Comparative Genomic and Genetic Functional Analysis of Industrial L-Leucine– and L-Valine–Producing *Corynebacterium glutamicum* Strains^S

Yuechao Ma¹, Qixin Chen¹, Yi Cui¹, Lihong Du¹, Tuo Shi¹, Qingyang Xu^{1,2,3}, Qian Ma^{1,2,3}, Xixian Xie^{1,2,3*}, and Ning Chen^{1,2,3*}

¹National and Local United Engineering Lab of Metabolic Control Fermentation Technology, Tianjin University of Science & Technology, Tianjin 300457, P.R. China

²Key Laboratory of Microbial Engineering of China Light Industry, Tianjin University of Science & Technology, Tianjin 300457, P.R. China

³Tianjin Engineering Lab of Efficient and Green Amino Acid Manufacture, Tianjin University of Science & Technology, Tianjin 300457, P.R. China

Received: May 13, 2018
Revised: September 7, 2018
Accepted: September 12, 2018

First published online
September 20, 2018

*Corresponding authors
X.X.
Phone: +86-22-60601251;
Fax: +86-22-60602298;
E-mail: xixianxie@tust.edu.cn
N.C.
Phone: +86-22-60601251;
Fax: +86-22-60602298;
E-mail: ningch@tust.edu.cn

^SSupplementary data for this
paper are available on-line only at
http://jmb.or.kr.

pISSN 1017-7825, eISSN 1738-8872

Copyright© 2018 by
The Korean Society for Microbiology
and Biotechnology

Corynebacterium glutamicum is an excellent platform for the production of amino acids, and is widely used in the fermentation industry. Most industrial strains are traditionally obtained by repeated processes of random mutation and selection, but the genotype of these strains is often unclear owing to the absence of genomic information. As such, it is difficult to improve the growth and amino acid production of these strains via metabolic engineering. In this study, we generated a complete genome map of an industrial L-valine-producing strain, *C. glutamicum* XV. In order to establish the relationship between genotypes and physiological characteristics, a comparative genomic analysis was performed to explore the core genome, structural variations, and gene mutations referring to an industrial L-leucine-producing strain, *C. glutamicum* CP, and the widely used *C. glutamicum* ATCC 13032. The results indicate that a 36,349 bp repeat sequence in the CP genome contained an additional copy each of *lrp* and *brnFE* genes, which benefited the export of L-leucine. However, in XV, the *kgl* and *panB* genes were disrupted by nucleotide insertion, which increase the availability of precursors to synthesize L-valine. Moreover, the specific amino acid substitutions in key enzymes increased their activities. Additionally, a novel strategy is proposed to remodel central carbon metabolism and reduce pyruvate consumption without having a negative impact on cell growth by introducing the CP-derived mutant H⁺/citrate symporter. These results further our understanding regarding the metabolic networks in these strains and help to elucidate the influence of different genotypes on these processes.

Keywords: *Corynebacterium glutamicum*, comparative genomics analysis, genotype, branched-chain amino acid, L-valine, L-leucine

Introduction

Corynebacterium glutamicum is a bacterial platform that allows for the efficient synthesis of amino acids and is also widely used to produce other commercially important compounds [1]. Industrial amino acid-producing bacteria are traditionally obtained by multiple rounds of random mutation and selection [43]. These mutants generally have

one or more of the following characteristics: loss of the ability to catabolize the target product, decreased transcriptional repression or feedback inhibition by the target product towards key enzymes, and/or high product export capacity [2]. However, multiple rounds of random mutagenesis can result in the accumulation of detrimental or unnecessary mutations [3], and these undesired side mutations may result in auxotrophy, growth deficiencies,

Fig. 1. Biosynthesis pathway of L-valine and L-leucine.

The dashed arrow lines indicate feedback inhibition by L-valine or L-leucine.

or low sugar consumption, which can directly or indirectly limit the efficiency of target product synthesis [4]. Because the genotype of these strains is often unclear, it is difficult to augment deficiencies in growth or amino acid production by way of metabolic engineering. Additionally, it may be difficult to identify the causative mutations from the multitude of mutations present in such strains.

Advances in genome sequencing have facilitated the study of genetic systems, and such analyses have allowed researchers to recognize previously unidentified genotypic patterns and improve our understanding of *C. glutamicum* [5, 27]. Introduction of relevant genes to wild-type strain can help define desirable synthesis pathways [6, 7]. However, the limited genomic information related to amino acid-producing strains is insufficient to allow researchers to engineer such pathways [8]. The metabolic network of branched-chain amino acid (BCAA) biosynthesis, which includes L-leucine, L-valine, and L-isoleucine, is an important focus regarding *C. glutamicum* engineering. The intracellular synthesis of BCAAs involves central carbon metabolism, BCAA synthesis, coenzyme supply, transcriptional regulation,

and a transport system [35]. The metabolic pathway and core genes for BCAA biosynthesis are shown in Fig. 1. In wild-type strains, the synthesis of BCAAs is strictly subjected to feedback regulation systems, and some mutagenized strains already have desired phenotypes that make their genotypes of high research value [9].

In this study, we determined the genome sequence for *C. glutamicum* strains that have been used in industrial production for more than ten years, including that of an L-leucine-producing strain (*C. glutamicum* CP) and an L-valine-producing strain (*C. glutamicum* XV), with these two strains independently obtained via mutagenesis of the same original strain. Subsequently, comparative genomic analysis was used to explore the genotypic differences between the two strains and the most widely-used reference strain, *C. glutamicum* ATCC 13032 [10]. Because the original parental strain was lost, ATCC 13032 was used as a reference strain, as it lacks the ability to efficiently synthesize BCAAs. Genetic functional validation experiments were performed to verify the results of the comparative genomic analysis, and the verified genotypes were used for

de novo construction of new L-valine- and L-leucine-producing strains based on ATCC 13032.

Materials and Methods

Bacterial Strain and Genome Extraction

The strains, plasmids, and primers used and constructed in this study are listed in Tables S1, S2, and S3. These strains are accessible at the China General Microbiological Culture Collection Center with the identifiers of CGMCC 11425 (*C. glutamicum* CP), and CGMCC 1.15672 (*C. glutamicum* XV). Bacteria were routinely cultivated at 32°C in a Luria-Bertani medium containing 10 g/l tryptone (Oxoid, China), 5 g/l yeast extract (Oxoid, China), and 10 g/l NaCl (Aladdin, China) with 5 g/l glucose (Aladdin, China) for 24 h. The genomic DNA from bacteria was extracted using the Promega Wizard Genomic DNA Purification Kit (Promega, USA), according to the manufacturer's instructions.

Genome Sequencing, Quality Control, Assembly, and Annotation

Sequencing using the PacBio RSII platform was performed at the Wuhan Institute of Biotechnology. Sequencing libraries (20-kb SMRTbell templates) were generated using the PacBio Template Prep Kit (PacBio, USA), following the manufacturer's instructions. Sequencing primers were annealed to the hairpins of the SMRTbell templates followed by binding with the P6 sequencing polymerase and MagBeads (Pacific Biosciences, USA). One cell was run on the PacBio RSII system using P6–C4 chemistry and a 240-min data-collection mode. For sequencing with the Illumina platform, sequencing libraries were generated using the Truseq Nano DNA HT Sample Preparation Kit (Illumina, USA), following the manufacturer's instructions. These libraries were sequenced using the Illumina HiSeq2500 platform at Beijing Novogene Bioinformatics Technology CO. Ltd., and 125-bp paired-end reads were generated with an insert size of around 350 bp. The raw reads obtained from the PacBio RSII platform were filtered using the SMRT Analysis v2.3.0 (<http://www.pacb.com/support/software-downloads/>), and the filtered reads were then assembled into complete genomes without gaps using the Celera Assembler v8.2 [11] under the default parameters. The raw reads of FASTQ format obtained from the Illumina platform were processed through a series of quality control (QC) steps to remove low-quality reads using the NGS QC Toolkit v2.3.3 [12]. Open reading frames and functional predictions were performed with Prokka v1.11 [13] based on the complete genome sequences. Predicted genes were annotated using the KEGG Orthology database online annotation service (<http://www.genome.jp/kegg/annotation/>). The sequence of promoters was predicted using BPROM software [14].

Comparative Genomic Analysis

Synteny plots were produced using the related reference contig arrangement tool r2cat [15] for the complete genome sequences of the two BCAA-producing strains and the reference strain *C. glutamicum* ATCC 13032 (NC_003450.3). The genomes were

aligned using the multiple genome alignment tool Mauve v2.3.1 [16] to identify the genome structural variations between the two BCAA-producing strains and ATCC 13032. The core genome and specific gene analysis were performed using the cluster analysis function of the pan-genomes analysis pipeline v1.12 [17] using the default parameters and then verified by BLASTP (E-value 10^{-6}). For detection of SNPs, clean reads obtained from the Illumina platform of XV were mapped to the genome of CP and ATCC 13032, using the Burrows–Wheeler Aligner tool v0.7.12 [18] and then were manually verified using Tablet v1.15 [19] using the default parameters. Variants to indel and SNP calling were performed using SAMtools v1.1.19 [20] and MUMmer v3.23 [21] under the default parameters. Finally, the variants were annotated and the effect was predicted using SnpEff v4.1 [22] under the default parameters.

Construction of Plasmids and Strains

General DNA manipulations were performed using Takara QuickCut Enzyme (TaKaRa, Dalian, China) for plasmid DNA linearization and using the ClonExpress II One Step Cloning Kit (Vazyme Biotech, China) for DNA ligation. The *ilvBN* gene was amplified by PCR from different strains and ligated into pXT01 (using primers 13/14). The *leuA* gene was amplified by PCR from different strains and ligated into pEC-XK99E [23] (using primers 15/16). The ATCC 13032-derived *leuB* and *leuCD* genes were amplified by PCR (using primers 17–20) and separately ligated into pXT01. The gene *CP_103* was amplified by PCR (using primers 23/24) from CP and ligated into pXT01. The expression vectors were separately introduced in *C. glutamicum* by electroporation as described [24].

Chromosomal DNA manipulations were achieved via a markerless system using the suicide vector pK18mobsacB as described previously [25]. In order to integrate XV-derived *ilvBN* into the genome of ATCC 13032, DNA fragments of the 5'- and 3'-region of *Cgl2610* and *Cgl1890* were amplified by PCR using genomic DNA of ATCC 13032 as the template (using primers 1–4 and 7–10), and, the sequence of the P_{trf} promoter and XV-derived *ilvBN* were amplified by PCR using plasmid pXT01+*ilvBN*^{XV} as the template (using primers 5/6 and 11/12). DNA fragments were ligated by overlapping PCR and ligated into pK18mobsacB. The resulting suicide vector was electroporated into ATCC 13032. In order to integrate *CP_103* into the genome of AN02, DNA fragments of the 5'- and 3'-region of *Cgl2103* were amplified by PCR using genomic DNA of AN02 as the template (using primers 67–70), and, the sequence of the P_{trf} promoter and *CP_103* were amplified by PCR using plasmid pXT01+*CP_103* as the template (using primers 71/72). DNA fragments were ligated by overlapping PCR and ligated into pK18mobsacB. The resulting suicide vector was electroporated into AN02.

Fermentation Experiments

To test the activity of mutant acetohydroxy acid synthase, strains were cultivated in 30 ml CGX minimal medium contains (per liter)

20 g of $(\text{NH}_4)_2\text{SO}_4$, 0.5 g of KH_2PO_4 , 0.5 g of K_2HPO_4 , 0.25 g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 mg of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 10 mg of $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 1 mg of $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.2 mg of CuSO_4 , 0.02 mg of NiCl_2 , and 0.2 mg of biotin [26]. After sterilization, glucose was added to a final concentration of 30 g/l and dry CaCO_3 was added to a final concentration of 20 g/l. Strains were cultivated in 500-ml flasks at 32°C and 200 rpm.

To test the maximal production of strains, semi-synthetic media were used. Seed cultures were cultivated in 500-ml flasks at 32°C of 30 ml medium containing (per liter) 30 g of glucose, 5 g of yeast extract (Oxoid, China), 30 ml corn steep liquor (Julong, China), 2 g of KH_2PO_4 and 1 g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ with shaking at 200 rpm. Then, 3 ml of seed culture in the logarithmic growth phase was transferred into 30 ml of fermentation medium containing (per liter) 80 g of glucose, 2 g of yeast extract (Oxoid, China), 30 ml corn steep liquor (Julong, China), 2 g of sodium citrate, 2 g of KH_2PO_4 , 1 g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 mg of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 10 mg of $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, and 10 mg of thiamine in 500-ml flasks at 32°C for 32 h. During the fermentation process, the pH was maintained at approximately 7.0 using 250 ml/l NH_4OH and the concentration of glucose in the culture was maintained at not lower than 20 g/l by feeding with glucose (800 g/l). When needed, 50 mM sodium citrate was added. All media components used in this study were procured from Aladdin (China) except for the annotated tryptone, yeast extract and corn steep liquor.

Analytical Procedures

Glucose concentration was measured by an enzyme electrode glucose sensor (SBA-40E; Biology Institute of Shandong Academy of Sciences, China). The amino acid contents were measured by an automatic amino acid analyzer (S433D; SYKAM, Germany) equipped with LCA K06 (4.6*150 mm) column and an ultraviolet (UV) detector. The column temperature was maintained at 57°C and the reactor temperature was maintained at 130°C. Citrate buffer A (70 mM, pH 3.45) and citrate buffer B (66 mM, pH 10.85) were run at 0.45 ml/min with gradient elution as a mobile phase, and 20 g/l ninhydrin was run at 0.20 ml/min as a derivative agent.

RNA was extracted from the cells during the fermentation process in the exponential phase (ATCC 13032 at 6 h, CP and XV at 8 h) and in stationary phase (ATCC 13032 at 12 h, CP and XV at 16 h) using TaKaRa MiniBEST Universal RNA Extraction Kit (TaKaRa, China). Fluorescence quantitative RT-PCR assay on the StepOne Real-Time PCR System (Thermo-Fisher Scientific, USA) was used to measure mRNA level. Primers used in fluorescence quantitative RT-PCR are listed in Table S3.

Statistical analysis was performed by using Welch's t test for appropriate comparison between the data measured. Significance was assumed at a *p* value of < 0.05.

Results

Comparative Genomic Analysis

Both the L-leucine hyperproducing strain *C. glutamicum*

CP and the L-valine hyperproducing strain *C. glutamicum* XV were previously generated by traditional mutagenesis from the same parent strain. As such, a comparative genomic analysis might aid in our understanding of the genetic mechanisms that contribute to the overproduction of BCAAs in these strains. In a previous study, the entire genome of CP was sequenced and assembled to a 3,342,897 bp chromosome (Genbank Accession No. CP012194.1) and a 4,447 bp endogenous plasmid (Genbank Accession No. CP012412.2) [28]. Here, the whole genome of XV was sequenced using the PacBio RSII and Illumina platforms, and the filtered reads were assembled into a 3,333,639 bp chromosome (Genbank Accession No. CP018175.1) and a 4,447 bp endogenous plasmid (Genbank Accession No. CP018174.1). Genes were predicted and annotated for both strains (Tables S4 and S5) along with their plasmids (Tables S6 and S7), and a comparative genomics analysis was performed using *C. glutamicum* ATCC 13032 as reference. The core genome consisted of 2,576 genes detected in the two strains with a high sequence homology to the reference strain. A total of 172 genes were detected in CP and XV with no homologous sequence present in ATCC 13032. More than half of these 172 genes (91 genes) coded for hypothetical proteins, and the remaining 81 genes were functionally annotated and determined to be transferases, transcriptional regulators, response regulator proteins, energy metabolism-related oxidoreductases, and membrane transporters. Of the 308 genes present in ATCC 13032 with no homologues in CP and XV, 184 belonged to the prophage CGP3 [29] and most of the other 124 genes encoded hypothetical proteins.

The genome structural variation was evident in the CP and XV strains (Fig. 2). Interestingly, there was a large repeat sequence (36,349 bp in length) identified in the CP genome, from 331,795 to 368,143 (Fig. 3), which contained additional copies of *leuA* (*leuA_2*), *lrp*, and *brnFE*. The *leuA_2* gene was split into *CP_345* and *CP_348* by the insertion of two transposase coding genes, *CP_346* and *CP_347*. Indels and single nucleotide polymorphisms (SNP) were identified from the data with a quality score ≥ 25 (Tables S8 and S9). The sequenced reads of XV were mapped to the genome of the reference strain and CP, respectively. According to the annotation of the KEGG Orthology database, mutant genes were identified that affected central carbon metabolism, key enzymes involved in amino acid synthesis, the supply of coenzymes, those affecting the reducing power of the cells, and membrane transport proteins.

Fig. 2. Genome structural variation analysis.

(A) Sequence alignment based on the whole genome sequences of the two BCAA-producing strains, *C. glutamicum* CP and *C. glutamicum* XV, and reference strain *C. glutamicum* ATCC 13032. (B) Syntenic blocks between the genomes. The obviously discontinuous region indicates the sequence of the CGP3 prophage [29], which is not present in CP or XV.

The Effect of Mutations on the Activity and Transcription of Key Enzymes

The acetohydroxy acid synthase (AHAS) encoded by the *ilvBN* genes is the first key enzyme in the BCAA synthesis pathway [30]. The sequence of the promoter region (438 bp upstream of the gene) of *ilvBN* was compared among the strains, but no mutations were identified. Meanwhile, there were five nonsynonymous mutations in the catalytic subunit gene, *ilvB*, and one in the regulatory subunit gene, *ilvN*, between the two BCAA-producing strains and ATCC 13032. Moreover, *ilvB*^{CP} and *ilvB*^{XV} respectively showed a single-residue mutation (G703A) and (C413T) leading to a single amino acid substitution (G235S) and (A138V) in the mutant AHAS (Table S10).

We cloned the different *livBN* sequences into the expression vector pXT01 of *C. glutamicum* and tested expression in ATCC 13032 using the titer of L-valine as the assessment

criteria. Compared to the control strain ATCC 13032 (pXT01), the titer of L-valine increased from 0.35 ± 0.03 g/l to 1.27 ± 0.37 g/l (ATCC 13032 (pXT01+*ilvBN*^{ATCC 13032})), 2.34 ± 0.15 g/l (ATCC 13032 (pXT01+*ilvBN*^{CP})), and 3.13 ± 0.07 g/l (ATCC 13032 (pXT01+*ilvBN*^{XV})), for the three strains expressing different *ilvBN* from ATCC 13032, CP, and XV, respectively, for 32 h in CGX minimal medium with 30 g/l glucose (Fig. 4). This result indicated that the activity of CP- and XV-derived AHAS were both higher than ATCC 13032 and the XV-derived AHAS had the highest activity. To construct a better L-valine-producing strain, two copies of the *ilvBN*^{XV} gene were successively integrated into the genome of ATCC 13032. L-valine production of the resultant strains was tested in semi-synthetic medium. The titers of L-valine were 18.58 ± 0.58 g/l (AN01, one copy of *ilvBN*^{XV}) and 20.49 ± 0.40 g/l (AN02, two copies of *ilvBN*^{XV}), respectively; however, the titers of L-leucine and L-isoleucine were not

Fig. 3. The large repeat sequence of the L-leucine-producing strain, *C. glutamicum* CP.

(A) Comparative analysis of the repeat sequence in the genome of CP compared to XV and ATCC 13032. Box arrows indicate predicted genes and their orientation, orange box arrows represent the *leuA* gene, and the other genes are color-coded to show the conserved sequences. Homologous genes are linked vertically with pink bars. (B) The specific *leuA* gene of CP was split into CP_345 and CP_348. (C) The *leuA* gene was verified by PCR using the primers illustrated in panel b (primers 15/16). The theoretical length of the amplification product of the *leuA* gene was 2197 bp, and the length of the sequence from CP_345 to CP_348 was 3490 bp.

increased.

For L-leucine synthesis, isopropylmalate synthase (IPMS), encoded by *leuA*, is the key enzyme. There were four nonsynonymous mutations between the two BCAA-producing strains and ATCC 13032 (Table S10). Besides, the mutant IPMS of XV showed a specific single amino acid substitution (E104K) (Table S10). Whereas, in the other genes relating to L-leucine synthesis (*i.e.*, *leuB*, *leuC*, and *leuD*), there was no mutation between CP and XV. We firstly overexpressed the *leuA* of ATCC 13032, CP, and XV

in AN02, but the highest titer of L-leucine was only 0.44 ± 0.05 g/l, which was obtained in AN02 (pEC-XK99E+*leuA*^{CP}), and the L-valine titer remained unchanged. We further tried to construct an artificial wild-type-derived *leuBCD* operon and used a double plasmid expression system to individually overexpress the *leuA* of different strains, respectively (Fig. 5A). Surprisingly, the titer of L-leucine was dramatically increased to 13.09 ± 0.45 g/l in AN02 [(pXT01+*leuBCD*), (pEC-XK99E+*leuA*^{CP})], indicating that simultaneous overexpression of *leuBCD* and *leuA*^{CP} promoted

Fig. 4. The expression analysis of the acetohydroxy acid synthase coding genes, *ilvBN*.

The BCAA production of the different strains expressing *ilvBN* genes during a 32 h flask fermentation in CGX minimal medium. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

L-leucine synthesis. Moreover, the titer of L-valine decreased to 3.04 ± 0.25 g/l (Fig. 5B). Whereas, the overexpression of *leuA*₂^{CP}, *leuA*^{XV}, and *leuA*^{ATCC 13032} did not promote L-leucine synthesis, even when *leuBCD* was overexpressed. This indicated that amino acid substitutions significantly increased the activity of the mutant IPMS of CP, which generated more precursor (*i.e.*, 2-isopropylmalate)

resulting in increased synthesis of L-leucine. However, the single amino acid substitution (E104K) caused the IPMS of XV to lose its activity thoroughly and the *leuA*₂^{CP} had no activity. Moreover, the transcriptional levels of *leuB* and *leuCD* were also identified as a limiting factor in L-leucine synthesis.

The large repeat sequence in the CP genome contained additional copies of *leuA* (*leuA*₂), *lrp*, and *brnFE*. The transcriptional levels of these genes were measured by fluorescence quantitative RT-PCR and normalized to the transcriptional level in the exponential phase of ATCC 13032 as a control (Fig. 6). At the same time, we also measured the transcriptional levels of *leuB* and *leuCD*. The transcriptional levels of these genes were obviously increased in CP except for the *leuA* gene. The measured *leuA* gene transcriptional level (ATCC 13032: CP: XV) was 1.00: 1.57: 1.85 in the exponential phase and 0.14: 1.02: 4.42 in the stationary phase, respectively. It should be noted that the primers we used to detect the *leuA* gene did not detect the incomplete *leuA*₂ gene. It was previously reported that the *leuA* gene is regulated by transcriptional attenuation [33]. Since there were no base changes identified in the promoter region, the transcription of *leuA* was repressed by the intracellular L-leucine of CP.

Fig. 5. The effect of the mutant *leuA* gene and the double-plasmid expression system on L-leucine synthesis.

(A) The double-plasmid expression system: the ATCC 13032-derived *leuBCD* genes were expressed from the pXT01 plasmid under the control of the P_{tuf} promoter; the *leuA* gene from the different strains was expressed from the pEC-XK99E plasmid under the control of the P_{trc} promoter. (B) The BCAA production of the different strains during a 32 h flask fermentation. Data are presented as means $\pm$ standard deviation of three independent experiments. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Fig. 6. The relative transcriptional levels of genes related to L-leucine synthesis.

The expression levels of the genes were measured using fluorescence quantitative RT-PCR from RNA extracted in the exponential and stationary phases during the fermentation process. Primers used are listed in Table S3. Data are presented as means $\pm$ standard deviation of three independent experiments. ** $p \leq 0.01$, *** $p \leq 0.001$.

Mutations and Their Effect on Central Carbohydrate Metabolism

The glucose consumption rate and cell growth rate of CP and XV were both lower than reference strain ATCC 13032, and this was most prominent in XV. These differences were likely caused by weak central carbon metabolism in CP and XV. There were no sequence variations identified in genes related to glucose uptake that differed between CP and XV. However, mutation of the *pgi* and *pfkA* (*Cgl1250*) genes of the glycolysis pathway probably affected the distribution ratio of glucose into the PPP [36]. Additionally, genes related to pyruvate biosynthesis (*Cgl2089*, *Cgl2910*, and *Cgl0552*), pyruvate carboxylation (*Cgl1585*, *Cgl2863*, *Cgl0689*, and *Cgl3007*), and other genes involved in organic acid synthesis from pyruvate (*ldh*, *Cgl0901*, *Cgl2752*, *aceE*, and *aceF*) all presented with mutations [37]. These variations likely weakened the catabolism of pyruvate and the synthesis of byproducts and, hence, promoted entry of more pyruvate into the BCAA biosynthesis pathway [38].

Further analysis showed that the α -ketoglutarate decarboxylase coding gene *kgl* had a frameshift mutation

due to a base insertion in the XV genome. This could result in blockage of the tricarboxylic acid (TCA) cycle at the 2-ketoglutarate node [39, 40]. As such, this may have significantly reduced the metabolic rate of the TCA cycle and caused accumulation of upstream metabolites, such as citrate. Although this would have a negative impact on cell growth, the weakened TCA cycle would also reduce the meaningless consumption of pyruvate and increase the conversion ratio from glucose to BCAA. This strategy could be introduced to the L-leucine-producing strain and even other strains that synthesize target products competing against the TCA cycle.

A Novel Mutant H⁺/Citrate Symporter and Its Impacts on Central Carbon Metabolism and BCAA Production

Several mutations were identified in the membrane transporters of CP and XV. The BCAA importer BrnQ of XV, a Na⁺-dependent symporter [41], was divided into two segments by a nucleotide insertion, which would result in reduced uptake of BCAAs. Meanwhile, efficient export of intracellular BCAAs would also promote reduced feedback

Fig. 7. The effect of the CP-derived H⁺/citrate symporter on cell growth and product synthesis.

(A) L-valine production. (B) Curves of cell density and glucose consumption. Data are presented as means ± standard deviation of three independent experiments. * $p \leq 0.05$.

inhibition and repression [31].

To test the relation of the gene variations involved in BCAA export, we overexpressed 23 mutant membrane transporters of CP and XV using the pXT01 plasmid and the corresponding wild-type genes as controls (Table S10). Overexpression of the CP-derived mutant H⁺/citrate symporter (L285F), encoded by *CP_103*, in AN02 resulted in an increase in L-valine production and a decline in glucose consumption. To identify how this mutant promoted BCAA synthesis, we first examined its activity on citrate transport. ATCC 13032, CP, XV, AN02, and the *CP_103* overexpressing strain AN02 (pXT01+*CP_103*) were cultivated in CGX minimal medium using citrate as the sole carbon and energy source. As a result, only CP and AN02 (pXT01+*CP_103*) could grow, though at only a low growth rate compared to using glucose as a carbon source. This indicated that the mutant H⁺/citrate symporter worked to transport extracellular citrate into the cell and was more active than the wild-type protein.

Integrating *CP_103* into the AN02 genome, the titer of L-valine increased 18.7% (from 20.49 ± 0.40 g/l to 24.33 ± 0.46 g/l) when citrate was added (Fig. 7A). Moreover, the overexpression of *CP_103* showed no negative impact on glucose consumption and even slightly increased cell growth (Fig. 7B). After being transported into the cell, citrate could be rapidly utilized in the TCA cycle. Due to a great deal of pyruvate being consumed to synthesize BCAA, the extra supply of citrate likely reduced the demand of the TCA cycle for pyruvate. Importantly, this strategy could be used for the construction of strains in which the target product requires pyruvate as a precursor.

Discussion

Traditional breeding efforts are time consuming and involve a considerable amount of work, which can hinder the ability to carrying out of scientific research. Compared with these approaches, metabolic engineering is more effective in this regard. However, making targeted genetic modifications that are beneficial is dependent on our understanding of gene functions and metabolic networks. In order to efficiently generate a viable strategy towards improving the performance of a given strain, existing strains with desired traits can be used as a starting point. Comprehensive analysis of the genetic and metabolic mechanisms of bacteria using a whole genome approach is now feasible. However, the problem faced by all researchers is that these bacteria can be difficult to identify and culture from natural environments. It is worth mentioning that traditional mutants, especially those currently used in industrial production, possess many properties unmatched by current, metabolically-engineered strains.

In this study, L-valine and L-leucine *Corynebacterium glutamicum* strains used in industrial production were employed to explore the relationship between the strains' genotypes and product synthesis mechanisms. The synthesis of BCAAs involves a variety of metabolic pathways and many of them also relate to the synthesis of other amino acids. Therefore, the information obtained from this work can be also used in the construction of many other amino acid producers. Our analysis focused on the SNPs that caused nonsynonymous mutations, which most likely determined the observed phenotypic differences. Variations located in

the intergenic region may also have effects, such as on transcription regulation, and this aspect will be the focus of future studies that combine sequence analysis and transcriptome data.

Using structural analogues of metabolites to screen for mutants in key enzymes with high activity or that serve to relieve feedback inhibition is a common strategy of traditional breeding. This study proved these structural analogues mainly changed the activity of the corresponding key enzymes, but did not affect their transcription. In CP and XV, the mutations in the key enzymes (*e.g.*, *ilvBN* and *leuA*) resulted in increased enzyme activity and benefited the synthesis of target product. Moreover, the large repeat sequence identified in CP appeared to play an important role in L-leucine export; however, how this repeat sequence was generated is currently unclear, but it probably depended on a genomic restructuring event that occurred during the mutagenesis process [42]. Since the overexpression of *brnFE* has been found to result in a significantly increased export rate of BCAAs [31, 32], the additional copies of *lrp* and *brnFE* in this repeat sequence benefited L-leucine synthesis. The mutations mentioned above have not been previously reported, and each of these unique genotypes can be considered as a new element to use in combination with more effective genotypes, with the goal of constructing a defined and efficient metabolic network for the optimal synthesis of target products.

In addition, precursor supply is another important factor influencing BCAA production. Pyruvate is the direct precursor of BCAA synthesis and is also the starting metabolite of the TCA cycle. For the flux distribution of pyruvate, there is a trade-off between BCAA synthesis and the TCA cycle. To prevent pyruvate converting to acetyl-CoA and entering into the TCA cycle in BCAA-production strains, weakening or deletion of *aceE* and *gltA* were considered. However, such genetic modifications commonly interfere with the cell's metabolic network and, hence, lead to a significant and irreversible negative impact on cell growth [30, 34, 35]. These engineered strains often have other disadvantages in industrial applications, such as a longer fermentation time and poor adaptation to culture conditions. In *C. glutamicum* XV, the TCA cycle is blocked at 2-ketoglutarate due to inactivation of α -ketoglutarate decarboxylase, which serves to drive more pyruvate into BCAA synthesis; however, this also greatly affects the growth of the strain. It might be worth considering an alternative approach to compensate for the insufficiency in TCA flux by directly utilizing intermediate metabolites of the TCA cycle. In this study, it was shown that mutation of

the CP-derived H^+ /citrate symporter served to transport extracellular citrate into the cell at a much higher efficiency, and this might be used to remodel central carbon metabolism. An increase of intracellular citrate, which can be rapidly utilized, can relieve the demand of the TCA cycle for pyruvate and, hence, lead to an increased intracellular availability of pyruvate. In addition to BCAA production, this strategy could be used in the construction of strains that accumulate products that utilize pyruvate as a precursor.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (31470211 and 31770053), Natural Science Foundation of Tianjin (17JCQNJC09500), and Tianjin Municipal Science and Technology Commission (17YFZCSY01050). The authors would like to thank Dr. Jibin Sun, Dr. Ping Zheng, and Xiaomeng Ni of the Tianjin Institute of Industrial Biotechnology of Chinese Academy of Sciences for their support in the comparative genomic analysis.

Conflict of Interest

The authors have no financial conflicts of interest to declare.

References

1. Lee JY, Na YA, Kim E, Lee HS, Kim P. 2016. The *Actinobacterium Corynebacterium glutamicum*, an Industrial Workhorse. *J. Microbiol. Biotechnol.* **26**: 807-822.
2. Kase H, Nakayama K. 1997. L-Isoleucine production by analog-resistant mutants derived from threonine-producing strain of *Corynebacterium glutamicum*. *Agric. Biol. Chem.* **41**: 109-116.
3. Vallino JJ, Stephanopoulos G. 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol. Bioeng.* **41**: 633-646.
4. Park JH, Lee SY, Kim TY, Kim HU. 2008. Application of systems biology for bioprocess development. *Trends Biotechnol.* **26**: 404-412.
5. Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, *et al.* 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9**: 189-197.
6. Ikeda M, Ohnishi J, Hayashi M, Mitsunashi S. 2006. A genome-based approach to create a minimally mutated *Corynebacterium glutamicum* strain for efficient L-lysine

- production. *J. Ind. Microbiol. Biotechnol.* **33**: 610-615.
7. Becker J, Zelder O, Häfner S, Schröder H, Wittmann C. 2011. From zero to hero--design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metab. Eng.* **13**: 159-168.
 8. Ikeda M, Nakagawa S. 2003. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* **62**: 99-109.
 9. Wu Y, Li P, Zheng P, Zhou W, Chen N, Sun J. 2015. Complete genome sequence of *Corynebacterium glutamicum* B253, a Chinese lysine-producing strain. *J. Biotechnol.* **207**: 10-11.
 10. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, et al. 2003. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J. Biotechnol.* **104**: 5-25.
 11. Denisov G, Walenz B, Halpern AL, Miller J, Axelrod N, Levy S, et al. 2008. Consensus generation and variant detection by Celera Assembler. *Bioinformatics* **24**: 1035-1040.
 12. Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**: e30619.
 13. Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.
 14. Salamov VSA, Solov'yev A. 2011. Automatic annotation of microbial genomes and metagenomic sequences, pp. 61-78. In Li RW (ed.), *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers, Hauppauge, N.Y.
 15. Husemann P, Stoye J. 2010. r2cat: synteny plots and comparative assembly. *Bioinformatics* **26**: 570-571.
 16. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**: 1394-1403.
 17. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**: 416-418.
 18. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.
 19. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, et al. 2010. Tablet--next generation sequence assembly visualization. *Bioinformatics* **26**: 401-402.
 20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
 21. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* **5**: R12.
 22. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**: 80-92.
 23. Kirchner O, Tauch A. 2003. Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. *J. Biotechnol.* **104**: 287-299.
 24. van der Rest ME, Lange C, Molenaar D. 1999. A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Appl. Microbiol. Biotechnol.* **52**: 541-545.
 25. Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, Pühler A. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**: 69-73.
 26. Cremer J, Eggeling L, Sahm H. 1991. Control of the lysine biosynthesis sequence in *Corynebacterium glutamicum* as analyzed by overexpression of the individual corresponding genes. *Appl. Environ. Microbiol.* **57**: 1746-1752.
 27. Wendisch, Volker F. 2006. Genetic regulation of *Corynebacterium glutamicum* metabolism. *J. Microbiol. Biotechnol.* **16**: 1010-1016.
 28. Gui Y, Ma Y, Xu Q, Zhang C, Xie X, Chen N. 2016. Complete genome sequence of *Corynebacterium glutamicum* CP, a Chinese l-leucine producing strain. *J. Biotechnol.* **220**: 64-65.
 29. Baumgart M, Unthan S, Rückert C, Sivalingam J, Grünberger A, Kalinowski J, et al. 2013. Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl. Environ. Microbiol.* **79**: 6006-6015.
 30. Vogt M, Haas S, Klaffl S, Polen T, Eggeling L, Ooyen JV, et al. 2014. Pushing product formation to its limit: metabolic engineering of *Corynebacterium glutamicum* for L-leucine overproduction. *Metab. Eng.* **22**: 40-52.
 31. Xie X, Xu L, Shi J, Xu Q, Chen N. 2012. Effect of transport proteins on L-isoleucine production with the L-isoleucine-producing strain *Corynebacterium glutamicum* YILW. *J. Ind. Microbiol. Biotechnol.* **39**: 1549-1556.
 32. Yin L, Shi F, Hu X, Chen C, Wang X. 2013. Increasing L-isoleucine production in *Corynebacterium glutamicum* by overexpressing global regulator *Lrp* and two-component export system *BrnFE*. *J. Appl. Microbiol.* **114**: 1369-1377.
 33. Pátek M. 2005. Regulation of Gene Expression, pp. 81-98. In Eggeling L, Bott M (eds.), *Handbook of Corynebacterium glutamicum*. CRC Press, Florida.
 34. Blombach B, Schreiner ME, Holátko J, Bartek T, Oldiges M, Eikmanns BJ. 2007. L-valine production with pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* **73**: 2079-2084.
 35. Chen C, Li Y, Hu J, Dong X, Wang X. 2015. Metabolic engineering of *Corynebacterium glutamicum* ATCC13869 for L-valine production. *Metab. Eng.* **29**: 66-75.
 36. Hua Q, Yang C, Baba T, Mori H, Shimizu K. 2003. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J. Bacteriol.* **185**: 7053-7067.

37. Inui M, Murakami S, Okino S, Kawaguchi H, Vertès AA, Yukawa H. 2004. Metabolic analysis of *Corynebacterium glutamicum* during lactate and succinate productions under oxygen deprivation conditions. *J. Mol. Microbiol. Biotechnol.* **7**: 182-196.
38. Peters-Wendisch P, Stolz M, Etterich H, Kennerknecht N, Sahm H, Eggeling L. 2005. Metabolic engineering of *Corynebacterium glutamicum* for L-serine production. *Appl. Environ. Microbiol.* **71**: 7139-7144.
39. Wagner T, Bellinzoni M, Wehenkel A, O'Hare HM, Alzari PM. 2011. Functional plasticity and allosteric regulation of α -ketoglutarate decarboxylase in central mycobacterial metabolism. *Chem. Biol.* **18**: 1011-1020.
40. Bunik VI, Fernie AR. 2009. Metabolic control exerted by the 2-oxoglutarate dehydrogenase reaction: a cross-kingdom comparison of the crossroad between energy production and nitrogen assimilation. *Biochem. J.* **422**: 405-421.
41. Tauch A, Hermann T, Burkovski A, Krämer R, Pühler A, Kalinowski J. 1998. Isoleucine uptake in *Corynebacterium glutamicum* ATCC 13032 is directed by the *brnQ* gene product. *Arch. Microbiol.* **169**: 303-312.
42. Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimarães PM, et al. 2009. An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. *BMC Genomics* **10**: 45.
43. Leuchtenberger W. 1996. Amino acids—technical production and use, pp. 465-502. In Rehm HJ, Reed G (eds.), *Biotechnology: Products of primary metabolism*, 2th Ed. Verlag-Chemie, Weinheim, Germany.