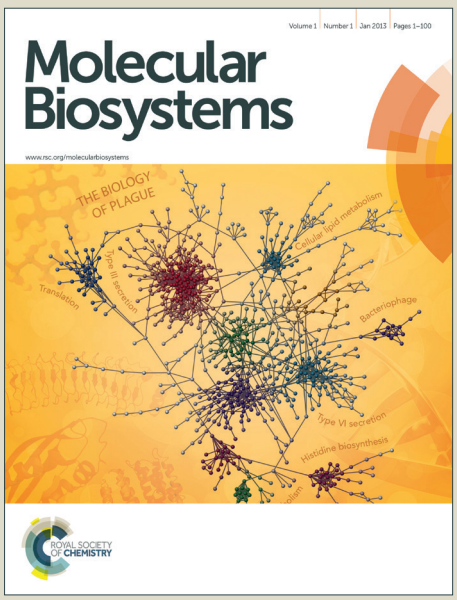
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Comparative transcriptome analysis between *csrA*-disruption *Clostridium acetobutylicum* and its parent strain

Yang Tan[1,3], Zi-Yong Liu[1], Zhen Liu[1], Hua-Jun Zheng[2] and Fu-Li Li[1,*]

1. Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China

2. Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China

3. University of Chinese Academy of Sciences, Beijing 100039, China

*Corresponding author: Fu-Li Li. Mailing address: Institute of Bioenergy and Bioprocess Technology, CAS, Songling Road No. 189, Qingdao 266101, China; Phone: +86 532-80662655; Fax: +86 532-80662778; Email: lifl@qibebt.ac.cn

Abstract

The genome of *Clostridium acetobutylicum* contains the gene encoding CsrA, a carbon storage regulator. We investigated the function of CsrA in *C. acetobutylicum* by insertionally inactivating the encoding gene, CA_C2209 using the ClosTron. Disruption of *csrA* obviously decreases the growth of the organism and reduces the yield of acetone, butanol and ethanol (ABEs). Like the *csrA* in *Escherichia coli*, RNA-seq and β-galactosidase analysis revealed that *csrA* in *C. acetobutylicum* was closely involved in regulating multiple pathways including; flagella assembly, oligopeptide transporting, iron uptake, and central carbon metabolism. It has also been newly demonstrated that *csrA* in *C. acetobutylicum* is related to the regulation of pathways involved in the phosphotransferase transporting systems, synthesis of riboflavin, and stage III of sporulation. This research represented first investigation on global regulation by CsrA in the strain belonging to Gram-positive bacteria through transcriptome analysis and provided the important theoretical evidence for improving solvent production by transcriptor engineering in *C. acetobutylicum*.

Keywords: *Clostridium acetobutylicum*; CsrA; RNA-seq; global regulation; butanol

Introduction

Originally identified in *Escherichia coli*, Csr (carbon starvation regulator) or its orthologue, Rsm (repressor of stationary-phase metabolites) is a post-transcriptional regulator which binds to conserved sequences in its target mRNAs and positively or negatively regulates the translation and/or stability of target transcripts. In this way, CsrA activates exponential growth phase processes while repressing several stationary growth phase processes.[1-3]

CsrA is found to be widely distributed among eubacteria.[4, 5] The CsrA pathway and the mechanism of regulation have been studied extensively in Gram-negative bacteria. In *E. coli,* CsrA is involved in regulating central carbon metabolism [5-10], biofilm formation [11-13], motility [8, 10, 14], cyclic di-GMP synthesis [15], peptide uptake [7], the stringent response [1] and quorum sensing.[16] In many other Gram-negative pathogenic bacteria, the regulating function of CsrA has also been studied, revealing that CsrA or RsmA not only regulates stationary-phase metabolism but also is an important regulator of virulence determinants including host cell invasion, quorum sensing, biofilm formation, iron acquisition, type III secretion systems, and outer membrane protein expression.[17-36] In *E. coli*, the CsrA regulation system has been manipulated to modulate the expression of target genes for metabolic engineering application.[9, 37-39] In the Gram-positive bacteria, CsrA has been reported to regulate the assembly of flagella in *Bacillus subtilis*.[40-42]

In *E. coli*, the regulating mechanism of CsrA has been elucidated. CsrA is a homodimeric RNA-binding protein that typically binds the 5' untranslated regions of target mRNA at sites characterized by a GGA sequence, which is often located within the loop of a short stem-loop structure.[43-45] These sites are often adjacent to or overlapping the ribosome binding site, thus inhibiting ribosome access to the

ribosome binding site and inhibiting translation initiation, which can either increase or decrease mRNA half-life.[3, 7, 13, 46-49] CsrA is regulated by two noncoding RNAs, CsrB and CsrC, containing multiple CsrA binding sites, which permit them to sequester and antagonize CsrA.[3, 29, 50-52] The expression of CsrB and CsrC is activated by the BarA/UvrY two-component system.[51, 53] However, in Gram-positive bacteria such as bacillus, there is no report about the presence of the noncoding RNAs like CsrB and CsrC, which interact with the CsrA.

Clostridium is a genus of Gram-positive bacteria, including many species for commercial use such as *Clostridium acetobutylicum, Clostridium thermocellum* and *Clostridium ljungdahlii,* for the production of biofuels and chemicals. The genus also comprises of many human pathogens such as *Clostridium botulinum*, *Clostridium difficile* and *Clostridium perfringens*. *C. acetobutylicum*, isolated from soil, has been used to develop an industrial starch-based acetone, butanol, and ethanol (ABEs) fermentation process.[54] The *C. acetobutylicum* ATCC 824 strain is a model organism for *Clostridium* and is used in a variety of molecular biology and metabolic engineering studies.[54-67]

CsrA homologues are widely distributed in clostridial genomes. However, until now there is no report about the function of CsrA in the genus. In this study, we investigated the global effects of CsrA (encoded by CA_C2209) in *C. acetobutylicum* by RNA-seq. This study represented the first attempt to investigate the global regulation of CsrA through transcriptome analysis in Gram-positive bacteria.

Materials and methods

Bacterial strains and media

Bacterial strains utilized in this study are listed in Table 1. *C. acetobutylicum* ATCC

824 and its mutant derivatives were grown in Clostridial growth medium (CGM) [68] for routine manipulations. P2 medium was used for fermentation.[69] *Escherichia coli* strains were grown in Luria-Bertani medium (LB). Appropriate antibiotics were used when needed (erythromycin, 5 μg/ml and thiamphenicol, 15 μg/ml for *C. acetobutylicum*; and chloramphenicol, 25 μg/ml and tetracycline, 10 μg/ml for *E. coli*).

Disruption of the *csrA* gene and its complementation in *C. acetobutylicum*

The plasmid pMTL007C-E2 was used for the construction of retargeted ClosTron plasmids.[70] Table S1 lists the primers for the disruption of *csrA* designed by the Perutka algorithm and used for plasmid construction. The plasmid pMTL007C-E2::Cac-csrA-97s was methylated in *E. coli* Top10 (pAN2) [70] and then electroporated into *C. acetobutylicum* ATCC 824 as described previously.[71] Following the introduction of the retargeted ClosTron plasmids into *C. acetobutylicum*, integrants were selected on the basis of acquisition of erythromycin resistance. The integrants were screened by PCR using the screening primers. Finally, all isolates were sequenced using the primers csrA-ID-for and csrA-ID-rev to confirm integration of the group II intron at the correct location in the genome and Southern Blot analysis was carried out to ensure that only a single insertion of the intron occurred. For complementation, a 494 bp fragment encompassing the gene and its promoter was amplified using primers Cac-csrA-NotI/F and Cac-csrA-NcoI/R. This fragment was cloned into pMTL85141.[72] The resulting construct was introduced into the Cac-*csrA*97s::CT.

Southern Blot analysis

Genomic DNA of the wild type strain and Cac-*csrA*97s::CT strain was subjected to *Hin*dIII restriction enzyme digestion. Southern hybridization was carried out

according to the instructions provided by the Roche DIG High Prime DNA Labeling (Roche, Shanghai, China). Probes designed to hybridize to the inserted intron of each mutant were generated by PCR using primers SalI-R1 and Cac-csrA-97s-EBS2.

Fermentation

Fermentation experiments were performed anaerobically in P2 medium and carried out in a 100 mL working volume in 250 mL serum bottles with three replicates. Growth was monitored by measuring the optical density at OD_{600nm}. The concentration of the fermentation products in culture supernatants were analyzed by High Performance Liquid Chromatography as previously described.[73]

RNA isolation and RNA-seq

When the wild type and Cac-*csrA*97s::CT were grown in late exponential growth phase with an OD_{600nm} of 8–9, 20 mL of cell cultures were collected by centrifugation at 4,000 ×*g* for 10 min at 4°C. All the samples were immediately frozen in liquid nitrogen and stored at –80°C before RNA extraction. Total RNA was extracted using RNAprep pure Cell/Bacteria Kit (Tiangen Biotech Co., Ltd, Beijing, China). The concentration of total RNA was determined using the NanoVue spectrophotometer (GE Healthcare Bioscience Corp., Piscataway, USA) and the RNA integrity value (RIN) was checked using the RNA 6000 Pico LabChip on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Ten microgram RNA used for Hiseq2000 (Illumina Inc., San Diego, USA) was treated with 25 U DNase I (Roche, Mannheim, Germany) at 37°C for 1 h to remove DNA. Bacterial 16S and 23S ribosomal RNA were removed with a MICROBExpress™ kit (Ambion Inc., Austin, TX, USA). The mRNA was fragmented ultrasonically and the fragmented mRNA was converted to a RNA-Seq library using the mRNA-Seq library construction kit

(Illumina Inc.) according to the manufacturer's instruction. RNA 2×100 bp paired-end sequencing was performed on the Illumina Hiseq2000, according to the manufacturer's protocol.

RNA-Seq data analysis

The generated 100 bp reads were mapped to the *C. acetobutylicum* ATCC 824 genome using Bowtie with default parameter, and those that did not align uniquely to the genome were discarded. [74] Raw read counts of each gene in the two samples were imported into DEGseq. Then, the MARS model (MA-plot-based method with the random sampling model) in the DEGseq package was used to calculate the expression abundance of each gene between the two samples. [75] DEGseq will transform the raw read counts into RPM （Reads Per Million reads）when calculating. FDR (false discovery rate) was used to determine the threshold of the *p*-value for this analysis. In both sets of comparisons (Cac-*csrA*97s::CT *vs.* the wild type), the expression of genes showed significant difference at FDR < 0.001 and |Normalized Fold_-change| ≥2. For the identification of the pathways that the DEGs were predicted to participate in, we mapped all DEGs to terms in the KEGG database and looked for significantly enriched KEGG terms.

Quantitative real-time PCR (qRT-PCR) verification

Total RNA used for RNA-seq was treated with the DNase I and quantitative RT-PCR was performed on RNA samples to confirm the absence of DNA contamination using 16S primers.[76] The cDNA synthesis was performed according to the instruction (Tiangen Biotech Co.). The genes and primers used for qRT-PCR are shown in Table S1. The relative gene expression data were analyzed using the $2^{-\Delta\Delta Ct}$ method as described.[77] The pullulanase gene (CA_C2679) was used as an internal control according to the previous report [78], and the RNA-seq result here also indicated that

CA_C2679 had a constant expression profile in both the WT and Cac-*csrA*97s::CT strains.

Software was used for the analysis of the data (Roche Applied Science, Indianapolis, USA). All quantitative PCRs were repeated in three biological and three technical replications.[79]

The construction of plasmids for expression analysis

The vector pIMP-GFP was digested with restriction endonuclease *Nhe*I and *Nco*I for the deletion of GFP. The *lac*Z gene was amplified from the genome DNA of *Thermoanaerobacterium thermosulfurigenes* with the primers, lacZ_NheI/F and lacZ_ NcoI/R. The generated PCR fragment was used to replace GFP in the plasmid pIMP-GFP for constructing the pIMP-lacZ plasmids. The pIMP-lacZ was subsequently digested with restriction endonuclease *Pst*I and *Nhe*I for the deletion of the *thl* promoter[80]. The corresponding promoters of selected genes for the assay of β-galactosidase were amplified from the genome DNA of *C. acetobutylicum* ATCC 824 with the primers, Cac-csrA-Not/F and Cac-csrA-Not/R (Table S1). The resulting fragment carrying *Pst*I and *Nhe*I restriction sites was ligated into the plasmid pIMP-*lac*Z digested by restricted endonucleases *Pst*I and *Nhe*I.

β-galactosidase assays

The β-galactosidase assay was performed spectrophotometrically with o-nitrophenyl-β-D-galactopyranoside, as described previously. [24]

Motility assays

The wild-type and Cac-*csrA*97s::CT strains were grown anaerobically overnight, and 10 μl of each was spotted onto semi-solid CGM agar (0.35%). Plates were incubated at 37°C anaerobically and the motility for both of the wild-type and the Cac-*csrA*97s::CT strain was assessed after 48 h culture.

Transmission electron microscopy

The wild-type and Cac-*csrA*97s::CT cells grown in CGM medium were placed on 400-mesh carbon coated copper grids, and incubated for 5 min for adsorption of the sample to the grid. The grids were then stained with 2% phosphotungstic acid. The samples were observed in a Hitachi H-7650 transmission electron microscopy (Hitachi, Tokyo, Japan).

Spore assays

Spore assays were performed as described by Steiner et al.[81] Briefly, C. acetobutylicum strains were grown in 5 ml CBM (Clostridial Basal Medium) [82], which contained 0.5% $CaCO_3$ and 5% glucose to enable sporulation. After 5 days growth, 200 µL sample of culture was heated to 80°C for 10 min. The samples were subjected to serial dilutions and 10 aliquots of the heat-treated cell suspension were spread onto CBM agar. After 24 h of incubation colonies were counted. This experiment is designed to quantify the number of heat resistant CFU (colony forming unit) as a measure for spores, which are able to germinate and grow after a heat treatment for 10 min at 80°C. Three experimental replicates for both the wild type and Cac-*csrA*97s::CT strain were performed.

Protein database search

The amino acids of the related enzymes in the study were obtained by performing BLASTp searches (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/BLAST) using the default parameters. The multiple sequence alignment was performed with the ClustalW program using the standard parameters. The phylogenetic tree was constructed with the neighbour-joining method with Mega 5.0 using the default parameters. Bootstrap values were calculated based

on 1000 replicates.

Nucleotide sequence accession numbers

All RNA-seq data for wild type and Cac-*csrA*97s::CT strains have been submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) under the accession numbers SRR1184237 and SRR1184230.

Result and discussion

Phylogenetic tree analysis of CsrA

CsrA proteins have been discovered to be small RNA-binding proteins, globally regulating primary and secondary metabolic pathways, biofilm formation, motility, virulence circuitry of pathogens, quorum sensing and stress response systems by binding to conserved mRNA targets and altering their translation and/or turnover.[2] Although CsrA and their homologues widely exist in the eubacterial domain, the functions that they regulate are quite diverse. In the genome of the genus *Clostridium*, CsrA homologues are widely distributed. The phylogenetic tree, shown in Fig 1, was constructed by Mega 5.0, using the deduced amino acid sequences of *csrA* genes derived from the genomes of *Clostridium acetobutylicum* and 16 other strains. Fig. 1 shows that the CsrAs from Gram-positive and Gram-negative strains are clustered in individual branches. In the Gram-positive branch, the CsrA of *C. acetobutylicum* shared 46.4% amino acid identity with the CsrA of *B subtilis*, which is reported to be involved in the assembly of flagella.[40] However, the two CsrAs are obviously divided in the phylogenetic tree. It is predicted that the regulatory function of CsrA in *C. acetobutylicum* will be different from CsrA in *B. subtilis*.

The disruption of *csrA* and comparison of the growth and fermenting profile between wild type and Cac-*csrA*97s::CT

Using the ClosTron system, *csrA* was insertionally inactivated and the mutant was confirmed to have a single intron insertion into the genome by Southern hybridization (Fig. S1). The growth profiles between wild type and Cac-*csrA*97s::CT strains were compared and it was shown that the growth of Cac-*csrA*97s::CT strain was retarded. The maximum OD_{600nm} value of Cac-*csrA*97s::CT strain was achieved at 40 hours, and that of wild-type strain was obtained at 34 hours (Fig. 2). Although the highest OD_{600nm} of the two strains reached OD_{600nm} ~8 , Cac-*csrA*97s::CT strain cell density decreased much faster. At the end of fermentation, the OD_{600nm} for Cac-*csrA*97s::CT strain was 2.2 and the wild type's remained 5.1.

It was also noted that the ABE final production showed significant difference between the wild type and the Cac-*csrA*97s::CT strain ($P<0.01$). The yields of butanol, acetone and ethanol in the Cac-*csrA*97s::CT strain are lower than the ones for wild type. The total ABEs in the Cac-*csrA*97s::CT strain is 24.6% lower than the one for the wild type strain (12.8 g/l *VS* 17.0g/L). There is no significant difference in the maximum yield of acetate and butyrate between the Cac-*csrA*97s::CT and wild type strain (acetate, 0.93 g/l *VS* 0.89 g/l. butyrate, 2.57 g/l *VS* 2.66 g/l) (Fig. 2). In conclusion, the results suggest that the *csrA* gene plays a very important role in maintaining the growth and production of ABEs in *C. acetobutylicum*.

Profiling the transcriptome of Cac-*csrA*97s::CT

RNA-seq was performed to analyze the global transcriptional regulation by CsrA. The statistical data for the transcriptomes of wild type and Cac-*csrA*97s::CT were summarized in Table 2. More than 3 Million high-quality base pairs reads were generated for both samples. In addition, saturation analysis showed that sequencing became saturated when the number of reads reached 1 Million for both of the samples (Fig. 3). In total, more than 3,800 protein coding genes were covered, which equate to

98.7% of the total 3,847 genes. It is demonstrated that quality of RNA sequencing is up to the requirement for the following analysis. Under the two conditions, a total of 552 differentially-expressed genes; including 240 upregulated genes and 312 downregulated genes were identified. Among these differentially expressed genes, there were a total of 216 genes (96 upregulated genes and 120 downregulated genes) which can be clustered into the specific pathways. Carbohydrate metabolism and amino acid metabolism included the most differentially expressed genes, 46 and 39 genes, respectively. Energy metabolism (16 genes), metabolism of cofactors and vitamins (13 genes), membrane transport (19 genes) and cell motility (20 genes) had the intermediate number of differentially expressed genes. The remaining 18 pathways contain less than 10 differentially expressed genes (Fig. 4).

Regulation of cell motility

It has been reported that CsrA is involved in the synthesis of flagella in *E. coli* [2], *Salmonella typhimurium* [24] and *B. subtilis*.[40] In the genome of *C. acetobutylicum,* the genes related to the assembly of flagella were clustered in the operon CA_C2139-CA_C2167 and the expression for most of these genes were downregulated in the Cac-*csrA*97s::CT strain (1.0–4.4-folds) (Table S2). To test the regulation of flagella synthesis by CsrA, we used a *lacZ* fusion in the flagella operon to measure β-galactosidase expression. Lower activity of β-galactosidase was observed in the Cac-*csrA*97s::CT strain (Fig. 5A), but functional complementation of *csrA* on a plasmid restored *lacZ* expression of the mutant to wild-type levels.

The motility analysis of wild type and the mutant showed that the Cac-*csrA*97s::CT strain lost its motility on the semi-solid agar plate (Fig. 6) with TEM demonstrating that the flagella is not present in the Cac-*csrA*97s::CT strain (Fig. 7). Like the CsrA in *E. coli*, *S. typhimurium* and *B. subtilis*, it is shown that CsrA in *C. acetobutylicum* is

also involved in the regulation of flagella synthesis.

Regulation of genes associated with the synthesis of riboflavin

Riboflavin, also known as vitamin B2, is the central component for the synthesis of the cofactors FAD and FMN. Vitamin B2 is required for a wide variety of cellular processes and is known to play a key role in energy metabolism, as well as the metabolism of fats, ketone bodies, carbohydrates, and proteins.[83] In the genome of *C. acetobutylicum,* one operon (CA_C0590-CA_C0593) encodes the proteins responsible for synthesis of vitamin B2. However, it was shown that the transcriptions of the genes within this operon were weakened (1.1–1.4-folds) in the Cac-*csrA*97s::CT strain (Table 3) and that the activity of β-galactosidase was shown to be higher in wild type cells than in the Cac-*csrA*97s::CT strain (Fig. 5B). This indicates that the pathway for vitamin B2 synthesis is more active in the wild type cells than in the Cac-*csrA*97s::CT strain. This result was also consistent with the retarded growth profile of the Cac-*csrA*97s::CT strain. To our knowledge, this is the first report that CsrA is involved in the regulation of vitamin B2 synthesis.

Membrane transporting system

In *E. coli*, CsrA regulates the expression of genes involved in oligopeptide transporting and iron uptaking.[2] In the genome of *C. acetobutylicum*, there are 6 operons encoding oligopeptide ABC transporters (CA_C3634–CA_C3638, CA_C3628–CA_C3633, CA_C3641–CA_C3644, CA_C0177–CA_C0180, CA_C2374–CA_C2377 and CA_C3179–CA_C3183). Among these operons, the expression of only one operon (CA_C3634–CA_C3638) was upregulated and for the rest, the expression trend was unaltered. In addition, the operon (CA_C1029-CA_C1032) responsible for iron uptake was affected by the disruption of *csrA,* and the genes within this operon were downregulated about 1.2–2.0-fold (Table S3). It is

shown that CsrA in *C. acetobutylicum* is also related to the expression of genes for the oligopeptide transporting and iron uptaking.

C. acetobutylicum is able to utilize various monosaccharides and disaccharides, starches [54]. The phosphotransferase system (PTS) plays a very important role in the utilization of these substrates. Compared with the wild type, the expression of two operons encoding the PTS, fructose (mannose)-specific transporter (CA_C1457–CA_C1463) and the galactitol/fructose specific transporter (CA_C2956–CA_C2958) were upregulated (1.4–3.3-folds), and the expression of the operon encoding fructose-specific transporter (CA_C0231–CA_C0234) were downregulated (1.3–1.9-folds) (Table S3) in the mutant. Analysis of the expression of the operon encoding the fructose (mannose)-specific transporter (CA_C1457–CA_C1463) through *lacZ* fusion and subsequent β-galactosidase production showed that the activity of β-galactosidase is about 5-folds higher in the Cac-*csrA*97s::CT strain than that in the wild type (Fig 5D). This indicated that CsrA is also involved in regulating the expression of genes encoding the PTS.

Stage III sporulation

C. acetobutylicum, which is isolated from soil, is an important industrial microorganism for production of solvents including; acetone, butanol, and ethanol (ABEs). The fermentation of sugars by clostridia is typically characterized by three different growth phases: exponential growth and acids formation, followed by stationary growth phase with reassimilation of acids and solvent production, and finally endospore formation. The role of the well-known sporulation regulator Spo0A of Gram-positive bacteria has been found to be crucial for the endospore initiation and solventogenesis, but in regards to the complex regulation networks, more regulatory factors might be involved.[84]

Spore formation has an impact on the production of ABEs. Comparative transcriptome analysis demonstrated that the expression of operons related to spore stage III sporulation (CA_C2086–CA_C2093) and spore coat proteins (CA_C1336–CA_C1338 and CA_C2906–CA_C2910) were significantly upregulated in the Cac-*csrA*97s::CT strain, and the fold change ranged from 1.7 to 2.3 (Table S4). Furthermore, β-galactosidase analysis for the expression of genes involved in the stage III of the sporulation cascade also demonstrated that the activity of β-galactosidase is much higher in the Cac-*csrA*97s::CT strain than in wild-type (Fig. 5C). However, the spore assay demonstrated that the number of spores in the Cac-*csrA*97s::CT strain did not show a significant difference compared to the wild-type strain (3.43×10^7 cfu/mL *Vs* 3.35×10^7 cfu/mL, respectively). These results showed that inactivition of *csrA* did not repress the formation of spores in *C. acetobutylicum*, but still suggests CsrA is involved in spore formation. The study about the regulation by CsrA has been almost limited to some strains belonging to the Gram-negative bacteria, which do not produce spore. In *C. acetobutylicum,* CsrA is involved in the expression of stage III of the sporulation cascade, this shows CsrA diverse regulation function in gram-positive bacteria compared with Gram-negative bacteria such as *E. coli*.

Carbon central metabolism

CsrA has been originally reported to influence the expression of genes related to the carbon central metabolism, and its effect has been verified in the *E. coli, P. aeruginosa* and *S. typhimurium*.[2] The expression of some genes involved in the central carbon metabolism has been significantly altered, which include the upregulated genes (CA_P0064, CA_C1347, CA_C1341 and CA_C1342) and downregulated genes (CA_C3021, CA_C0712, CA_C2458, CA_C2018, CA_C2612

and CA_C2973) (Table S5). It is predicted that CsrA also participates in the regulation of genes involved in the central carbon metabolic pathway in *C. acetobutylicum*.

In the *E. coli*, CsrA regulates the expression of genes involved in glycogen anabolism and catabolism. However, *C. acetobutylicum* does not synthesize glycose because the pathway for glycose synthesis is incomplete and the genome of *C. acetobutylicum* only contains two genes for glucose-1-phosphate adenylyltransferase (CA_C2237) and glycogen synthase (CA_C2239), and their expression did not show significant difference between the wild type and Cac-*csrA*97s::CT strain. In addition, CsrA is also involved in starch degradation, as there are two genes encoding alpha-amylase (CA_P0098 and CA_P0168) in the genome of *C. acetobutylicum*, the expression of CA_P0098 is downregulated in the Cac-*csrA*97s::CT and CA_P0168 showed stable expression profile. The assay for the activity of total amylase for the Cac-*csrA*97s::CT and WT was conducted as BASSAM et al. described[85]. It was shown that the activity in wild type strain is significantly higher than that in the Cac-*csrA*97s::CT strain (4.3 μ/mg *vs* 2.1 μ/mg, $P<0.01$).

Non-coding RNA

It has been previously shown that small non-coding RNA plays a very important role in global regulation of gene expresson.[86] CsrA regulates the expression of genes by influencing the stability of mRNA[2] and it is predicted that CsrA also participates in the expression of sRNA. Based on the methods in this study, it has been shown that there are in total 14 sRNAs found in both wild type and Cac-*csrA*97s::CT strain (Table 4). Four of these sRNAs show similar functions with Lysine (Lysine riboswitch), tmRNA (transfer-messenger RNA), 6S RNA, and OLE (Ornate Large Extremophilic RNA) families and a further ten sRNAs show no similarity with any

sRNA in the database. Among these 14 sRNAs, the expression of five (sRNA_01, sRNA_04, sRNA_05, sRNA_11 and sRNA_13) has been upregulated in the Cac-*csrA*97s::CT strain and the fold change ranged from 1.1 to 2.1 (Table 4). It is predicted that the CsrA also influences the physiology of the strain through regulating the expression of sRNA.

Verification of RNA-seq Data by qRT-PCR

The qRT-PCR was used to verify the reliability of RNA-seq expression data. In total, 10 genes involved in the flagella assembly, synthesis of riboflavin, membrane transporting system, sporulation and carbon central metabolism were selected for verification of RNA-seq data. Although the magnitude of gene changing folds varied between the two data sets, the qRT-PCR results displayed similar trends of up- or downregulation similar to those observed in RNA-seq results, supporting the validity of the RNA-seq data (Table S6).

Conclusion

Using *C. acetobutylicum* as the model organism, this study firstly investigated the global regulation of CsrA through transcriptome analysis in this Gram-positive bacterium. The Cac-*csrA*97s::CT mutant and wild type strain showed difference in physiology, such as a retarded growth profile, less ABE production and non-motility. Through comparative transcriptomes analysis and β-galactosidase activity we identified the pathways closely involved in CsrA regulation, which included cell motility, synthesis of riboflavin, membrane transporting system, sporulation and carbon central metabolism. This study enables us to more deeply understand the multitude of regulatory functions CsrA has in Gram-positive bacteria.

Acknowledgement

This study was supported by grants from the National Natural Science Foundation of China (30970049, 11311140293) and Distinguished Middle-Aged and Young Scientist Encourage and Reward Foundation of Shandong Province to ZL (BS2013CL029).

We thank David Walker, University of Massachusetts Amherst, for proofreading the manuscript.

Reference

1. A. N. Edwards, L. M. Patterson-Fortin, C. A. Vakulskas, J. W. Mercante, K. Potrykus, D. Vinella, M. I. Camacho, J. A. Fields, S. A. Thompson, D. Georgellis, M. Cashel, P. Babitzke and T. Romeo, *Mol Microbiol*, 2011, 80, 1561-1580.
2. T. Romeo, C. A. Vakulskas and P. Babitzke, *Environ Microbiol*, 2013, 15, 313-324.
3. T. Romeo, *Mol Microbiol*, 1998, 29, 1321-1330.
4. D. White, M. E. Hart and T. Romeo, *Gene*, 1996, 182, 221-223.
5. J. Mercante, K. Suzuki, X. Cheng, P. Babitzke and T. Romeo, *J Biol Chem*, 2006, 281, 31832-31842.
6. O. Revelles, P. Millard, J. P. Nougayrede, U. Dobrindt, E. Oswald, F. Letisse and J. C. Portais, *PLoS One*, 2013, 8, e66386.
7. A. K. Dubey, C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo and P. Babitzke, *J Bacteriol*, 2003, 185, 4450-4460.
8. N. A. Sabnis, H. Yang and T. Romeo, *J Biol Chem*, 1995, 270, 29096-29104.
9. M. Tatarko and T. Romeo, *Curr Microbiol*, 2001, 43, 26-32.
10. B. Wei, S. Shin, D. LaPorte, A. J. Wolfe and T. Romeo, *J Bacteriol*, 2000, 182, 1632-1640.
11. D. W. Jackson, K. Suzuki, L. Oakford, J. W. Simecka, M. E. Hart and T. Romeo, *J Bacteriol*, 2002, 184, 290-301.
12. K. Agladze, X. Wang and T. Romeo, *J Bacteriol*, 2005, 187, 8237-8246.
13. X. Wang, A. K. Dubey, K. Suzuki, C. S. Baker, P. Babitzke and T. Romeo, *Mol Microbiol*, 2005, 56, 1648-1663.

14. B. L. Wei, A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke and T. Romeo, *Mol Microbiol*, 2001, 40, 245-256.

15. K. Jonas, A. N. Edwards, R. Simm, T. Romeo, U. Romling and O. Melefors, *Mol Microbiol*, 2008, 70, 236-257.

16. H. Yakhnin, C. S. Baker, I. Berezin, M. A. Evangelista, A. Rassin, T. Romeo and P. Babitzke, *J Bacteriol*, 2011, 193, 6162-6170.

17. S. Bhatt, A. Anyanful and D. Kalman, *J Bacteriol*, 2011, 193, 4516-4522.

18. A. Brencic and S. Lory, *Mol Microbiol*, 2009, 72, 612-632.

19. A. Chatterjee, Y. Cui, P. Chakrabarty and A. K. Chatterjee, *Mol Plant Microbe Interact*, 2010, 23, 1316-1323.

20. A. Chatterjee, Y. Cui, Y. Liu, C. K. Dumenyo and A. K. Chatterjee, *Appl Environ Microbiol*, 1995, 61, 1959-1967.

21. J. A. Fields and S. A. Thompson, *J Bacteriol*, 2008, 190, 3411-3416.

22. J. A. Fields and S. A. Thompson, *BMC Microbiol*, 2012, 12, 233.

23. P. R. Kulkarni, X. Cui, J. W. Williams, A. M. Stevens and R. V. Kulkarni, *Nucleic Acids Res*, 2006, 34, 3361-3369.

24. S. D. Lawhon, J. G. Frye, M. Suyemoto, S. Porwollik, M. McClelland and C. Altier, *Mol Microbiol*, 2003, 48, 1633-1645.

25. C. W. Sze and C. Li, *Infect Immun*, 2011, 79, 1270-1279.

26. J. W. Williams, A. L. Ritter and A. M. Stevens, *FEMS Microbiol Lett*, 2012, 329, 28-35.

27. C. Altier, M. Suyemoto and S. D. Lawhon, *Infect Immun*, 2000, 68, 6790-6797.

28. G. M. Knudsen, J. E. Olsen, S. Aabo, P. Barrow, I. Rychlik and L. E. Thomsen, *Microbiology*, 2013, 159, 1497-1509.

29. D. R. Fortune, M. Suyemoto and C. Altier, *Infect Immun*, 2006, 74, 331-339.

30. K. Heurlier, F. Williams, S. Heeb, C. Dormond, G. Pessi, D. Singer, M. Camara, P. Williams and D. Haas, *J Bacteriol*, 2004, 186, 2936-2945.

31. D. A. Higgins, M. E. Pomianek, C. M. Kraml, R. K. Taylor, M. F. Semmelhack and B. L. Bassler, *Nature*, 2007, 450, 883-886.

32. H. Mulcahy, J. O'Callaghan, E. P. O'Grady, C. Adams and F. O'Gara, *Infect Immun*, 2006, 74, 3012-3015.

33. E. P. O'Grady, H. Mulcahy, J. O'Callaghan, C. Adams and F. O'Gara, *Infect Immun*, 2006, 74, 5893-5902.

34. G. Pessi, F. Williams, Z. Hindle, K. Heurlier, M. T. Holden, M. Camara, D. Haas and P. Williams, *J Bacteriol*, 2001, 183, 6676-6683.

35. Y. Irie, M. Starkey, A. N. Edwards, D. J. Wozniak, T. Romeo and M. R. Parsek, *Mol Microbiol*, 2010, 78, 158-172.

36. H. Mulcahy, J. O'Callaghan, E. P. O'Grady, M. D. Macia, N. Borrell, C. Gomez, P. G. Casey, C. Hill, C. Adams, C. G. Gahan, A. Oliver and F. O'Gara, *Infect Immun*, 2008, 76, 632-638.

37. S. M. Yoo, D. Na and S. Y. Lee, *Nat Protoc*, 2013, 8, 1694-1707.

38. D. Na, S. M. Yoo, H. Chung, H. Park, J. H. Park and S. Y. Lee, *Nat Biotechnol*, 2013, 31, 170-174.

39. A. E. McKee, B. J. Rutherford, D. C. Chivian, E. K. Baidoo, D. Juminaga, D. Kuo, P. I. Benke, J. A. Dietrich, S. M. Ma, A. P. Arkin, C. J. Petzold, P. D. Adams, J. D. Keasling and S. R. Chhabra, *Microb Cell Fact*, 2012, 11, 79.

40. H. Yakhnin, P. Pandit, T. J. Petty, C. S. Baker, T. Romeo and P. Babitzke, *Mol Microbiol*, 2007, 64, 1605-1620.

41. S. Mukherjee, H. Yakhnin, D. Kysela, J. Sokoloski, P. Babitzke and D. B. Kearns, *Mol Microbiol*, 2011, 82, 447-461.

42. S. Mukherjee, P. Babitzke and D. B. Kearns, *J Bacteriol*, 2013, 195, 297-306.

43. M. Y. Liu and T. Romeo, *J Bacteriol*, 1997, 179, 4639-4642.

44. A. K. Dubey, C. S. Baker, T. Romeo and P. Babitzke, *RNA*, 2005, 11, 1579-1587.

45. M. Schubert, K. Lapouge, O. Duss, F. C. Oberstrass, I. Jelesarov, D. Haas and F. H. Allain, *Nat Struct Mol Biol*, 2007, 14, 807-813.

46. P. Babitzke and T. Romeo, *Curr Opin Microbiol*, 2007, 10, 156-163.

47. C. S. Baker, I. Morozov, K. Suzuki, T. Romeo and P. Babitzke, *Mol Microbiol*, 2002, 44, 1599-1610.

48. M. Y. Liu, H. Yang and T. Romeo, *J Bacteriol*, 1995, 177, 2663-2672.

49. N. Majdalani, C. K. Vanderpool and S. Gottesman, *Crit Rev Biochem Mol Biol*, 2005, 40, 93-113.

50. M. Y. Liu, G. Gui, B. Wei, J. F. Preston, L. Oakford, U. Yuksel, D. P. Giedroc and T. Romeo, *J Biol Chem*, 1997, 272, 17502-17510.

51. K. Suzuki, X. Wang, T. Weilbacher, A. K. Pernestig, O. Melefors, D. Georgellis, P. Babitzke and T. Romeo, *J Bacteriol*, 2002, 184, 5130-5140.

52. T. Weilbacher, K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke and T. Romeo, *Mol Microbiol*, 2003, 48, 657-670.

53. R. G. Chavez, A. F. Alvarez, T. Romeo and D. Georgellis, *J Bacteriol*, 2010, 192, 2009-2012.

54. J. Nolling, G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin and D. R. Smith, *J Bacteriol*, 2001, 183, 4823-4838.

55. C. M. Cooksley, Y. Zhang, H. Wang, S. Redl, K. Winzer and N. P. Minton, *Metab Eng*, 2012, 14, 630-641.

56. S. Dusseaux, C. Croux, P. Soucaille and I. Meynial-Salles, *Metab Eng*, 2013, 18, 1-8.

57. X. Hou, W. Peng, L. Xiong, C. Huang, X. Chen and W. Zhang, *J Biotechnol*, 2013, 166, 25-33.

58. H. Janssen, C. Grimmler, A. Ehrenreich, H. Bahl and R. J. Fischer, *J Biotechnol*, 2012, 161, 354-365.

59. K. Jia, Y. Zhang and Y. Li, *PLoS One*, 2012, 7, e38815.

60. D. Lehmann and T. Lutke-Eversloh, *Metab Eng*, 2011, 13, 464-473.

61. C. Ren, Y. Gu, Y. Wu, W. Zhang, C. Yang, S. Yang and W. Jiang, *BMC Genomics*, 2012, 13, 349.

62. N. Y. Saad, B. Schiel, M. Braye, J. T. Heap, N. P. Minton, P. Durre and H. D. Becker, *J Biol Chem*, 2012, 287, 20382-20394.

63. C. Xue, J. Zhao, C. Lu, S. T. Yang, F. Bai and I. C. Tang, *Biotechnol Bioeng*, 2012, 109, 2746-2756.

64. K. Kovacs, B. J. Willson, K. Schwarz, J. T. Heap, A. Jackson, D. N. Bolam, K. Winzer and N. P. Minton, *Biotechnol Biofuels*, 2013, 6, 117.

65. J. R. Ventura, H. Hu and D. Jahng, *Applied microbiology and biotechnology*, 2013, 97, 7505-7516.

66. D. Lehmann, D. Honicke, A. Ehrenreich, M. Schmidt, D. Weuster-Botz, H. Bahl and T. Lutke-Eversloh, *Applied microbiology and biotechnology*, 2012, 94, 743-754.

67. J. Lee, H. Yun, A. M. Feist, B. O. Palsson and S. Y. Lee, *Applied microbiology and biotechnology*, 2008, 80, 849-862.

68. D. P. Wiesenborn, F. B. Rudolph and E. T. Papoutsakis, *Appl Environ Microbiol*, 1988, 54, 2717-

2722.

69. S. H. Baer, H. P. Blaschek and T. L. Smith, *Appl Environ Microbiol*, 1987, 53, 2854-2861.

70. J. T. Heap, S. A. Kuehne, M. Ehsaan, S. T. Cartman, C. M. Cooksley, J. C. Scott and N. P. Minton, *J Microbiol Methods*, 2010, 80, 49-55.

71. L. D. Mermelstein, N. E. Welker, G. N. Bennett and E. T. Papoutsakis, *Biotechnology (N Y)*, 1992, 10, 190-195.

72. J. T. Heap, O. J. Pennington, S. T. Cartman and N. P. Minton, *J Microbiol Methods*, 2009, 78, 79-85.

73. W. L. Zhang, Z. Y. Liu, Z. Liu and F. L. Li, *Lett Appl Microbiol*, 2012, 55, 240-246.

74. B. Langmead, C. Trapnell, M. Pop and S. L. Salzberg, *Genome Biol*, 2009, 10, R25.

75. L. Wang, Z. Feng, X. Wang and X. Zhang, *Bioinformatics*, 2010, 26, 136-138.

76. D. Metcalf, S. Sharif and J. S. Weese, *Anaerobe*, 2010, 16, 439-443.

77. Q. Ren, Z. L. Xu, X. W. Wang, X. F. Zhao and J. X. Wang, *Fish Shellfish Immunol*, 2009, 26, 787-798.

78. K. V. Alsaker, C. J. Paredes and E. T. Papoutsakis, *Biotechnol Bioproc E*, 2005, 10, 432-443.

79. J. Liu, Y. Tan, X. Yang, X. Chen and F. Li, *Journal of bioscience and bioengineering*, 2013, 116, 460-464.

80. G. Z. Cui, W. Hong, J. Zhang, W. L. Li, Y. Feng, Y. J. Liu and Q. Cui, *J Microbiol Methods*, 2012, 89, 201-208.

81. E. Steiner, J. Scott, N. P. Minton and K. Winzer, *Appl Environ Microbiol*, 2012, 78, 1113-1122.

82. R. W. O'Brien and J. G. Morris, *Journal of general microbiology*, 1971, 68, 307-318.

83. A. Bacher, S. Eberhardt, M. Fischer, K. Kis and G. Richter, *Annu Rev Nutr*, 2000, 20, 153-167.

84. C. J. Paredes, K. V. Alsaker and E. T. Papoutsakis, *Nature reviews. Microbiology*, 2005, 3, 969-978.

85. B. A. Annous and H. P. Blaschek, *Appl Environ Microbiol*, 1990, 56, 2559-2561.

86. Y. Chen, D. C. Indurthi, S. W. Jones and E. T. Papoutsakis, *MBio*, 2011, 2, e00340-00310.

87. J. T. Heap, O. J. Pennington, S. T. Cartman, G. P. Carter and N. P. Minton, *J Microbiol Methods*, 2007, 70, 452-464.

Table 1 Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics[a]	Source[b]
Clostridium acetobutylicum ATCC 824	Wild type	ATCC
Thermoanaerobacterium thermosulfurigenes	Wild type	Lab stock
E. coli TOP10		Invitrogen
Cac-*csrA*97s::CT	*C. acetobutylicum* 824 *csrA* mutant	This study
pMTL007C-E2	Clostridial expression vector for expression of ClosTron containing ErmRAM. Cm^R	Heap et al [70]
pAN2	Plasmid containing Φ3T I methyltransferase gene of B. subtilis phage Φ3*t*I, used for methylation of E. coli DNA	Heap et al [87]
pMTL007C-E2:: Cac-*csrA*-97s	ClosTron plasmid retargeted to *C. acetobutylicum* CA_C0951 gene	This study
pIMP-GFP	Clostridial expression vector carrying the GFP (green fluorescence protein)	Cui et al. [80]
pMTL85141	Clostridial expression vector for expression of ClosTron 80000 series	Heap et al. [72]

[a] Cm^R, chloramphenicol/thiamphenicolresistance gene.

[b] ATCC, American type culture collection.

Table 2 The statistical analysis of sequencing for wild type and Cac-*csrA*97s::CT strain

Sample	WT	Cac-*csrA*97s::CT
Number of total reads	4,523,933	3,974,575
High-quality total reads	4,101,850	3,227,752
Total base pairs	678,589,950	596,186,250
High-quality base pairs	615,277,508	484,162,854

Table 3 Expression profiles of genes involved in synthesis of riboflavin. FDR = false discovery rate

Accession	Annotation	FDR	Folds
CA_C0590	riboflavin biosynthesis protein RIBD (pirimidine deaminase and pirimidine reductase)	2.63E-28	-1.4
CA_C0591	riboflavin synthase subunit alpha	4.67E-21	-1.4
CA_C0592	riboflavin biosynthes protein RIBA (gtpcyclohydrolase/3,4-dihydroxy-2-butanone 4-phosphate synthase)	2.89E-44	-1.2
CA_C0593	6,7-dimethyl-8-ribityllumazine synthase	7.34E-17	-1.2

Table 4 Expression profiles of genes involved in non-coding RNAs. FDR = false discovery rate

sRNA	Start	End	Length (nt)	Rfam annotation	Accession	FDR	Folds
sRNA_01	224783	224321	463	-		0	1.792368
sRNA_02	693623	693236	388	-		2.52E-34	0.722605
sRNA_03	704080	704396	317	Lysine	RF00168	1.32E-22	-0.57627
sRNA_04	832957	833360	404	tmRNA	RF00023	0	1.149772
sRNA_05	933531	933690	160	-		3.28E-50	1.896093
sRNA_06	1376789	1377037	249	6S	RF00013	6.93E-30	-0.37464
sRNA_07	1432251	1432338	88	-		5.19E-13	-0.42318
sRNA_08	1914403	1914468	66	-		0.001428	0.401765
sRNA_09	2173478	2173432	47	-		0.038092	-0.22798
sRNA_10	2182117	2181530	588	OLE	RF01071	1.08E-57	0.500337
sRNA_11	2620008	2619702	307	-		8.26E-51	2.078447
sRNA_12	2812081	2811846	236	-		1.84E-23	0.912897
sRNA_13	3111265	3111434	170	-		2.73E-89	1.902201
sRNA_14	3752919	3752849	71	-		0.033011	0.30677

Figure legends

Fig. 1 A phylogenetic tree derived from the CsrA amino acid sequences extracted from NCBI database for *C. acetobutylicum* ATCC 824 and related bacteria. The tree was constructed using the neighbour-joining method based on 1,000 resamplings. *C. acetobutylicum* ATCC 824 CsrA is shown in bold.

Fig. 2 The growth and fermentation profiles of *C. acetobutylicum* ATCC 824 wild type and Cac-*csrA*97s::CT mutant in the P2 medium.

Fig. 3 Saturation analysis of sequencing for *C. acetobutylicum* ATCC 824 wild type and Cac-*csrA*97s::CT by RNA-seq.

Fig. 4 Numbers of differentially expressed genes by KEGG pathway categories. 1，Carbohydrate metabolism；2，Energy metabolism；3，Lipid metabolism；4，Nucleotide metabolism；5，Amino acid metabolism；6，Metabolism of other amino acids；7，Glycan biosynthesis and metabolism；8，Metabolism of cofactors and vitamins；9，Metabolism of terpenoids and polyketides；10，Biosynthesis of other secondary metabolites；11，Xenobiotics biodegradation and metabolism；12，Enzyme families；13，Transcription；14，Translation；15， Folding, sorting and degradation；16，Replication and repair；17，RNA family；18，Membrane transport；19，Signal transduction；20，Signaling molecules and interaction；21，Transport and catabolism；22，Cell motility；23，Cell growth and death；24，Cell communication.

Fig. 5 β-Galactosidase production from *lacZ* fusions to *fla* (Flagella), *rib* (Riboflavin), *spo* (Sporulation) and *pts* (PTS) were used to measure expression in *C. acetobutylicum* ATCC 824 wild type (a), Cac-*csrA*97s::CT mutant (b) and the complemented *csrA* mutant (c). Error bars represent the standard error of the mean.

Fla, the operon CA_C2139-CA_C2167 for the assembly of flagella; *rib,* the operon CA_C0590-CA_C0593 for the riboflavin synthesis; *spo,* the operon CA_C2086–CA_C2093 for the stage III sporulation. *Pts,* the operon CA_C1457–CA_C1463 for the PTS system fructose (mannose)-specific transporter.

Fig. 6 The analysis of mobility for *C. acetobutylicum* ATCC 824 wild type and Cac-*csrA*97s::CT strain semi-solid CGM agar.

Fig. 7 Transmission electron micrographs of negatively stained *C. acetobutylicum* ATCC 824. A, Cac-*csrA*97s::CT strain; B, wild type strain.

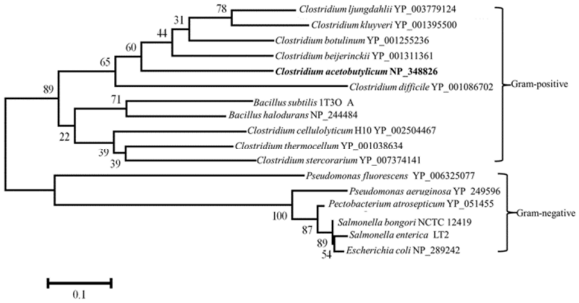


Fig 1

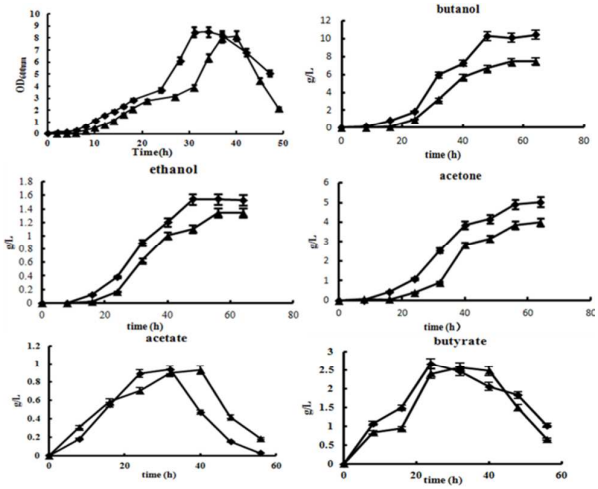


Fig 2

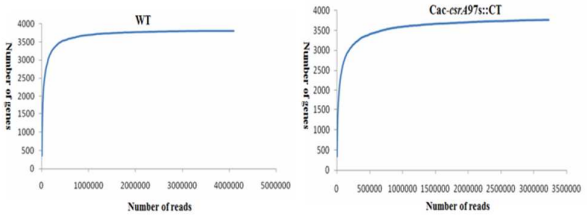


Fig 3

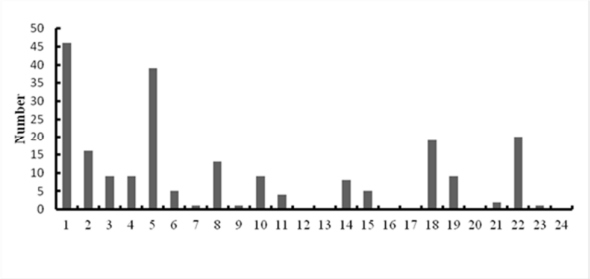


Fig 4

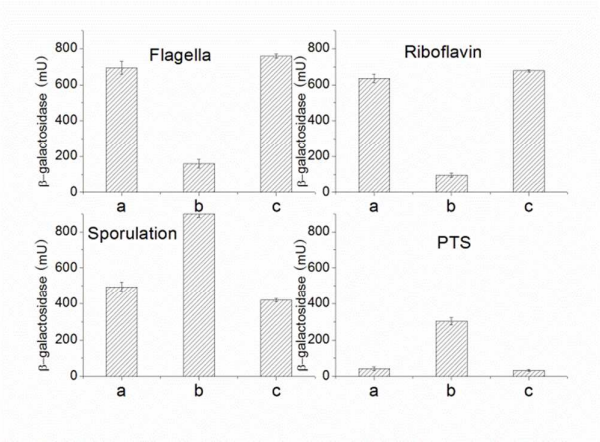


Fig 5

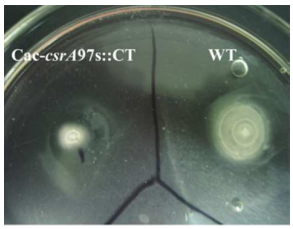


Fig 6

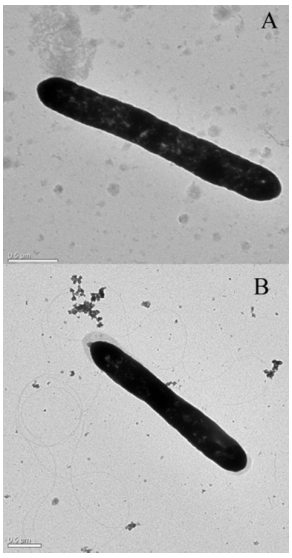


Fig 7