7
pKQ1_pyrF3	GGC	ATTATTTTTCTCAGTATTTTGGCCTCCTCTTTATACG	48.3 ± 11.5
pKQ1_ldh-H12	TTA	CATACAAGATTTCTGGACTTCCATGGGGCAGAGTTTT	6.2 ± 4.4
pKQ1_argR-H12	TTA	TAGAGGAGCTTAAGCTAAATAAAATTATAGCTACACA	3.1 ± 1.9

Endogenous Type I-B CRISPR-based deletion of ldh in SCUT27

After verifying the activity of the native CRISPR system, we constructed the gene editing plasmid pKQ1_ldh-HA12 (Fig. 3a) targeting the gene ldh (V518_0188) for knockout, as the resultant mutant could easily be identified by its product without lactate and showed obvious advantage for growth and ethanol production based on our previous report [3]. A spacer targeting the chromosomal ldh locus was chosen by the same process as described above. Besides, a repair template was introduced for CRISPR/Cas mediated homology directed repair at the downstream of the CRISPR mini-array. It consisted of two homologous arms (HAs) of ~750 bp each and a KpnI digestion site instead of 805 bp in ldh locus to inactivate its function (Fig. S2a).

The transformation of pKQ1_ldh-HA12 into SCUT27 resulted in 12-65 colonies on MTCK plate (~6.2 CFU/ug DNA). Colony PCR was carried out with randomly picked colonies to screen the ldh deletion mutants. Twelves colonies were picked and verified by PCR. As shown in Fig. 3b, the PCR products in three kinds of cases were observed: Nine of 12 colonies (75%) had a clean deletion genotype (Δldh) with 1616 bp, two colonies had a wild-type genotype with 2421 bp and one colony showed the mixed genotype. For the wild type, it was supposed that some site mutation appeared on the key Cas proteins leading to the ineffective function of editing system. For the mixture type, it hasn't been given clear explanations. Overall, with the facilitation of native CRISPR system, desirable mutants could be reproducibly obtained with high edited rate varying from 58.3 to 100 % within 2-4 days after transformation. The colony PCR product was sent for DNA sequencing (Fig. S2a), showing the target locus was edited as we designed. We also cultured the mutant and took the fermentation sample for HPLC with the result of non-lactate product and the same characteristics as SCUT27/Δldh::ermR in our lab (Table S5).

Promotion of endogenous CRISPR system with various promoters for sgRNA expression

Two metrics were considered to evaluate the usefulness of the genome editing tool. One metric is the total number of transformants (the transformation efficiency), the other is the fraction of correct transformants (true-editing efficiency). The number of correct transformants is the product of the above two metrics. Although the Type I−B system in SCUT27 showed efficiency at CRISPR mediated killing and high edited rate was observed in several times, few transformants were obtained after transformation of SCUT27 with editing plasmids. The low efficiency of transformation is possibly due to the high toxicity of CRISPR-induced DNA cutting with sgRNA expression under P*kan*. To further explore the efficiency of CRISPR-mediated DNA cleavage and transformation efficiency, we tested other three constitutive promoters (P*adhE*, P*clo1313_1194* and P*cat1*) for regulating the expression of sgRNA to see which expressing extent was better for editing. The expressing strength of the promoters were characterized by enzymatic activity downstream of them in overexpression vectors (Fig. S3a).

It showed that P*clo1313_1194* as a strong promoter in *C. thermocellum* did not work in SCUT27 (Fig. 4a). The strength of P*adhE* was weaker than P*kan*, and the efficiency of transformation by P*adhE* (~5.5 CFU/µg DNA) was very close to that by P*kan*, as well as the similar editing efficiency (~75.0%) (Fig. 4a, Table S3). The strength of P*cat1* was obviously stronger than P*kan*. However, the transformation and editing efficiency of P*cat1* were both lower than those of P*kan* (Fig. 4a, Table S3). We speculated that stronger expression of sgRNA caused much stronger cleavage and led to higher toxicity to cell, especially in the initial period after electro-transformation, possibly leading to an increased proportion of escape mutants with wild type phenotype as a result.

Notably, through the enzyme activity testing, we found the activity value of *ldh* complement by pIKM1 (P*ldh*-*ldh*) was nearly equal to that of wild type. So, we speculated that the plasmid pIKM1 in our strain mostly had only one copy, which was consistent with Walker's resequencing date in *C. thermocellum*, although the plasmid is originally known to have 10-1000 copy number in *C. thermocellum* [9]. This may could give an explanation to the low number of colonies on MTCK plate after electro-transformation. As shown in Fig. 5, with a constitutive promoter to express sgRNA, the cell would die due to the chromosome breaking under the case of no homology directed repairing, no matter how many plasmids transformed into the cell. However, under the case of homology directed repaired, if only one plasmid was transformed into the cell, the cell would also die owing to the plasmid-breaking (loss of the resistance to kanamycin); only the cells with several plasmids transformed would be survival (SCUT27/Δ*ldh* we got was in this case).

Thus, we looked for an inducible promoter to control the expression of sgRNA and restrict the cutting by the CRISPR system on the initial stage of the transformation. As shown in Fig. 5, with an inducible promoter, the cell containing one plasmid would be survival on MTCK plates without inducer, due to the repressed expression of sgRNA. Then, the inducer was added to activate the cleavage. The colonies with one plasmid under the case of homology directed repaired could be picked on the non-selective plates, while others would be death as no homology directed repairing happened. And in this case, the mutants could be obtained with plasmid-curing in theory.

To verify our guess, here we used a thermostable riboswitch element [27] and constructed the adenine-controlled inducible promoter P*kan*-RS$_{pbuE}$ to control the expression of sgRNA in order to avoid the cutting at the initial stage after transformation. The results of enzyme activity assays showed the riboswitch-mediated inducible promoter was strictly controlled by inducer adenine (Fig. 4a). No enzyme activity was observed when the medium didn't add extra adenine, while 0.04 U/mg enzyme activity was observed with 1 mM adenine in medium. With the sgRNA expression under P*kan*-RS*pbuE*, the number of transformants was obviously increased on MTCK plates (approximately 42.0 CFU/μg DNA), and colonies were picked and cultured in medium with 1 mM adenine for inducing. However, the clones on MTCA plates were mostly false-positive and showed lower editing efficiency (Fig. 4a), even after five times passaging for enrichment. We speculated that the induced promoter here for the expression of CRISPR mini-array might be not strong enough to make ssDNA break thus leading to an increased number of wild type strains or escape mutants. And the mutants we got in this way still contained the plasmids, indicating these mutants had several plasmids initially and cells with a single plasmid maybe difficult to be edited with weak induced promoter. Therefore, utilization of a thermostable and strong inducible promoter for controlling the expression of sgRNA seems a key to this genetic modification tool.

Other attempts to improve the efficiency of transformation and editing with the plasmids

The highly efficient genome editing system (cutting of ssDNA) resulted in few alive colonies on plates, which was likely due to low homologous recombination [9]. It has been noted in CRISPR/Cas genome editing for Clostridial organisms that serial transferring (1:20 dilution) and sub-culturing in liquid medium could enrich the desirable homologous recombination and increase the probability of gene editing [16]. As our previous protocol without extra enrichment, we speculated that it might be also required due to insufficient homology-directed repair in SCUT27. Thus, after the electric shock, we cultivated the bacteria solution and transferred with several times to increase the opportunity of homology repairing. After 5, 10 and 20 rounds of serial passaging in MTCK medium, the results showed that the editing efficiency was between 15.0 - 58.3%, lower than that without serial transferring, while a large number of true-edited mutants could be obtained. It seemed that the escaped mutants increased more rapidly than the true-positive mutants during the serial transferring.

In this study, we also adjusted the amounts of plasmids from 5 μg to 100 μg in 300 mL buffer of electric shock with no prominent improvement (data not shown). Besides, we explored the effects of the various lengths of HAs on the efficiency of transformation and editing with the endogenous CRISPR/Cas system in our strain. The length of HAs in the plasmid was set to 250 bp, 500 bp, 750 bp and 1000 bp. As shown in Fig. 4b, the transformation efficiency was improved as the length of HAs increased. The highest editing efficiency (~ 75% correct edited) was observed when the length of HAs was up to 750 or 1000 bp. And we tried to respectively add 5 and 20 μg HAs of 1 kb (purified PCR products) mixed with 20 μg plasmids into the buffer during the electro-transformation, in order to strength the homologous recombination [28]. The results showed that the addition of repair template (HAs) had an obviously positive effect on transformation efficiency while the editing efficiency of the transformants was relatively low as a compensation (Fig. 4b).

Plasmid curing and argR deletion in mutant Δldh for higher ethanol production

As described above, single gene deletion was achieved with high efficiency using endogenous CRISPR/Cas system and we further explored this system for multiplex genome editing in SCUT27. Considering the limitation of number of both selective makers and thermostable shuttle vectors available in SCUT27, plasmid curing was a must for continuous editing. The editing plasmids pKQ1_*ldh*-H12 in SCUT27/Δ*ldh* needed to be eliminated for acquiring marker-free mutants. So, the mutants were firstly transferred into medium without kanamycin for 5-20 generations transferring in order to lose the plasmid spontaneously. However, no colony without plasmid was picked, indicating the plasmid was difficult to be eliminated probably due to its stable replication region. Then some chemical and physical methods were applied for plasmid curing. To weaken cell wall or damage cell membrane, curing agents like isonicotinic acid hydrazide of 4μg/mL or 0.002% sodium dodecyl sulfate (SDS) (the maximum non-lethal concentration for SCUT27 we tested) were added into cultures in early exponential phase [29, 30]. However, almost all of the picked colonies had retained the plasmids even after several generations. Physical processing such as repeated freezing-thawing, sublethal temperature, electroporation or integrating several methods together were also tried but no one worked efficiently.

Here we developed a novel way for efficiently screening cells of plasmid curing and recycling the editing plasmid for multi-genes editing, based on the thymidine kinase (*tdk*) as a negative selection marker [31]. The 5-fluoro-2'-deoxyuridine (FUDR) is the agent for negative selection as shown in Fig. S4a. The toxicity of FUDR for SCUT27 and SCUT27/Δ*tdk* was tested. As shown in Fig. S4b, under the concentration of 50 μg/mL FUDR, no colonies of wild type grew while lots of colonies of Δ*tdk* were observed, indicating *tdk* could be used as a selectable marker. Besides, deletion of *tdk* showed no influence on growth and properties of fermentation for SCUT27 (Fig. S4d, Table S4). Thus, SCUT27/Δ*tdk* could be used as a starting strain for engineering, unnecessary for reintroducing *tdk* at last. As shown in Fig. 6, SCUT27/Δ*tdk* was firstly picked by homologous recombination and verified by PCR (Fig. S4c), and then the mutant Δ*tdk*/Δ*ldh* was obtained via electro-transformation with pKQ1_Δ*ldh*-H12::*tdk*. Through successive transferring (three generations here) and spreading on MTCF plate, colonies of Δ*tdk*/Δ*ldh* without plasmids were picked with a positive rate around 25% (Fig. S5), showing a significant improvement compared with the previous blind screening. The mutant of plasmid curing was prepared for next round editing.

Here, arginine repressor (*argR*), which had been reported on previous research in our lab [4], was chosen as the target gene for the second editing with pKQ1_Δ*argR*-H12::*tdk*. The mutant with *argR* inactivation had a good performance on growth, ethanol yield and energy level, and its ability to utilize xylose and lignocellulosic hydrolysates had been enhanced [4, 5]. Thus, the editing of *argR* based on SCUT27/Δ*ldh* was valuable to explore for better performance. With the same operation above, the mutant SCUT27/Δ*tdk*/Δ*ldh*/Δ*argR* (termed as Δ*ldh*/Δ*argR* below) was successfully obtained (Fig. S2b, S6), indicating the way for multiplex genome editing was feasible.

Analysis on glucose and xylose utilization ability of the mutants in serum bottles

In most of the lignocellulose hydrolysates, glucose and xylose are the two major fermentable sugars [32]. Here, three kinds of glucose/xylose ratios (1:0, 2:1, 0:1) were investigated by flask fermentation with wild type, Δ*ldh*, Δ*argR*, and Δ*ldh*/Δ*argR*. As shown in Fig. 7 and Table S5, compared with wild type strain under different carbon sources, the maximum OD_{600} of Δ*ldh*, Δ*argR* and Δ*ldh*/Δ*argR* had respectively increased by 95.80-129.44%, 37.64-81.60% and 130.63-135.44%, respectively, with the ethanol production increased by 178.74-218.14%, 102.97-211.54% and 272.96-319.97% as well as acetic acid production, respectively, increased by 215.07-242.21%, 73.80-146.87% and 182.98-195.90%. Besides, the sugar uptake rate of SCUT27/Δ*ldh*, SCUT27/Δ*argR* and SCUT27/Δ*ldh*/Δ*argR* had increased by 79.62-98.68%, 57.02-66.55% and 94.31-147.56%, respectively. The results suggested that knockout of both *ldh* and *argR* strengthened the substrate utilization and the growth of the strain along with the increased production of ethanol. The carbon distribution of SCUT27/Δ*ldh*/Δ*argR* notably flowed to ethanol with the yield of ethanol (g/g sugar) increased by 69.61-121.62%, 11.42-26.74% and 0.83-31.33% compared with wild type strain, SCUT27/Δ*ldh* and SCUT27/Δ*argR*, respectively, under different carbon sources. Especially with xylose, compared with single gene-deletion mutants, SCUT27/Δ*ldh*/Δ*argR* had a more outstanding performance in sugar utilization ability and obviously shortened the lag phase of the xylose utilization caused by carbon catabolite repression under the mixed sugars, suggesting that SCUT27/Δ*ldh*/Δ*argR* could be a potential candidate for ethanol production with various pretreated lignocellulosic hydrolysates.

Besides, as shown in Table S5, SCUT27/Δ*ldh* gained in this study has the similar characterization with the previous mutant SCUT27/Δ*ldh::ErmR* in our lab (2), indicating the editing way by using the CRISPR/Cas system was reliable as same as the traditional way of homologous recombination with a selective marker.

Fermentation kinetic of Δ ldh /Δ argR with various lignocellulosic hydrolysates

An ideal microorganism for lignocellulosic bioethanol production should not only utilize various sugars efficiently but also tolerate high temperature (at least 40 °C) and inhibitors [33]. *Saccharomyces cerevisiae* as a well-known ethanol fermentation strain could not grow well above 40 °C. Though some thermotolerant yeast such as *Kluyveromyces marxianus* or *Ogataea polymorpha* were found, the fermentation traits of them were far inferior to *S. cerevisiae* [34]. Besides, yeasts showed unsatisfactory performance on ethanol fermentation with xylose as sole carbon source due to the cofactor imbalance in cell [35, 36] or the inefficient XI pathway for ethanol production owing to the thermodynamic limit with an unfavorable equilibrium between the xylose and xylulose [37]. Thus, it's urgent to obtain an ideal microorganism for ethanol production especially with xylose. Here, we focused on the ethanol fermentation from xylose with SCUT27/Δ*tdk*/Δ*ldh*/Δ*argR*, a candidate for ethanol fermentation with xylose-rich lignocellulosic hydrolysates.

Hydrolysates of rice straw (RSH), sorghum straw (SSH), peanut straw (PSH), wheat straw (WSH) and soybean straw (OSH) pretreated with diluted H_2SO_4 were chosen to assess the ethanol production of Δ*ldh*/Δ*argR*. Xylose was the main sugar in the hydrolysates of the most lignocellulosic biomass with partial glucose and cellobiose. Here each hydrolysate had been diluted with water at initial xylose

concentration about 15 g/L. As shown in Fig. 8, within the expectation, the improved ethanol production had been obtained by Δ*ldh*/Δ*argR* along with improved sugar utilization ability with all hydrolysates. The ethanol production and yield of the mutant Δ*ldh*/Δ*argR* had been greatly improved about 147.42–739.40% and 112.67–267.89% respectively when compared with wild type. The ethanol titers of RSH, SSH, PSH, WSH and OSH were up to 9.84 g/L, 10.25 g/L, 9.27 g/L, 10.45 g/L and 9.70 g/L in serum bottles, respectively. And the ethanol yields were 0.59 g/g, 0.53 g/g, 0.57 g/g, 0.61 g/g and 0.60 g/g, respectively, all of which were higher than 0.35-0.40 g/g ethanol yield with pure sugars and also beyond the theoretical maximal yield of ethanol 0.51g/g glucose or xylose. We speculated that trace amounts of other substrates such as cellobiose, arabinose and some protein might exist in hydrolysates and had not been calculated in the ethanol yield [38, 39]. Nearly all the xylose and glucose in hydrolysates except RSH could be exhausted by SCUT27/Δ*ldh*/Δ*argR*, which might attribute to the enhanced activity of xylose isomerase and xylulokinase as well as higher energy level of cells for xylose transport [4]. Both wild type strain and Δ*ldh*/Δ*argR* had the strong ability to tolerate lignocellulose-derived inhibitors in hydrolysates such as weak acids, furan derivatives and phenolic compounds [40], and the ability of Δ*ldh*/Δ*argR* was much stronger, especially under the rice and peanut straw hydrolysates (Fig. 8), possibly due to both DnaK-DnaJ-GrpE system and the GroEL-GroES chaperonin up-regulated simultaneously [5], the potential improved ability to eliminate ROS against the inhibitors [5] and improved ATP level to synthesis heat shock protein, pump cytoplasmic protons and transform inhibitors [41]. Results above suggested that Δ*ldh*/Δ*argR* was a promising bioethanol producer with prominent advantages for dealing with harsh environment from various hydrolysates.

Discussion

Over the last decade, CRISPR/Cas system as adaptive immune system in bacteria or archaea, has been repurposed for genome editing and transcriptional regulation in various species [42]. However, the majority of the application are based on the type II CRISPR/Cas9 system, which is hard to be met for thermophiles due to the demand of thermostability and toxicity of Cas9 (poor transformation efficiency) [43]. After the failure attempt with the type II CRISPR/Cas9 system, we set our sights on the endogenous Type I CRISPR/Cas system, which had been harnessed in a few species of bacteria for efficient genome editing including transcriptional regulation [9, 15–17, 25, 44]. Compared with the type II CRISPR/Cas9 system, one of the advantages for genome editing with native CRISPR/Cas system is that the employment of the longer spacer sequence (~37 nt vs. 20 nt for CRISPR/Cas9) can abate the potential off-target effect as the targeting of the crRNA is more specific [14] and thus avoid undesired mutations on the other sites of the chromosome.

In this study, we developed and successfully repurposed the Type I-B CRISPR/Cas system as an efficient genome editing tool in SCUT27. Due to the deficiency of spacers in SCUT27, *in silico* analysis of CRISPR array was also made in *T. thermosaccharolyticum* M0795 which is a closely-related species with SCUT27, showing the same direct repeat sequences of them and a potential PAM site that they may share together. The way for finding PAM was also successfully applied in *Clostridium tyrobutyricum* [45]. The subsequent toxicity assay demonstrated the bioinformatically predicted PAMs identified from M0795

could effectively get employed to SCUT27 with the activity of the endogenous CRISPR/Cas system. With the constitutive promoter P*kan* to drive the transcription of the synthetic CRISPR mini-array on the plasmid without homology repairing templates, the transformation efficiency with the target sgRNA (for *pyrF*) was ~21-fold lower than that with non-target sgRNA (Table 2). For the genome editing with the endogenous CRISPR/Cas system, *ldh* was successfully edited under the repairing templates as we designed. And the desirable mutants could reproducibility obtained with a high editing efficiency of 58.3-100% (Fig. 3, Table S3).

Although the high editing efficiency was displayed with CRISPR/Cas system in SCUT27, the overall transformation efficiency was low for practical use as a genome editing tool. Thus, several promoters including a riboswitch-controlled inducible promoter were chosen to replace P*kan*, exploring the editing effects under the various extent of sgRNA expression. However, it seemed that no significant improvement was discovered with the replacement of the promoters (Fig. 4a), and the transformation efficiency were decreased when the strong promoter P*cat1* was used (Table S3), possibly due to its intense toxicity to cells. In *Clostridium tyrobutyricum*, lactose inducible promoter was utilized for CRISPR array expression and limit the initial toxicity of native system to gain a high editing efficiency [16], however, the riboswitch-controlled promoter P*kan-RS$_{pbuE}$* induced by adenine as a thermostable inducible promoter here seemed not strong enough to transcribe sgRNA leading ssDNA break, with a low editing efficiency but more transformants (Fig. 4a, Table S3). Based on the results above for the deletion of *ldh*, P*kan* seems like a good option with the high edited rate.

It was also worthwhile to point out that through the enzyme activity test for characterization of promoter strength, we made a speculation that the plasmid pIKM1 in our strain mostly had only one copy (Fig. 4a), which was also reported at Walker's resequencing date in *C. thermocellum*, but not having 10-1000 copy numbers in the original report [9]. This may could give an explanation to the low transformation efficiency in this study, as shown in Fig. 5. With a constitutive promoter to express sgRNA and homology directed repaired, the cell with single copy of editing plasmid would also die as the result of plasmid breaking (loss of the resistance to kanamycin); only the cells with several plasmids transformed would be survival. Thus, the riboswitch-based inducible promoter P*kan-RS$_{pbuE}$* here was used to control the expression of sgRNA and restrict the cutting by the CRISPR system on the initial stage of the transformation (Fig. 5). However, the results showed that the clones picked were mostly of false-positive with a low editing efficiency (Fig. 4a), even after five times passage for enrichment, indicating this induced promoter for the expression of CRISPR mini-array may too weak to make ssDNA break thus leading to a lot of escape mutants. Therefore, the application of a thermostable and strong inducible promoter for controlling the expression of sgRNA seems a key to performing a more convenient genetic modification.

The highly efficient genome editing system with few alive colonies on plates was also likely due to low homologous recombination [9]. It has been noted in CRISPR/Cas genome editing for Clostridial organisms that serial transferring and sub-culturing in liquid medium could enrich the desirable homologous recombination and increased the probability of gene editing [16]. In this way, the true-edited mutants were enriched but with the low edited rate between 15.0% - 58.3% after 5-20 rounds of

passaging. It seemed that the escaped mutants increased more rapidly than the true-positive mutants during the serial transferring. The effects of the various lengths of HAs and addition with a moderate amount of extra HAs in electro-transformation were also tested in order to overcome the potential limitation in homologous recombination. The results showed that utilization of long homologous arms and addition of HAs had an obviously positive effect on transformation efficiency (Fig. 4b). However, the edited rate was significantly lower with the extra HAs addition. Therefore, depending on the different genome editing purposes, one can make a tradeoff between the transformation efficiency and genome editing efficiency by the addition of HAs or use of various promoters.

To get the engineered strains with better performance, multiplex genome editing was a significant and necessary technology. In *C. tyrobutyricum*, the artificial CRISPR mini-array on one plasmid with two spacers for deletion of two genes (*spo0A* and *pyrF*) was constructed. However, the editing efficiency of double deletion vector was significantly lower than that of single deletion vector at the 8th generations during the sub-culturing [16]. Furthermore, as the number of editing loci increased, the size of the single plasmid (containing HAs) would be greater, which increase the difficulty for vector construction and transformation. And the mini-array could not be easily gained by overlap PCR when the number of spacers was more than one (the same sequence of direct repeats in CRISPR array would disturb the PCR of fragments for plasmid construction). As a result, the multi-genes editing CRISPR mini-array could only get obtained on costly DNA synthesis.

In order to edit multiple loci in genome cheaply and efficiently, successive editing with one plasmid including one spacer would be a better choose. In our strain, under the condition with limited amounts of positive selection markers available, plasmid curing was inevitable for successive editing. Some attempts were then performed to eliminate the plasmids in mutants, however, it seemed difficult as the mutants without plasmids were hard to screen out, restricting the application of CRISPR/Cas genome editing system. Here, a novel way for efficiently screening cells without plasmids was developed and made the multiplex genome editing come true in SCUT27. As shown in Fig. 6, the negative selection marker *tdk* inserted into the editing plasmid and its selection agent FUDR were used for the recycling of the editing plasmid and the resistant marker *kanR*. Based on the starting strain SCUT27/*Δtdk*, SCUT27/*Δtdk/Δldh* was firstly acquired after the first around of editing. Then with successive transferring, the plasmid-free mutants appeared and screened out on plates with 5-FUDR. The mutants of plasmid curing were further edited for the deletion of *argR* with the native CRISPR/Cas system. Here we got the mutants SCUT27/*Δtdk/Δldh/ΔargR* for fermentation profiles testing. The framework of plasmid with diverse sgRNA and HAs could be successively delivered into the cells for multiple genome editing.

To date, this is the first success for successive gene-editing in *T. aotearoense* strains with type I-B CRISPR/Cas system. And the mutant SCUT27/*Δtdk/Δldh/ΔargR* showed a better performance in sugar utilization ability with the improved ethanol production and yield, especially under xylose (Fig. 7). Compared with Δ*ldh* and Δ*argR*, the ethanol yield (g/g sugar) of*Δldh/ΔargR* was notably increased by 21.72.42-26.74% and 17.97-31.33% under the xylose or the glucose/xylose mixture (Fig. 7, Table S5). Under the five kinds of lignocellulosic hydrolysates, the ethanol production and yield of the mutant Δ*ldh/*

ΔargR had been up to 9.27-10.45 g/L and 0.53-0.61 g/g sugar, and greatly improved about 147.42–739.40% and 112.67–267.89% respectively when compared with SCUT27 (Fig. 8). Δldh/ΔargR also showed strong ability to tolerate lignocellulose-derived inhibitors in hydrolysates such as weak acids, furan derivatives and phenolic compounds [40], especially under the rice and peanut straw hydrolysates (Fig. 8). The great performance of the mutant was possibly due to both DnaK-DnaJ-GrpE system and the GroEL-GroES chaperonin up-regulated simultaneously [5]. And the mutant had the potential ability to eliminate ROS against the inhibitors and improve ATP level with more energy to synthesis heat shock protein, pump cytoplasmic protons and transform inhibitors to respond the stress [5, 41]. Results above suggested that Δldh/ΔargR was a promising bioethanol producer with prominent advantages for dealing with harsh environment from various hydrolysates.

The applications of endogenous CRISPR/Cas system for genome editing in SCUT27 could be potentially larger than the just generation of deletion mutants. This technique could be readily applied for introducing other types of mutations, i.e., fragment insertions and even point mutations. For the point mutation, the homologous arms on the editing plasmid should be designed to introduce changes in the functional PAMs to a nonfunctional motif. Alternatively, substitutions could be introduced into a seed region, the first 8 nucleotides of the protospacer, crucial for CRISPR targeting [20, 26]. Besides, the CRISPR interference (CRISPRi), which allows repression of the expression of target genes, had already been performed with the subtype I CRISPR/Cas system [26, 46] rather than only with dCas9 [47, 48]. The subtype I-B CRISPR-Cas system of *Haloferax volcanii* lacking cas3 and cas6 genes could be used for gene repression with various extents [46]. Type I-E CRISPR/Cas system with the deletion of Cas3 in *Escherichia coli* was also co-opted for programmable transcriptional repression, and yielded the strongest repression targeting promoter regions [44]. This strategy offers a simple approach to convert many endogenous Type I systems into transcriptional regulators, expanding the available toolkit for CRISPR-mediated genetic control while creating new opportunities for genome-wide screens and pathway engineering.

Conclusion

The protocol of utilizing endogenous Type I-B CRISPR/Cas system for multiplex genome editing in this study extends the existing genetic toolbox in SCUT27 and may be adapted to other microorganisms that carry active endogenous CRISPR/Cas systems and limited by available selective markers and plasmid. And the deletion of both *ldh* and *argR* by native CRISPR/Cas system here is a successful strategy to engineer SCUT27 for enhanced ethanol production with lignocellulosic hydrolysates.

Methods

Strains and culture condition

The strains used and engineered in this study are all listed in Table 1. *Escherichia coli* DH5α used for plasmid construction and propagation were selected and grown at 37°C aerobically in LB broth or on

solid LB agar plate supplemented with 50 µg/mL kanamycin or 100 µg/mL ampicillin as required for plasmid maintenance. SCUT27 and the mutants were cultivated with optimized MTC medium (pH 7.0) [3, 49] at 55 °C and 150 rpm anaerobically in 100 mL serum bottles. For selection of transformants, solid MTC medium containing 50µg/mL kanamycin (MTCK), 1 mM adenine (MTCA) or 50 µg/mL 5-fluoro-2'-deoxyuridine (MTCF) were used when required, and were put into anaerobic chamber (Shellab, USA) for pre-deoxygenation about 12 h before using. The lignocellulosic hydrolysates were prepared as the previous study [4] and were added with 20 g/L $CaCO_3$ for adjustment of pH during the fermentation. The concentration of xylose in hydrolysates was diluted around 15 g/L.

Plasmid construction

The plasmids constructed in this study are listed in Table 1 and oligonucleotides used are listed in Table S1. All plasmids were constructed using standard molecular biology technique. The fragments were obtained by the PCR using DNA polymerase KOD (Toyobo Co., Ltd). Restriction endonucleases used in this experiment were purchased from New England Biolabs (Ipswich, MA). The plasmid of editing or overexpression were constructed by double enzyme digestion and Gibson assembly using a recombinase named ClonExpress® MultiS One Step Cloning Kit (Vazyme, Nanjing, China). The recombinant plasmids were transformed into *E. coli* DH5α competent cell (Weidi, Shanghai, China) for amplification, and confirmed by colony-PCR with Rapid Taq DNA polymerase (Vazyme, Nanjing, China) and sequencing (Sagan Co., Shanghai, China). Plasmid isolation and purification was performed using TIANGEN kits.

When screening of promoters, P*kan* was cloned from upstream of kanamycin resistant gene on pIKM1. P*cat1* was amplified from *cat1* (CTK_C06520) promoter region in *C. tyrobutyricum* ATCC 25755. P$_{Clo1313\_1194}$ was amplified from *Clo1313_1194* promoter region in *C. thermocellum* [50]. The fragment of adenine-responsive riboswitch RS*pbuE* was amplified from *Bacillus subtilis* [51], and the construction of P*kan*-RS*pbuE* was shown in Fig. S3b. P*adhE* and gene *ldh* were cloned from SCUT27. Every promoter was fused with ORF of *ldh* through Splicing by Overlap Extension (SOE) PCR, and then was cloned into MCS of pIKM1 by Gibson Assembly, generating vectors of pIKM1-P-*ldh* for LDH expression in SCUT27/Δ*ldh::ermR*. Their responding LDH enzyme activity was tested for characterization of their strength. As a control, 225 bp upstream of *ldh* CDS region was amplified as the potential promoter for *ldh* complementation.

For genome engineering by CRISRP/Cas9 in SCUT27, plasmids pLY1 and pLY2 were constructed containing the thermostable *cas9* gene from *G. stearothermophilus*, respectively under the control of P*adhE* and P*kan*-RS$_{pbuE}$ to regulate the toxicity by Cas9 expression. In addition, a sgRNA expressing module (the motif as described by [10]) targeting gene *ldh* under the control of the constitutive promoter P*cat1* were inserted into these two plasmids as well as a repairing template for *ldh* deletion through homologous recombination. The resultant plasmids pLY1_*ldh*-HA12 and pLY2_*ldh*-HA12 were then transformed into SCUT27.

The construction of all editing plasmids with native CRISRP mini-array was based on plasmid pKQ1 which was inserted into the promoter of kanamycin resistance gene P*kan* and native putative terminator fragment *Ter* (298 bp of sequence located at the 3' end of the largest CRISPR array in SCUT27) into the multiple cloning sites (MCS) of pIKM1. To obtain the killing or editing vectors with mini-array, promoter P*kan* and terminator *Ter* respectively with the sequence of direct repeat (DR) and partial targeting spacer at one end of them were amplified and then through SOE PCR formed ~950 bp CRISPR expression cassette (P*kan*-DR-Spacer-DR-Ter) followed by Gibson Assembly with the *BamHI* and *EcorRI* sites on linear vector according to the manufacturer's guidelines. The repaired templates each containing two homologous arms flanking *ldh* (V518_0188) or *argR* (V518_1864) of wild type was amplified from 250-1000 bp and introduced into the *SalI* and *EcoRI* restriction sites of pKQ1 to mediate the recombination. For plasmid recycling, the fragment *tdk* with the putative promoter and terminator (amplified with primers *tdk*-in-F/R) was inserted into upstream of P*kan* on editing plasmid.

Transformations and mutant screening

All plasmids were ultimately isolated from *E. coli* DH5α and were transformed into SCUT27 (OD_{600} ~ 0.8) and recovered in MTC overnight as previously described [3, 30]. Clones that could grow on MTCK plates were picked into 2 mL centrifuge tube containing 1.0 mL, 50 μg/mL kanamycin MTC medium and then cultured at 55 °C for 24 h in anaerobic chamber (Shellab, USA). Then, they were transferred into 20 mL serum bottles containing 10 mL medium for 24 h at 55 °C and re-transferred into 100 mL serum bottles containing 50 mL medium for enriching the mutants (adenine or FUDR was added when required). A series of sub-culturing (5% v/v inoculum) was carried out to enrich the strains with desirable homologous recombination or plasmid curing, before plating onto plates for selection. The last step was to identify the mutants (deletion of *ldh* and *argR*, plasmid curing or overexpression) by colony PCR, and sequencing of the PCR amplicon when required, with primers designed of which one annealed to the upstream of repair template and the other annealed to the downstream of repair template, as shown in Table S1.

Fermentation kinetics and analytical methods

The flask batch fermentations of SCUT27 and mutants were performed at 55 °C and 150 rpm in 100 mL serum bottles containing 50 mL of MTC medium with various carbon sources such as 25 g/L glucose, 25 g/L xylose or 25 g/L glucose/xylose mixture (mass ratio = 2:1) to mimic lignocellulosic hydrolysates, as well as kinds of lignocellulosic hydrolysates. During the fermentation, the samples were taken at every 4 h intervals until the end to detect the OD_{600} of strains, residual sugar, ethanol, acetic acid and lactic acid concentration in the broth. Cell density was monitored by spectrophotometer (PERSEE T6, Beijing, China) at 600 nm. The concentration of glucose and xylose, and production of ethabol and other metabolites were measured by HPLC (Waters, Milford, USA). All experiments in this study were performed independently at least three times, and data are expressed as mean and standard deviation (SD). To compare statistical difference between the groups of experiment data, student's t test was used.

Enzyme assays

SCUT27 and the mutants of lactate dehydrogenase (LDH) overexpressing were cultured anaerobically in modified MTC medium at 55 °C with 20 g/L glucose. When the OD_{600} reached to ~ 0.8, 5 mL of cell culture was collected by centrifugation and the cell pellet was washed twice with PBS (potassium phosphate buffer) and resuspended in 400 μL of PBS for crushing by ultrasonication on ice. The supernatant of the cells was obtained by centrifugation at 12000×g at 4 °C for 5 min, and the protein concentration was measured by the Pierce TM BCA Protein Assay kit (Thermo Fisher Scientific, USA). The activity of LDH was measured in the direction of pyruvate reduction according to our previous study [52]. The volume of lactate dehydrogenase reaction system was 210 μL, containing 100 μL of 5 mM NADH, 100 μL of 22.7 mM pyruvate and 10 μL cell extract. The reaction was performed at 55 °C for 15 min and NADH consumed was measured at 340 nm. Specific activities were denoted as U/mg protein of cell extract and one unit of enzyme activity corresponds to one unit value decline of A340 nm per minute as the previous study reported [52].

Single guide RNA design and CRISPR/Cas toxicity assays

By identifying the PAM site (TTA) in the locus of target gene, the spacer of single guide RNA was chosen with 37 nt sequence adjacent downstream of PAM, as shown in Table 2. The length was designed as same as the average of spacers in native CRISPR arrays. In toxicity assay of native CRISPR system, a sequence of 37 bp on the sense strand in *pyrF* (V518_1373) locus with a motif of 'GGC' instead of predicted PAM 'TTA' was identified as a control and used for the non-target spacer. P*kan* was chosen for mini-array expression, as is not present in genome of SCUT27 thus not a target for unwanted homologous recombination.

The toxicity assay was used to access functional DNA cutting by the Type I−B CRISPR system in SCUT27, and was characterized by bacterial survival rate (transformation efficiency). Transformation of wild type strain by plasmids with target sgRNA or non-target sgRNA targeting to the *pyrF* coding region (with the predicted PAM or without a predicted PAM) were employed as described above. All transformations were done in biological duplicates and the toxicity (cells killed rate) was calculated using the equation: (CFU of non-target sgRNA - CFU of target sgRNA)/(CFU of non-target sgRNA) * 100%. The editing plasmid used here was modified to include a repair template consisting of two homology arms (HAs), with deletion of a fragment and insertion of a restriction enzyme cutting site. The effect of CRISPR editing efficiency was determined using the equations: X = (CFU of true-edited on MTCK plates)/ (CFU on MTCK plates with plasmid) * 100%.

Statistical analysis

Statistical tests were performed using Statistical Package for the Social Sciences (SPSS ver. 19; IBM Corporation, USA). The data presented in the manuscript and Additional file are the means and standard deviations (**$p \leq 0.01$; *$0.01 \leq p \leq 0.05$; t test).

Abbreviations

CRISPR: Clustered regularly interspaced short palindromic repeat; Cas: CRISPR-associated; sgRNA: single guide RNA; PAM: protospacer-adjacent motif; HAs: homologous arms; PyrF: orotidine-5'-phosphate decarboxylase; LDH: lactate dehydrogenase; ArgR: arginine repressor; Tdk: thymidine kinase; LCB: Lignocellulosic biomass. Δ: deletion; *E. coli*: *Escherichia coli*; HPLC: high-performance liquid chromatography; NADPH and NADP: reduced and oxidized form of nicotinamide adenine dinucleotide phosphate, respectively; OD_{600}: optical density at wavelength (λ) 600 nm; ORF: open reading frame; WT: wild-type.

Declarations

Acknowledgements

We thank Professor Mingjun Zhu for the gift of strain *C. thermocellum* DSM1313.

Authors' contributions

KQD and JFW conceived and designed the experiments; KQD performed experiments and wrote the manuscript; HXF, XLG, JFW contributed reagents or analytical tools, analyzed data, and drafted the manuscript; CYQ helped to design experiments and drafted the manuscript. All authors read and approved the final manuscript.

Funding

Availability of data and material

Data will be made available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have read and approved this manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China. [2] Guangdong Provincial Key Laboratory of Fermentation and Enzyme Engineering, South China University of Technology, Guangzhou 510006, China

References

1. Singhvi MS, Gokhale DV: **Lignocellulosic biomass: Hurdles and challenges in its valorization**. *Appl Microbiol Biotechnol* 2019, **103**:9305–9320.

2. Ai H, Zhang J, Yang M, Yu P, Li S, Zhu M, Dong H, Wang S, Wang J: **Draft Genome Sequence of an Anaerobic, Thermophilic Bacterium**, *Thermoanaerobacterium aotearoense* **SCUT27, Isolated from a Hot Spring in China**. *Genome Announc* 2014, **2**.

3. Cai Y, Lai C, Li S, Liang Z, Zhu M, Liang S, Wang J: **Disruption of lactate dehydrogenase through homologous recombination to improve bioethanol production in** *Thermoanaerobacterium aotearoense*. *Enzyme Microb Technol* 2011, **48**:155–161.

4. Qu C, Chen L, Fu H, Wang J: **Engineering** *Thermoanaerobacterium aotearoense SCUT27* **with argR knockout for enhanced ethanol production from lignocellulosic hydrolysates**. *Bioresource Technology* 2020, **310**:123435.

5. Qu C, Dai K, Fu H, Wang J: **Enhanced ethanol production from lignocellulosic hydrolysates by** *Thermoanaerobacterium aotearoense* **SCUT27/ΔargR1864 with improved lignocellulose-derived inhibitors tolerance**. *Renewable Energy* 2021, **173**:652–661.

6. Barrangou R, Marraffini Luciano A: **CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity**. *Molecular Cell* 2014, **54**:234–244.

7. Sorek R, Lawrence CM, Wiedenheft B: **CRISPR-mediated adaptive immune systems in bacteria and archaea**. *Annu Rev Biochem* 2013, **82**:237–266.

8. Westra ER, van Houte S, Gandon S, Whitaker R: **The ecology and evolution of microbial CRISPR-Cas adaptive immune systems**. *Philos Trans R Soc Lond B Biol Sci* 2019, **374**:20190101.

9. Walker JE, Lanahan AA, Zheng T, Toruno C, Lynd LR, Cameron JC, Olson DG, Eckert CA: **Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium** *Clostridium thermocellum*. *Metab Eng Commun* 2020, **10**:e00116.

10. Mougiakos I, Mohanraju P, Bosma EF, Vrouwe V, Finger Bou M, Naduthodi MIS, Gussak A, Brinkman RBL, van Kranenburg R, van der Oost J: **Characterizing a thermostable Cas9 for bacterial genome editing and silencing**. *Nature communications* 2017, **8**:1647–1647.

11. Horwitz Andrew A, Walter Jessica M, Schubert Max G, Kung Stephanie H, Hawkins K, Platt Darren M, Hernday Aaron D, Mahatdejkul-Meadows T, Szeto W, Chandran Sunil S, Newman Jack D: **Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas**. *Cell Systems* 2015, **1**:88–96.

12. Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J: **Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana**

using guide RNA and Cas9. *Nat Biotechnol* 2013, **31**:688–691.

13. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: **Multiplex genome engineering using CRISPR/Cas systems**. *Science* 2013, **339**:819–823.

14. Shah SA, Erdmann S, Mojica FJ, Garrett RA: **Protospacer recognition motifs: mixed identities and functional diversity**. *RNA Biol* 2013, **10**:891–899.

15. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP: **Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium**. *Sci Rep* 2016, **6**:25666.

16. Zhang J, Zong W, Hong W, Zhang ZT, Wang Y: **Exploiting endogenous CRISPR-Cas system for multiplex genome editing in** *Clostridium tyrobutyricum* **and engineer the strain for high-level butanol production**. *Metab Eng* 2018, **47**:49–59.

17. Zhou X, Wang X, Luo H, Wang Y, Wang Y, Tu T, Qin X, Su X, Bai Y, Yao B, et al: **Exploiting heterologous and endogenous CRISPR-Cas systems for genome editing in the probiotic** *Clostridium butyricum*. *Biotechnol Bioeng* 2021, **118**:2448–2459.

18. Liu L, Yang D, Zhang Z, Liu T, Hu G, He M, Zhao S, Peng N: **High-efficiency genome editing based on endogenous CRISPR-Cas system enhances cell growth and lactic acid production in** *Pediococcus acidilactici*. *Appl Environ Microbiol* 2021:AEM0094821.

19. Grissa I, Vergnaud G, Pourcel C: **CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats**. *Nucleic Acids Research* 2007, **35**:W52-W57.

20. Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina-Bedoya M, Fortier LC, Severinov K, Dupuy B, Soutourina O: **Function of the CRISPR-Cas System of the Human Pathogen** *Clostridium difficile*. *mBio* 2015, **6**:e01112-01115.

21. Biswas A, Gagnon JN, Brouns SJ, Fineran PC, Brown CM: **CRISPRTarget: bioinformatic prediction and analysis of crRNA targets**. *RNA Biol* 2013, **10**:817–827.

22. Fischer S, Maier LK, Stoll B, Brendel J, Fischer E, Pfeiffer F, Dyall-Smith M, Marchfelder A: **An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA**. *J Biol Chem* 2012, **287**:33351–33363.

23. Mendoza BJ, Trinh CT: **In Silico Processing of the Complete CRISPR-Cas Spacer Space for Identification of PAM Sequences**. *Biotechnol J* 2018, **13**:e1700595.

24. Leenay RT, Beisel CL: **Deciphering, Communicating, and Engineering the CRISPR PAM**. *J Mol Biol* 2017, **429**:177–191.

25. Maikova A, Kreis V, Boutserin A, Severinov K, Soutourina O: **Using an Endogenous CRISPR-Cas System for Genome Editing in the Human Pathogen** *Clostridium difficile*. *Appl Environ Microbiol* 2019, **85**.

26. Li M, Gong L, Cheng F, Yu H, Zhao D, Wang R, Wang T, Zhang S, Zhou J, Shmakov SA, et al: **Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems**. *Science* 2021, **372**.

27. Marcano-Velazquez JG, Lo J, Nag A, Maness PC, Chou KJ: **Developing Riboswitch-Mediated Gene Regulatory Controls in Thermophilic Bacteria**. *ACS Synth Biol* 2019, **8**:633–640.

28. Shaw AJ, Covalla SF, Hogsett DA, Herring CD: **Marker removal system for** *Thermoanaerobacterium saccharolyticum* **and development of a markerless ethanologen**. *Appl Environ Microbiol* 2011, **77**:2534–2536.

29. El-Mansi M, Anderson KJ, Inche CA, Knowles LK, Platt DJ: **Isolation and curing of the** *Klebsiella pneumoniae* **large indigenous plasmid using sodium dodecyl sulphate**. *Res Microbiol* 2000, **151**:201–208.

30. Mai V, Lorenz WW, Wiegel J: **Transformation of** *Thermoanaerobacterium sp. strain* **JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance**. *FEMS Microbiology Letters* 1997, **148**:163–167.

31. Shao X, Zhou J, Olson DG, Lynd LR: **A markerless gene deletion and integration system for** *Thermoanaerobacter ethanolicus*. *Biotechnology for Biofuels* 2016, **9**.

32. Fu H, Yang S-T, Wang M, Wang J, Tang IC: **Butyric acid production from lignocellulosic biomass hydrolysates by engineered Clostridium tyrobutyricum overexpressing xylose catabolism genes for glucose and xylose co-utilization**. *Bioresource Technology* 2017, **234**:389–396.

33. Du C, Li Y, Zhao X, Pei X, Yuan W, Bai F, Jiang Y: **The production of ethanol from lignocellulosic biomass by** *Kluyveromyces marxianus* **CICC 1727-5 and** *Spathaspora passalidarum* **ATCC MYA-4345**. *Applied Microbiology and Biotechnology* 2019, **103**:2845–2855.

34. Radecka D, Mukherjee V, Mateo RQ, Stojiljkovic M, Foulquié-Moreno MR, Thevelein JM: **Looking beyond Saccharomyces: the potential of non-conventional yeast species for desirable traits in bioethanol fermentation**. *FEMS Yeast Res* 2015, **15**.

35. Neirinck LG, Maleszka R, Schneider H: **The requirement of oxygen for incorporation of carbon from d-xylose and d-glucose by** *Pachysolen tannophilus*. *Archives of Biochemistry and Biophysics* 1984, **228**:13–21.

36. Morales-Rodriguez R, Perez-Cisneros ES, de Los Reyes-Heredia JA, Rodriguez-Gomez D: **Evaluation of biorefinery configurations through a dynamic model-based platform: Integrated operation for bioethanol and xylitol co-production from lignocellulose**. *Renewable Energy* 2016, **89**:135–143.

37. Park H, Jeong D, Shin M, Kwak S, Oh EJ, Ko JK, Kim SR: **Xylose utilization in** *Saccharomyces cerevisiae* **during conversion of hydrothermally pretreated lignocellulosic biomass to ethanol**. *Applied Microbiology and Biotechnology* 2020, **104**:3245–3252.

38. Lai Z, Zhu M, Yang X, Wang J, Li S: **Optimization of key factors affecting hydrogen production from sugarcane bagasse by a thermophilic anaerobic pure culture**. *Biotechnology for Biofuels* 2014, **7**:119.

39. Lopez-Hidalgo AM, Hernandez VEB, De Leon-Rodriguez A: **Scale-up of hydrogen and ethanol co-production by an engineered** *Escherichia coli*. *Fuel* 2021, **300**.

40. Baral NR, Shah A: **Microbial inhibitors: formation and effects on acetone-butanol-ethanol fermentation of lignocellulosic biomass**. *Applied Microbiology and Biotechnology* 2014, **98**:9151–

9172.

41. Gheshlaghi R, Scharer JM, Moo-Young M, Chou CP: **Metabolic pathways of clostridia for producing butanol**. *Biotechnology Advances* 2009, **27**:764–781.

42. Barrangou R, Horvath P: **A decade of discovery: CRISPR functions and applications**. *Nature Microbiology* 2017, **2**:17092.

43. Wang Y, Zhang Z-T, Seo S-O, Lynn P, Lu T, Jin Y-S, Blaschek HP: **Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in** *Clostridium beijerinckii* **as an Example**. *ACS Synthetic Biology* 2016, **5**:721–732.

44. Luo ML, Mullis AS, Leenay RT, Beisel CL: **Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression**. *Nucleic Acids Research* 2015, **43**:674–681.

45. Bao T, Feng J, Jiang W, Fu H, Wang J, Yang ST: **Recent advances in n-butanol and butyrate production using engineered** *Clostridium tyrobutyricum*. *World J Microbiol Biotechnol* 2020, **36**:138.

46. Stachler A-E, Marchfelder A: **Gene Repression in Haloarchaea Using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas I-B System***. *Journal of Biological Chemistry* 2016, **291**:15226–15242.

47. Qi Lei S, Larson Matthew H, Gilbert Luke A, Doudna Jennifer A, Weissman Jonathan S, Arkin Adam P, Lim Wendell A: **Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression**. *Cell* 2013, **152**:1173–1183.

48. Shen A: **A Xylose-Inducible Expression System and a CRISPR Interference Plasmid for Targeted Knockdown of Gene Expression in** *Clostridioides difficile*. *J Bacteriol* 2019, **201**.

49. Qu C, Chen L, Li Y, Fu H, Wang J: **The redox-sensing transcriptional repressor Rex is important for regulating the products distribution in** *Thermoanaerobacterium aotearoense* **SCUT27**. *Appl Microbiol Biotechnol* 2020, **104**:5605–5617.

50. Olson DG, Maloney M, Lanahan AA, Hon S, Hauser LJ, Lynd LR: **Identifying promoters for gene expression in** *Clostridium thermocellum*. *Metabolic Engineering Communications* 2015, **2**:23–29.

51. Marcano-Velazquez JG, Batey RT: **Structure-guided mutational analysis of gene regulation by the** *Bacillus subtilis* **pbuE adenine-responsive riboswitch in a cellular context**. *J Biol Chem* 2015, **290**:4464–4475.

52. Qu C, Chen L, Fu H, Wang J: **Engineering** *Thermoanaerobacterium aotearoense* **SCUT27 with argR knockout for enhanced ethanol production from lignocellulosic hydrolysates**. *Bioresour Technol* 2020, **310**:123435.

53. Crooks GE, Hon G, Chandonia J-M, Brenner SE: **WebLogo: a sequence logo generator**. *Genome research* 2004, **14**:1188–1190.

Figures
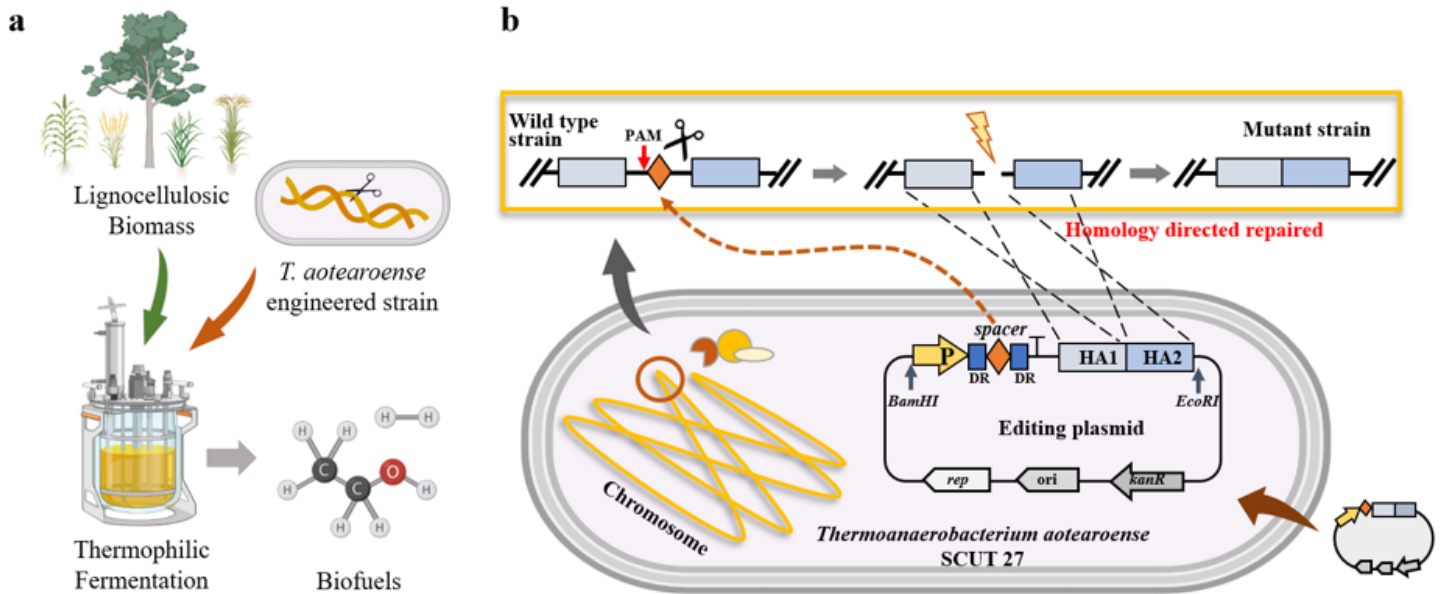
Figure 1

The prospect and CRISPR/Cas genome editing system of SCUT27. a With an optimal growth temperature over 50 °C and some advantages over mesophilic bacteria, SCUT27 is a promising candidate for producing biofuels by using lignocellulosic biomass. b Schematic of the endogenous Type I-B CRISPR/Cas system in SCUT27 functional by electro-transformation with editing vectors. Cleavage at target chromosome region results from the expression of CRISPR mini-array under a constitutive or inducible promoter on plasmid, forming a ribonucleoprotein (RNP) complex with native Cas proteins. The editing plasmid carrying repair templates allows recombination between plasmid and chromosome before or during the CRISPR interference.
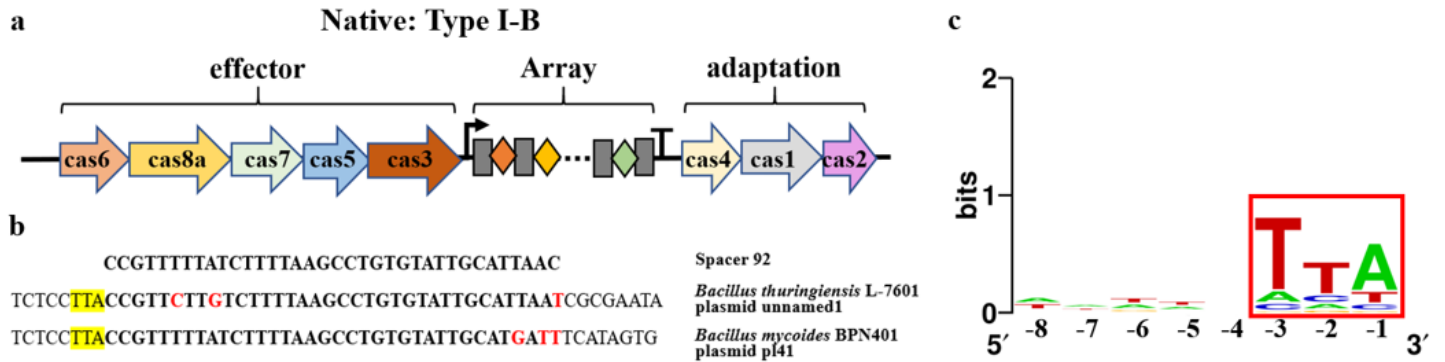
Figure 2

Characterization of the Type I-B CRISPR/Cas system in SCUT27. a Structure of the only intact Type I−B CRISPR/Cas loci in the genome of SCUT27, which possesses a representative Type I-B Cas operon including cas6-cas8a-cas7-cas5-cas3 (effector), cas4-cas1-cas2 (adaptation) and the longest array

containing 55 distinct spacers (diamonds) separated by 30-nt direct repeats (rectangles). b Identification of putative protospacers via in silico analysis in T. thermosaccharolyticum M0795. Spacer-protospacer mismatches are indicated with red color. Eight nt of the 5'- and 3'- end adjacent sequences are provided. Putative PAM are indicated in yellow. c PAM identification for the CRISPR system in M0795. The alignment of regions flanking protospacers was used to create the sequence logo by WebLogo [53]. The 5' adjacent motifs at positions -8 to -1 of the potential protospacers is indicated and 'PAM' of 3-nucleotide is shown in red box.
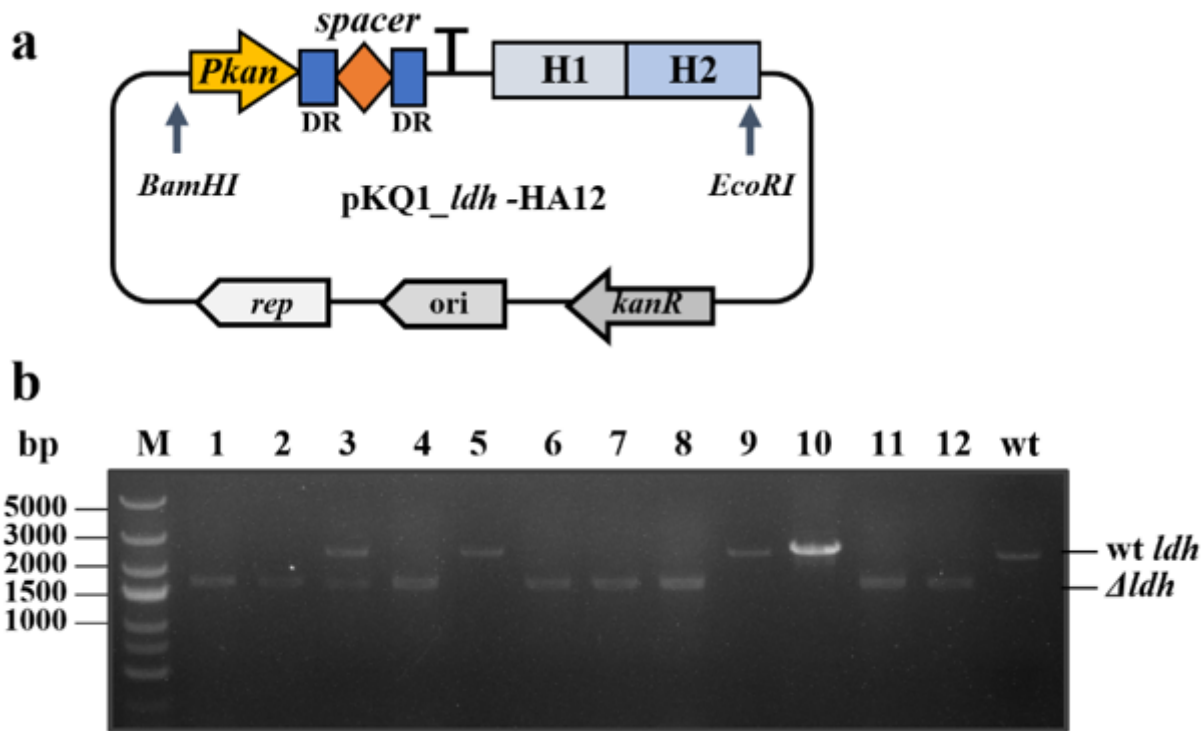
Figure 3

The construction and results of the editing plasmid pKQ1_ldh-HA12 for ldh deletion. a Schematic of pKQ1_ldh-HA12, comprised of a synthetic CRISPR expression cassette and a ldh editing template. b Gel electrophoresis analysis for ldh deletion mutants (1,616 bp PCR bands) and SCUT27 wild type (2,421 bp). PCR products of SCUT27/Δldh (lane 1, 2, 4, 6, 7, 8, 11, 12) and wild type with primer ldh-c-F and ldh-c-R distal to the entire recombination region.
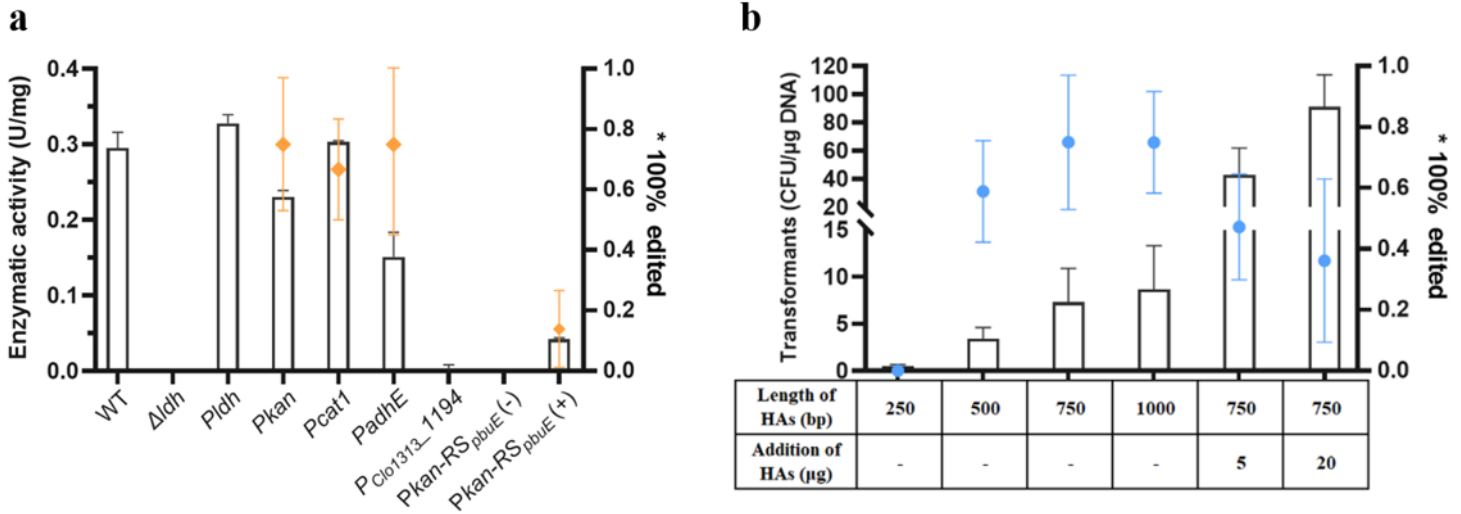
Figure 4

Promotion for native CRISPR genome editing system. a Relative expressing strength of various promoters and editing efficiency with sgRNA expression. The left y-axis shows the activity of LDH (bars) and the right y-axis shows the editing efficiency (orange points) with the same sgRNA_ldh. Δldh was as negative control, wild type and Δldh::Pldh-ldh were as positive controls. (+) means culturing with 1 mM adenine; (-) means culturing without adenine. b The efficiency of transformation and CRISPR editing with various length or extra addition of HAs. The right y-axis shows transformation efficiency (bars) and the left y-axis shows editing efficiency (blue points). "-" means without HAs addition. The data above are the means and standard deviations of three replicates.
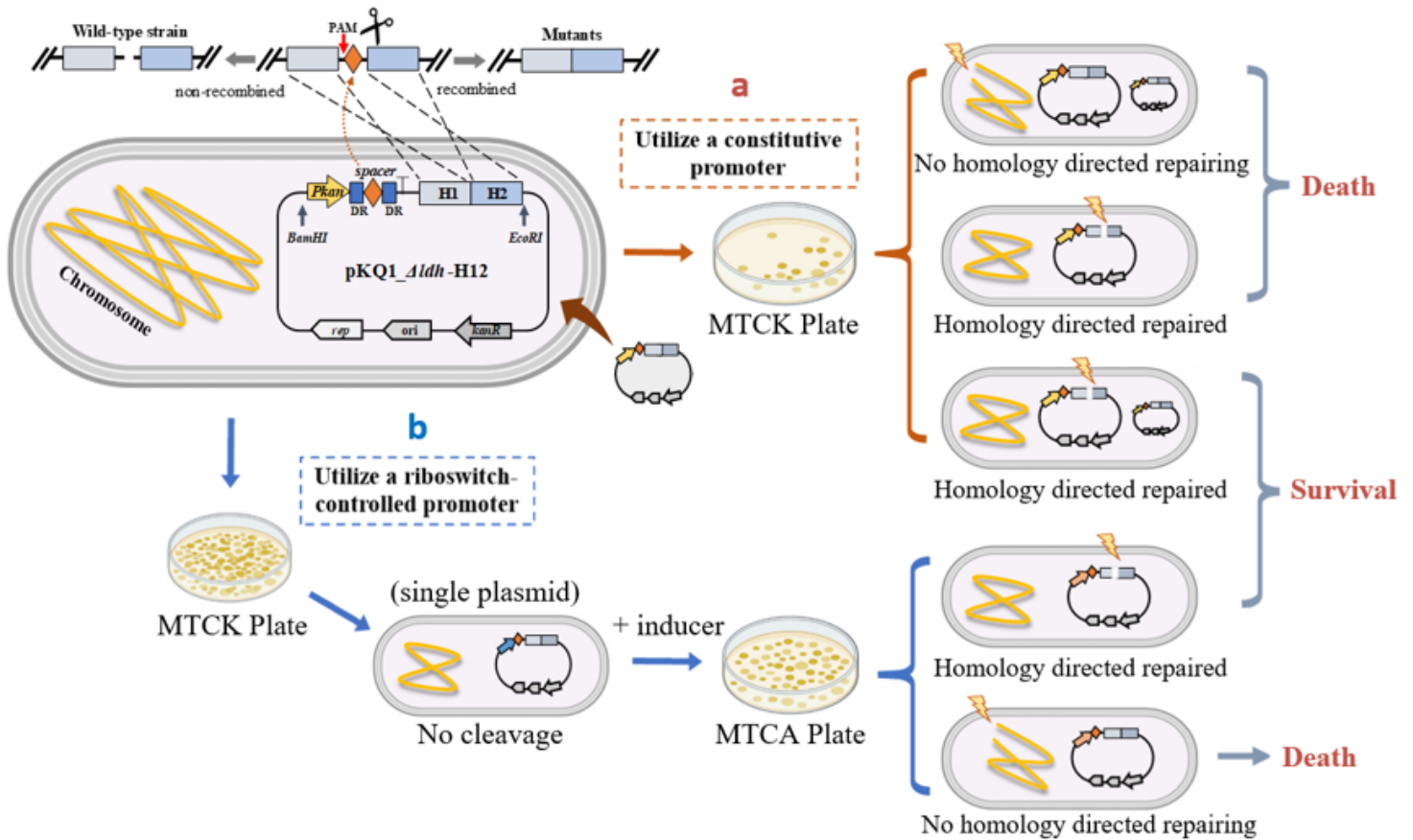
Figure 5

Different processes and results of the transformation by editing plasmids with two kinds of promoters. The expression of sgRNA on plasmids were respectively controlled by constitutive promoter Pkan (a) and adenine-induced riboswitch-based promoter Pkan-RSpbuE (b). In the former case, self-cleavage was performed by immediate expression of the CRISPR mini-array after transformation. On MTCK plates, the transformants with no homology directed repairing, or homology directed repaired but containing just one plasmid, were dead; Only the transformants with several plasmids and homology directed repaired could be survival. In the latter case, the transformants with one plasmid could be survival without inducer on MTCK plate. Then after induction, the transformants of homology directed repaired could be survival on non-selective plate while that with no homology directed repairing were dead.
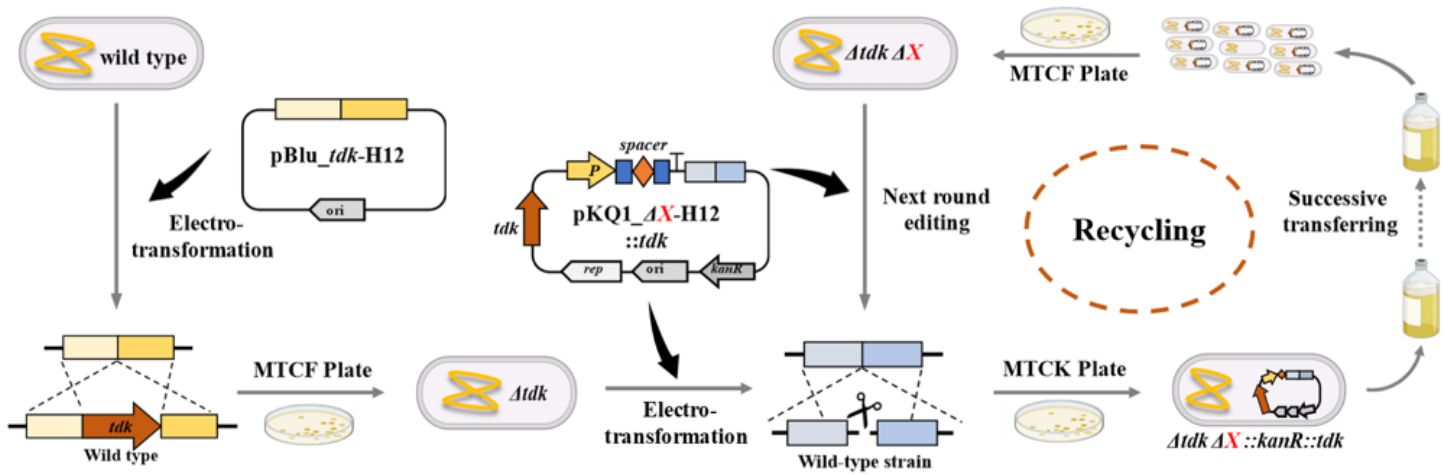
Figure 6

The general workflow of multiplex genome editing system in SCUT27. The suicide vector pBlu_tdk-H12 was used for recombination event to delete tdk in wild-type strain. The shuttle vector pKQ1_ΔX-H12::tdk (X in this study was ldh or argR) was used for CRISPR-mediated recombination events for deleting gene X in SCUT27/Δtdk or other engineered strains without plasmids for successive editing.
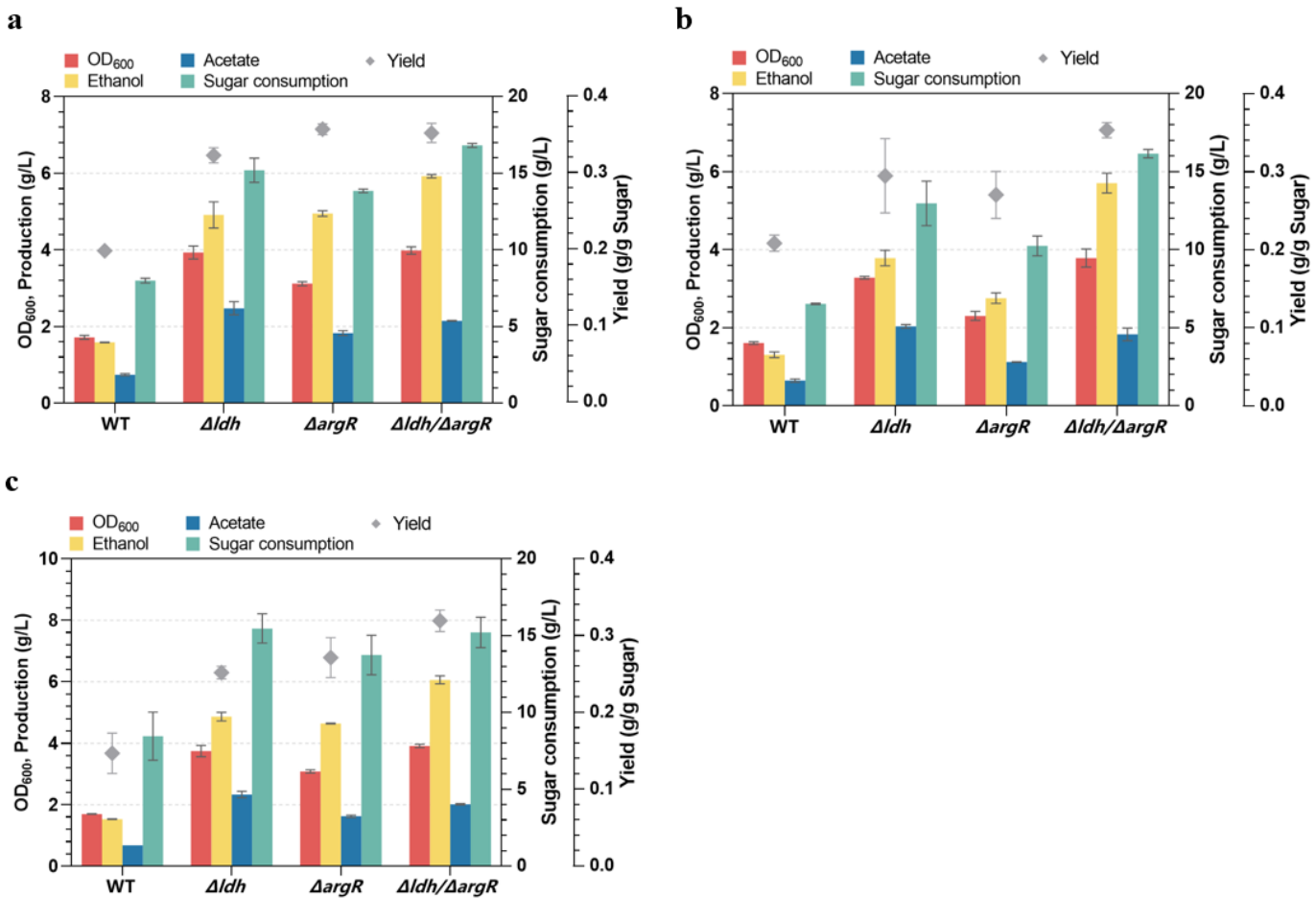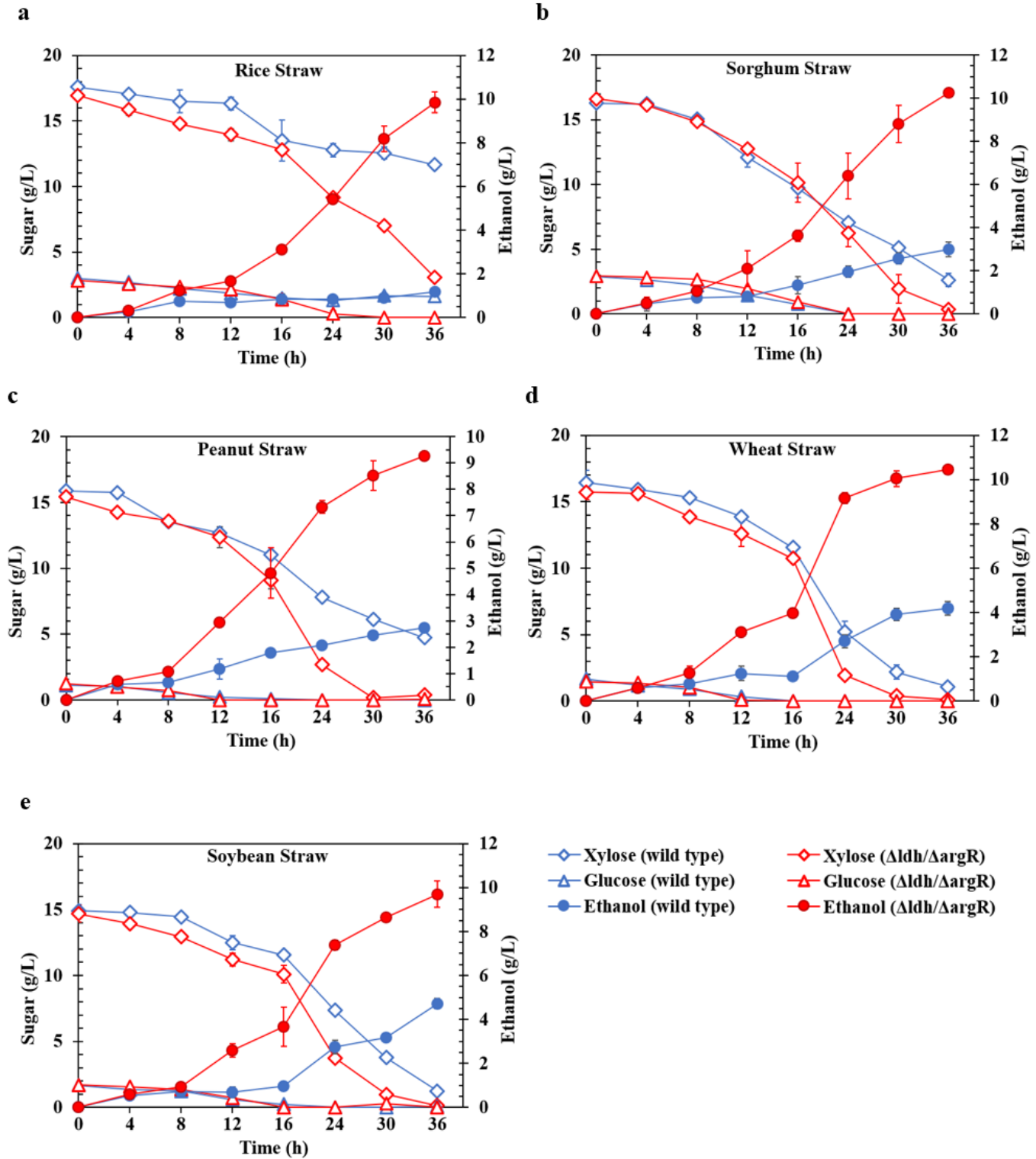

Figure 7

The fermentation property of SCUT27 wild-type strain, Δldh, ΔargR and Δldh/ΔargR with ~25g/L of different carbon sources (a. glucose; b. xylose; c. mixed glucose: xylose = 2:1).

Figure 8

Fermentation profiles by SCUT27 and SCUT27/Δldh/ΔargR under dilute acid pretreated biomass hydrolysates in serum bottles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- crisprsupplement.docx