Synthetic and Systems Biotechnology 9 (2024) 43–54

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

A toolbox for genetic manipulation in intestinal *Clostridium symbiosum*

Pengjie Yang [a,b,1], Jinzhong Tian [a,c,1], Lu Zhang [d], Hui Zhang [e], Gaohua Yang [a,f], Yimeng Ren [d], Jingyuan Fang [d,*], Yang Gu [a,**], Weihong Jiang [a,***]

[a] CAS-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
[b] University of Chinese Academy of Sciences, Beijing, 100049, China
[c] Xianghu Laboratory, Hangzhou, 311231, China
[d] NHC Key Laboratory of Digestive Diseases, Division of Gastroenterology and Hepatology, Shanghai Institute of Digestive Disease, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, 145 Middle Shandong Road, Shanghai, 200001, China
[e] Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518120, China
[f] The Wallenberg Laboratory, Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg, Bruna Straket 16, Gothenburg, 41345, Sweden

ARTICLE INFO

Keywords:
Gut *Clostridium symbiosum*
Toolbox
Gene overexpression
Gene deletion
Gene regulation

ABSTRACT

Gut microbes are closely related with human health, but remain much to learn. *Clostridium symbiosum* is a conditionally pathogenic human gut bacterium and regarded as a potential biomarker for early diagnosis of intestinal tumors. However, the absence of an efficient toolbox that allows diverse genetic manipulations of this bacterium limits its in-depth studies. Here, we obtained the complete genome sequence of *C. symbiosum* ATCC 14940, a representative strain of *C. symbiosum*. On this basis, we further developed a series of genetic manipulation methods for this bacterium. Firstly, following the identification of a functional replicon pBP1 in *C. symbiosum* ATCC 14940, a highly efficient conjugative DNA transfer method was established, enabling the rapid introduction of exogenous plasmids into cells. Next, we constructed a dual-plasmid CRISPR/Cas12a system for genome editing in this bacterium, reaching over 60 % repression for most of the chosen genes as well as efficient deletion (>90 %) of three target genes. Finally, this toolbox was used for the identification of crucial functional genes, involving growth, synthesis of important metabolites, and virulence of *C. symbiosum* ATCC 14940. Our work has effectively established and optimized genome editing methods in intestinal *C. symbiosum*, thereby providing strong support for further basic and application research in this bacterium.

1. Introduction

The human intestinal tract is not only an important digestive organ but also a favorable habitat for microorganisms. On average, the intestine of a healthy adult is colonized by over 1000 microbial species [1], which can affect host health via different mechanisms. Gut microbes have been known to play vital roles in physiological activities of host, such as inflammation, barrier homeostasis, metabolism, and hematopoiesis [2]. Thus, the roles of gut microbes in human's physiological processes and disease occurrence and development have been attracting extensive attention.

Firmicutes are the most major component of the gut microbiota, in which 95 % belongs to clostridia [3]. Specific to intestinal *Clostridium*, multiple species have been found to be associated with human diseases such as bacteremia, anorexia, diabetes, autism spectrum disorders, and obesity. *C. symbiosum* has been reported as an opportunistic pathogen causing bacteremia [4–6]. Currently, research indicates that *C. symbiosum* may be associated with the development of inflammatory bowel disease (IBD) [7] and coronary artery disease [8]. Of note, recent studies revealed significantly increased abundance of *C. symbiosum* in

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: jingyuanfang@sjtu.edu.cn (J. Fang), ygu@cemps.ac.cn (Y. Gu), wjiang@cemps.ac.cn (W. Jiang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.synbio.2023.12.005
Received 28 August 2023; Received in revised form 8 December 2023; Accepted 24 December 2023
Available online 29 December 2023

the stools of patients with colonic adenomas or early and advanced colorectal cancer, suggesting that this bacterium can serve as a tumor biomarker for the early diagnosis of colorectal cancer [9].

A comprehensive and thorough understanding of causal relationship between *C. symbiosum* and the development of colorectal cancer will require efficient genetic manipulation of this bacterium. The first step towards this goal has been made recently with three atypical *C. symbiosum* strains (*C. symbiosum* DSM 29356, WAL-14163, and WAL-14673) tested, in which gene overexpression and CRISPRi-based gene repression were achieved in WAL-14163 [10]. This work strongly supports a continued development and optimization of the toolbox in *C. symbiosum*, enabling its expansion to more important strains.

In this study, the first complete genome of the representative *C. symbiosum* strain, ATCC 14940, is provided. Subsequently, a CRISPR-Cas system employing Cas12a was constructed, achieving efficient gene deletion and expressional regulation in the ATCC 14940 strain. Furthermore, a series of promoters with wide strength range in this bacterium were identified, enabling the controlled overexpression of target genes. Based on this comprehensive toolbox, multiple genes associated with crucial characteristics of *C. symbiosum* ATCC 14940 were identified. Together, this work effectively expands and optimizes the existing toolbox in intestinal *C. symbiosum* strains.

2. Materials and methods

2.1. Strains, cell lines, media, and reagents

The *E. coli* strains TOP10, HB101, S17-1, and their derived strains were cultured in Luria-Bertani (LB) broth or on LB agar plate at 37 °C supplemented with chloramphenicol (12.5 μg/mL) or erythromycin (500 μg/mL) when needed. The *E. coli* strain TOP10 was used for plasmid construction. HB101 and S17-1 were used as the conjugation donors. *C. symbiosum* ATCC 14940 was anaerobically cultivated at 37 °C using the Brain-Heart Infusion medium (37 g/L brain heart infusion, 5 g/L yeast extract, 5 g/L K_2HPO_4, 0.5 g/L l-cysteine hydrochloride, 0.2 % (v/w) resazurin,5 mg/L hemin, 1 mg/L menadione). When required, antibiotics were added to growth media at the working concentrations: thiamphenicol (10 μg/mL), erythromycin (1 μg/mL), polymyxin B (50 μg/mL).

HCT116 human colorectal adenocarcinoma (CRC) cell line was obtained from American Type Culture Collection (ATCC). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL, Grand Island, NY, USA) supplemented with 10 % fetal bovine serum (FBS, YOUSHI, Wuhan, China) and incubated at 37 °C under a humidified incubator (37 °C, 5 % CO_2). Cell passage was treated with trypsin-EDTA (Gibco, Burlington, Canada). For the co-culture of *C. symbiosum* and HCT116 cells, approximately 4×10^5 HCT116 cells were grown in a 6-well plate and then infected with *C. symbiosum* for 4 h (multiplicity of infection = 100) under anaerobic condition at 37 °C. Then, the cells were washed with PBS and cultured in complete medium prepared for subsequent experiments.

KOD plus Neo and KOD FX DNA polymerase (Toyobo, Osaka, Japan) were used for PCR amplification. Plasmid construction was performed using the ClonExpress MultiS One Step Cloning Kit (Vazyme Biotech Co., Ltd., Nanjing, China) or restriction enzymes (Thermo Fisher Scientific, Vilnius, Lithuania) and ligase (Takara, Dalian, China). Extraction of plasmid and DNA purification were performed by kits (Axygen Biotechnology Company Limited, Hangzhou, China). All the primers used in this study were synthesized by BioSune (Biosune, Shanghai, China).

2.2. Plasmid construction

The plasmids and primers used in this study are listed in Tables S1 and S2. The starting plasmids used for constructing the other plasmids in this study were pMTL82151 and pMTL82254.

The plasmid pMTL-P_{1339}-ccpA used for gene overexpression in *C. symbiosum* ATCC 14940 was constructed as follows: the DNA fragment of the promoter P_{1339} was obtained via PCR amplification using the primers P_{1339}–F/P_{1339}–R and the genomic DNA of *C. acetobutylicum* ATCC 824 as template. The *ccpA* gene was obtained via PCR amplification using the primers ccpA-F/ccpA-R and the genomic DNA of *C. symbiosum* ATCC 14940 as the template. Subsequently, the DNA fragment (P_{1339}-ccpA) was yielded via overlapping PCR using the primers P_{1339}–F/ccpA-R, and then linked with the linear plasmid pMTL82151 vector (digested with *BamH*I and *Sac*I) using ClonExpress II One Step cloning kit (Vazyme, Nanjing, China), generating the target plasmid pMTL-P_{1339}-ccpA. The pMTL-P_{1339}-ackA plasmid was constructed via the same steps except that gene sequences were changed accordingly.

The plasmid pMTLddcas12a was generated as follows: the DNA fragments corresponding to the promoter P_{01440} and P_{thl} were obtained via PCR amplification using the primers P_{01440}–F/P_{01440}–R and P_{thl}-F/P_{thl}-R, and the genomic DNA of *C. ljungdahlii* and *C. acetobutylicum* as the template, respectively. The ddCas12a was codon-optimized based on the genomic sequences of *C. beijerinckii* NCIMB 8052 and *C. acetobutylicum* ATCC 824. The gene fragment of ddCas12a was obtained via PCR amplification using the primers ddcas12a-F/ddcas12a-R. The P_{01440} fragment was assembled with the P_{thl} fragment via overlapping PCR using the primers P_{01440}–F/P_{thl}-R. The obtained two DNA fragments, *Fn*ddCas12a and P_{thl}-P_{01440}, were further linked with the linear pMTL82151 (digested with restriction enzymes *BamH*I and *Not*I) using the ClonExpress MultiS One Step Cloning Kit, generating the target plasmid pMTLddcas12a. The pMTLddcas12a-birAi-TS plasmid was constructed as follows: a DNA fragment containing the crRNA spacer and repeat sequence was first obtained by PCR amplification using the primers birAT-crRNA-F/birAT-crRNA-R. Then, this DNA fragment was used as a template for PCR amplification using the primers universal-F/universal-R. The generated DNA fragment birAT-crRNA, which contained the crRNA, terminator, and repeat sequence, was assembled with the linear plasmid pMTLddcas12a (digested with restriction enzymes *BamH*I and *Nco*I) using the ClonExpress II One Step cloning kit, yielding the target plasmid pMTLddcas12a-birAi-TS. The pMTLddcas12a-birAi-NTS and pMTLddcas12a-ackAi plasmids were constructed via the same steps except that the crRNA expression cassettes were changed accordingly.

The plasmid pMTLcas12a was constructed as follows: The DNA fragment cas12a-M was obtained via PCR amplification using the primers M-F/M-R and the plasmid pMTLddcas12a as the template with the introduction of a point mutation (A1006E) in the Cas12a protein. The obtained DNA fragment was assembled with the linear plasmid pMTL-ddcas12a (digested with restriction enzyme *Swa*I) using the ClonExpress II One Step cloning kit, generating the target plasmid, pMTLcas12a.

The plasmid pMTLcas12a-ΔccpA was constructed as follows: the two homologous arms flanking the open reading frame of *ccpA* (obtained by PCR amplification using the primers ccpA-up-F/ccpA-up-R and ccpA-down-F/ccpA-down-R) and the crRNA (obtained by PCR amplification using the primers ccpA-crRNA-F/ccpA-crRNA-R) were assembled by overlapping PCR, yielding a large DNA fragment ccpAcrRNA-ccpAup-ccpAdown. Thereafter, this DNA fragment was assembled with the linear plasmid pMTLcas12a (digested with *Xho*I and *BamH*I) using the ClonExpress II One Step cloning kit, generating the target plasmid pMTLcas12a-ΔccpA. The pMTLcas12a-Δ02043 plasmid was constructed via the same steps.

The plasmid pMTL-P_{01440}-cas12a was constructed as follows: the DNA fragment P_{01440}-cas12a was obtained via PCR amplification using the primers cas12a-F/P_{01440}-*Bam*HI-R and the pMTLcas12a plasmid as the template. The obtained DNA fragment P_{01440}-cas12a was linked with the linear pMTL82254 (digested with *BamH*I and *Not*I) using the ClonExpress II One Step cloning kit, generating the target plasmid pMTL-P_{01440}-cas12a. The derivative plasmids, i.e., pMTL-P_{fba2}-cas12a, pMTL-P_{gpm}-cas12a, and pMTL-P_{thl}-cas12a, were constructed via the same

steps.

The plasmid pMTL-P_{thl}-ccpAcrRNA was constructed as follows: the DNA fragment P_{thl}-ccpAcrRNA-ccpAup-ccpAdown was obtained via PCR amplification using the primers P_{thl}-crRNA-F/ccpAdown-*Sal*I-R and plasmid pMTLcas12a-ΔccpA as the template. Then, the obtained DNA fragment P_{thl}-ccpAcrRNA-ccpAup-ccpAdown was linked with the linear plasmid pMTL82151 (digested with *BamH*I and *Sac*I) using ClonExpress II One Step cloning kit, generating the target plasmid pMTL-P_{thl}-ccpAcrRNA. The derivative plasmids, i.e., pMTL-P_{fba2}-ccpAcrRNA, pMTL-P_{gpm}-ccpAcrRNA, and pMTL-P_{01440}-ccpAcrRNA, were constructed via the same steps.

The plasmid pMTL82151-P_{1339}-LacZ was constructed as follows: the DNA fragment corresponding to *LacZ* was obtained via PCR amplification using the primers LacZ-F/LacZ-R and the pIMP1-P_{thl}-LacZ plasmid [11] as the template. The DNA fragment corresponding to the P_{1339} promoter was obtained via PCR amplification using the primers P_{1339}-lacZ-F/-P_{1339}-lacZ-R and the genomic DNA of *C. acetobutylicum* ATCC 824 as the template. These two DNA fragments were assembled via overlapping PCR using the primers P_{1339}-lacZ-F/LacZ-R. The obtained P_{1339}-lacZ fragment was further linked with the linear plasmid pMTL82151 (digested with *BamH*I and *Sac*I) using ClonExpress II One Step cloning kit, generating the target plasmid pMTL82151-P_{1339}-LacZ. The other plasmids for assessing promoter strength (pMTL82151-P_{thl}-LacZ, pMTL82151-P_{01440}-LacZ, pMTL82151-P_{fba2}-LacZ, pMTL82151-P_{gpm}-LacZ, pMTL82151-P_{pfkA2}--LacZ, pMTL82151-P_{pgi}-LacZ, pMTL82151-P_{eno}-LacZ, pMTL82151-P_{py-k}-LacZ, pMTL82151-P_{pgk}-LacZ, pMTL82151-P_{gap}-LacZ, and pMTL8 2151-P_{fbaA}-LacZ) were constructed via the same steps.

2.3. RNA purification and real-time quantitative PCR (RT-qPCR)

C. symbiosum cells were collected when they reached an optical density (OD_{600}) of 1.0 and then stored at −80 °C. Total RNA extraction and RT-qPCR were conducted as reported previously [12]. The 16S rDNA was employed as the internal control. The primers used for RT-qPCR are listed in Supplementary Table S2.

2.4. C. symbiosum conjugation

The culture of *C. symbiosum* ATCC 14940 was anaerobically cultured at 37 °C in BHI medium until reaching OD_{600} of 0.4–1.2. The donor strain of *E. coli* (containing the shuttle plasmid) was cultivated in LB broth supplemented with appropriate antibiotics until reaching OD_{600} of 0.4–1.2. Subsequently, 5 mL of the *E. coli* culture was subjected to centrifugation (at 3,000 × *g* for 10 min). The obtained cell pellets were gently resuspended in 1 mL of PBS buffer and subjected to centrifugation again. The obtained cell pellets were then gently suspended in the *C. symbiosum* ATCC 14940 culture (200 μL). The mixed culture was spotted onto the BHI agar plates and incubated at 37 °C for 2−8 h. The colonies on the agar plates were collected and resuspended in 1 mL of PBS, and then plated onto the BHI agar plates containing the appropriate antibiotics (50 μg/mL polymyxin B, 10 μg/mL thiamphenicol, or 1 μg/mL erythromycin). Colonies appeared after 3−4 d of incubation at 37 °C.

2.5. Fermentation

The *C. symbiosum* strains were cultivated anaerobically in BHI medium containing the appropriate antibiotic. In brief, 250 μL of the frozen stock was inoculated into BHI medium (5 mL) and then incubated anaerobically at 37 °C. When the OD_{600} of grown cells reached 1.0, 0.5 mL of the inoculum was transferred into 9.5 mL of BHI medium for cultivation at 37 °C. The culture was sampled every 2 h. All the procedures were performed in an anaerobic chamber (Whitley A35 Anaerobic Workstation, Don Whitley Scientific Limited, Bingley, West Yorkshire, UK).

2.6. Analytical methods

The measurements of cell growth and acetate concentration were performed according to previously reported methods [11]. Briefly, cell growth was determined based on the absorbance of the culture at A_{600} (OD_{600}) using a spectrophotometer (U-1800, Hitachi, Japan). The concentration of acetate was determined as described previously using a gas chromatography (7890 A, Agilent, Wilmington, DE, USA) equipped with a flame ionization detector and a capillary column (EC-Wax, Alltech, Lexington, KY, USA) [11].

2.7. Cell proliferation assays

Cell proliferation was evaluated using a Cell Counting Kit-8 (CCK-8, DOJINDO, Kumamoto, Japan) according to the manufacturer's instructions. In brief, the HCT116 cells were seeded into 96-well culture plates at a density of 3,000 cells per well and cultivated in a CO_2 (5 %, v/v) incubator at 37 °C overnight. 10 μl of CCK-8 solution was added to each well every 24 h for another 2h incubation at 37 °C. The absorbance was measured at 450 nm.

2.8. β-Galactosidase assays

C. symbiosum cells were grown at 37 °C under anaerobic condition and harvested by centrifugation (6,000 × *g* for 10 min) when the grown cells reached the optical density (OD_{600}) of 1.0. Cells were resuspended in 500 μl of B-PER reagent (Thermo Scientific Pierce, USA) with vigorous vortexing for 3 min. The resulting cell lysates were heated at 60 °C for 30 min to remove heat unstable proteins and then subjected to centrifugation (12,000 × *g* for 30 min). The supernatants were used for the β-galactosidase assays as previously reported [13].

2.9. DNA sequencing and genome assembly

The *C. symbiosum* cells were collected when the optical density (OD_{600}) reached 1.0 by centrifugation (10,000 × *g* for 10 min) and then immediately frozen in liquid nitrogen. Next, genomic DNA was extracted by using a DNA extraction kit (Shandong Sparkjade Biotechnology Co., Ltd., China).

Genome sequencing was carried out using the Pacific Biosciences (PacBio) RS II sequencing platform. In brief, genomic DNA was sheared into ~10 kb fragments using a Covaris g-TUBE shearing device (Covaris, Woburn, Massachusetts, USA). DNA fragments were purified, end-repaired, and ligated to SMRTParkbell hairpin adapters using the PacBio SMRTbell library preparation kit (Pacific Biosciences, Menlo Park, China, USA). SMRTbell DNA libraries were constructed according to the manufacturer's protocol (Pacific Biosciences, Menlo Park, CA, USA). The library quality analysis and quantification were determined using the Qubit 2.0 Fluorometer (Bio-Medlab, China) and the Agilent 2100 Bioanalyzer system (Agilent Technologies, SantaClara, CA, USA). The SMRT sequencing was accomplished using the PacBioRSII system (Pacific Biosciences, Menlo Park, CA, USA) according to the standard protocol. The obtained continuous PacBio long-reads generated from SMRT sequencing runs were adopted for *de novo* assembly using the program Spades (version 3.14.1), yielding one circular chromosome.

2.10. Gene prediction and functional annotation

RepeatMasker (v4.0.6) was used to identify and annotate repeats [42]. The tandem repeats were further identified using Tandem Repeats Finder software (v4.07b) [43] with the following parameters: Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50.

Augustus (v3.2.1) [44] was used for the *de novo* gene prediction. For homology-based prediction, protein sequences of *C. symbiosum* ATCC 14940 were downloaded from the NCBI database, and then aligned to the repeats of a soft-masked genome using TBlastN (v2.2.29)[45] with a

cut-off e-value of 1e−5. The results from the *de novo* and homology prediction were integrated using EVidenceModeler (v1.1.1) [46]. The annotated genes were also subjected to a BLAST analysis in the Virulence Factor Database (VFDB; http://www.mgc.ac.cn/VFs/main.htm) using the Diamond software with the default parameters to identify virulence genes [14].

3. Results and discussion

3.1. Complete genome sequence of C. symbiosum ATCC 14940

C. symbiosum ATCC 14940 is the model strain of intestinal *C. symbiosum*. Its draft genome sequence has been presented by the National Institutes of Health (NIH) (GenBank accession No.514700). However, the complete genome sequence of this strain has not yet been reported so far. To fill this gap and obtain the detailed genomic information, we re-sequenced this stain and finished *de novo* assembly. The sequencing data showed a circular chromosome of 5,022,615 bp with a G + C content of 47.8 % (Fig. 1 and Table 1). A total of 4,496 coding sequences were predicted, including 4,419 protein-coding sequences, 12 rRNA and 65 tRNA genes (Genome Warehouse in National Genomics Data Center: GWHBEHX01000000). No significant difference was observed in the genome size and GC content between *C. symbiosum* ATCC14940 and the *C. symbiosum* LT0011 (a reference genome providing by NCBI) (Table S4).

3.2. Efficient DNA transformation for C. symbiosum

It is known that bacteria process a series of defense systems that provide protection against foreign DNA, in which the restriction-modification (R-M) system normally plays a major role [15]. During the conjugative DNA transfer, plasmid DNA is transferred in a single-stranded form; thus, it is likely to be minimally affected by the R-M system [16]. Therefore, we attempted to establish the DNA transformation method of *C. symbiosum* ATCC 14940 via conjugational transfer using *E. coli* as the donor strain.

Table 1
The genome features of *Clostridium symbiosum* ATCC 14940.

Features	Results
Sequencing technology	PacBio, Illumina
Genome topology	Circular
Total size (bp)	5,022,615
GC content (%)	47.80
Contig N50 (bp)	977,160
Scaffold number	1
Protein-coding genes	4,419
tRNA genes	65
rRNA genes	12

We firstly assessed the resistance of *C. symbiosum* ATCC 14940 to different antibiotics and found that the supplementation of 10 μg/mL thiamphenicol or 1 μg/mL erythromycin in the medium could completely inhibit cell growth (Table 2). Thus, the *Clostridium perfringens catP* gene that can give *Clostridium* strains the resistance to thiamphenicol [17] was used in the following plasmid construction. Next, four pMTL80000 plasmids, i.e., pMTL82151, pMTL83151, pMTL84151, and pMTL85151, which contain the abovementioned *catP* gene as well as different clostridial replicons (pBP1, pCB102, pCD6, and pIM13, respectively) [18], were tested in conjugative plasmid transfer of *C. symbiosum* ATCC 14940 (Fig. 2A). Here, two types of the *E. coli* strains were adopted as the donor strains, namely a standard strain S17-1 and a methylation-deficient strain HB101 (pRK2013). Concurrently, polymyxin B, which can inhibit the growth of these *E. coli* stains but has no effect on *C. symbiosum* ATCC 14940, was used for the selection of transconjugants. The four plasmids were firstly transferred into the *E. coli* strain S17-1 and strain HB101 (details shown in Materials and Methods). Encouragingly, both the *E. coli* S17-1 and HB101 strain that contained the plasmid pMTL82151 generated transconjugants, in which S17-1 produced an average of 30 transconjugants per conjugation (equivalent to a conjugation efficiency of 8×10^{-8}) while HB101 only produced an average of 12 transconjugants (equivalent to a conjugation efficiency of 3×10^{-8}), suggesting that *E. coli* S17-1 is a better conjugal

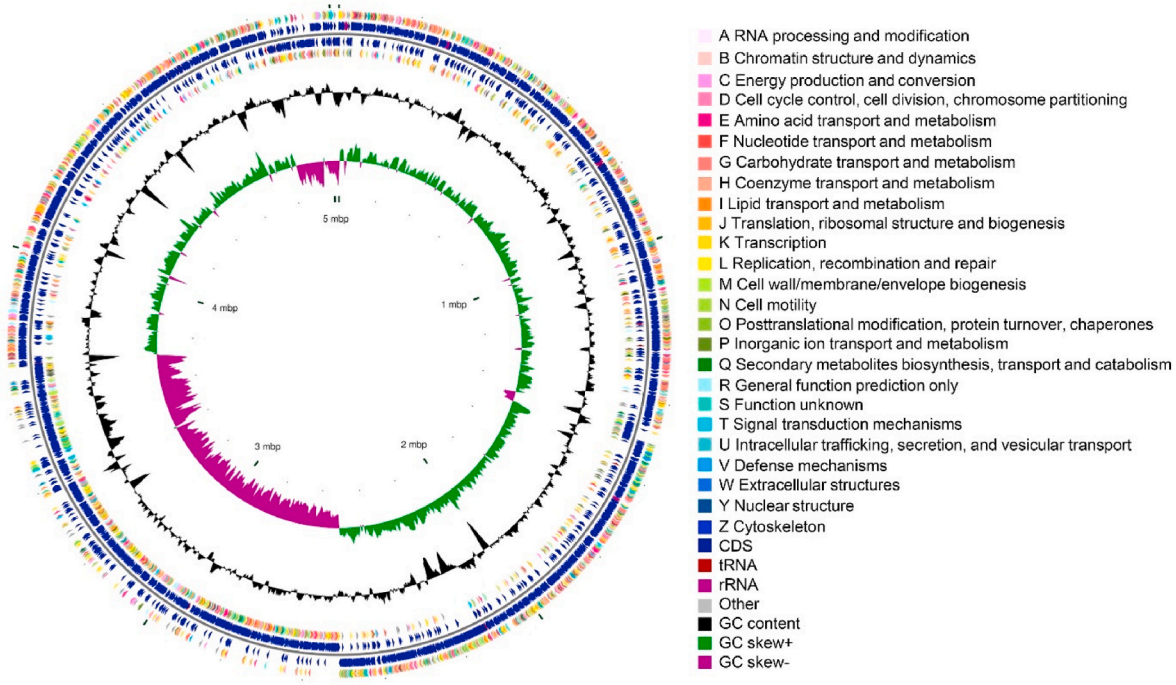

Fig. 1. Circular genome map for *Clostridium symbiosum* ATCC 14940. From outside to inside: circle 1, CDSs on the forward strand colored according to COG category; circle 2, CDSs, rRNA and tRNA on forward strand; circle 3, CDSs, rRNA, and tRNA on the reverse strand; circle 4, CDSs on the reverse strand colored according to COG category; circle 5, GC content; circle 6, GC skew.

Table 2
Antibiotic sensitivity test of *C. symbiosum*.

Antibiotics	Concentration (mg/L)	Sensitivity
Thiamphenicol	1	+
	5	–
	10	–
	20	–
Erythromycin	0.1	–
	0.5	–
	10	–
Spectinomycin	50	+++
	100	+
	200	–
Ampicillin	50	–
	100	–
Kanamycin	50	++
	100	+
	150	–
Streptomycin	50	+
	100	–
Apramycin	50	++
	100	+
	150	–
Tetracycline	5	–
	10	–
Polymyxin B	100	+++
	200	+++

–: High sensitivity; +: moderate sensitivity; ++: low sensitivity; +++: insensitive.

donor strain (Fig. 2B).

Next, the conjugation transfer process was further optimized. We firstly adjusted the co-culturing time of *E. coli* S17-1 and *C. symbiosum* ATCC 14940. The combined cultures were incubated stationarily at 37 °C for 2, 4, 6, or 8 h. As shown in Fig. 2C, the 6-h incubation offered higher conjugation transfer efficiency over the other three time-intervals. We further examined the influence of the growth state of the donor and recipient strains on the conjugation efficiency. The highest plasmid transfer efficiency was obtained from the culture of *E. coli* S17-1 and *C. symbiosum* ATCC 14940 grown to OD_{600} of 1.0 and 0.4, respectively (Fig. 2D and E). Finally, a robust conjugation protocol was presented, reaching 5×10^3 transconjugants per milliliter of recipient cells (equal to a conjugation efficiency of 2×10^{-5}) for *C. symbiosum* ATCC 14940.

3.3. Gene overexpression in C. symbiosum

Having established an efficient conjugative plasmid transfer method for *C. symbiosum* ATCC 14940, we attempted to test gene overexpression in this bacterium. Three heterologous strong promoters, P_{thl} and P_{1339} from *C. acetobutylicum* ATCC 824 [13,19] and P_{01440} from *C. ljungdahlii* DSM 13528 [20], were chosen for an initial test. These promoters were placed upstream of a reporter gene *lacZ* in the pMTL82151 plasmid for strength evaluation by measuring β-galactosidase activity. As shown in Fig. 3A, all the promoters generated high β-galactosidase activities, demonstrating that they are active in *C. symbiosum* ATCC 14940.

Subsequently, these promoters were further used to overexpress some genes that are potentially associated with specific phonotypic outcomes of *C. symbiosum*. Here, CcpA (catabolite control protein A), a global transcriptional regulator involved in the regulation of cell growth and substrate metabolism in Gram-positive bacteria including *Clostridium* species [21,22], was chosen for testing. According to the genome annotation, the gene (CSYM_04033) that encodes CcpA in *C. symbiosum* was found and then cloned into the pMTL82151 plasmid under the control of the promoter P_{1339}. The yielding plasmid was introduced into *C. symbiosum* through conjugative transfer for gene overexpression (Fig. 3B). As expected, a significantly increased transcriptional level of *ccpA* was detected in the *ccpA*-overexpressing strain compared with the control strain (Fig. 3C), indicating that this gene was sufficiently expressed with the promoter P_{1339}. Interestingly, the growth rate of the *ccpA*-overexpressing strain was lower than that of the control strain (Fig. 3D), which was different from the findings in the industrial C. *acetobutylicum* [12], in which *ccpA* overexpression could promote cell growth. Therefore, this result indicates the functional diversity of CcpA in different bacteria.

3.4. CRISPR interference (CRISPRi) in C. symbiosum using the DNase-deactivated Cas12a (ddCas12a)

Next, we attempted to further develop a CRISPRi system in *C. symbiosum* ATCC 14940, aiming to achieve the efficient repression of target genes. Compared with Cas9, Cas12a has a less toxic effect and requires only a single RNA molecule, the crRNA, which presents a more minimalistic system [23,24], and thus, Cas12a rather than Cas9 was employed in the construction of the CRISPRi system. Using the pMTL82151 plasmid as the backbone, a DNase-deactivated Cas12a from *Francisella novicida* (dd*Fn*Cas12a) was inserted into the plasmid under the control of the abovementioned promoter P_{01440} (Fig. 4A). A previously reported functional gene, *birA*, was adopted as the target to test our CRISPRi system. BirA is known to be a bifunctional protein acting as both a biotin ligase and a repressor on the biotin synthesis operon in bacteria [25,26].

To target the *birA* gene (CSYM_01417) in *C. symbiosum* ATCC 14940, two crRNA sequences complementary to different regions (one on the template DNA strand and another on the non-template DNA strand) of *birA*, were designed and then separately inserted into the plasmid under the control the abovementioned promoter P_{thl} (Fig. 4A). As expected, after introducing the plasmid into the strain, both the two dd*Fn*Cas12a-crRNA constructs could efficiently inhibit the expression of *birA*, reaching 80 % and 60 % repression for the template and non-template DNA strand, respectively (Fig. 4B). Meanwhile, both of the resulting two strains exhibited slower growth rate compared to the control strain (Fig. 4C), indicating the impact of *birA* repression on cell metabolism. This phenotypic change was similar with the previous finding in *C. ljungdahlii* that the inactivation of *birA* could impair the cell growth [27], which was largely due to the damage of BirA's biotin protein ligase activity on the biotinylation of acetyl-CoA carboxylase (ACC) and pyruvate carboxylase (PYC), two crucial enzymes for the basic metabolism in cells [27]. However, it should be noted that the repression degree on *birA* based on the above dd*Fn*Cas12a-crRNA construct was still incomplete, requiring a further promotion (Fig. 4B). A feasible strategy is to use a dual-crRNA construct targeting the same gene, which could lead to stronger transcriptional inhibition [28]. According to the previous reports, the efficiency of ddCpf1 (ddCas12a) in CRISPRi-based gene repression was higher using the crRNAs targeting the template strand relative to those targeting the non-template strand [29]. Our study also had a similar finding. In addition, the repression efficiency was found to be higher when the targeted site of crRNA was closer to the transcription start site [30]. Therefore, we designed the crRNAs that targeted the template DNA strand within 100 bp relative to the translation initiation code "ATG".

3.5. Construction and optimization of CRISPR-Cas12a-based gene deletion in C. symbiosum

After achieving CRISPRi, we attempted to further develop gene deletion in *C. symbiosum* ATCC 14940 using the CRISPR–Cas12a system. An expression cassette containing the *Fn*Cas12a protein driven by the P_{01440} promoter and a crRNA sequence (targeting the *ccpA* gene mentioned above) driven by the P_{thl} promoter was integrated into the plasmid. Meanwhile, two homologous arms (1,000 bp for both) flanking *ccpA* for repairing double strand break were also inserted into the plasmid (Fig. 5A). The yielding plasmid pMTLcas12a-ΔccpA was introduced into cells, and the transformant colonies occurred on the agar plates were used for PCR screening to examine the desired deletion

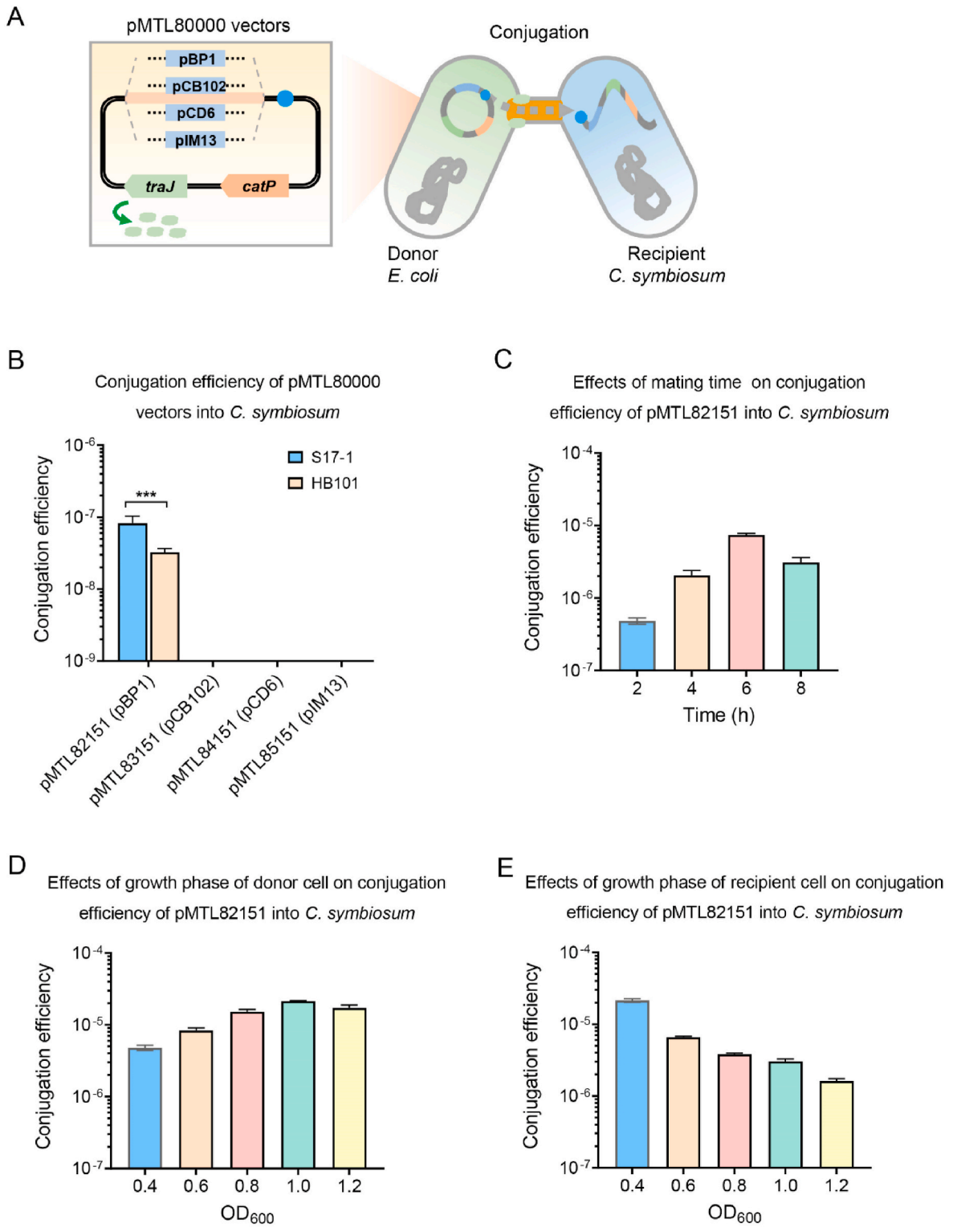

Fig. 2. Development of a conjugative DNA transfer system for *C. symbiosum* ATCC 14940. (A) Schematic diagram of bacterial conjugation. (B) Effect of the donor strain and shuttle plasmid on conjugation efficiency. (C) Effect of co-incubation time on conjugation efficiency. (D) Effect of the growth state (OD_{600}) of the donor bacterium (*E. coli* S17-1) on conjugation efficiency. (E) Effect of the growth state of the recipient bacteria (*C. symbiosum* ATCC 14940) on conjugation efficiency. Conjugation efficiency was calculated as transconjugant CFU/total *C. symbiosum* CFU. Data are represented as mean ± SD ($n = 3$).

events. Encouragingly, of the total 32 colonies, evidence for *ccpA* deletion was obtained for five cases, in which one colony harbored pure gene-deleted cells while the other four colonies harbored mixed wild-type and gene-deleted cells (Fig. 5B and Table 3). The deletion events were further confirmed by sequencing (Fig. 5C). Furthermore, we investigated the influence of *ccpA* deletion on the growth of *C. symbiosum* ATCC 14940, and found that the mutant strain (Δ*ccpA*) exhibited an improved growth compared with the control strain (Fig. 5D); when *ccpA* was re-introduced into the mutant strain via an expression plasmid, the growth of the resulting strain (Δ*ccpA*-P_{1339}-

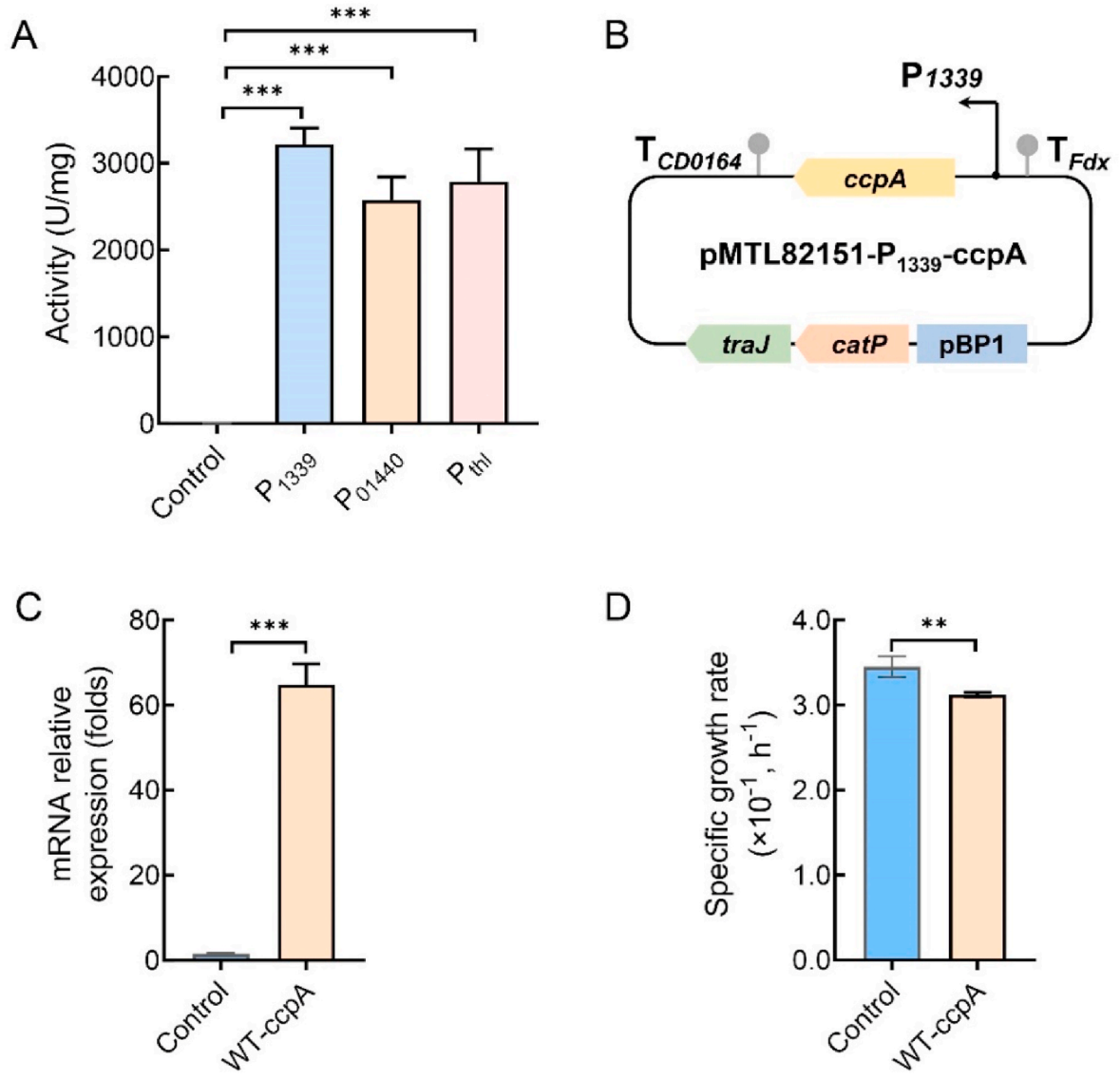

Fig. 3. Establishment of the gene expression system in *C. symbiosum* ATCC 14940. (A) Determination of the β-galactosidase activities that represent the strength of the P_{1339}, P_{01440}, and P_{thl} promoters. Control: no promoter driving the expression of *lacZ*. (B) A schematic diagram of the pMTL-P_{1339}-ccpA plasmid for the overexpression of *ccpA*. P_{1339}: the reported promoter from *C. acetobutylicum* ATCC 824; *traJ*: encoding an *oriT*-recognizing protein for initiation of transfer DNA replication; *catP*: chloramphenicol resistance gene; pBP1: the Gram-positive replicon; TCD0164 and TFdx: terminators. (C) Comparison of the transcriptional levels of *ccpA* in *ccpA*-overexpressing and wild-type strains. (D) The specific growth rates of the wild-type and *ccpA*-overexpressing strains. Control: the wild-type strain contained the plasmid pMTL82151; WT-ccpA: the *ccpA*-overexpressing strain. Data are represented as mean ± SD ($n = 3$). Statistical analysis was performed by a two-tailed Student's *t*-test. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, no significance.

ccpA) was partially restored (Fig. 5D). These results confirmed that the growth change caused by *ccpA* deletion was reliable.

Although we have achieved CRISPR-Cas12a-based gene deletion in *C. symbiosum* ATCC 14940, the efficiency of the conjugative transfer and gene deletion was unacceptable (Table 3), thereby requiring a further promotion. Considering that the conjugative plasmid transfer efficiency is usually associated with the size of plasmid [31,32], we thus allocated the elements of the above CRISPR-Cas construct (Cas12a, crRNA, and homologous arms) to two plasmids: (i) pMTL-cas12a carrying Cas12a; (ii) pMTL-crRNA carrying crRNA and homologous arms (Fig. 6A). Additionally, for efficient expression of Cas12a and crRNA, multiple endogenous promoters in *C. symbiosum* ATCC 14940 were identified and subjected to activity assay (Fig. 6B). Among them, two stronger promoters (P_{fba2} and P_{gpm}), together with the aforementioned P_{thl} and P_{01440} promoters, were chosen for driving the expression of Cas12 or crRNA, generating 16 pairs of plasmids (Fig. 6A). These plasmids were then transferred into *C. symbiosum* ATCC 14940, yielding the transformed cells that harbored eight different plasmid pairs (the other eight plasmid pairs did not give transformed cells due to unknown factors). A number of transformed cells were subjected to PCR screening to detect the desired double-crossover recombination events. The results showed that the deletion efficiencies of these dual-plasmid systems for the *ccpA* gene varied greatly (ranging from 30 % to 90 %) with the highest efficiency obtained by the combination of pMTL-P_{fba2}-cas12a and pMTL-P_{fba2}-crRNA (Fig. 6C). Interestingly, P_{fba2} was not the strongest within the abovementioned four promoters; thus, it seems that an appropriate expression level of the Cas12a protein and crRNA was crucial for the editing efficiency in *C. symbiosum*. Obviously, compared with one plasmid that carried all the genetic elements (Fig. 6C), the dual-plasmid system exhibited higher editing efficiency. Actually, the advantage of the dual-plasmid system over single plasmid in CRISPR-Cas12a-based genome editing has also been found in some other *Clostridium* species, mainly attributing to the smaller plasmid size of the former [33–35].

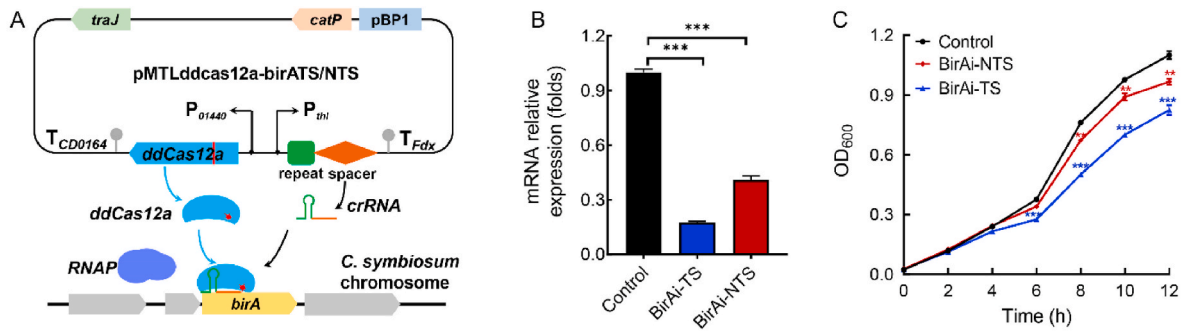

Fig. 4. Development of the CRISPRi system in *C. symbiosum* ATCC 14940. (A) Schematic diagram showing the workflow of *birA* repression using the CRISPRi system. P_{01440} and P_{thl}: the promoters for the expression of ddCas12a and crRNA, respectively; RNAP: RNA polymerase. (B) The transcriptional changes of *birA* after introducing the CRISPRi plasmid into *C. symbiosum* ATCC 14940. (C) The effect of the *birA* repression on the growth of *C. symbiosum* ATCC 14940. Control: the strain containing the plasmid carrying ddCas12a but no crRNA; birAi-TS: the strain harboring the CRISPRi plasmid with crRNA targeting the template strand of *birA*; birAi-NTS: the strain harboring the CRISPRi plasmid with crRNA targeting the non-template strand of *birA*. Data are represented as mean ± SD ($n = 3$). Statistical analysis was performed by a two-tailed Student's *t*-test. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, no significance.

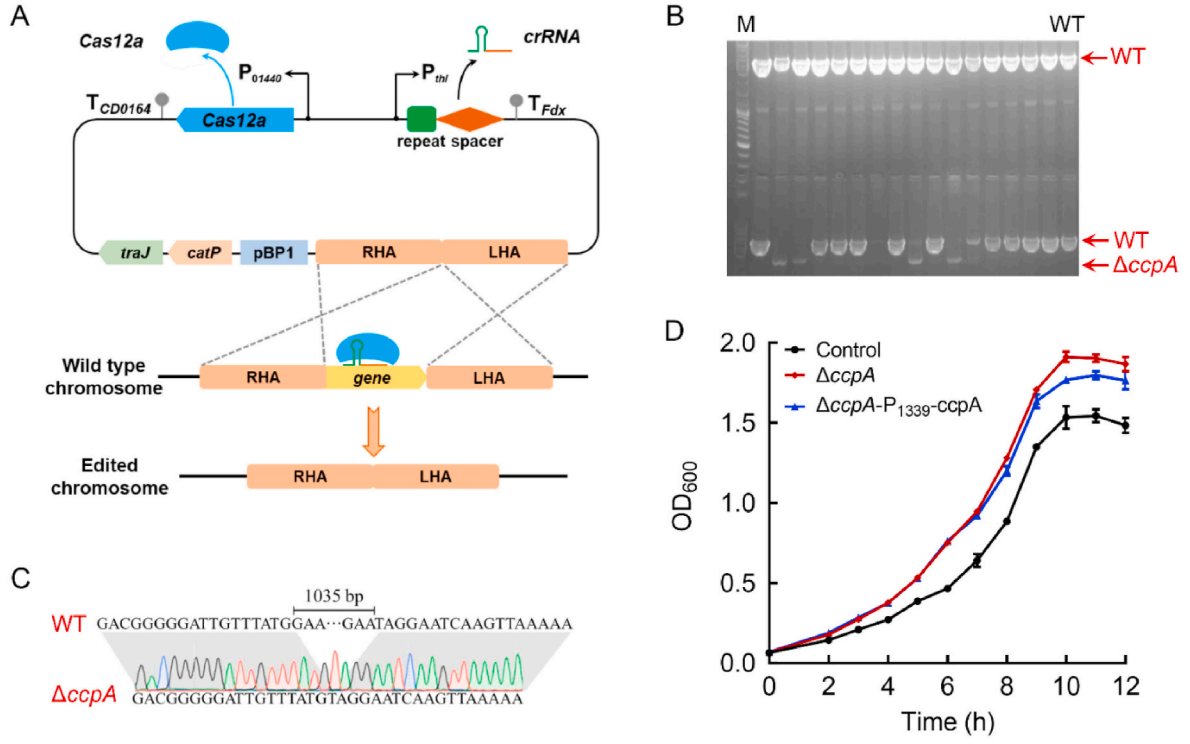

Fig. 5. CRISPR-Cas12a-basded gene deletion in *C. symbiosum* ATCC 14940 using the single-plasmid system. (A) Schematic diagram of the gene deletion system. P_{01440}: the reported promoter from *C. ljungdahlii* DSM13528; P_{thl}: the reported promoter from *C. acetobutylicum* ATCC 824; RHA: right homologous arm; LHA: left homologous arm. (B) PCR screening of the *ccpA* deletion events. (C) Confirmation of the *ccpA* deletion by sequencing. (D) The growth of the *ccpA*-deleted, Δ*ccpA*-P_{1339}-ccpA, and control strains. Control: the wild-type strain containing the plasmid pMTL82151; Δ*ccpA*: the *ccpA*-deleted strain containing the plasmid pMTL82151; Δ*ccpA*-P_{1339}-ccpA: the *ccpA*-deleted strain complemented with the plasmid-carried *ccpA* expression. Data are represented as mean ± SD ($n = 3$).

3.6. Identification of bacterial virulence factors via the genetic toolbox

The establishment of an efficient genetic toolbox enabled in-depth studies to be carried out in *C. symbiosum*. Short-chain fatty acids are crucial metabolites of gut microbial for maintaining intestinal homeostasis [36]. For example, acetate produced by intestinal bacteria can induce apoptosis by inhibiting cell proliferation, thereby inhibiting the occurrence and progression of colorectal cancer [37,38]. To investigate the influence of acetic acid synthesis on *C. symbiosum*, the *ackA* gene (CSYM_02545), encoding acetate kinase (Fig. 7A), was repressed through the CRISPRi system (Fig. S1), yielding the strain (ackAi) to check phenotypic changes. As shown in Fig. 7B and C, the ackAi strain exhibited improved growth compared with the control strain, although the acetic acid formation decreased after *ack* repression. This finding is interesting, because it is quite different from the previous report regarding other *Clostridium* species, in which the disruption of acetic synthesis normally impaired cell growth [11]. We also overexpressed *ackA* in the wild-type strain, and the resulting strain (WT-ackA) showed increased acetic acid formation but impaired growth (Fig. 7E and F), which were just opposite to the above results from *ackA* repression (Fig. 7B and C). Additionally, the ackAi strain showed higher promotion effect on the proliferation of colon cancer cells compared with the control strain (Fig. 7D), indicating a crucial role of acetate in inhibiting cell proliferation. Taken together, all these findings revealed the phenotypic changes generated by altering *ackA* expression in *C. symbiosum* ATCC 14940.

Table 3
Efficiency of *ccpA* deletion in *C. symbiosum*.

Plasmids	Elements	Results (M/P/W/T)[a]	Efficiency[b] (%)	Conjugation efficiency
Control	Cas12a + crRNA	none	–	–
pMTLcas-ΔccpA	Cas12a + crRNA + Homologous arms	4/1/27/32	15.6 %	$2.48 \pm 0.28 \times 10^{-8}$

[a] M: number of colonies that harbored mixed wild-type and genedeleted cells; P: number of colonies that harbored pure gene-deleted cells; W: number of colonies that harbored pure wild-type cells; T: total number of colonies used for PCR screening.
[b] Efficiency: probability of deletion events occurring, calculated as (M + P)/T × 100 %.

As aforementioned, *C. symbiosum* is a potential pathogenic bacterium related with colorectal cancer [9], in which virulence genes may play crucial roles. Based on the genomic data, a large number of putative virulence genes were found in *C. symbiosum* ATCC 14940 (Fig. S2). Among them, CSYM_02043 and CSYM_03550, two potentially crucial virulence genes encoding a glucose-1-phosphate thymidylyltransferase (catalyzing the conversion of glucose-1-phosphate to dTDP-D-glucose) and a rearrangement hotspot (RHS) repeats-containing protein, respectively, were picked out and subjected to the deletion. It has been known that peptidoglycan containing L-rhamnose in enteric pathogens can serve as an antigen to activate the immune system of intestinal epithelial cells, thereby leading to a series of inflammatory reactions if the host's immune response is excessive [39]. Additionally, RHS repeat-containing proteins are known for their diverse functions and roles in intercellular communication and cell-cell contact [40,41]. Through the above-mentioned dual-plasmid CRISPR-Cas12a editing system, the CSYM_02043 and CSYM_03550-deleted strains were rapidly obtained with 97 % and 95 % deletion efficiency, respectively (Figs. S3 and S4), further confirming the utility of this editing system in *C. symbiosum*. The following phenotypic analyses showed that the deletion of CSYM_02043 and CSYM_03550 had the opposite effect on cell growth (Fig. 7G and I), but both led to the reduced promotion effect on the proliferation of colorectal cancer cell HCT116 compared with the wild type strain (Fig. 7H and J). This finding suggests that these two gene are closely associated with the growth and pathogenesis of *C. symbiosum* ATCC 14940.

4. Conclusions

Efficient genome editing of enteric pathogens is essential for the comprehensive discovery of genes associated with important physiological processes as well as pathogenesis in these organisms. In this study, we established a toolbox in *C. symbiosum* ATCC 14940, a conditionally pathogenic bacterium, achieving efficient gene overexpression and CRISPR-Cas12a-based gene deletion and repression. Based on this toolbox, multiple genes involved in synthesis of important metabolites and pathogenesis were identified, exemplifying the utility of this toolbox. The tools developed in this study fill a major gap in the existing genetic tools for intestinal *C. symbiosum*, providing opportunities for understanding and engineering this important intestinal bacterium.

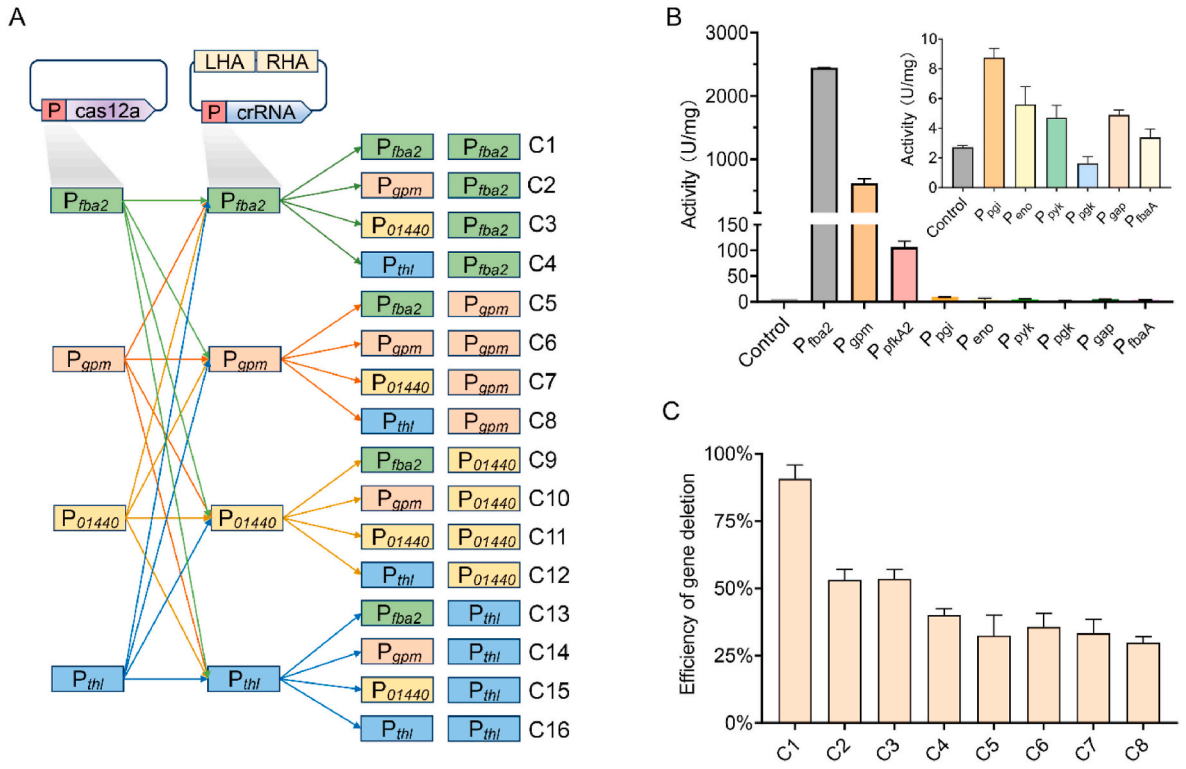

Fig. 6. CRISPR-Cas12a-basded gene deletion in *C. symbiosum* ATCC 14940 using a dual-plasmid system. (A) Adjustment of the expression of Cas12a and crRNA with different promoters. (B) The strength of the tested endogenous promoters in *C. symbiosum* ATCC 14940. The LacZ activity was used to indicate promoter strength. Control: no promoter driving the *lacZ* expression. (C) Editing efficiencies of the transformed cells that harbored different plasmid pairs. Data are represented as mean ± SD ($n = 3$).

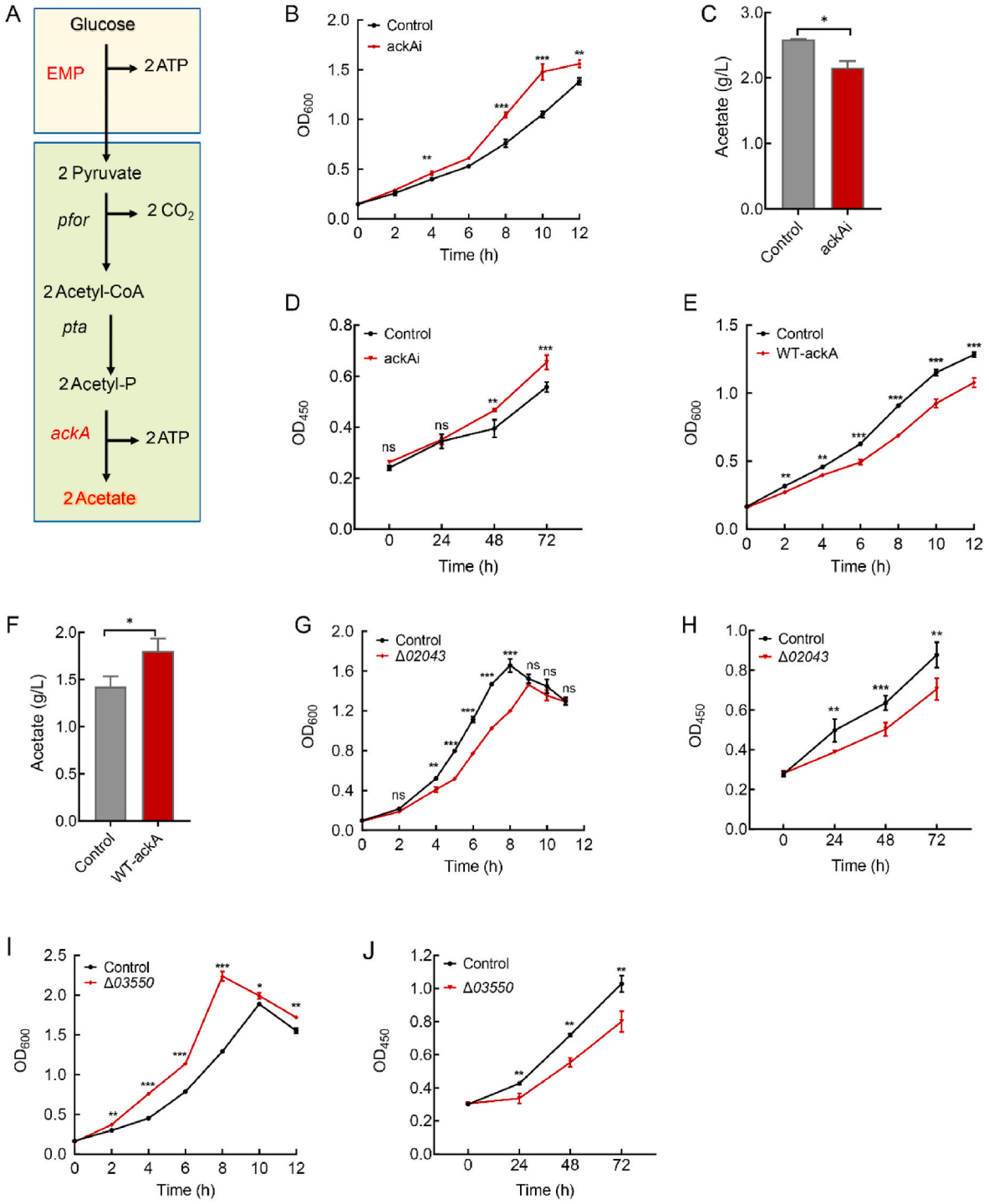

Fig. 7. The influence of the downregulation of *ack* and the deletion of CSYM_02043/CSYM_03550 on the performance of *C. symbiosum* ATCC 14940. (A) Schematic diagram of the acetate synthesis pathway. (B–D) The changes in growth, acetate synthesis, and virulence of *C. symbiosum* ATCC 14940 after the repression of *ackA*. Control: the wild-type strain harboring the plasmid that carried ddCas12a but no crRNA; ackAi: the wild-type strain containing the CRISPRi plasmid pMTLddcas12a-ackAi. (E, F) The changes in the growth and acetate synthesis after *ackA* overexpression. Control: the wild-type strain containing the plasmid pMTL82151; WT-ackA: the wild-type strain containing the pMTL82151-P_{1339}-ackA plasmid. (G, H) The changes in the growth and virulence of *C. symbiosum* ATCC 14940 after CSYM_02043 deletion. Control: the wild-type strain; Δ*02043*: the CSYM_02043-deleted strain. (I, J) The changes in the growth and virulence of *C. symbiosum* ATCC 14940 after CSYM_03550 deletion. Control: the wild-type strain; Δ*03550*: the CSYM_03550-deleted mutant. Data are represented as mean ± SD ($n = 3$). Statistical analysis was performed by a two-tailed Student's *t*-test. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, no significance.

CRediT authorship contribution statement

Pengjie Yang: Investigation, Validation, Writing – original draft. **Jinzhong Tian:** Methodology. **Lu Zhang:** Resources. **Hui Zhang:** Formal analysis. **Gaohua Yang:** Investigation. **Yimeng Ren:** Resources. **Jingyuan Fang:** Conceptualization, Resources. **Yang Gu:** Conceptualization, Writing – review & editing, Supervision, Project administration. **Weihong Jiang:** Conceptualization, Writing – original draft, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (2018YFA0901500), Science and Technology Commission of Shanghai Municipality (21DZ1209100), DNL Cooperation Fund, CAS (DNL202013), and Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-KJGG-016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2023.12.005.

References

[1] Rajilić-Stojanović M, de Vos WM. The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev 2014;38(5):996–1047. https://doi.org/10.1111/1574-6976.12075.
[2] Afzaal M, Saeed F, Shah YA, Hussain M, Rabail R, Socol CT, et al. Human gut microbiota in health and disease: unveiling the relationship. Front Microbiol 2022;13:999001. https://doi.org/10.3389/fmicb.2022.999001.
[3] Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms 2019;7(1):14. https://doi.org/10.3390/microorganisms7010014.
[4] Elsayed S, Zhang KY. Bacteremia caused by *Clostridium symbiosum*. J Clin Microbiol 2004;42(9):4390–2. https://doi.org/10.1128/JCM.42.9.4390-4392.2004.
[5] Decousser J, Bartizel C, Zamni M, Fadel N, Doucet Populaire F. *Clostridum symbiosum* as a cause of bloodstream infection in an immunocompetent patient. Anaerobe 2007;13(3–4):166–9. https://doi.org/10.1016/j.anaerobe.2007.04.003.
[6] Toprak NU, Özcan ET, Pekin T, Yumuk PF, Soyletir G. Bacteraemia caused by *Clostridium symbiosum*: case report and review of the literature. Indian J Med Microbiol 2014;32(1):92. https://doi.org/10.4103/0255-0857.124343.
[7] Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory bowel disease. Nature reviews. Nat Rev Microbiol 2019;17(8):497–511. https://doi.org/10.1038/s41579-019-0213-6.
[8] Liu HH, Tian R, Wang H, Feng SQ, Li HY, Xiao Y, et al. Gut microbiota from coronary artery disease patients contributes to vascular dysfunction in mice by regulating bile acid metabolism and immune activation. J Transl Med 2020;18(1):382. https://doi.org/10.1186/s12967-020-02539-x.
[9] Xie YH, Gao QY, Cai GX, Sun XM, Zou TH, Chen HM, et al. Fecal *Clostridium symbiosum* for noninvasive detection of early and advanced colorectal cancer: test and validation studies. EBioMedicine 2017;25:32–40. https://doi.org/10.1016/j.ebiom.2017.10.005.
[10] Jin WB, Li T-T, Huo D, Qu S, Li XV, Arifuzzaman M, et al. Genetic manipulation of gut microbes enables single-gene interrogation in a complex microbiome. Cell 2022;185(3):547–62. https://doi.org/10.1016/j.cell.2021.12.035. e522.
[11] Huang H, Chai C, Li N, Rowe P, Minton NP, Yang S, et al. CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. ACS Synth Biol 2016;5(12):1355–61. https://doi.org/10.1021/acssynbio.6b00044.
[12] Zhang L, Liu YQ, Yang YP, Jiang WH, Gu Y. A novel dual-cre motif enables two-way autoregulation of CcpA in *Clostridium acetobutylicum*. Appl Environ Microbiol 2018;84(8):e00114. https://doi.org/10.1128/AEM.00114-18. 00118.
[13] Girbal L, Mortier-Barriere I, Raynaud F, Rouanet C, Croux C, Soucaille P. Development of a sensitive gene expression reporter system and an inducible promoter-repressor system for *Clostridium acetobutylicum*. Appl Environ Microbiol 2003;69(8):4985–8. https://doi.org/10.1128/AEM.69.8.4985-4988.2003.
[14] Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 2005;33(suppl_1):D325–8. https://doi.org/10.1093/nar/gki008.
[15] Waller MC, Bober JR, Nair NU, Beisel CL. Toward a genetic tool development pipeline for host-associated bacteria. Curr Opin Microbiol 2017;38:156–64. https://doi.org/10.1016/j.mib.2017.05.006.
[16] Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, et al. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Mol Microbiol 2002;46(2):439–52. https://doi.org/10.1046/j.1365-2958.2002.03134.x.
[17] Abraham LJ, Rood JI. Identification of Tn4451 and Tn4452, chloramphenicol resistance transposons from *Clostridium perfringens*. J Bacteriol 1987;169(4):1579–84. https://doi.org/10.1128/jb.169.4.1579-1584.1987.
[18] Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for *Clostridium* shuttle plasmids. J Microbiol Methods 2009;78(1):79–85. https://doi.org/10.1016/j.mimet.2009.05.004.
[19] Zhang Y, Xu S, Chai CS, Yang S, Jiang WH, Minton NP, et al. Development of an inducible transposon system for efficient random mutagenesis in *Clostridium acetobutylicum*. FEMS Microbiol Lett 2016;363(8):fnw065. https://doi.org/10.1093/femsle/fnw065.
[20] Tan Y, Liu JJ, Chen XH, Zheng HJ, Li FL. RNA-seq-based comparative transcriptome analysis of the syngas-utilizing bacterium *Clostridium ljungdahlii* DSM 13528 grown autotrophically and heterotrophically. Mol Biosyst 2013;9(11):2775–84. https://doi.org/10.1039/c3mb70232d.
[21] Zhang L, Liu YQ, Zhao R, Zhang C, Jiang WH, Gu Y. Interactive regulation of formate dehydrogenase during CO_2 fixation in gas-fermenting bacteria. mBio 2020;11(4):e00650. https://doi.org/10.1128/mbio.00650-20. 00620.
[22] Yang YP, Zhang L, Huang H, Yang C, Yang S, Gu Y, et al. A flexible binding site architecture provides new insights into CcpA global regulation in Gram-positive bacteria. mBio 2017;8(1). https://doi.org/10.1128/mBio.02004-16. 10.1128/mbio. 02004-02016.
[23] Wen Z, Qian F, Zhang J, Jiang Y, Yang S. Genome editing of *Corynebacterium glutamicum* using CRISPR-cpf1 system. Methods Mol Biol 2022;2479:189–206. https://doi.org/10.1007/978-1-0716-2233-9_13.
[24] Paul B, Montoya G. CRISPR-Cas12a: functional overview and applications. Biomed J 2020;43(1):8–17. https://doi.org/10.1016/j.bj.2019.10.005.
[25] Weaver LH, Kwon K, Beckett Dmatthews BW. Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. Proc Natl Acad Sci USA 2001;98(11):6045–50. https://doi.org/10.1073/pnas.111128198.
[26] Soares da Costa TP, Tieu W, Yap MY, Pendini NR, Polyak SW, Pedersen DS, et al. Selective inhibition of biotin protein ligase from *Staphylococcus aureus*. J Biol Chem 2012;287(21):17823–32. https://doi.org/10.1074/jbc.M112.356576.
[27] Zhang C, Nie XQ, Zhang H, Wu YW, He HQ, Yang C, et al. Functional dissection and modulation of the BirA protein for improved autotrophic growth of gas-fermenting *Clostridium ljungdahlii*. Microb Biotechnol 2021;14(5):2072–89. https://doi.org/10.1111/1751-7915.13884.
[28] Tian JZ, Yang GH, Gu Y, Sun XQ, Lu YH, Jiang WH. Developing an endogenous quorum-sensing based CRISPRi circuit for autonomous and tunable dynamic regulation of multiple targets in *Streptomyces*. Nucleic Acids Res 2020;48(14):8188–202. https://doi.org/10.1093/nar/gkaa602.
[29] Zhang X, Wang J, Cheng Q, Zheng X, Zhao G, Wang J. Multiplex gene regulation by CRISPR-ddCpf1. Cell discovery 2017;3:17018. https://doi.org/10.1038/celldisc.2017.18.
[30] Meliawati M, Schilling C, Schmid J. Recent advances of Cas12a applications in bacteria. Appl Microbiol Biotechnol 2021;105(8):2981–90. https://doi.org/10.1007/s00253-021-11243-9.
[31] Heinemann JA, Sprague Jr GF. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature 1989;340(6230):205–9. https://doi.org/10.1038/340205a0.
[32] Heinemann JA, Sprague Jr GF. Transmission of plasmid DNA to yeast by conjugation with bacteria. Methods Enzymol 1991;194:187–95. https://doi.org/10.1016/0076-6879(91)94016-6.
[33] Wasels F, Jean-Marie J, Collas F, López-Contreras AM, Ferreira N. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. J Microbiol Methods 2017;140:5–11. https://doi.org/10.1016/j.mimet.2017.06.010.
[34] Diallo M, Hocq R, Collas F, Chartier G, Wasels F, Wijaya HS, et al. Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*. Methods 2020;172:51–60. https://doi.org/10.1016/j.ymeth.2019.07.022.
[35] Guo CJ, Allen BM, Hiam KJ, Dodd D, Van Treuren W, Higginbottom S, et al. Depletion of microbiome-derived molecules in the host using *Clostridium* genetics. Science 2019;366(6471):eaav1282. https://doi.org/10.1126/science.aav1282.
[36] Wells JM, Rossi O, Meijerink M, van Baarlen P. Epithelial crosstalk at the microbiota–mucosal interface. Proc Natl Acad Sci USA 2011;108(supplement_1):4607–14. https://doi.org/10.1073/pnas.1000092107.
[37] Marques C, Oliveira CS, Alves S, Chaves SR, Coutinho OP, Côrte-Real M, et al. Acetate-induced apoptosis in colorectal carcinoma cells involves lysosomal membrane permeabilization and cathepsin D release. Cell Death Dis 2013;4(2):e507. https://doi.org/10.1038/cddis.2013.29.

[38] Fang YK, Yan C, Zhao Q, Xu JM, Liu ZZ, Gao J, et al. The roles of microbial products in the development of colorectal cancer: a review. Bioengineered 2021;12 (1):720–35. https://doi.org/10.1080/21655979.2021.1889109.
[39] Mistou MY, Sutcliffe IC, Van Sorge NM. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in gram-positive bacteria. FEMS Microbiol Rev 2016;40(4):464–79. https://doi.org/10.1093/femsre/fuw006.
[40] Koskiniemi S, Lamoureux JG, Nikolakakis KC, t'Kint de Roodenbeke C, Kaplan MD, Low DA, et al. Rhs proteins from diverse bacteria mediate intercellular competition. Proc Natl Acad Sci USA 2013;110(17):7032–7. https://doi.org/10.1073/pnas.1300627110.
[41] Jamet A, Nassif X. New players in the toxin field: polymorphic toxin systems in bacteria. mBio 2015;6(3):e00285. https://doi.org/10.1128/mBio.00285-15. 00281.
[42] Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform 2009 Mar;Chapter 4:4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.
[43] Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 1999;27(2):573–80. https://doi.org/10.1093/nar/27.2.573.
[44] Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 2008; 24(5):637–44. https://doi.org/10.1093/bioinformatics/btn013.
[45] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10. https://doi.org/10.1016/S0022-2836(05) 80360-2.
[46] Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 2008;9(1):R7. https://doi.org/10.1186/gb-2008-9-1-r7.