TargeTron technology applicable in solventogenic clostridia: Revisiting 12 years' advances

Zhiqiang Wen[1], Minrui Lu[1], Rodrigo Ledesma-Amaro[2], Qi Li[3], Mingjie Jin[1§], Sheng Yang[4,5*]

[1] School of Environmental and Biological Engineering, Nanjing University of Science & Technology, Nanjing, 210094, China

[2] Department of Bioengineering, Imperial College London, London SW7 2AZ, UK

[3] College of Life Sciences, Sichuan Normal University, Longquan, Chengdu 610101, China

[4] Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

[5] Huzhou Center of Industrial Biotechnology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, 200032, China

*Corresponding author: ShengYang, E-mail: syang@sibs.ac.cn

[§]Co-corresponding author: Mingjie Jin, E-mail: jinmingjie@njust.edu.cn

Abstract: *Clostridium* has great potential in industrial application and medical research. But low DNA repair capacity and plasmids transformation efficiency severely delayed development and application of genetic tools based on homologous recombination (HR). TargeTron is a gene editing technique dependent on the mobility of group II introns, rather than homologous recombination, which made it very suitable for gene disruption of Clostridium. The application of TargeTron technology in *Clostridium* was academically reported in 2007 and this tool has been introduced in

Received: 23/06/2019; Revised: 20/08/2019; Accepted: 21/08/2019

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/biot.201900284.

various clostridia as it is easy to operate, time-saving, and reliable. TargeTron has made great progress in solventogenic *Clostridium* in the aspects of acetone-butanol-ethanol (ABE) fermentation pathway modification, important functional genes identification, and xylose metabolic pathway analysis & reconstruction. In the review, we revisited 12 years' advances of TargeTron technology applicable in solventogenic Clostridium, including its principle, technical characteristics, application and efforts to expand its capabilities, or to avoid potential drawbacks. Some other technologies as putative competitors or collaborators are also discussed. We believe that TargeTron combined with CRISPR/Cas- assisted gene/base editing and gene-expression regulation system will make a better future for clostridial genetic modification.

Key words: *Clostridium*, group II intron, gene disruption, metabolic engineering, functional gene identification

Abbreviations: CRISPR, clustered regulatory interspaced short palindromic repeat; **EBS**, exon-binding site; **Ery**, erythromycin; **FOA**,5-fluoroorotic acid; **IBS**, intron-binding site; **IEP**, intron-encoded protein; **ORF**, open reading frame; **RAM**, retrotransposition-activated marker; **RNP**, ribonucleoprotein; **RT**, reverse transcriptase; **ABE**, acetone-butanol-ethanol; **FRT**, flipase recognition target; **FLP**, flipase; **R&M** system: restriction & modification system; **CCP**, catabolite control protein; **CCR**, carbon catabolite repress; **EN**, endonuclease; **PCR**, polymerase chain reaction; **HR**, homologous recombination; **NHEJ**, non-homologous end joining; **Cre**, cyclization recombination enzyme; **loxP**, locus of X-overP1.

1 Introduction

Clostridium is an important genus of prokaryotes which is comprised of approximately 200 distinct species [1, 2]. Most clostridia are Gram-positive, obligately anaerobic and can form endospores under advertise environment. clostridia can be divided into pathogenic and nonpathogenic species. The nonpathogenic species, such as *Clostridium acetobutylicum, C. beijerinckii*, *C. tyrobutyricum*, *C. thermocellum*, *C. cellulovorans*, *C. diolis*, and *C. ljungdahlii* have great industry potential. They can be adapted to produce bulk chemicals such as ethanol, butanol, butyrate and 1,3-propanediol from starch, glycerol, lignocellulosic biomass, or syngas [3-5]. The pathogenic clostridia which compass *C. perfringens, C. botulinum, C. diffcile* and *C. tetani* are valuable in medical studies and clinical practice [6, 7]. Genetic tools based on homologous recombination were hampered in *Clostridium* due to low plasmids transformation efficiency and inefficient DNA repair, which resulted in lacking of efficient genetic tools for targeted gene deletion or insertion [8-10].

TargeTron is a genome editing technology dependent on site-specific insertion of mobile group II introns rather than homologous recombination [11-14]. It is very suitable for genetic modification of bacteria that was considered intractable to conventional genetic tools. Mesophilic-TargeTron and thermo-TargeTron technology applicable to clostridia have been developed based on group IIA intron Ll. LtrB as well as group IIB intron TeI3c and TeI4c, respectively [15-17]. They have been applied in

many clostridia for metabolic engineering and identification of functional genes because of manipulation convenience and high reliability [9, 18, 19]. TargeTron technology was seemingly preferred in most cases compared to other genetic tools based on homologous recombination such as allelic coupled-exchange and even the emerging CRISPR/Cas9 system. However, there are some native drawbacks in TargeTron technology such as low frequency of off-target and putative polar effect [9-11, 18].

Recently, a CRISPR/Cas- assisted base editing system applicable in clostridia has been developed [20]. It also does not depend on homologous recombination, and is as programmable as TargeTron. Moreover, it has potential to edit multiple genes targets at one time. Besides, gene expression regulation system [21, 22], and even the emerging RNA editing [23] based on CRISPR/Cas system also challenge TargeTron in design and operation convenience. Are these techniques competitors or collaborators to TargeTron?

The review presents developments, improvements and applications of clostridial TargeTron technology in the last 12 years, as summarized in Figure 1. How to improve and make use of this technology in post-genome-editing era are also discussed.

2 Group II intron

The group II intron exists widely in bacteria, fungi, archaea and plants, and is considered as the ancestor of cleavage introns and retrotransposons of prokaryotes[24]. It mainly encompasses a catalytically active RNA (a ribozyme) and an intron-encoded protein (IEP) with reverse transcriptase activity[14, 25]. The intron folds into a conserved secondary structure with Mg^{2+} binding to the active site to achieve self-splicing [25]. The secondary structure of the RNA has six domains (Domain I to Domain VI, hereinafter abbreviated as D I- DVI) [26, 27], as indicated in Figure 2A. During splicing, DVI can attack the 5′ splice site of intron RNA, resulting in lariat RNA formation. The complex of RT and intron lariat RNA can recognize specific DNA sites and integrate intron RNA into a DNA strand, which is called "retrohoming"[28-30].

The process of “retrohoming” is shown as Figure 2B. It involves many proteins, the most important of which is the reverse transcriptase encoded by the intron. IEP binds specifically to intron RNA and stabilizes its catalytic structure [27]. Most IEPs have four domains: N-terminal reverse transcriptase (RT) domain; X domain; C-terminal DNA-binding domain DNA (D) and endonuclease (EN). The X domain and RT domain are responsible for cDNA synthesis [29, 31]. The RT domain contains the same conserved sequence as the fingers/palm and thumb domain of the retrovirus (RT1-7). Notably, since the N-terminal extension and the insertion between the blocks, the RT of the group II intron is larger than that of the retrovirus RTs [27, 31]. EN domain worked under the guide of DIV domain. The latter binds to the target DNA and the former cleaves the DNA strand for subsequent reverse transcription[32]. Interestingly, not all bacterial IEPs have EN domain [29, 31].

The group II introns is divided into three sub-classes - IIA, IIB, IIC, which are similar in secondary structure but distinguished by the interaction between exon region and the active site [33]. The IBS1, IBS2 (intron binding site 1, 2)and σ' of 5' exon region binds to EBS1, EBS2 (exon-binding site 1, 2) and δ of DI domain respectively in IIA [30]; while in IIB, δ' is combined with EBS3 of DI

domain[34]. In contrast, there are only two interactions, IBS1 and EBS1, IBS3 and EBS3, involved in IIC intron [35]. Besides, IIC intron may also recognize the stem-loop structure of the 5' exon transcription terminator (*attC* site) [35]. All interactions above also occur in the process of intron RNA recognizing target DNA.

EBS1, EBS2, and δ of the DI domain in intron RNA can recognize IBS1, IBS2, and δ' on the target DNA. Furthermore, RT recognizes sequences upstream and downstream of the IBS/δ region. Accordingly, it is totally possible to reprogram the intron RNA to target genes of interest. Karberg et al. demonstrated the method was feasible, and then developed the TargeTron technology based on the *LI.LtrB* of the group IIA from *Lactococcus lactis* [13, 36-38]. The optimal IEP insertion site was predicted by algorithm derived from the sequence recognition rules, and the EBS1, EBS2 and δ were modified using the programmed polymerase chain reaction (PCR) primers. They successfully achieved the disruption of DExH/D-box protein encoding gene and DNA helicase gene in *Escherichia coli*, which paved the way for the TargeTron technology development. Afterwards, adaption and optimization of TargeTron based on various group II introns begun to emerge and extended to other microorganisms.

3 The application of TargeTron system in *Clostridium*

The efficiencies of TargeTron is so high that the positive mutants can be easily identified by colony PCR. It even does not require selectable marker. Besides, insertion site is only related to the designed RNA instead of the host. These feature makes it particularly suitable for strains intractable to genetic modification due to low homologous recombination efficiency, for instance, *Clostridium*. TargeTron technology has been successfully introduced into many clostridia and plays an irreplaceable role in gene function identification and metabolic engineering.

In 2007, Heap et al. and Shao et al. individually developed a TargeTron technology applicable for mesophilic *Clostridium* [15, 16]. The mesophilic TargeTron system is based on the *LI.LtrB* of the group IIA from *Lactococcus lactis.* It should be noted that *Ll.LtrB*-based TargeTron technology is suitable for almost all bacteria, even some eukaryotes [11, 19, 33]. Its plasmid system and DNA integration mechanism are shown in Figure 3A.

The RNP contains a self-splicing lariat intron RNA and its encoded transcriptase-active protein IEP. IEP recognizes 4-6 bases on double-stranded DNA and anneals the double-stranded DNA, helping EBS1, EBS2 and δ on the intron RNA recognize the corresponding IBS1, IBS2 and δ' region on the target DNA [29, 33]. The intron RNA then reversely splices the DNA strand by its ribozyme function, and the IEP cleaves the antisense strand as a template for cDNA synthesis. The cDNA is integrated into the genome by the host's DNA repair mechanism. Positive ratio of mesophilic TargeTron ranges from 1% to 100%. Shao et al. successfully interrupted two genes in *C. acetobutylicum* [15] and Heap et al. realized multiple genes disruption in *C. acetobutylicum, C. difficile, C. botulinum* and *C. sporogenes* [16]. However, multiple insertion sites make the insertion efficiency varied from a few percent to 60%, which greatly hindered the screening and isolation of the positive mutants. Inspired by the Zhong et al. [39], Heap et al. introduced the

retrotransposition-activated marker (RAM) to improve TargeTron technology, which is then universally applied in *Clostridium* [40].

Thermophilic *Clostridium* such as *C. thermocellum*, *C. thermosaccharolyticum* are promising platforms for the industrial production of bioethanol or biobutanol, but mesophilic TargeTron is not effective in them [17, 18]. Mohr et al. thus developed a TargeTron technology applicable in Thermophiles [17]. Its plasmid system and DNA integration mechanism are shown in Figure 3B. Key genetic elements of the technique are *TeI3c* of group IIB intron and *TeI4c*RT from thermophilic cyanobacteria *Thermosynechococcus elongatus*. *TeI3c* is a group II intron naturally without RT ORF, but the RT ORF from other introns (*TeI4c*) can be inserted into it. Interestingly, *TeI4c*RT can drive *TeI3c* more efficiently than its own intron *TeI4c* [41]. The IEP of this system only recognizes two adenylate residues upstream the IBS. The EBS1, EBS2, EBS3 of *Tel3c*, which are characteristic in IIB intron, recognize the IBS1, IBS2, IBS3 of both target DNA and 3' exon of the intron RNA. The targeting efficiency of Thermo-TargeTron ranged from 67% to 100%, and the positive transformants are also very easy to screen.

DNA recognition rules of mesophilic group IIA intron *Ll. LtrB* and thermophilic group IIB intron *TeI3c* as well as the corresponding plasmids system are compared in Figure 3. It shows that the different intron RNA has different recognition rules. Besides, the number of base-pairs recognized by RT is also critical. Compared to *LtrA* RT, *Tel4C* RT can work by recognizing only 2 base-pairs. Thermo-TargeTron is therefore preferred for targeting short ORFs or small non-coding RNAs [17]. In addition, *Tel3c*-based TargeTron can also work at room temperature, which means it may be applied to mesophilic clostridia though there is no relevant report yet.

The principles of TargeTron technologies mentioned above determine their characteristics. They disrupt target gene by site-specifically inserting intron RNA. The integration site of intron is independent to the host. It is only related to the programed RNA and can be predicted by algorithms based on the sequence recognition rules. Therefore, the mesophilic TargeTron technology based on Group IIA intron Ll. LtrB as well as thermo-TargeTron technology based on group IIB intron TeI3c and TeI4c have been introduced into many clostridia.

4 Development of TargeTron in *Clostridium*

Owing to the advantages of independently on homologous recombination, low requirements for transformation efficiency, transformants screening by colony PCR, operation convenience, time-saving, stabilization and reliability, TargeTron technology was greatly welcomed as soon as it appeared and was quickly adapted to other *Clostridium* species. We summarized the study milestones of TargeTron development, improvement and application, as shown in Figure 4.

Minton and colleagues have carried out a lot of basic and groundbreaking research work in TargeTron including technology development and improvement, genetic elements standardization, and websites development for insertion site prediction of intron RNA (http://www.clostron.com). They introduced TargeTron technology to *C. acetobutylicum, C. difficile, C. botulinum,* and *C. sporogenes* for the first time [16]. To avoid off-target, they constructed intron plasmids containing RAM (retrotransposition-activated marker) [16] . In another study, they extended TargeTron to *C.*

beijerinckii, and improved the RAM by adding FLP recombinase recognition target (FRT) to the intron RNA for the purpose of eliminating selectable marker using flipase [40]. Besides, they expanded function of TargeTron from disruption to targeted integration. Hybrid intron RNA composing of heterologous phage DNA was successfully constructed and integrated into the host's genome [40].

Almost simultaneously with Minton and colleagues, Shao et al. independently developed the application of TargeTron in *C. acetobutylicum* ATCC824 [15]. Subsequently, Hu et al. and Wen et al. respectively adapted Shao's TargeTron system to *C. acetobutylicum* EA2018 and *C. cellulovorans* DSM 743B [21, 42]. More recently, Wen et al. developed a genetic system of *C. diolis* DSM 15140 and extends this technology to the strain (data unpublished) , which naturally utilizes glycerol to produce 1,3-propanediol with high titer [43].

TargeTron technology also are applied in other solventogenic clostridia such as *C. phytofermentans*[44], *C. butylicum* [45] and *C. tyrobutyricum* [46]. Recently, this technology has been extended to cellulolytic clostridia including *C. cellulolyticum* [47], *C. cellulovorans* [21, 48], and gas-fermenting clostridia *C. autoethanogenum* [49, 50]. Besides, TargeTron technology specifically dedicated to thermophilic clostridia (for example, *C. thermocellum*) is developed by Mohr et al. [17], which avoids limited efficiency of mesophilic TargeTron in thermophiles.

The diversity of *Clostridium* lead to diverse applications of TargeTron. It has been used in pathogenic *Clostridium* such as *C. difficile, C. botulinum,* and *C. sporogenes* [16, 40, 51], and industrial microorganism such as solventogenic, cellulolytic, and gas-fermenting clostridia [9, 10, 18]. Particularly, many progresses have been made in the enhancement of solvents-producing pathways, the xylose metabolic pathways reconstruction, identification of functional genes such as resistance genes, global or local regulatory factors, and so on [21, 46, 48, 52-60]. Basically, TargeTron provides an efficient approach and paved the way for complex genetic modification in many clostridia.

5 Application of TargeTron technology in solventogenic clostridia

Clostridial metabolic pathway inactivation and functional gene identification mainly relied on single exchange and in-frame deletion based on homologous recombination before TargeTron technology was developed. The efficiency of electroporation or conjugative transformation is very low because of the thick cell wall and complicated restriction & modification system (R&M system) of *Clostridium* [9, 61]. What's worse, the inefficient homologous recombination capacity makes the allelic-exchange time-consuming and laborious to obtain positive transformants. The TargeTron technology almost completely rewrote the pattern and greatly accelerated the research progress. TargeTron has been used for genetic studies of various medical and industrial microorganisms, especially in solventogenic clostridia, because they have great potential in commodity chemicals and biofuels production.

The solventogenic clostridia, such as *C. acetobutylicum* [62], *C. beijerinckii* [63], etc., can use a wide range of substrates for acetone-butanol-ethanol (ABE) fermentation, which is a typical bi-phase fermentation [64, 65]. At the acidogenic phase, the main products are acetic acid and butyric acid; At

the solventogenic phase, the organic acid accumulated at the acidogenic phase are re-assimilated and converted into solvents such as ethanol, butanol and acetone [66]. Except for CoA-dependent ABE synthesis pathway, glycolytic pathway, pentose phosphate pathway and phosphoketolase pathway are also critical to ABE synthesis. Besides, ABE fermentation is usually accompanied by physiological phenomena such as sporulation, quorum sensing and solvents tolerance, thus some global or specific regulatory factors, and genes responsible for sporulation and stress resistance involve in ABE fermentation [66-68]. To enhance ABE output, it is necessary to identify the function of key genes in prior to metabolic engineering.

Taking common solventogenic clostridia (*C. acetobutylicum* and *C. beijerinckii*) for example, we briefly depict the application of TargeTron technology in ABE metabolic pathway modification, regulatory genes function identification, pentose metabolic pathways analysis and reconstruction.

5.1 ABE pathway modification

TargeTron provides an effective gene disruption approach for Clostridium. It was firstly applied to inactivate the acetate or butyrate synthesis pathways in order to eliminate by-products, and then gradually extended to other metabolic pathways to increase solvents titer and yield. Some genes of ABE pathway such as *adhE1, adhE2, bdhA, bdhB, ptb, ack, pta, ack, adc, ctfA/B* and *hydA* in *C. acetobutylicum* or *C. beijerinckii* have been disrupted by TargeTron to study their roles on ABE production (Figure 5).

Figure 5 and Table 2 depict the genes disrupted in the ABE metabolic pathway along with the phenotypic changes of the mutants. *Pta, ptb, ack*, and *buk* are genes involved in the acetate and butyrate synthesis, which are usually inactivated to reduce acid production and drive more carbon source to ethanol/butanol [15, 54, 64, 69-72].

Adc and *ctfA*, *ctfB* are essential genes for acetone production; *ctfA* and *ctfB* also catalyze the re-assimilation of acetic acid and butyric acid into acetyl-CoA and butyryl-CoA. The *ctfA/B* deficient mutant produced less amount of acetone and solvents, and took up less acids [54]. Interrupting *adc* did not completely abolish acetone formation, which may be attributed to the spontaneous decarboxylation of acetoacetate [54]. The dramatically high ratio of acetic acid production by the *ctfA/B* deficient mutant to wild type indicated that *ctfA/B* gene played a major role in acetic acid re-assimilation, while butyric acid is mainly generated from the reaction catalyzed by *ptb* and *buk* gene [64, 69].

AdhE1, adhE2, bdhA, bdhB are critical enzymes for alcohol pathways [73, 74]. The solvent production of *AdhE1* interrupted mutant was nearly eradicated, while the *adhE2, bdhA, bdhB* interrupted mutant unchanged. Interestingly, in the *adhE1* deficient mutant, *adhE2* can compensate for the lack of *adhE1* to some extent [73].

In addition to metabolic pathway genes, some genes involved in reducing power consumption and regeneration have also been interrupted, attempting to reshape the driving force for ABE fermentation. Jang et al. successfully knocked out the *hydA* gene in a *pta-buk-ctfAB* deficient strain. It is the only report about *hydA* gene disruption by TargeTron [58]. In another study, Liu et al. identified

a *nuoG* (NADH-quinone oxidoreductase) encoding gene *Cbei_4410* in *C. beijerinckii* [65]. The intron mutant harbored more reducing power, leading to higher glucose consumption rates and butanol production than the wild type [65].

Generally, TargeTron is highly effective in reducing by-product formation and increasing the solvent production. However, there are also some unexpected phenotypes reported. For example, knocking out *ptb* almost completely eliminates butyric acid production [54, 69], but does not certainly lead to an increase in butanol. The *ptb* deficient mutants produce ethanol as main product, but butanol formation is comparable to the wild type [69]. It is difficult to determine whether these unpredictable phenotypes are due to gene deletions or to the off-target or polar effects of TargeTron. Off -target sometimes occurred in TargeTron. Cooksley et al. confirmed that knockout of *ptb* in *C. acetobutylicum* ATCC 824 caused frameshift mutation of *thl* gene, affecting the transformation of acetyl-CoA to butyryl CoA, the precursor of butanol [54]. In another study, knocking out *buk1* caused partial deletion of the upstream *ptb* gene [60]. Therefore, it is necessary to confirm the uniqueness of the intron insertion by southern blotting, reverse PCR sequencing or even genome re-sequencing [54]. In addition, gene complementation analysis, as well as experimental design to avoid polar effects, are also essential [11].

Another issue should be noted is that inactivation efficiency of TargeTron are genes specific. Some genes such as *ack* and *adc* are easily to be inactivated in *C. beijerinckii* and *C. acetobutylicum*. However, it is very difficult to obtain the mutants with inactivated *ptb* or *hydA* using TargeTron [58]. Moreover, some genes like *thl*, *bcd* or *crt* in the CoA dependent (carbon chain elongation) pathway, have never been successfully knocked out by TargeTron to date, except that some unexpected non-specific insertion (off-target) or frame-shift occurred occasionally [54, 60]. These suggested some putative drawbacks in TargeTron.

5.2 Identification of functional genes

To adapt to the complex and hostile environment, the physiology and metabolism of solventogenic clostridia are strictly and complicatedly regulated by a variety of transcription factors or two-component systems [67]. They are very good model microorganism to study the function of regulating network. Because comparative genomics is not yet able to accurately predict all genes, the function of some key genes still needs to be confirmed by integrated genetic and biochemical technologies. TargeTron are commonly adapted for gene function identification due to its simple operation and reliable results. The identification process is to interrupt putative functional genes and then observe change of physiology or metabolism phenotype. Subsequently, gene complementation is performed to recovery the phenotype. Finally, some other genetic and biochemical tests were combined to confirm the function of the gene [75-78]. TargeTron provides important clues for revealing the regulatory factors function by disrupting putative regulatory genes and thereby resulting in phenotypic changes.

ccpA gene encodes a catabolite control protein (CCP), which can regulate carbon catabolite repress (CCR) by binding to specific sequence or sites (binding motif or cre site) of its target genes [42, 52, 59]. CCP can regulate (mainly depress) expression of genes involved in xylose metabolism

[79]. Inactivation of *ccpA* can release its inhibition of some xylose metabolic pathway genes, thereby promoting xylose metabolism.

The *CAC3037* gene in *C. acetobutylicum* is considered a putative *ccpA* gene, but its function has not been confirmed. Ren et al. [52] and Yang et al. [80] investigated the growth, sugar utilization and solvent-producing phenotype of *ccpA* disrupted strain. These phenotypic changes supplied important clues for further analysis of CcpA regulatory functions, and even identification of nucleotide targets and binding motifs. Above method also was applied to the functional analysis of other important regulatory genes, such as *Rex* (*CAC2713*) [77, 81], *celR* (*CAC0382*) [82], *AbrB* (*AbrB0310* and *AbrB3647*) [83], *csrA* (*CAC2209*) [84], etc. Besides, TargeTron plays a key role in identification of some other important functional genes such as the stress resistance genes *CAC3319* and *Cbei3304* [85, 86] or the quorum sensing gene *agr ABC* (*CAC0078-0081*) [87]. Generally, TargeTron has become an important genetic tool for gene function identification.

5.3 Analysis and reconstruction of xylose metabolic pathway

Solventogenic clostridia cannot utilize pentose efficiently due to the carbon catabolite repression (CCR) [88-90], which seriously reduce the solvent yield from lignocellulosic biomass [91], because pentose proportion (mainly xylose and arabinose) is second only to glucose in lignocellulosic hydrolysates [92]. It is thus very necessary to improve the efficiency of pentose uptake and utilization. The first step towards that goal is to understand transporters, metabolic pathways and metabolic regulation of xylose and arabinose in *Clostridium*, and then to modify them, as shown in Figure 6.

Gu et al [76] identified the xylose isomerase gene (*xylA-II, CAC2610*, xylose isomerase) and xylose kinase gene (*xylB, CAC2612*) functions in *C. acetobutylicum* using TargeTron combined with other genetic and biochemical methods. Based on their previous studies and comparative genomics predictions, xylose metabolic pathway was reconstructed in *C. acetobutylicum* and even extended to Firmicutes. Subsequently, Hu et al. [42] and Xiao et al. [93] adapted the strategy and identified putative transcriptional regulator XylR like *CEA_G2622* and *Cbei_2385* involved in xylose metabolism in *C. acetobutylicum* EA 2018 and *C. beijerinckii* NCIMB 8052 respectively. Recently, Sun et al. [94] demonstrated a hexa-protein module XylFII-LytS/YesN-XylFGH associated with xylose utilization in *C. beijerinckii.* They further confirmed that it is a new "three-component" xylose response and regulation systems whose molecular mechanisms is then deeply analyzed.

Except for pentose phosphate pathway, phosphoketolase pathway also contributes to pentose metabolism. Liu et al. [95] revealed how *C. acetobutylicum* metabolizes xylose with both PPP and PK pathway by isotope tracing and TargeTron techniques. They knocked out the *xfp* (*CAC1343*, encoding xylulose-5-P/fructose-6-P phosphoketolase) gene by TargeTron and unexpectedly found no significant changes in xylose metabolism. Interestingly, this gene is closely related to the phosphoketolase metabolic pathway of arabinose, which was verified by Servinsky et al. [96]. In another study, Zhang et al. [78] analyzed the functions of arabinose metabolism regulation genes *araR* (*CAC_1340*) and ribulose kinase *araK* (*CAC_1344*) in *C. acetobutylicum* with aid of TargeTron, and then reconstructed arabinose metabolic pathway.

The pentose metabolism-related functional genes mentioned above basically mapped the framework of the pentose transport, metabolism and regulation in solventogenic clostridia (Fig. 6). It provides targets for subsequent metabolic engineering, which also can be performed via TargeTron.

5.4 Other applications

TargeTron has also been adapted to interrupt restriction & modification system (R&M system) to increase plasmids transformation efficiency. The R&M system acts like the immune system in prokaryotes and differs in different *Clostridium* species [9, 97]. The restriction endonuclease in this system would recognize and degrade the invading DNA without proper unmethylation [98]. Therefore, it is necessary to perform plasmids pre-methylation, although it is time-consuming and tedious to some extent. The R&M system in *Clostridium* has been an obstacle for plasmids transformation with high efficiency. Identification and inactivation of the restriction endonuclease genes in the R&M system can greatly improve the transformation efficiency. Dong et al. confirmed and knocked out gene *CAC1502* encoding the type II restriction endonuclease *Cac824I* in *C. acetobutylicum*. The mutant could accept unmethylated plasmid as efficient as methylated plasmid [99]. Similarly, in *C. celluloyticum*, inactivating *MspI* gene that encodes endonuclease allowed an efficient transformation of unmethylated TargeTron plasmid to the *MspI*-deficient mutant. The genome of the resulting double knockout mutant remained stable after 100 generations [47].

Due to easy operation as well as high reliability, TargeTron is now a primary genetic tool for gene disruption in *Clostridium*. Currently, the application range of TargeTron has gradually been extended from the ABE metabolic pathway to other engineering fields like xylose metabolism modification and cofactor regulation. TargeTron can be used not only for the identification of unknown genes function, but also for the re-cognization of the known genes.

6 Functional expansion and performance improvement of TargeTron in *Clostridium*

The wide development of TargeTron highlights its important value in industry and academia. However, there have been still drawbacks, some of which originate from its working principle or mechanism. For example, intron insertion may inactivate the gene downstream the integration site, which is usually called polar effects [11]. In addition, TargeTron can only inactivate the target gene, but cannot delete the entire OFR, thus not completely knock-out the gene. Off-target is another drawback resulting from the continuous expression of TargeTron elements or sequence similarity [14]. Therefore, many efforts have been made to improve TargeTron technology in the aspects of reducing off-target [16, 100] and polarity effects [40, 101], exogenous DNA insertion [40], plasmid curing [102], marker recycle [40], etc. to expand the function or eliminate the disadvantages (Table 3).

6.1 Exogenous DNA integration

It is well acknowledged that TargeTron is mainly used for gene inactivation. Less noticed is that the stem-loop structure in DIV of intron RNA can carry exogenous DNA. It indicates that TargeTron can also serve as a vector used for heterologous gene expression [38, 103]. In 2007, Chen et al. integrated intron carrying the *SIV P27* gene into open reading frame (ORF) of the *pfoA* gene in *C. perfringens* to produce antibodies targeted to the Gut-Associated Lymphoid Tissue. The engineered *C. perfringens* caused no antibiotic resistance, because antibiotic was not indispensable to screen mutants when TargeTron is adapted [103]. In another study, Heap et al. disrupted the *pyrE* gene by inserting intron RNA harboring λ phage DNA. Although the integration efficiency is 10^{-4} of that of the unmodified TargeTron, the feasibility of the method is confirmed. It is important to notice that the size of cargo DNA should be less than 1 kb [40].

6.2 off-target

As previously mentioned, the insertion efficiency of the TargeTron is high enough to screen mutant without marker. However, it also results in a putative disadvantage of off-target due to sequence similarity or continuous expression of TargeTron elements [16, 100]. To decrease non-specific insertion, Heap et al. introduced a retrotransposition-activated marker (RAM), which is a transposition activation tag which consists of group II introns, resistance genes and a self-cleavage group Ⅰ introns [16]. Before group II introns work, group Ⅰ introns insert into the resistance gene. Once group II introns inserted into the target site, group Ⅰ introns self-splice from the resistance gene. Positive mutants can then be screened by the corresponding antibiotic [16]. The strategy can efficiently filter out non-specific insertions thus reducing off-target frequency.

Another strategy to decrease off-target frequency is down-regulating transcription of intron RNA and IEP [100]. Zhang et al. developed an arabinose-induced TargeTron system ARAi for *C. cellulolyticum*, which composed of an AraR operon from *C. acetobutylicum* and an anti-selection marker mazE/F. The inducible TargeTron system was designed to target *mspI* and *cipC* genes. Transformants were cultured in a non-resistant medium, and arabinose was added to control the expression of intron RNA and IEP. The result showed that the off-target rate was greatly reduced from 100% to 0.

6.3 Avoiding unexpected genotype

In addition to off-target, TargeTron also accidentally causes some unexpected genotypes and phenotypes, such as incomplete phenotype defects, phenotypic recovery, polar effects and large DNA fragment deletions. Homologous recombination tools have been integrated with TargeTron to avoid these potential unexpected phenotypes.

According the specific site of insertion in the locus, gene disruption caused by TargeTron is divided into conditional and unconditional inactivation [40, 104]. When intron RNA is inserted into

the top or sense strand of the target gene, the promoter of the target gene may drive the transcription of the intron RNA, and the intron may be removed from the mRNA due to self-cleavage with the aid of the (endogenous or exogenously introduced) RT enzyme, which may result in conditional transcription and translation of the target gene; however, if the intron RNA is inserted at the end of the target gene or the antisense strand, the mRNA is unconditionally inactive, because it contains an intron that cannot be removed. For conditional inactivation, phenotypic iteration may occur. This situation can be avoided by inserting the intron to the antisense strand of target DNA. However, the insertion efficiency is sometimes lower than inserting into the sense strand [18, 40, 104].

The choice of insertion site is critical to avoid conditional interruptions. The locus at gene tail in sense strand, or in antisense strand, is usually preferred [18, 40, 104]. Depressing expression level of RT can also reduce the frequency of phenotypic recovery in conditional knockouts, but it may reduce knockout efficiency [100]. In addition, some homologous recombination tools were introduced to remove all or part of inserted intron RNA [18], making self-splicing ineffective. Recently, Enyeart et al. introduced Cre/lox system for deletion of partial introns RNA [105]. The lox site is designed in DIV of intron RNA, and then integrated with the intron to the host chromosome. The intron sequence is removed after induced expression of Cre recombinase. This strategy has been proved in *Escherichia coli*, but has not been verified in Clostridium.

TargeTron can disrupt transcription, but do not cause deletion of the target gene, whose function may not be completely eliminated and even may be recovered in the case of conditional knockout [18, 19, 33]. Besides, TargeTron leaves a “scar” (the inserted intron RNA) in the host chromosome [40]. A potential drawback is that it may cause a polar effect if the target gene is part of a co-transcribed gene cluster (or operon) [9-11, 18]. The intron inserted in upstream genes may affect expression of downstream genes [40, 101]. Basically, bringing in a native promoter in domain IV of the intron RNA may relieve some of the polar effects [11, 19]; another approach is cloning the operon (without the target gene) into a plasmid to compensate for any polar effect [11].

Another putative risk is that homologous recombination may occur between intron RNAs in mutants that have undergone multiple rounds modifications by TargeTron. That may result in a deletion of large DNA fragment within two intron RNAs [101]. Although the probability is very low, it is still necessary to delete part of target genes and remove the "scar". Jia et al. [101] developed the group II intron-anchored gene deletion method by combining homologous recombination with the TargeTron technology. The resistance gene and the target gene homology arm were designed to be part of the intron RNA, and were inserted into the genome site-specifically. Then, colony PCR was performed to screen for mutants generated from homologous arm exchange, during which, the target gene sequence and part of the intron RNA sequence were deleted. In this way, they successfully deleted the *CAC1493-1494* and *ctfAB* genes of *C. acetobutylicum* [101]. This strategy achieves the deletion of target genes. Besides, it only left intron RNA with very small size in genome, which significantly reduces the possibility of homologous recombination between intron RNAs. Compared to the classical allelic-exchange, the strategy of introducing a homologous arm to the genome by TargeTron greatly improved the screening efficiency of the mutant with single-exchange, accelerating the acquisition of the mutant with "double-exchange", although it still cannot achieve a marker-free knockout.

Flipase recognition target (FRT) and FLP recombinase genes were also introduced into TargeTron system for recycle of resistance markers. Heap et al. for the first time improved the *ermB* RAM [40], which was flanked with two repeated FLP (flipase) recognition target (FRT) sites. The plasmid harboring the FLP recombinase gene was transferred to the first intron inserted mutant to remove the *ermB* marker. Using this method, an *agrA/spo0A* double mutant was successfully obtained.

6.4 Plasmids curing

Vectors carrying TargeTron elements are commonly replicating plasmids, which are stable in the host under selective pressure. It facilitates TargeTron work and thereby improves the insertion efficiency. Once a positive mutant obtained, the plasmids need to be cured, because they may hinder plasmids used for next round operation due to plasmid incompatibility and cause genotype recovery via RT expression. Conventional plasmid curing is performed by sub-culturing the mutant in a non-resistant medium until the plasmid is spontaneously discarded [15]. However, this method requires a lot of screening work, which is very time-consuming and laborious. In addition, although the reason is still unclear, according to our experience, plasmid curing become more difficult when the host has undergone successive gene disruption.

To overcome the barrier, Cui et al. developed a *pyrF*-based screening method for plasmid curing [102]. *PyrF* is an important gene in pyrimidine synthesis pathway and a target for the toxic substance 5-fluoroorotic acid (FOA). A *pyrF* gene-deficient mutant by homologous recombination was adapted as chassis cell, which can resist 5-FOA. A TargeTron plasmid harboring *pyrF* gene cassette was transformed to target genes of interest. Once the positive mutant was confirmed, it was transferred to agar medium with 5-FOA as a selective pressure. Strains that have cured plasmid can be easily screened. Successive gene disruption of *Cel48F* or *CipC*, as well as efficient plasmid curing were achieved in *C. cellulolyticum*. Similarly, some other counter-screening markers, such as *tdc*, *mazF* etc. [106, 107], can also be introduced into the TargeTron's plasmid system to establish screening system for successive TargeTron modification. Moreover, the principle of plasmid incompatibility can also be introduced to cure plasmid according to our experimental results in *C. beijerinckii* (data unpublished).

Although TargeTron has made great progress in the function extension and efficiency improvement by combining with new synthetic biological tools such as recombinases, counter selection markers, inducible promoters, etc., it must be acknowledged that TargeTron can still be improved. For example, an algorithm based on intron RNA recognition rule (http://www.targetrons.com/) can provide multiple possible insertion sites and rank their corresponding intron sequences to guide TargeTron design. The ranking, however, is not absolutely reliable. Therefore, in addition to the optimal recommended insertion site, several alternative insertion sites are usually attempted until a desired mutant is obtained [40]. Although the off-target rate of TargeTron has been greatly reduced in recent years, confirmatory analysis such as gene complementation, southern blotting, reverse PCR sequencing, or even genome resequencing is still indispensable to exclude non-site-specific insertion [54]. Generally, alternative plan and rigorous verification are necessary guarantee for TargeTron in practice.

7 Emergence of new technologies, competitors or collaborators?

Although TargeTron has been greatly improved and widely used in clostridia, it cannot deal with all cases of genetic manipulations. Due to the great value of *Clostridium* in medicine and industry, various efforts have been made to outline a technical guide or roadmap for genetic system development [8, 9], or to enrich synthetic biology toolkit [108], which have already been extensively reviewed [7, 10, 109]. Table 1 summarized several mainstream genetic manipulation tools applicable in clostridia. Their principles and characteristics are compared to each other. The characteristics and application scenarios of these genetic tools are closely related to DNA repair methods (homologous or non-homologous recombination), gene targeting (random or precise), genetic modification results (scar-less and scar-residue), and purpose (gene inactivation or heterologous gene expression). These make them differ in the aspects of plasmids design, plasmids construction (operation), mutant screening, experimental cycle, targeting efficiency, unexpected genotype and application.

Generally speaking, compared to other genetic tools, TargeTron is superior in many aspects except the unexpected phenotype. However, some other technologies are complementary to TargeTron in function. For example, RNA interference as well as the recent emergence of CRISPR/dCas9 and small RNA-based gene down-regulation techniques provide approaches to inhibit gene expression at the transcriptional level in clostridium, which is what TargeTron can't achieve. As for the transposon, it has advantages in constructing a randomly mutagenized library. although it cannot be adapted to target genes of interest [110-112].

In addition, TargeTron can only modify genes by insertion of intron, while allelic-coupled exchange via single/double-crossover is a precise genetic tool based on homologous recombination [106, 107, 113]. It can carry out in-frame deletion, insertion or point mutation. The only drawback is that there is a very low frequency to obtain the positive single-crossover or double-crossover mutants, due to inefficient homologous recombination and low plasmids transformation efficiency in *Clostridium*; Although counter-selection markers such as *mazF* and *tdc* [106, 107] as well as endonuclease I-SceI [74] were applied to facilitate the screening of mutants yield in the second exchange, it still suffers a long operation cycle and low first single-crossover efficiency [113-115]. CRISPR-Cas system has been adapted to increase the frequency of mutants with single or double-crossover [22, 89, 116]. DNA double strands break created by Cas9 or DNA single strand break created by Cas9 nicakse can filter out most false transformants that did not complete DNA repair by homologous recombination (HR) or non-homologous end joining (NHEJ). CRISPR-Cas system is much more powerful than allelic-coupled exchange, but it still relies on homologous recombination capacity of host and sometimes leads much lower transformation efficiency [22, 89, 116]. In contrast to the CRISPR-Cas9 system, TargeTron is not toxic to the host and has almost no requirement of plasmids transformation efficiency.

More recently, Li. et al. developed a base editing system based on cytidine deaminase (rat Apobec1) and uracil DNA glycosylase inhibitor (UGI) as well as CRISPR-Cas9^{D10A} system [20]. Apobec1 and UGI can efficiently convert specific C·G nucleotide base pairs in the CRISPR-Cas9 targeting window sequence to T·A, which can create a missense mutation or null mutations in a gene. It is as precise as Cas9- mediated genome editing, but does not have to cleavage DNA. Therefore, it

requires no DNA repair templates. It is very easy to design and theoretically programmable to target all genes in Clostridium. In preparing the paper, Strecker et al. developed a RNA-guided DNA insertion method with CRISPR-associated transposases in *Escherichia coli* [117], which can be introduced to genus *Clostridium* and facilitate the development of programmable genetic manipulation tools based on non-homologous recombination. As an emerging gene editing technique, CRISPR-Cas system has shown its great potential in clostridial genetic manipulation, but the homologous recombination capacity of *Clostridium* itself still needs further be improved by introducing genetic parts like RecT [69].

Some of above genetic tools present alternatives to the TargeTron, while others complement it by presenting enhanced abilities for specific purposes. We tend to think that they are not competitors, but collaborators. They jointly promote the engineering and study of *Clostridium*. As an example, in the recent work of Wen and colleagues [21], CRISPR/dCas9 and TargeTron were combined to increase butyrate production in *C. cellulovorans*. TargeTron was used to knock out *ack* and *ldh* to save carbon towards butyrate, while dCas9 was adapted to down-regulate expression level of an essential gene *hydA* to reduce reducing power loss, which can almost impossibly be achieved via TargeTron. In another recent study, an efficient site-specific chromosomal integration through a dual integrase cassette exchange strategy was developed. In this approach, the CRISPR/Cas9 system has been used to insert *attB* sites in the chromosome to facilitate integrase recognition [118]. Alternatively, for some *Clostridium* without available CRISPR/Cas9 system, the insertion of *attB* sites can be accomplished by TargeTron, which may greatly expand the scope of the chromosomal integration tool. Generally, TargeTron is still one of the most powerful genetic manipulation tools specifically dedicated to *Clostridium*. The combination of TargeTron with other genetic tools can complement each other, and together foster complex genetic modification in *Clostridium* more effectively.

8 Conclusion and outlook

The TargeTron technology based on mobility of group II introns is a particularly suitable technology to engineer *Clostridium*. It is easy and efficient to design and operate, and the mutant can be directly identified by colony PCR. Meanwhile, some putative drawbacks of TargeTron, such as low-frequent off-target, polar effects, unexpected genotype, etc., have been partially overcome by introducing inducible promoters, counter-screening markers, or recombinase systems. The development and improvement of TargeTron technology has greatly facilitated metabolic pathway modification and gene function identification in solventogenic *Clostridium*.

Although emerging genetic manipulation tools such as the CRISPR/Cas9 system and its assisted base editing systems or gene expression regulation systems pose powerful alternatives, TargeTron still has advantages in some aspects. From the perspective of gene manipulation, they are collaborators, rather than competitors to each other. It is foreseeable that the combination and innovation of TargeTron technology with other synthetic biology tools as well as genetic components will become a trend to serve a better clostridial study.

Acknowledgments

This work is supported by the grant from the National Natural Science Foundation of China (21706133, 21825804, 31670094, 31971343), the Funds for Creative Research Groups of China (31921006), the Fundamental Research Funds for the Central Universities (30918011310), Key Laboratory of Biomass Chemical Engineering of Ministry of Education, Zhejiang University (2018BCE003).

Conflict of interest

The authors declare no financial or commercial conflict of interest.

9 References

[1] Dürre, P., *Handbook on clostridia*, CRC press 2005.

[2] Wiegel, J., Tanner, R., Rainey, F. A., *An Introduction to the Family Clostridiaceae*, 2006.

[3] Cho, C., Jang, Y.-S., Moon, H. G., Lee, J., Lee, S. Y., Metabolic engineering of clostridia for the production of chemicals. *Biofuels, Bioproducts and Biorefining* 2015, *9*, 211-225.

[4] Ren, C., Wen, Z., Xu, Y., Jiang, W., Gu, Y., Clostridia: a flexible microbial platform for the production of alcohols. *Current opinion in chemical biology* 2016, *35*, 65-72.

[5] Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C., Papoutsakis, E. T., Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Current Opinion in Biotechnology* 2012, *23*, 364-381.

[6] Hatheway, C. L., Toxigenic clostridia. *Clinical Microbiology Reviews* 1990, *3*, 66.

[7] Shen, A., Expanding the Clostridioides difficile Genetics Toolbox. *Journal of Bacteriology* 2019, *201*, e00089-00019.

[8] Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., *et al.*, A roadmap for gene system development in Clostridium. *Anaerobe* 2016, *41*, 104-112.

[9] Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A., Chou, C. P., Technical guide for genetic advancement of underdeveloped and intractable Clostridium. *Biotechnology advances* 2014, *32*, 623-641.

[10] Joseph, R. C., Kim, N. M., Sandoval, N. R., Recent Developments of the Synthetic Biology Toolkit for Clostridium. *Frontiers in Microbiology* 2018, *9*.

[11] Cui, X., Davis, G., Mobile group II intron targeting: applications in prokaryotes and perspectives in eukaryotes. *Frontiers in Bioscience* 2007, *12*, 4972-4985.

[12] Rodriguez, S. A., Davis, G., Klose, K. E., Targeted gene disruption in Francisella tularensis by group II introns. *Methods* 2009, *49*, 270-274.

[13] Karberg, M., Guo, H. T., Zhong, J., Coon, R., *et al.*, Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol* 2001, *19*, 1162-1167.

[14] Lehmann, K., Schmidt, U., Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol* 2003, *38*, 249-303.

[15] Shao, L., Hu, S., Yang, Y., Gu, Y., *et al.*, Targeted gene disruption by use of a group II intron (targetron) vector in Clostridium acetobutylicum. *Cell Research* 2007, *17*, 963-965.

[16] Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., Minton, N. P., The ClosTron: A universal gene knock-out system for the genus Clostridium. *Journal of Microbiological Methods* 2007, *70*, 452-464.

[17] Mohr, G., Hong, W., Zhang, J., Cui, G.-z., *et al.*, A Targetron System for Gene Targeting in Thermophiles and Its Application in Clostridium thermocellum. *Plos One* 2013, *8*.

[18] Liu, Y.-J., Zhang, J., Cui, G.-Z., Cui, Q., Current progress of targetron technology: Development, improvement and application in metabolic engineering. *Biotechnology Journal* 2015, *10*, 855-865.

[19] Enyeart, P. J., Mohr, G., Ellington, A. D., Lambowitz, A. M., Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mobile DNA* 2014, *5*.

[20] Li, Q., Seys, F. M., Minton, N. P., Yang, J., *et al.*, CRISPR-Cas9(D10A) nickase-assisted base editing in the solvent producer Clostridium beijerinckii. *Biotechnology and bioengineering* 2019.

[21] Wen, Z., Minton, N. P., Zhang, Y., Li, Q., *et al.*, Enhanced solvent production by metabolic engineering of a twin-clostridial consortium. *Metabolic engineering* 2017, *39*, 38-48.

[22] Li, Q., Chen, J., Minton, N. P., Zhang, Y., *et al.*, CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnol J* 2016, *11*, 961-972.

[23] Jing, X., Xie, B., Chen, L., Zhang, N., *et al.*, Implementation of the CRISPR-Cas13a system in fission yeast and its repurposing for precise RNA editing. *Nucleic Acids Research* 2018.

[24] Candales, M. A., Adrian, D., Hood, K. S., Tony, L., *et al.*, Database for bacterial group II introns. *Nucleic Acids Research* 2012, *40*, D187.

[25] Toor, N., Hausner, G., Zimmerly, S., Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *Rna* 2001, *7*, 1142-1152.

[26] Navtej, T., Keating, K. S., Taylor, S. D., Anna Marie, P., Crystal structure of a self-spliced group II intron. *Science* 2008, *320*, 77-82.

[27] Blocker, F. J., Mohr Gconlan, L. H., Qi, L., Belfort, M., Lambowitz, A. M., Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *Rna-a Publication of the Rna Society* 2005, *11*, 14.

[28] Yang, J., Zimmerly, S., Perlman, P. S., Lambowitz, A. M., Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* 1996, *381*, 332-335.

[29] Mohr, G., Smith, D., Belfort, M., Lambowitz, A. M., Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes & Development* 2000, *14*, 559.

[30] Singh, N. N., Lambowitz, A. M., Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference 1 ☆. *Journal of Molecular Biology* 2001, *309*, 0-386.

[31] Cui, X., Matsuura, M., Wang, Q., Ma, H., Lambowitz, A. M., A Group II Intron-encoded Maturase Functions Preferentially In Cis and Requires Both the Reverse Transcriptase and X Domains to Promote RNA Splicing. *Journal of Molecular Biology* 2004, *340*, 0-231.

[32] Joseph, S. F., Lambowitz, A. M., Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *Journal of Molecular Biology* 2002, *324*, 933-951.

[33] Lambowitz, A. M., Zimmerly, S., Group II Introns: Mobile Ribozymes that Invade DNA. *Csh Perspect Biol* 2011, *3*.

[34] Costa, M., Michel, F., Westhof, E., A three-dimensional perspective on exon binding by a group II self-splicing intron. *Embo Journal* 2000, *19*, 5007–5018.

[35] Navtej, T., Robart, A. R., Joshua, C., Steven, Z., Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Research* 2006, *34*, 6461-6471.

[36] Yao, J., Zhong, J., Lambowitz, A. M., Gene targeting using randomly inserted group II introns (targetrons) recovered from an Escherichia coli gene disruption library. *Nucleic Acids Research* 2005, *33*, 3351-3362.

[37] Yao, J., Zhong, J., Fang, Y., Geisinger, E.*, et al.*, Use of targetrons to disrupt essential and nonessential genes in Staphylococcus aureus reveals temperature sensitivity of Ll.LtrB group II intron splicing. *Rna* 2006, *12*, 1271-1281.

[38] Plante, I., Cousineau, B., Restriction for gene insertion within the Lactococcus lactis Ll.LtrB group II intron. *Rna* 2006, *12*, 1980-1992.

[39] Zhong, J., Karberg, M., Lambowitz, A. M., Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research* 2003, *31*, 1656-1664.

[40] Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T.*, et al.*, The ClosTron: Mutagenesis in Clostridium refined and streamlined. *Journal of Microbiological Methods* 2010, *80*, 49-55.

[41] Mohr, G., Ghanem, E., Lambowitz, A. M., Mechanisms used for genomic proliferation by thermophilic group II introns. *PLoS Biol* 2010, *8*, e1000391.

[42] Hu, S., Zheng, H., Gu, Y., Zhao, J.*, et al.*, Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in Clostridium acetobutylicum EA 2018. *Bmc Genomics* 2011, *12*.

[43] Otte, B., Grunwaldt, E., Mahmoud, O., Jennewein, S., Genome Shuffling in Clostridium diolis DSM 15410 for Improved 1,3-Propanediol Production. *Applied and Environmental Microbiology* 2009, *75*, 7610-7616.

[44] Tolonen, A. C., Chilaka, A. C., Church, G. M., Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Molecular Microbiology* 2009, *74*, 1300-1313.

[45] Cai, G., Jin, B., Saint, C., Monis, P., Genetic manipulation of butyrate formation pathways in Clostridium butyricum. *Journal of Biotechnology* 2011, *155*, 269-274.

[46] Yu, M. R., Zhang, Y. L., Tang, I. C., Yang, S. T., Metabolic engineering of Clostridium tyrobutyricum for n-butanol production. *Metabolic engineering* 2011, *13*, 373-382.

[47] Cui, G.-z., Hong, W., Zhang, J., Li, W.-l., *et al.*, Targeted gene engineering in Clostridium cellulolyticum H10 without methylation. *Journal of Microbiological Methods* 2012, *89*, 201-208.

[48] Wen, Z., Ledesma-Amaro, R., Lin, J., Jiang, Y., Yang, S., Improved n-Butanol Production from Clostridium cellulovorans by Integrated Metabolic and Evolutionary Engineering. *Applied and Environmental Microbiology* 2019, *85*.

[49] Liew, F., Henstra, A. M., Kopke, M., Winzer, K., *et al.*, Metabolic engineering of Clostridium autoethanogenum for selective alcohol production. *Metabolic engineering* 2017, *40*, 104-114.

[50] Liew, F., Henstra, A. M., Winzer, K., Kopke, M., *et al.*, Insights into CO2 Fixation Pathway of Clostridium autoethanogenum by Targeted Mutagenesis. *mBio* 2016, *7*.

[51] Heap, J. T., Pennington, O. J., Cartman, S. T., Minton, N. P., A modular system for Clostridium shuttle plasmids. *Journal of Microbiological Methods* 2009, *78*, 79-85.

[52] Ren, C., Gu, Y., Hu, S., Wu, Y., *et al.*, Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in Clostridium acetobutylicum. *Metabolic engineering* 2010, *12*, 446-454.

[53] Lehmann, D., Lutke-Eversloh, T., Switching Clostridium acetobutylicum to an ethanol producer by disruption of the butyrate/butanol fermentative pathway. *Metabolic engineering* 2011, *13*, 464-473.

[54] Cooksley, C. M., Zhang, Y., Wang, H., Redl, S., *et al.*, Targeted mutagenesis of the Clostridium acetobutylicum acetone-butanol-ethanol fermentation pathway. *Metabolic engineering* 2012, *14*, 630-641.

[55] Xiao, H., Li, Z., Jiang, Y., Yang, Y., *et al.*, Metabolic engineering of D-xylose pathway in Clostridium beijerinckii to optimize solvent production from xylose mother liquid. *Metabolic engineering* 2012, *14*, 569-578.

[56] Dusseaux, S., Croux, C., Soucaille, P., Meynial-Salles, I., Metabolic engineering of Clostridium acetobutylicum ATCC 824 for the high-yield production of a biofuel composed of an isopropanol/butanol/ethanol mixture. *Metabolic engineering* 2013, *18*, 1-8.

[57] Bhandiwad, A., Shaw, A. J., Guss, A., Guseva, A., *et al.*, Metabolic engineering of Thermoanaerobacterium saccharolyticum for n-butanol production. *Metabolic engineering* 2014, *21*, 17-25.

[58] Jang, Y. S., Im, J. A., Choi, S. Y., Lee, J. I., Lee, S. Y., Metabolic engineering of Clostridium acetobutylicum for butyric acid production with high butyric acid selectivity. *Metabolic engineering* 2014, *23*, 165-174.

[59] Wu, Y., Yang, Y., Ren, C., Yang, C., *et al.*, Molecular modulation of pleiotropic regulator CcpA for glucose and xylose coutilization by solvent-producing Clostridium acetobutylicum. *Metabolic engineering* 2015, *28*, 169-179.

[60] Yoo, M., Croux, C., Meynial-Salles, I., Soucaille, P., Metabolic flexibility of a butyrate pathway mutant of Clostridium acetobutylicum. *Metabolic engineering* 2017, *40*, 138-147.

[61] Aune, T. E., Aachmann, F. L., Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl Microbiol Biotechnol* 2010, *85*, 1301-1313.

[62] Chen, L. J., Wu, Y. D., Xue, C., Bai, F. W., Improving fructose utilization and butanol production by Clostridium acetobutylicum via extracellular redox potential regulation and intracellular metabolite analysis. *Biotechnology journal* 2017, *12*, 1700198.

[63] Seo, S. O., Janssen, H., Magis, A., Wang, Y.*, et al.*, Genomic, transcriptional, and phenotypic analysis of the glucose derepressed Clostridium beijerinckii mutant exhibiting acid crash phenotype. *Biotechnology journal* 2017, *12*, 1700182.

[64] Jang, Y.-S., Lee, J. Y., Lee, J., Park, J. H.*, et al.*, Enhanced Butanol Production Obtained by Reinforcing the Direct Butanol-Forming Route in Clostridium acetobutylicum. *mBio* 2012, *3*.

[65] Liu, J., Guo, T., Shen, X., Xu, J.*, et al.*, Engineering Clostridium beijerinckii with the Cbei_4693 gene knockout for enhanced ferulic acid tolerance. *Journal of Biotechnology* 2016, *229*, 53-57.

[66] Duerre, P., Fermentative butanol production - Bulk chemical and biofuel, in: Wiegel, J., Maier, R. J., Adams, M. W. W. (Eds.), *Incredible Anaerobes: From Physiology to Genomics to Fuels* 2008, pp. 353-362.

[67] Yang, Y., Nie, X., Jiang, Y., Yang, C.*, et al.*, Metabolic regulation in solventogenic clostridia: regulators, mechanisms and engineering. *Biotechnology advances* 2018, *36*, 905-914.

[68] Durre, P., Physiology and Sporulation in Clostridium. *Microbiology spectrum* 2014, *2*, TBS-0010-2012.

[69] Lehmann, D., Radomski, N., Luetke-Eversloh, T., New insights into the butyric acid metabolism of Clostridium acetobutylicum. *Applied Microbiology and Biotechnology* 2012, *96*, 1325-1339.

[70] Wang, Y., Li, X., Milne, C. B., Janssen, H.*, et al.*, Development of a Gene Knockout System Using Mobile Group II Introns (Targetron) and Genetic Disruption of Acid Production Pathways in Clostridium beijerinckii. *Applied and Environmental Microbiology* 2013, *79*, 5853-5863.

[71] Dong, H., Tao, W., Gong, F., Li, Y., Zhang, Y., A functional recT gene for recombineering of Clostridium. *J Biotechnol* 2014, *173*, 65-67.

[72] Cho, C., Choe, D., Jang, Y. S., Kim, K. J.*, et al.*, Genome analysis of a hyper acetone- butanol- ethanol (ABE) producing Clostridium acetobutylicum BKM19. *Biotechnology journal* 2017, *12*, 1600457.

[73] Yoo, M., Croux, C., Meynial-Salles, I., Soucaille, P., Elucidation of the roles of adhE1 and adhE2 in the primary metabolism of Clostridium acetobutylicum by combining in-frame gene deletion and a quantitative system-scale approach. *Biotechnology for Biofuels* 2016, *9*.

[74] Zhang, N., Shao, L., Jiang, Y., Gu, Y.*, et al.*, I-SceI-mediated scarless gene modification via allelic exchange in Clostridium. *J Microbiol Methods* 2015, *108*, 49-60.

[75] Ren, C., Gu, Y., Wu, Y., Zhang, W.*, et al.*, Pleiotropic functions of catabolite control protein CcpA in Butanol-producing Clostridium acetobutylicum. *Bmc Genomics* 2012, *13*.

[76] Gu, Y., Ding, Y., Ren, C., Sun, Z.*, et al.*, Reconstruction of xylose utilization pathway and regulons in Firmicutes. *Bmc Genomics* 2010, *11*.

[77] Zhang, L., Nie, X., Ravcheev, D. A., Rodionov, D. A.*, et al.*, Redox-Responsive Repressor Rex Modulates Alcohol Production and Oxidative Stress Tolerance in Clostridium acetobutylicum. *Journal of Bacteriology* 2014, *196*, 3949-3963.

[78] Zhang, L., Leyn, S. A., Gu, Y., Jiang, W.*, et al.*, Ribulokinase and Transcriptional Regulation of Arabinose Metabolism in Clostridium acetobutylicum. *Journal of Bacteriology* 2012, *194*, 1055-1064.

[79] Lorca, G. L., Chung, Y. J., Barabote, R. D., Weyler, W.*, et al.*, Catabolite repression and activation in Bacillus subtilis: dependency on CcpA, HPr, and HprK. *J Bacteriol* 2005, *187*, 7826-7839.

[80] Yang, Y., Zhang, L., Huang, H., Yang, C.*, et al.*, A Flexible Binding Site Architecture Provides New Insights into CcpA Global Regulation in Gram-Positive Bacteria. *mBio* 2017, *8*.

[81] Wietzke, M., Bahl, H., The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in Clostridium acetobutylicum. *Applied Microbiology and Biotechnology* 2012, *96*, 749-761.

[82] Nie, X., Yang, B., Zhang, L., Gu, Y.*, et al.*, PTS regulation domain-containing transcriptional activator CelR and sigma factor sigma(54) control cellobiose utilization in Clostridium acetobutylicum. *Molecular Microbiology* 2016, *100*, 289-302.

[83] Xue, Q., Yang, Y. P., Chen, J., Chen, L.*, et al.*, Roles of three AbrBs in regulating two-phase Clostridium acetobutylicum fermentation. *Applied Microbiology and Biotechnology* 2016, *100*, 9081-9089.

[84] Tan, Y., Liu, Z. Y., Liu, Z., Zheng, H. J., Li, F. L., Comparative transcriptome analysis between csrA-disruption Clostridium acetobutylicum and its parent strain. *Mol Biosyst* 2015, *11*, 1434-1442.

[85] Xu, M., Zhao, J., Yu, L., Tang, I. C.*, et al.*, Engineering Clostridium acetobutylicum with a histidine kinase knockout for enhanced n-butanol tolerance and production. *Applied Microbiology and Biotechnology* 2015, *99*, 1011-1022.

[86] Liu, J., Lin, Q., Chai, X., Luo, Y., Guo, T., Enhanced phenolic compounds tolerance response of Clostridium beijerinckii NCIMB 8052 by inactivation of Cbei_3304. *Microbial Cell Factories* 2018, *17*.

[87] Steiner, E., Scott, J., Minton, N. P., Winzer, K., An agr Quorum Sensing System That Regulates Granulose Formation and Sporulation in Clostridium acetobutylicum. *Applied and Environmental Microbiology* 2012, *78*, 1113-1122.

[88] Mitchell, W. J., Sugar uptake by the solventogenic clostridia. *World Journal of Microbiology & Biotechnology* 2016, *32*.

[89] Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., Chou, C. P., Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus Clostridium. *Applied and Environmental Microbiology* 2016, *82*, 6109-6119.

[90] Bruder, M., Moo-Young, M., Chung, D. A., Chou, C. P., Elimination of carbon catabolite repression in Clostridium acetobutylicum-a journey toward simultaneous use of xylose and glucose. *Applied Microbiology and Biotechnology* 2015, *99*, 7579-7588.

[91] Balagurunathan, B., Jonnalagadda, S., Tan, L., Srinivasan, R., Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis. *Microbial Cell Factories* 2012, *11*.

[92] Yanase, H., Sato, D., Yamamoto, K., Matsuda, S.*, et al.*, Genetic engineering of Zymobacter palmae for production of ethanol from xylose. *Appl Environ Microbiol* 2007, *73*, 2592-2599.

[93] Xiao, H., Gu, Y., Ning, Y., Yang, Y., *et al.*, Confirmation and Elimination of Xylose Metabolism Bottlenecks in Glucose Phosphoenolpyruvate-Dependent Phosphotransferase System-Deficient Clostridium acetobutylicum for Simultaneous Utilization of Glucose, Xylose, and Arabinose. *Applied and Environmental Microbiology* 2011, *77*, 7886-7895.

[94] Sun, Z., Chen, Y., Yang, C., Yang, S., *et al.*, A novel three-component system-based regulatory model for D-xylose sensing and transport in Clostridium beijerinckii. *Molecular Microbiology* 2015, *95*, 576-589.

[95] Liu, L., Zhang, L., Tang, W., Gu, Y., *et al.*, Phosphoketolase Pathway for Xylose Catabolism in Clostridium acetobutylicum Revealed by C-13 Metabolic Flux Analysis. *Journal of Bacteriology* 2012, *194*, 5413-5422.

[96] Servinsky, M. D., Germane, K. L., Liu, S., Kiel, J. T., *et al.*, Arabinose is metabolized via a phosphoketolase pathway in Clostridium acetobutylicum ATCC 824. *Journal of Industrial Microbiology & Biotechnology* 2012, *39*, 1859-1867.

[97] Mermelstein, L., Papoutsakis, E., In vivo methylation in Escherichia coli by the Bacillus subtilis phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of Clostridium acetobutylicum ATCC 824. *Applied and Environmental Microbiology* 1993, *59*, 1077-1081.

[98] Thomas, C. M., Nielsen, K. M., Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology* 2005, *3*, 711.

[99] Dong, H., Zhang, Y., Dai, Z., Li, Y., Engineering Clostridium Strain to Accept Unmethylated DNA. *Plos One* 2010, *5*.

[100] Zhang, J., Liu, Y.-J., Cui, G.-Z., Cui, Q., A novel arabinose-inducible genetic operation system developed for Clostridium cellulolyticum. *Biotechnology for Biofuels* 2015, *8*.

[101] Jia, K., Zhu, Y., Zhang, Y., Li, Y., Group II Intron-Anchored Gene Deletion in Clostridium. *Plos One* 2011, *6*.

[102] Cui, G.-Z., Zhang, J., Hong, W., Xu, C., *et al.*, Improvement of ClosTron for successive gene disruption in Clostridium cellulolyticum using a pyrF-based screening system. *Applied Microbiology and Biotechnology* 2014, *98*, 313-323.

[103] Chen, Y., Caruso, L., McClane, B., Fisher, D., Gupta, P., Disruption of a toxin gene by introduction of a foreign gene into the chromosome of Clostridium perfringens using targetron-induced mutagenesis. *Plasmid* 2007, *58*, 182-189.

[104] Frazier, C. L., Joseph, S. F., Lambowitz, A. M., Mills, D. A., Genetic manipulation of Lactococcus lactis by using targeted group II introns: generation of stable insertions without selection. *Applied & Environmental Microbiology* 2003, *69*, 1121.

[105] Enyeart, P. J., Chirieleison, S. M., Dao, M. N., Perutka, J., *et al.*, Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* 2013, *9*.

[106] Al-Hinai, M. A., Fast, A. G., Papoutsakis, E. T., Novel System for Efficient Isolation of Clostridium Double-Crossover Allelic Exchange Mutants Enabling Markerless Chromosomal Gene Deletions and DNA Integration. *Applied and Environmental Microbiology* 2012, *78*, 8112-8121.

[107] Cartman, S. T., Kelly, M. L., Heeg, D., Heap, J. T., Minton, N. P., Precise manipulation of the Clostridium difficile chromosome reveals a lack of association between the tcdC genotype and toxin production. *Appl Environ Microbiol* 2012, *78*, 4683-4690.

[108] Little, G. T., Willson, B. J., Heap, J. T., Winzer, K., Minton, N. P., The butanol producing microbe Clostridium beijerinckii NCIMB 14988 manipulated using forward and reverse genetic tools. *Biotechnology journal* 2018, *13*, 1700711.

[109] McAllister, K. N., Sorg, J. A., CRISPR genome editing systems in the genus Clostridium: a timely advancement. *Journal of bacteriology* 2019.

[110] Cartman, S. T., Minton, N. P., A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile. *Appl Environ Microbiol* 2010, *76*, 1103-1109.

[111] Lanckriet, A., Timbermont, L., Happonen, L. J., Pajunen, M. I.*, et al.*, Generation of single-copy transposon insertions in Clostridium perfringens by electroporation of phage mu DNA transposition complexes. *Appl Environ Microbiol* 2009, *75*, 2638-2642.

[112] Vidal, J. E., Chen, J., Li, J., McClane, B. A., Use of an EZ-Tn5-based random mutagenesis system to identify a novel toxin regulatory locus in Clostridium perfringens strain 13. *PLoS One* 2009, *4*, e6232.

[113] Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K.*, et al.*, Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res* 2012, *40*, e59.

[114] Desai, R. P., Papoutsakis, E. T., Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum. *Appl Environ Microbiol* 1999, *65*, 936-945.

[115] Tummala, S. B., Welker, N. E., Papoutsakis, E. T., Design of antisense RNA constructs for downregulation of the acetone formation pathway of Clostridium acetobutylicum. *J Bacteriol* 2003, *185*, 1923-1934.

[116] Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K. J.*, et al.*, Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. *Journal of Biotechnology* 2015, *200*, 1-5.

[117] Strecker, J., Ladha, A., RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 2019.

[118] Huang, H., Chai, C., Yang, S., Jiang, W., Gu, Y., Phage serine integrase-mediated genome engineering for efficient expression of chemical biosynthetic pathway in gas-fermenting Clostridium ljungdahlii. *Metabolic engineering* 2019, *52*, 293-302.

[119] Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R.*, et al.*, High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Appl Environ Microbiol* 2011, *77*, 8288-8294.

[120] Xu, T., Li, Y. C., Shi, Z., Hemme, C. L.*, et al.*, Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase. *Applied and Environmental Microbiology* 2015, *81*, 4423-4431.

[121] Cho, C., Lee, S. Y., Efficient gene knockdown in Clostridium acetobutylicum by synthetic small regulatory RNAs. *Biotechnology and bioengineering* 2017, *114*, 374-383.

[122] Jang, Y.-S., Han, M.-J., Lee, J., Im, J. A.*, et al.*, Proteomic analyses of the phase transition from acidogenesis to solventogenesis using solventogenic and non-solventogenic Clostridium acetobutylicum strains. *Applied Microbiology and Biotechnology* 2014, *98*, 5105-5115.

[123] Liu, J., Guo, T., Wang, D., Shen, X., *et al.*, Enhanced butanol production by increasing NADH and ATP levels in Clostridium beijerinckii NCIMB 8052 by insertional inactivation of Cbei_4110. *Applied Microbiology and Biotechnology* 2016, *100*, 4985-4996.

[124] Zimmerly, S., Guo, H., Perlman, P. S., Lambowitz, A. M., Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 1995, *82*, 545-554.

Table 1. Comparison of genetic manipulation tools adapted in Clostridium.

Genetic tools	Principles and outcome(s)	Plasmid design	Operation	Mutant screening	Cycle	Targeting	Unexpected-genotype	Application	Ref.
TargeTron	Gene disrupted by intron RNA	☆☆☆	☆☆☆	☆☆☆	☆☆☆	☆☆☆	☆☆	☆☆☆	[15, 16]
RNAi	Gene knockdown by anti-sense RNA	☆☆☆	☆☆	☆	☆☆	☆☆	☆	☆	[114, 115]
Transposon	Random Transposon disruption	☆☆☆	☆☆☆	☆☆☆	☆☆	/	/	☆	[110-112]
Allelic-coupled exchange	HR-mediated in-frame deletion or insertion	☆☆	☆	☆	☆	☆☆☆	☆☆☆	☆☆☆	[113, 119]
Counter-marker-assisted. allelic-exchange	Counter-marker facilitate to screen mutants with double exchange	☆☆	☆	☆☆	☆☆	☆☆☆	☆☆☆	☆☆☆	[106, 107]
I-SceI-mediated allelic-exchange	I-SceI accelerates second single cross exchange	☆☆	☆	☆☆	☆☆	☆☆☆	☆☆☆	☆☆☆	[74]
Oligo-mediated Point mutation	Point mutation	☆☆	☆	☆	☆	☆☆☆	☆☆☆	☆	[71]
CRISPR/Cas 9 system	Cas9-mediated double-strands break and DNA repair	☆☆	☆☆	☆☆	☆☆☆	☆☆☆	☆☆☆	☆☆☆	[22, 89, 116, 120]

Gene knockdown by small regulatory RNAs (sRNA)	sRNA-mediated gene knockdown	☆☆	☆☆	☆☆	☆☆☆	☆☆	☆☆	☆☆	[121]
Base editing system	Cytidine deaminase and uracil DNA glycosylase inhibitor can efficiently convert specific C·G nucleotide base pairs to T·A	☆☆☆	☆☆☆	☆☆	☆☆☆	☆☆	☆☆☆	☆☆	[20]
Integrase-mediated chromosome integration	Phage serine integrase-mediated chromosome integration	☆☆	☆☆	☆☆	☆☆	☆☆☆	☆☆☆	☆☆	[118]

☆☆☆, good/easy; ☆☆, moderate; ☆, bad/difficult.

Table 2 Summary of disrupted genes involved in ABE fermentation.

Disrupted genes	Function	Strains	Effects	Ref.
Ack	*acetate kinase*	*C. acetobutylicium*	Increase ethanol output	[54, 69]
Adc	*acetoacetate decarboxylase*	*C. acetobutylicium*	Decrease butanol output	[54]
AdhE1	*alcohol/aldehyde dehydrogenases 1*	*C. acetobutylicium*	Decrease ABE output	[54, 73]
AdhE2	*alcohol/aldehyde dehydrogenases 2*	*C. acetobutylicium*	No effect on ABE output	[54, 73]
BdhA/B	*butanol dehydrogenases A/B*	*C. acetobutylicium*	No effect on ABE output	[54]
Buk	*butyrate kinase*	*C. acetobutylicium,*	Decrease butanol output	[15, 54, 60, 70]
		C. beijerinckii	Increase total solvents output	
CtfA/B	*CoA transferase*	*C. acetobutylicium*	Decrease butanol output	[54]
HydA	*hydrogenase*	*C. acetobutylicium*	Increase butyrate outputs lightly	[122]
Pta	*phosphotransacetylase*	*C. acetobutylicium*	Increase butanol output	[54, 64, 70]
		C. beijerinckii	Increase butyrate and lactate outputs	
Ptb	*phosphotransbutyrylase*	*C. acetobutylicium*	Increase ethanol output	[54, 60, 64, 69]
Spo0A	*sporulation transcription factor*	*C. acetobutylicium*	Decrease butanol and total solvents output	[16, 40]
Cbei_4110	*NADH-quinone oxidoreductase*	*C. beijerinckii*	Increase butanol output	[123]

Table 3 Improvement of TargeTron technology

Improvements	Strains	Strategy	Description	Ref.
Combined with HR or recombinases	*C.acetobutylicum*	Combined with FLP/FRT	Remove resistance marker from intron	[40]
		Combined with HR	Completely delete CAC1493-1494 and ctfAB	[101]
Plasmids curing	*C.cellulolyticum*	A pyrF-based screening system	Successive genes disruption；	[102]
	C.cellulolyticum	Inducible AraR system and mazE/F	efficient plasmids curing	[100]
	C. cellulovorans	Principle of plasmid incompatibility	Efficient plasmids replacement	
Avoiding of unexpected genotype	*C.acetobutylicum*	Gene complementation; Southern bolt	Confirmation of site-specific insertion	[54]
Functional gene integration	*C.perfringens*	Inserting an exogenous DNA sequence into an intron RNA	Integration and overexpression of SIV-P27	[103]
	C. sporogenes		Integration of λ-phage DNA	[40]
Off-target	*C. sporogenes*	RAM	Decreasing off-target frequency	[16]
	C.cellulolyticum	AraR-inducible system		[100]

Figure legends

Figure 1. Scheme of TargeTron technology development, improvement and application in Clostridium. The mesophilic TargeTron developed by Shao et al., relies on the EBS1, EBS2 as well as reverse transcriptase (RT) to recognize target DNA and complete intron insertion. TargeTron technology has been applied in many solventogenic, cellulolytic and pathogenic clostridia. Simultaneously, it is constantly being improved, including reducing off-target, introducing auxiliary genetic components (inducible promoters, counter-screening markers, or recombinase systems), and being used in gene integration. Development and improvement of the technology are interactive and mutually progressive. Together they promote the application in clostridial metabolic engineering and gene function identification.

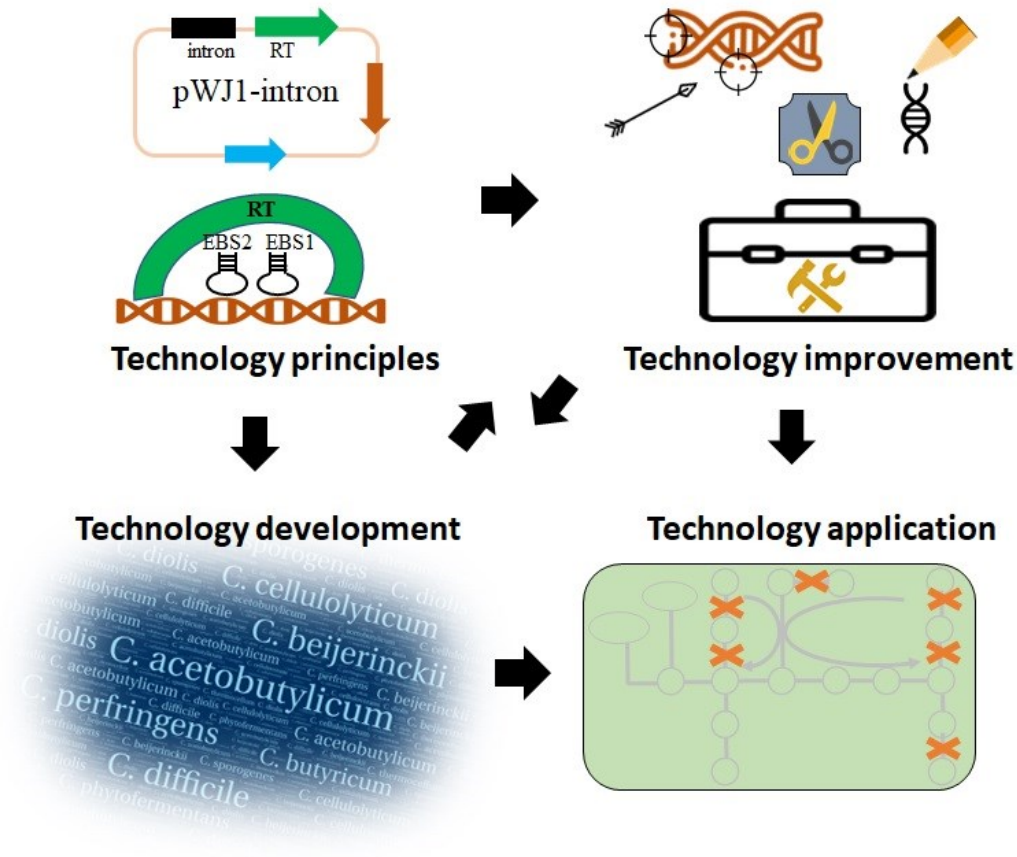


Figure 2 The secondary structure of group II intron RNA (left) and the process of "retrohoming" (right). D Ⅰ is the largest domain, which contains the 5' and 3' exon binding sites (such as EBS1, EBS2 and even EBS3) and the branch site nucleophile. Crystal structure implies that D Ⅰ binds to the 5' and 3' exon as well as D Ⅴ (a metal-ion-binding domain) to form an active center; DII and DIII are closely related to the folding and catalytic efficiency [33]. DIV encodes the an open reading frame (ORF) of reverse transcriptase (RT) [25, 31]. DVI contains a branch point nucleotide (adenosine nucleotide). The process of "retrohoming" can be briefly as follows. When a gene inserted by the group II introns is transcribed, the intron RNA self-splices from the large RNA fragment with the aid of IEP and folded into a catalytical structure. Meanwhile, the spliced lariat intron RNA together with IEP form a ribonucleoprotein (RNP) complex. The RNP recognizes the target DNA sequence, and then the intron RNA is inserted into the target site by reverse splicing. The IEP exerts endonuclease activity to excise the complementary DNA strand, leaving an overhang of 3'-end as a template. It subsequently catalyzes the synthesis of cDNA complementary to the intron RNA. The cDNA is finally integrated into another DNA strand by the repair mechanism or homologous recombination [28-30, 124].

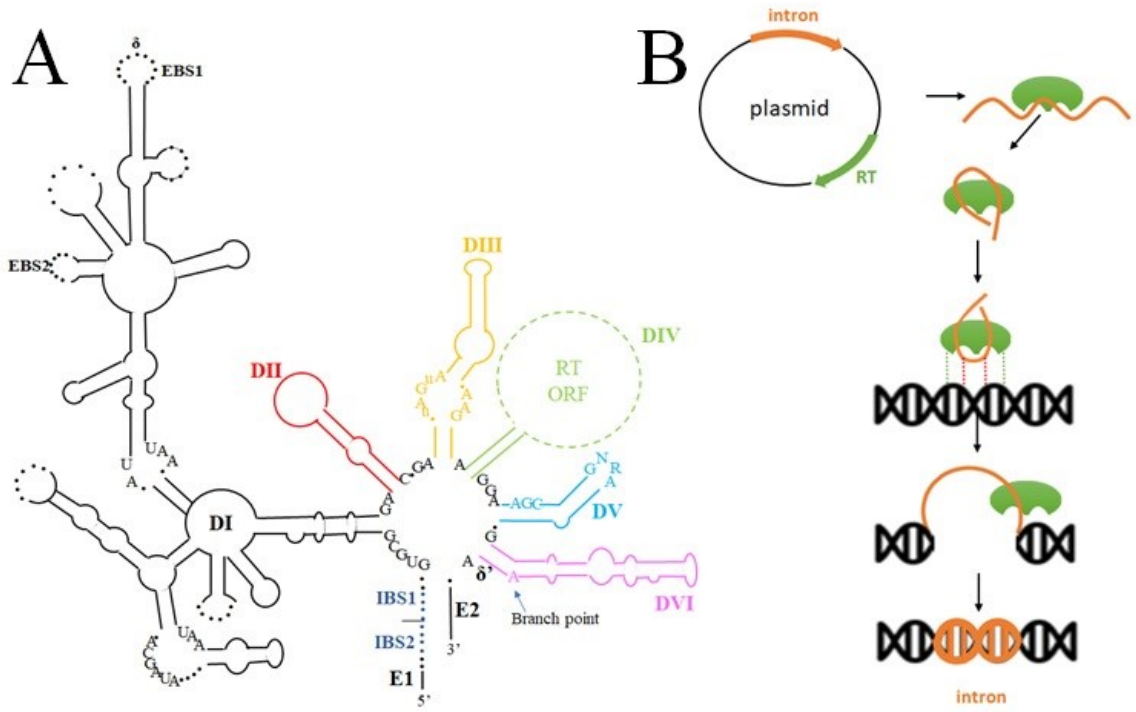


Figure 3. Plasmids system and the DNA recognition mechanism of mesophilic TargeTron and thermo-TargeTron. To coordinate the region of E1 and E2 (5' and 3' exons), a reverse complementary sequence is adapted and IEP (or RT) and intron are arranged beneath it to match the target sequence. The mesophilic TargeTron adapts the EBS1, EBS2, and δ to recognize the target DNA, while the thermo-TargeTron has no δ site but EBS3. The IEP of mesophilic TargeTron recognizes 6 base-pairs (with yellow background), while thermo-TargeTron only recognizes 2 base-pairs of the target DNA because the high temperature can anneal DNA without RT. IS, intron-insertion site; RT, reverse transcriptase; bla, ampicillin resistance; malsR, macrolide–lincosamide–streptogramin resistance; cat, chloramphenicol resistance; CloE1 origin, origin of replication applicable in *E. coli*; pCB102, repB, gram-positive origin of replication applicable in *Clostridium*.

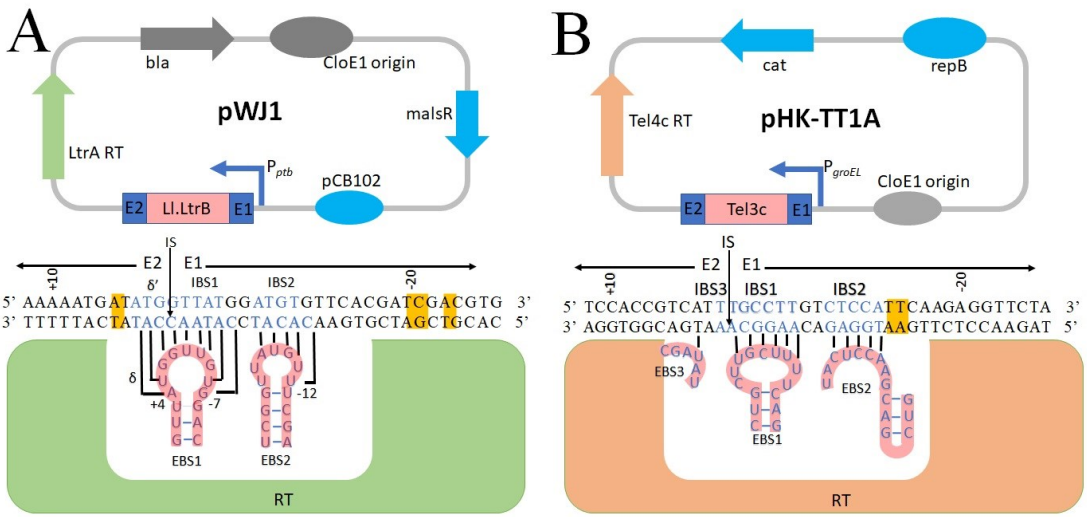


Figure 4 The progress of TargetTron technology development, improvement and application in Clostridium during past 12 years. TargeTron technology is suitable for solventogenic, cellulolytic and medical clostridia. The only obstacle of development is unavailable genetic system. With the successful genetic systems development of *C. autoethanogenum*, C. *cellulovorans* and C. *diolis*, TargetTron has also been successfully developed. Application is the purpose of development, so it has been running through the last 12 years. However, very few studies focus on technology improvement, especially after 2015, which may be attributed to the competition of new genetic manipulation tools such as CRISPR/Cas system in recent years.

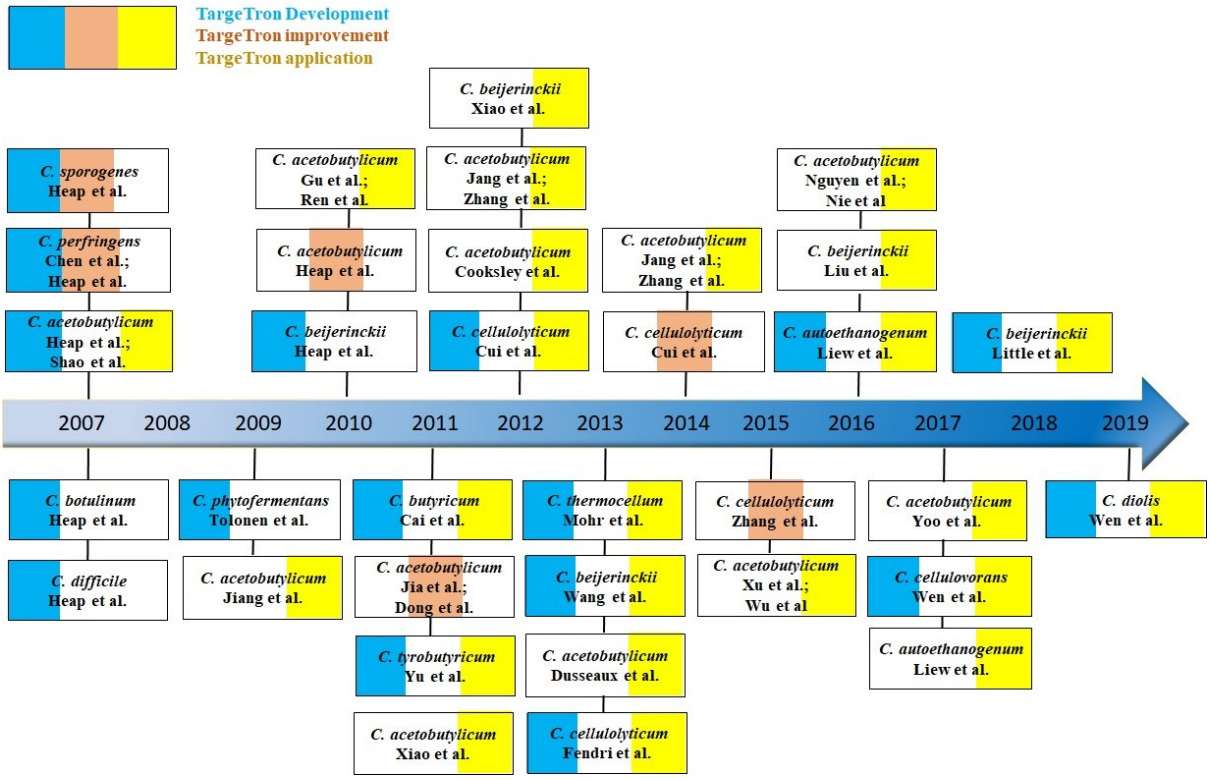


Figure 5 Target genes of TargeTron in acetone-butanol-ethanol pathway. Genes in *C. acetobutylicum* (CA), *C. beijerinckii* (Cbei) are highlighted in blue and green, respectively. *hydA*, ferredoxin hydrogenase; *ack*, acetate kinase; *pta*, phosphotransacetylase; *adhE* or *adhE1/2*, acetaldehyde/ethanol dehydrogenase; *bdhA/B*, butanol dehydrogenase; *edh*, alcohol dehydrogenase; *thl*, acetyl-CoA acetyltransferase; *hbd*, beta-hydroxybutyryl-CoA dehydrogenase; *crt*, crotonase; *bcd*, butyryl-CoA dehydrogenase; *ctfAB*, acetate/butyrate-acetoacetate COA-transferase; *adc*, acetoacetate decarboxylase; *buk*, butyrate kinase; *ptb*, phosphate butyryl-transferase.

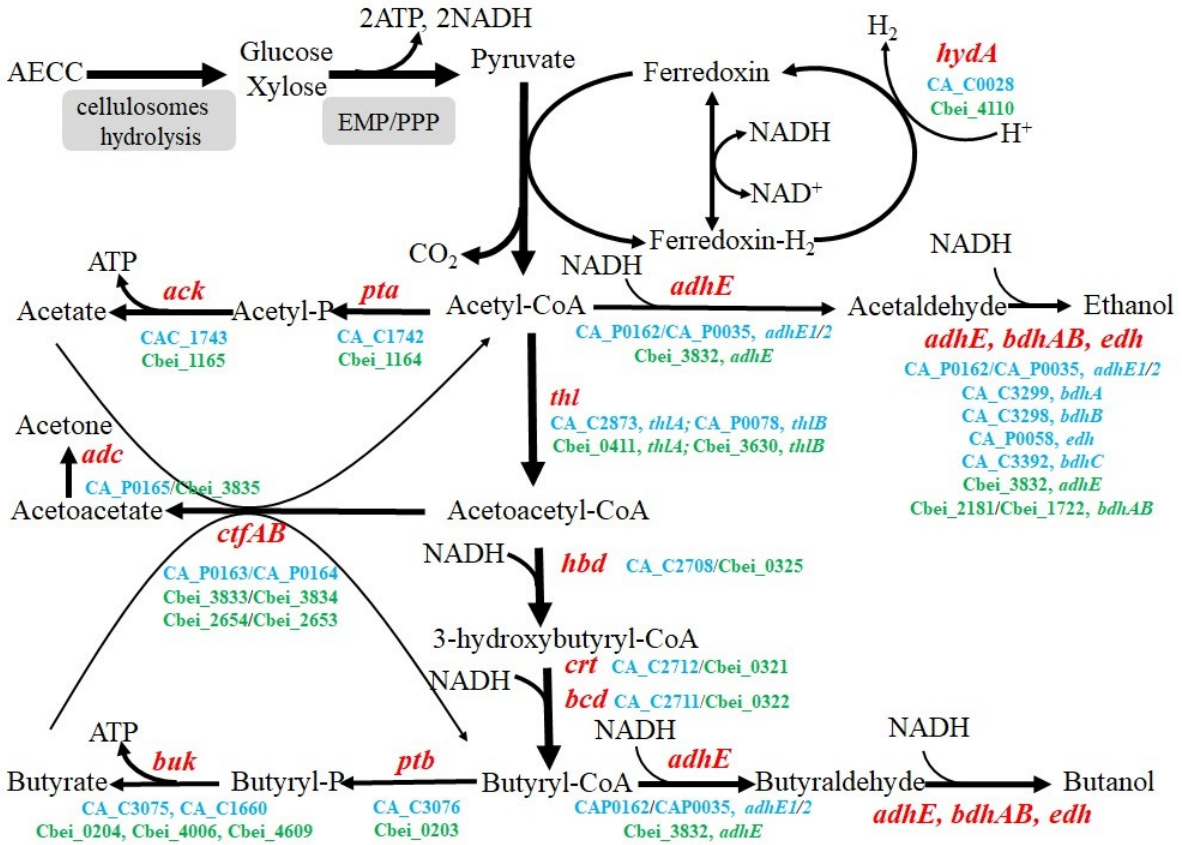
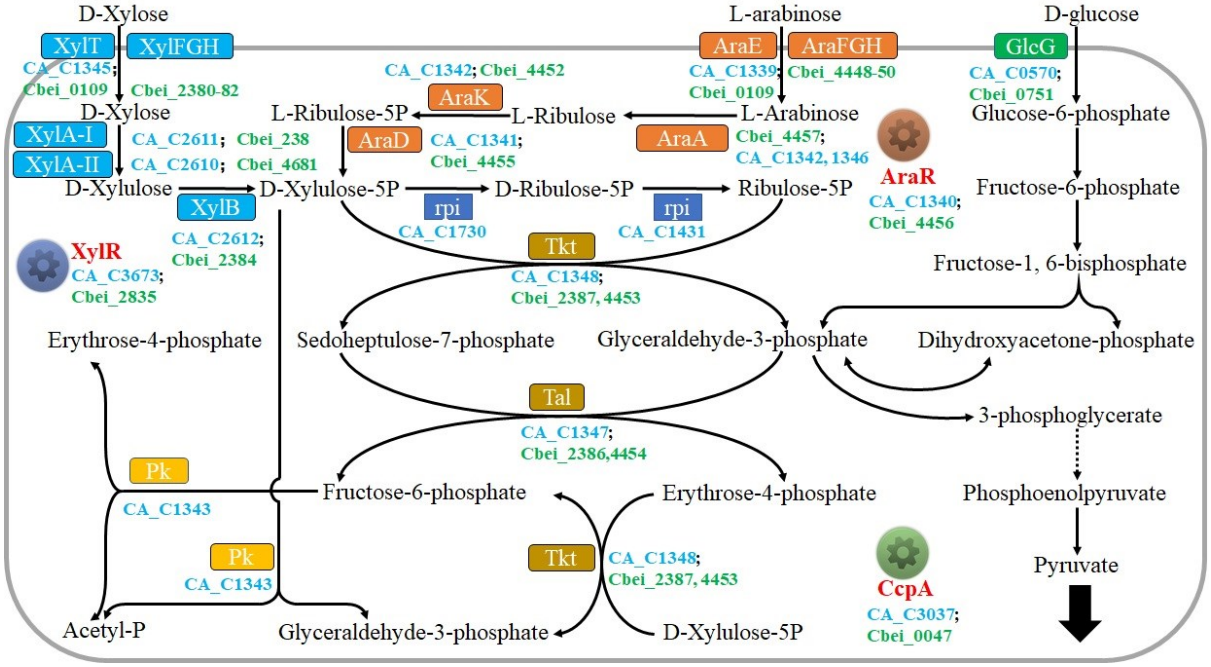


Figure 6 TargeTron used for the analysis and reconstruction of pentose utilize pathway in Clostridium. Genes in *C. acetobutylicum* (CA), *C. beijerinckii* (Cbei) are highlighted in blue and green, respectively. Transcriptional factors XylR (xylose metabolism regulation gene), AraR (arabinose metabolism regulation gene) and CcpA (encoding catabolite control protein A) that control pentose sugar utilization pathways are shown in circles with different colors; genes involved in transport, pentose-phosphate pathway and phosphoketolase pathway of xylose, arabinose are shown in by matching different background colors. XlyT, xylose proton-symporter; XylFGH, ABC transporter system for xylose; XylA-I, xylose isomerase; XylA-II, xylose isomerase; XylB, xylulokinase; AraE, arabinose-proton symporter; AraFGH, ABC transport system for arabinose; AraK, ribulokinase; AraA-I, L-arabinose isomerase; AraD, L-ribulose-5-phosphate 4-epimerase; GlcG, glucose-specific PTS transporter; Tkt, transketolase; Tal, transaldolase; Rpi, 5-phosphate ribose isomerase; Pk, phosphoketolase; Rpi, 5-phosphate ribose isomerase.



Sheng Yang obtained his Ph.D. from Shanghai Institute of Biochemistry, Chinese Academy of Sciences (CAS). He joined Shanghai Institutes for Biological Sciences in 2000 and became a professor in 2006. He is interested in the continuous expanding and improving the tool kit and uses it as a platform to develop biocatalysts together with industrial partners. The particular interests are biocatalysts for conversion of renewable resources to chemicals and biofuels with conventional chassis microorganisms, such as *Escherichia coli*, *Corynebacterium glutamicum* and *Saccharomyces cerevisiae* as well as non-conventional microbial hosts such as *Yarrowia lipolytica* and clostridia.

Zhi-qiang Wen is currently an assistant professor in School of Environmental and Biological Engineering, Nanjing University of Science & Technology. He earned his PhD in Biochemical engineering at Zhejiang University in 2014. After two years' postdoctoral training at **C**hinese **A**cademy of **S**cience (CAS), he joined Nanjing University of Science & Technology in 2016. His current research focuses on microbial genetic techniques development, synthetic biotechnology and metabolic engineering, as well as biorefinery and bioenergy. He has authored more than 10 papers in the fields of clostridial metabolic engineering and CRISPR-Cas system development.

In the review, we revisited 12 years' advances of TargeTron technology applicable in solventogenic clostridia, including its principle, technical characteristics, development, application and improvement. We also discussed its putative competitors or collaborators, for example, CRISPR-Cas9 system. We believe TargeTron combined with CRISPR-Cas-assisted gene/base editing and gene expression regulation system makes a better future for Clostridial genetic modification.