Journal of Microbiology (2013) Vol. 51, No. 6, pp. 877-880
DOI 10.1007/s12275-013-3236-0
Copyright ⓒ 2013, The Microbiological Society of Korea

NOTE

Next-Generation Sequencing-Based Transcriptome Analysis of L-Lysine-Producing *Corynebacterium glutamicum* ATCC 21300 Strain[§]

Hong-Il Kim[1], Jae-Young Nam[1], Jae-Yong Cho[2], Chang-Soo Lee[1], and Young-Jin Park[1*]

[1]*Department of Biomedical Chemistry, Konkuk University, Chung-Ju 380-701, Republic of Korea*
[2]*Department of Pharmaceutical Engineering, Sangji University, Wonju 220-702, Republic of Korea*

(Received April 23, 2013 / Accepted July 15, 2013)

In the present study, 151 genes showed a significant change in their expression levels in *Corynebacterium glutamicum* ATCC 21300 compared with those of *C. glutamicum* ATCC 13032. Of these 151 genes, 56 genes (2%) were up-regulated and 95 genes (3%) were down-regulated. RNA sequencing analysis also revealed that 11 genes, involved in the L-lysine biosynthetic pathway of *C. glutamicum*, were up- or down-regulated compared with those of *C. glutamicum* ATCC 13032. Of the 151 genes, 10 genes were identified to have mutations including SNP (9 genes) and InDel (1 gene). This information will be useful for genome breeding of *C. glutamicum* to develop an industrial amino acid-producing strain with minimal mutation.

***Keywords*:** transcriptome, mutation, genome, *Corynebacterium glutamicum*, RT-PCR

Corynebacterium glutamicum is widely used for biotechnological production of industrially important amino acids, such as glutamate and lysine (Kelle *et al*., 2005). Mutagenesis and screening are well-established procedures for developing an industrial strain (Rowlands, 1984). In a previous study, the ATCC 13287 strain of *C. glutamicum* produced L-lysine with conversion yields of up to 26%, and Kyowa Hakko presented a process resulting in 53.2 g/L L-lysine-HCl with 29% conversion in a batch process with *C. glutamicum* strain ATCC 21300 (Kelle *et al*., 2005). However, this classical approach based on random mutations often results in the introduction of detrimental or unnecessary mutations into the genome (Kelle *et al*., 2005). Thus, in-depth information on the metabolic pathways of these mutant strains and their regulation is necessary to further increase the yields of highly productive strains. Consequently, genome-wide profiling is required for a systematic and comprehensive study of the genetic regulation, as well as genetic variation, of *C. glutamicum* strains. Various approaches have been adopted for pathway analysis, including quantitative assessment of metabolic fluxes, combination metabolite pool with integration of enzyme analysis, and DNA microarrays (Sahm *et al*., 2000; Loos *et al*., 2001; Muffler *et al*., 2002). Recently, next-generation sequencing (NGS)-based sequencing technology has provided a cost-efficient and time-saving method for genome-wide analysis. In this study, we aimed to conduct a genome-wide comparison of the global transcriptome between the enhanced L-lysine-producing *C. glutamicum* ATCC 21300 strain and the wild-type *C. glutamicum* ATCC 13032 strain to obtain an insight into the physiology of *C. glutamicum*.

The ATCC 21300 and ATCC 13032 strains of *C. glutamicum* were obtained from the Korean Collection for Type Culture (KCTC) in Daejeon, Korea, and were cultured in brain-heart infusion medium (Difco, USA) at 30°C.

Total RNA was isolated from *C. glutamicum* as described by Jahn *et al*. (2008). Total RNA was further processed using the RNeasy system (Qiagen) with DNase on-column treatment according to the manufacturer's instructions for RNA extraction. Library preparation and sequencing were performed with 1 μg of each total RNA using the HiSeq2000 sequencing system according to the manufacturer's standard protocol (Illumina, Inc., USA). Image analysis, base calling, and quality score calibration were evaluated using SolexaQA software (Cox *et al*., 2010). Reads (paired-ends) were exported in the FASTQ format and deposited in the National Center for Biotechnology Information (NCBI) database under accession numbers SRP021063 (*C. glutamicum* ATCC 21300) and SRP021064 (*C. glutamicum* ATCC 13032). In order to control the quality of raw data, the SolexaQA package was used to verify the short reads of *C. glutamicum* ATCC 21300 (a total of 11,899,914 reads) and *C. glutamicum* ATCC 13032 (a total of 14,925,459 reads). The resulting short reads were individually mapped to the *C. glutamicum* ATCC 13032 genome (NCBI accession no. NC_003450) using Bowtie aligner (http://bowtie-bio.sourceforge.net/index.shtml). The number of mapped reads (a total of 11,154,978 and 13,873,927

*For correspondence. E-mail: yjpark@kku.ac.kr; Tel.: +82-43-840-3572; Fax: +82-43-851-8209
[§]Supplemental material for this article may be found at http://www.springerlink.com/content/120956.

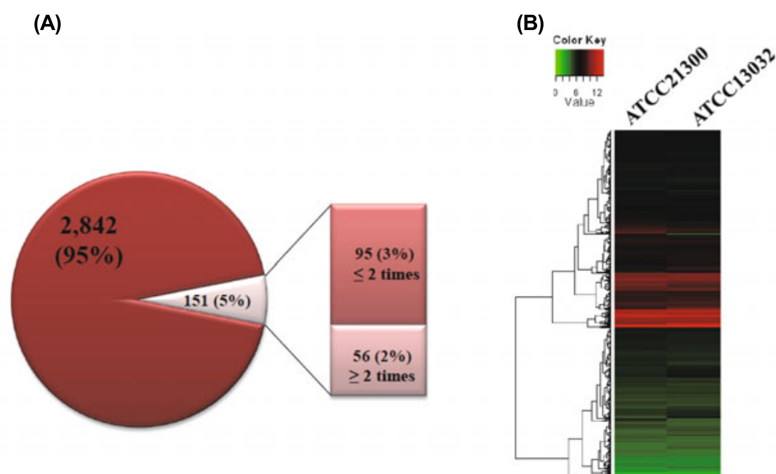

Fig. 1. Global transcriptome comparison between *C. glutamicum* ATCC 21300 and ATCC 13032. (A) Total number of the 2 times up- or down-regulated genes in *C. glutamicum* ATCC 21300 genome. (B) Heat map of the global transcriptomes of *C. glutamicum* ATCC 21300 and ATCC 13032 (log2 based normalization). Detailed information is shown in Supplementary data Table S2.

reads from ATCC 21300 and ATCC 13032, respectively) was normalized by calculating the reads per kilobase per million mapped reads (RPKM) and interpreted as expression level.

RT-PCR was performed as described previously (Lee *et al.*, 2012) using the gene-specific primer (Supplementary data Table S1) and PCR conditions: initial 5-min denaturation step at 94°C, followed by 25 cycles of PCR (94°C, 15 sec; 60°C, 15 sec; 72°C, 30 sec) and final extension for 10 min at 72°C.

Figure 1 shows an overall comparison of the global transcriptome between *C. glutamicum* ATCC 21300 and ATCC 13032 strains. We found a significant change in the expression levels of 151 genes of *C. glutamicum* ATCC 21300 (2 times more or less of RPKM) compared with those of *C. glutamicum* ATCC 13032. Of these 151 genes, 56 genes (2%) were up-regulated, and 95 genes (3%) were down-regulated. In addition, 71 of the 151 genes were annotated as genes of unknown function (hypothetical protein) (Supplementary data Table S2). Interestingly, some of these genes (encoding hypothetical protein) showed highly different expression levels (up to 42 times more or 14 times less) from those of *C. glutamicum* ATCC 13032 (Supplementary data Table S2). Although their functions are yet to be identified, this result suggests that some of these genes will be useful for developing enhanced amino acid-producing *C. glutamicum* strains. Genetic changes, such as mutations in the homoserine dehydrogenase gene (hom^{V59A}), aspartokinase gene ($lysC^{T311I}$), glucose 6-phosphate dehydrogenase gene (*zwf*), phospho-

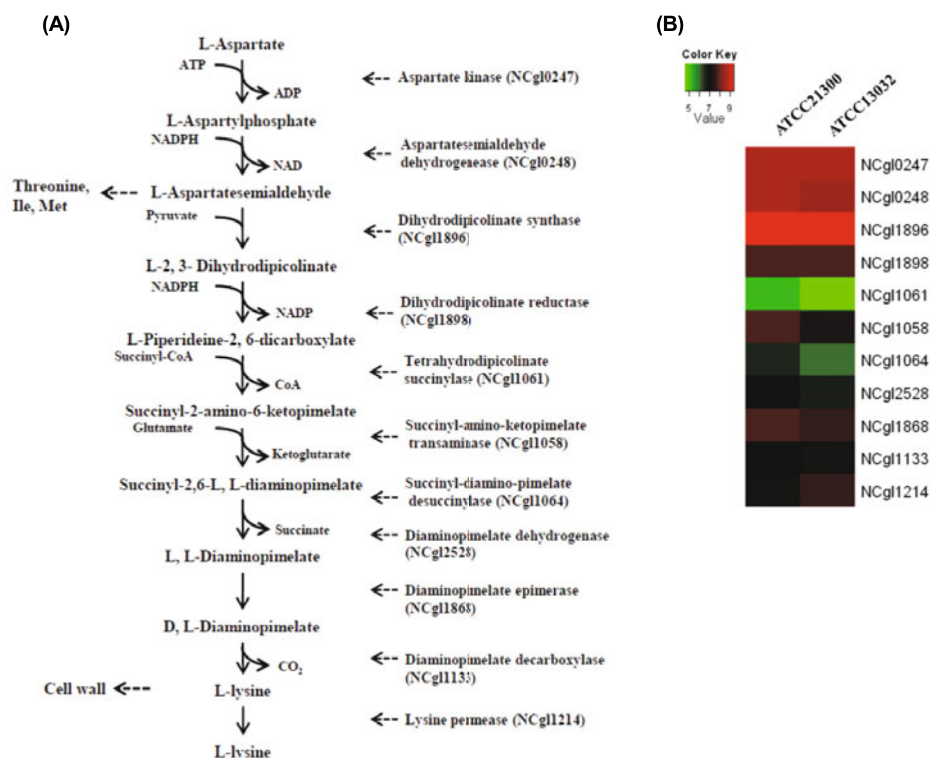

Fig. 2. Overview of L-lysine biosynthesis in *C. glutamicum* (A) and heat map of genes involved in the lysine biosynthesis pathway in *C. glutamicum* ATCC 21300 and ATCC 13032 strains (B). Detailed information is shown in Supplementary data Table S3.

glucoisomerase (*pgi*), and fructose 1,6-bisphosphatase gene (*fbp*), have been previously shown to be useful mutations for increasing the L-lysine production of *C. glutamicum* (Ikeda and Nakagawa, 2003; Lee, 2010). In addition, the reduction of by-product formation and redirection of central carbon metabolism are considered promising targets for strain development (Moritz *et al.*, 2000; Peters-Wendisch *et al.*, 2001; Koffas *et al.*, 2002). Analysis of the anaplerotic enzymes such as phosphoenolpyruvate carboxylase (encoded by the *ppc* gene) and pyruvate carboxylase (encoded by the *pyc* gene) demonstrated that the anaplerotic CO_2 incorporation via pyruvate carboxylase is a major bottleneck for amino acid production in *C. glutamicum* (Peters-Wendisch *et al.*, 2001). In this study, the expression level of the *ppc* gene encoding phosphoenolpyruvate carboxylase (NCgl1523) was increased in *C. glutamicum* ATCC 21300 compared with that in *C. glutamicum* ATCC 13032. Peters-Wendisch *et al.* (2001) reported that overexpression of the *pyc* gene, and thus an increase in pyruvate carboxylase activity, in an L-lysine-producing strain of *C. glutamicum* DG52-2 (isolated mutant derived from *C. glutamicum* ATCC 13032) resulted in approximately 50% higher L-lysine accumulation in the culture supernatant. However, the expression level of the *pyc* gene encoding pyruvate carboxylase (NCgl0659) was decreased in *C. glutamicum* ATCC 21300 (approximately 2 times) compared with that in *C. glutamicum* ATCC 13032. This might be due to the genome-wide mutational differences between these two strains. Ikeda *et al.* (2006) reported that a certain global regulatory mechanism is involved in the industrial levels of L-lysine production. These results also indicate that conventional production strains generated by random mutagenesis and selection will be replaced more and more by carefully designed strains with defined set of mutations. We also previously identified genome-wide mutations (SNP and InDel) in the *C. glutamicum* ATCC 21300 genome by comparative analysis with the *C. glutamicum* ATCC 13032 genome (Lee *et al.*, 2012). However, SNP or InDel mutation was not identified in either the *ppc* (NCgl 1523) or *pyc* (NCgl0659) gene of *C. glutamicum* ATCC 21300 (Supplementary data Table S2).

A total of 11 genes are involved in the L-lysine biosynthetic pathway of *C. glutamicum* (Fig. 2A). Our analyses revealed that the expression of 11 genes of *C. glutamicum* ATCC 21300 were up- or down-regulated compared with those of *C. glutamicum* ATCC 13032 (Fig. 2B and Supplementary data Table S3). However, none of these 11 genes were involved in the 151 genes, which showed a significant change in their expression levels in *C. glutamicum* ATCC 21300. A previous report (Hartman *et al.*, 2003) showed that overexpression of the two genes *dapF* (NCgl1868) and *dapC* (NCgl1058) encoding for diaminopimelate epimerase and succinyl-aminoketopimelate transaminase, respectively, in an industrial *C. glutamicum* strain resulted in increased L-lysine production. Likewise, the *dapF* (NCgl 1058) and *dapC* (NCgl1868) genes of *C. glutamicum* ATCC 21300 also showed increased expression levels compared with those of *C. glutamicum* ATCC 13032 (Supplementary data Table S3). The aspartate kinase encoded by *lysC* (NCgl0247) gene, slightly increased its expression in *C. glutamicum* ATCC 21300, was considered very early to be the key enzyme for the fermentative L-lysine production and a number of mutations are now localized that influence the allosteric control of the enzyme (Kelle *et al.*, 2005). The enhancement of dihydrodipicolinate synthase encoded by the *dapA* (NCgl1896) gene, slightly increased its expression in *C. glutamicum* ATCC 21300, is also considered a promising target for strain improvement strategies (Kelle *et al.*, 2005). With the discovery of the export mechanism for L-lysine, this area became another focus for further strain improvement. By overexpression of the *lysE* (NCgl1214) gene, slightly decreased its expression in *C. glutamicum* ATCC 21300, the excretion rate for L-lysine has been enhanced fivefold compared to the wild-type strain (Kelle *et al.*, 2005). However, SNP or InDel mutation was not identified in those 11 genes in *C. glutamicum* ATCC 21300 (Supplementary data Table S3).

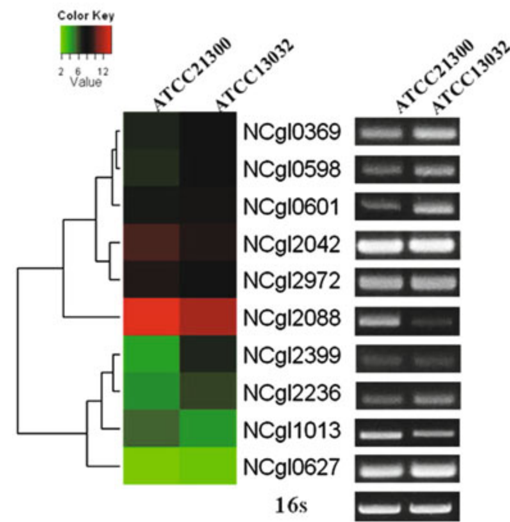

Fig. 3. RT-PCR analysis and heat map of 10 genes in *C. glutamicum* ATCC 21300 and ATCC 13032 strains. Mutations (SNP or InDel) were identified from the 10 genes of *C. glutamicum* ATCC 21300 strain when compared with those of *C. glutamicum* ATCC 13032 strain. Detailed information is shown in Supplementary data Table S4. NCgl0369, major facilitator superfamily permease; NCgl0598, phytoene synthase; NCgl0601, MarR family transcriptional regulator; NCgl2042, ribosome-associated heat shock protein; NCgl2972, hypothetical protein; NCgl2088, hypothetical protein; NCgl2399, gluconate kinase; NCgl2236, hypothetical protein; NCgl1013, phosphoglycerate mutase; NCgl0627, hypothetical protein; 16S, 16S rRNA (control).

Classical breeding of *C. glutamicum* by random mutagenesis has resulted in a huge variety of industrially useful mutants. However, strains developed through undirected mutagenesis have several disadvantages including sensitivity to higher temperature or pH (Kelle *et al.*, 2005). In addition, Ohnishi *et al.* (2002) reported that a limited number of mutations on a wild-type *C. glutamicum* background were effective to develop an industrial amino acid-producing strain. Furthermore, the increase of L-lysine yield is accompanied by a rather complex change of central metabolic pathway gene expression (Ohnishi *et al.*, 2002). As described earlier, 151 genes showed a significant change in their expression levels in *C. glutamicum* ATCC 21300 compared with those of *C. glutamicum* ATCC 13032. Of these, 10 genes were identified to have SNP (9 genes) or InDel (1 gene) mutation (Supple-

mentary data Table S4). None of these mutations that are related to a substantial change in expression level have been reported previously. Further, the RT-PCR results corroborated our observation and revealed difference in the expression of these genes between *C. glutamicum* ATCC 21300 and *C. glutamicum* ATCC 13032 strains (Fig. 3), except 3 genes (NCgl2399, NCgl2042, NCgl2972).

Many genes involved in amino acid biosynthesis have been cloned on multicopy plasmids, to achieve their amplification in the cells and to remove the supposed bottlenecks in biosynthesis pathway. To overcome repression of genes and inhibition of enzyme activity by the accumulated end product, depressed genes or genes coding for deregulated enzymes from mutant strains isolated during the era of classical breeding have usually been used for cloning. Although this simple approach is not considered a general strategy to obtain a producer strain, it has often been successful. With the introduction of methods for global transcriptional profiling including DNA microarray and RNA-seq new targets for modulating gene expression might be rapidly indentified for the further development of producer strains. In the present study, the HiSeq2000 sequencing system (Illumina, Inc.) was used to compare the transcriptomes of *C. glutamicum* ATCC 21300 and *C. glutamicum* ATCC 13032 strains. NGS technology-based global transcriptome analysis will help facilitate the identification of beneficial genes from each genome strain. Moreover, analysis of the expression profiles together with genome-wide genetic variations (SNP or InDel) will provide new and valuable insights for the development of future strategies of metabolic engineering, as well as genome breeding, for high-level industrial production.

This work was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ008099032011), Rural Development Administration, Republic of Korea.

References

Cox, M.P., Peterson, D.A., and Biggs, P. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485.

Hartman, M., Tauch, A., Eggeling, L., Bathe, B., Möckel, B., Pühler, A., and Kalinowski, J. 2003. Identification and characterization of the last two unknown genes, *dapC* and *dapF* in the succinylase branch of the L-lysine biosynthesis of *Corynebacterium glutamicum*. *J. Biotechnol.* **104**, 199–211.

Ikeda, M. and Nakagawa, S. 2003. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes *Appl. Microbiol. Biotechnol.* **62**, 99–109.

Ikeda, M., Ohnishi, J., Hayashi, M., and Mitsuhashi, 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.

Jahn, C.E., Charkowski, A.O., and Willis, D.K. 2008. Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J. Microbiol. Methods* **75**, 318–324.

Kelle, R., Hermann, T., and Bathe, B. 2005. L-lysine production. pp. 465–488. *In* Eggeling, L. and Bott, M. (eds.), Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, USA.

Koffas, M.A., Jung, G.Y., Aon, J.C., and Stephanopoulos, G. 2002. Effect of pyruvate carboxylase overexpression on the physiology of *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.* **68**, 5422–5428.

Lee, J.K. 2010. Carbon metabolism and its global regulation in *Corynebacterium glutamicum*. *Kor. J. Microbiol. Biotechnol.* **38**, 349–361.

Lee, C.S., Nam, J.Y., Son, E.S., Kwon, O.C., Han, W., Cho, J.Y., and Park, Y.J. 2012. Next-generation sequencing-based genome-wide mutation analysis of L-lysine-producing *Corynebacterium glutamicum* ATCC 21300 strain. *J. Microbiol.* **50**, 860–863.

Loos, A., Glanemann, C., Willis, L.B., O'Brien, X.M., Lessard, P.A., Gerstmeir, R., Guillouet, S., and Sinskey, A.J. 2001. Development and validation of *Corynebacterium* DNA microarrays. *Appl. Environ. Microbiol.* **67**, 2310–2318.

Moritz, B., Striegel, K., De Graff, A.A., and Sham, H. 2000. Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux *in vivo*. *Eur. J. Biochem.* **267**, 3442–3452.

Muffler, A., Bettermann, S., Haushalter, M., Hörlein, A., Neveling, U., Schramm, M., and Sorgenfrei, O. 2002. Genome-wide transcription profiling of *Corynebacterium glutamicum* after heat shock and during growth on acetate and glucose. *J. Bacteriol.* **98**, 255–268.

Ohnishi, J., Mitsuhashi, S., Hayashi, M., Ando, S., Yokoi, H., Ochiai, K., and Ikeda, M. 2002. A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant. *Appl. Microbiol. Biotechnol.* **58**, 217–223.

Peters-Wendisch, P.G., Schiel, B., wendisch, V.F., Katsoulidis, E., Mockel, B., Sahm, H., and Eikmanns, B.J. 2001. Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by *Corynebacterium glutamicum*. *J. Mol. Microbiol. Biotechnol.* **3**, 295–300.

Rowlands, R.T. 1984. Industrial strain improvement: mutagenesis and random screening procedures. *Enzyme Microb. Technol.* **6**, 3–10.

Sahm, H., Eggeling, L., and de Graaf, A.A. 2000. Pathway analysis and metabolic engineering in *Corynebacterium glutamicum*. *Biol. Chem.* **381**, 899–910.