The Journal of Microbiology (2012) Vol. 50, No. 5, pp. 860–863
DOI 10.1007/s12275-012-2109-2

NOTE

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

Chang-Soo Lee[1], Jae-Young Nam[1], Eun-Suk Son[1], O-chul Kwon[1], Woorijarang Han[1], Jae-Yong Cho[2], and Young-Jin Park[1]*

[1]*Department of Applied Biochemistry, Konkuk University, Chung-Ju 380-701, Republic of Korea*
[2]*Division of Animal Science and Biotechnology, Sangji University, Wonju 220-702, Republic of Korea*

(Received February 28, 2012 / Accepted May 24, 2012)

In order to identify single nucleotide polymorphism and insertion/deletion mutations, we performed whole-genome re-sequencing of the enhanced L-lysine-producing *Corynebacterium glutamicum* ATCC 21300 strain. In total, 142 single nucleotide polymorphisms and 477 insertion/deletion mutations were identified in the ATCC 21300 strain when compared to 3,434 predicted genes of the wild-type *C. glutamicum* ATCC 13032 strain. Among them, 110 transitions and 29 transversions of single nucleotide polymorphisms were found from genes of the ATCC 21300 strain. In addition, 11 genes, involved in the L-lysine biosynthetic pathway and central carbohydrate metabolism, contained mutations including single nucleotide polymorphisms and insertions/deletions. Interestingly, RT-PCR analysis of these 11 genes indicated that they were normally expressed in the ATCC 21300 strain. This information of genome-wide gene-associated variations will be useful for genome breeding of *C. glutamicum* in order to develop an industrial amino acid-producing strain with minimal mutation.

Keywords: mutation, SNP, InDel, genome, *Corynebacterium glutamicum*

Bacteria belonging to the genus *Corynebacterium* have useful biotechnological applications due to their ability to produce and secrete a number of industrially important amino acids, such as glutamate and lysine, and nucleotides that are useful in industrial production processes. A pioneering study, which proposed the introduction of fermentation in the industrial production of L-amino acids, was conducted by Kinoshita and colleagues at Kyowa Hakko (Japan) in 1957 (Kinoshita *et al.*, 1957; Udaka, 1960). Through later mutagenesis and screening programs, lysine-producing mutants were discovered and the basis for their use in lysine production was established. For example, *C. glutamicum* ATCC 13287 strain, auxotrophic for homoserine and claimed in 1961 (Kelle *et al.*, 2005), produced L-lysine with conversion yields up to 26%. Kyowa Hakko presented a process resulting in 53.2 g/L L-lysine-HCl with 29% conversion in a batch process with *C. glutamicum* ATCC 21300 strain, auxotrophic for threonine and leucine (Kelle *et al.*, 2005).

However, the classical method based on random mutations and selection often results in the introduction of detrimental or unnecessary mutations into the genome. Classically derived mutants are generally inferior to their wild-type strains with regard to industrially important properties such as growth, sugar consumption, and stress tolerance, which restricts the establishment of highly productive industrial processes. A methodology that can overcome these limitations involves genome breeding. Through this approach, specific mutations are systematically introduced into the genome, thus enabling the development of a defined mutant that carries only biotechnologically useful mutations (Ohnishi *et al.*, 2002; Ikeda *et al.*, 2005). To systematically and comprehensively study the genetic variation of *Corynebacterium glutamicum* strains, genome-wide profiling is required. Massively parallel sequencing technology (Bentley, 2006; von Bubnoff, 2008) has recently provided a cost-efficient and time-saving method for identifying genome-wide genetic variations. The complete genome sequence of the representative wild-type strain of *C. glutamicum* ATCC 13032 has been determined (Nakagawa, 2002; Kalinowski *et al.*, 2003) and can be used as a reference for comparative genome analysis to identify gene-associated variations including SNPs and InDels. Thus, the objective of this study was to conduct high-throughput genome sequencing of the enhanced L-lysine-producing *C. glutamicum* ATCC 21300 strain in order to identify gene-associated SNPs and InDels for the development of more efficient amino acid-producing bacteria.

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.

*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.

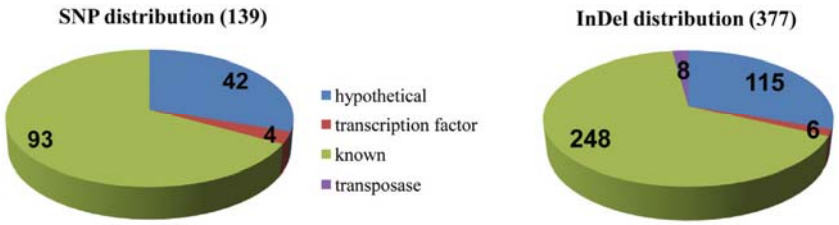

Fig. 1. **SNP and InDel distribution in *C. glutamicum* ATCC 21300 genome.**

Genomic DNA was isolated from *C. glutamicum* as described by Eikmanns *et al.* (1994). Sequencing was conducted commercially in Macrogen Inc. (Korea). Briefly, 1 μg of genomic DNA was used to prepare the Next-Generation Sequencing (NGS) library using TruSeq^{TM} DNA sample preparation kit (Illumina, Inc., USA) according to the manufacturer's instructions. One microliter of genomic DNA (gDNA) was fragmented by nebulization. The 3′ ends of the end-repaired DNA were adenylated, and gel electrophoresis was conducted to isolate 300–400 bp fragments, which were then purified and ligated with adapters. The prepared gDNA library was sequenced with 100-bp paired end reads on a single flow cell lane on the Hiseq 2000 platform (Illumina, Inc.). Image analysis, base calling, and quality score calibration were evaluated using the Illumina Pipeline Software CASAVA v1.8.0 according to the manufacturer's instructions. Reads were exported in the FASTQ format and deposited in the DDBJ Sequence Read Archive (SRA) under the accession number DRR001643. The resulting short-reads (total 28,508,632 reads) were directly mapped to the 3,434 predicted genes of *C. glutamicum* ATCC 13032 genome (NCBI Accession No. BX927147) using BWA (Burrows-Wheeler Aligner) (Li and Durbin, 2009) and GATK (The Genome Analysis Toolkit) (McKenna *et al.*, 2010) base quality score recalibration was applied using standard hard filtering parameters. Sequence variations (SNPs and InDels) were identified using the Samtools software v. 0.1.12a (Li *et al.*, 2009).

First-strand cDNA synthesis was performed using an Omniscript RT KIT (QIAGEN, USA) with 500 ng of total RNA in a total volume of 20 μl containing 20 pmol of the reverse primer. The reaction mixtures were incubated at 37°C for 60 min. After cDNA synthesis, 3 μl of the cDNA was used for PCR in 50-μl reaction mixtures (containing 10 pmol of the gene-specific primer, 10 mM dNTPs, 1 unit Taq DNA polymerase (Toyobo, Japan), and 10× Taq buffer supplied by the manufacturer). Each reaction included an initial 5-min denaturation step at 94°C, followed by 25 cycles of PCR (94°C, 15 sec; 55°C, 15 sec; 72°C, 30 sec) and final extension for 10 min at 72°C. Subsequently, 10 μl of each reaction mixture was separated on a 1.0% agarose gel.

The overall distribution of verified mutations including SNPs and InDels is shown in Fig. 1. A total of 516 mutations (139 SNPs and 377 InDels) were identified in the *C. glutamicum* ATCC 21300 genome compared to the predicted genes of the wild-type ATCC 13032 strain (Supplementary data Tables S1 and S2). Among them, 341 (66% of all mutations) mutations were located in genes with known function. Of the 139 SNPs, 79.1% (110 SNPs) and 26.6% (29 SNPs) were transition mutations (13 A→G, 51 G→A, 1 T→C, and 45 C→T) and transversion mutations (4 A→T, 2 T→A, 7 A→C, 4 G→T, 1 T→G, 1 G→C, and 10 C→G), respectively. Genome-wide genetic variation analysis indicated that *C. glutamicum* ATCC 21300 strains have a larger number of SNPs involving a single transition from G (C) to A (T) (approximately 70%; 96 SNPs). In addition, 5 genes including *ilvA* (cg2334, encoding threonine dehydratase), HSP90 family protein (cg0123), *infB* (cg2176, encoding translation initiation factor IF-2), *putA* (cg0129, encoding proline dehydrogenase), and *psp1* (cg2069, encoding a hypothetical protein) have SNPs at 2 or more sites (Supplementary data Table S1). However, in 4 genes including HSP90 family protein $^{V249M, V339M}$, *infB* ^{R81P}, *putA* ^{R1025C}, and *psp1* $^{D103E, E118D}$, the SNPs resulted in specific amino acid substitutions. Of the 377 InDels, 218 (57.8%) and 19 genes showed 1 guanine insertion and 2 or more (up to 5 nucleotides) nucleotide insertions,

Table 1. **SNP and InDel distribution in carbohydrate metabolism and L-lysine biosynthetic pathway of *C. glutamicum* ATCC 21300**

Locus tag	Gene symbol	Gene description (NC_006958 at NCBI)	Orientation	CDS (bp)	Mutation position	Mutation type	
cg1105	*lysI*	L-lysine permease	-	1503	3′- 1,275 bp - 5′	+G	InDel
cg1787	*ppc*	phosphoenolpyruvate carboxylase	-	2757	3′ - 2,757 bp - 5′	-T	
cg2117	*ptsI*	phosphoenolpyruvate:sugar phosphotransferase system enzymeI	-	1704	3′ - 1,126 bp - 5′	+C	
cg1643	*gnd*	6-phosphogluconate dehydrogenase	-	1476	3′ - 698 bp - 5′	+T	
cg0766	*icd*	isocitrate dehydrogenase	-	2214	3′ - 1,677 bp - 5′	+G	
cg0791	*pyc*	pyruvate carboxylase	+	3420	5′ - 2,149 bp - 3′	+G	
cg1778	*zwf*	glucose-6-phosphate 1-dehydrogenase	+	1542	5′ - 742 bp - 3′	+T	
		glucose-6-phosphate 1-dehydrogenase	+	1542	5′ - 1,035 bp - 3′	+T	
cg3107	*adhA*	Zn-dependent alcohol dehydrogenase	-	1035	3′ - 144 bp - 5′	+G	
cg1280	*kgd*	alpha-ketoglutarate decarboxylase	-	3663	3′ - 1,997 bp - 5′	+T	
cg2291	*pyk*	pyruvate kinase	-	1431	3′ - 263 bp - 5′	T → G (*pyk*^{H390P})	SNP
cg2466	*aceE*	pyruvate dehydrogenase subunit E1	+	2766	5′ - 1,030 bp - 3′	G → A(*aceE*^{G312S})	

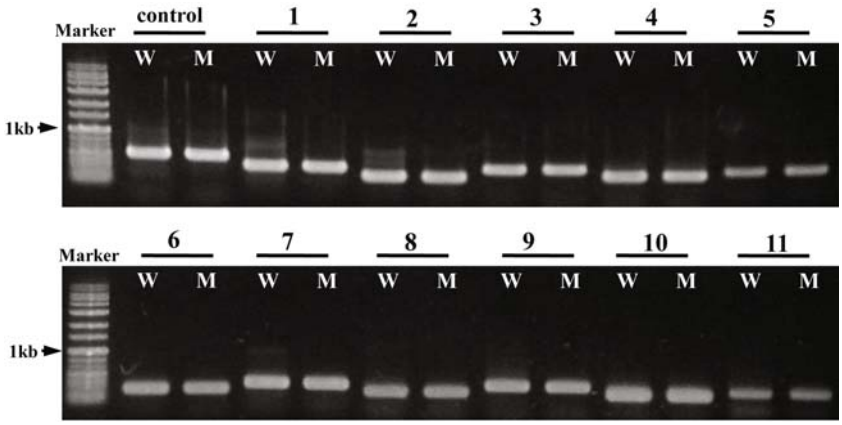

Fig. 2. RT-PCR assays. W, *C. glutamicum* ATCC 13032; M, *C. glutamicum* ATCC 21300; Control, 16s rRNA; 1, 6-phosphogluconate dehydrogenase (*gnd*); 2, α-ketoglutarate decarboxylase (*kgd*); 3, glucose-6-phosphate 1-dehydrogenase (*zwf*); 4, isocitrate dehydrogenase (*icd*); 5, L-lysine permease (*lysI*); 6, phosphoenolpyruvate carboxylase (*ppc*); 7, phosphoenolpyruvate:sugar phosphotransferase system ENZYMEI (*ptsI*); 8, pyruvate carboxylase (*pyc*); 9, pyruvate dehydrogenase subunit E1 (*aceE*); 10, pyruvate kinase (*pyk*); 11, Zn-dependent alcohol dehydrogenase (*adhA*). Markers, 1 kb-ladder.

respectively. Of the 11 genes with specific nucleotide deletions, *ftsY* (cg2262, encoding signal recognition particle GTPase) had the highest level of deletion (GGTTCCTCAA CAATTGCT) (Supplementary data Table S2). These results suggest that the nucleotide changes caused by the InDel lead to a frameshift mutation and thus resulted in a different translated product.

The genome sequences obtained from a classically derived L-lysine hyper-producer *C. glutamicum* strain were compared with the wild-type sequences already available in the database, resulting in the identification of many mutations. Several of these genetic changes, such as mutations in the homoserine dehydrogenase gene (*hom* ^{V59A}), aspartokinase gene (*lysC* ^{T311I}), and pyruvate carboxylase gene (*pyc* ^{P458S}), have previously shown to be useful mutations for increasing L-lysine production (Ikeda and Nakagawa, 2003). In addition, Lee (2010) reported another gene including glucose 6-phosphate dehydrogenase (G6PD, *zwf*), phosphoglucoisomerase (*pgi*), and fructose 1,6-bisphosphatase (*fbp*) useful for L-lysine production of *C. glutamicum*. In our study, we confirmed the previous findings as well as identified novel genetic mutations in important genes involved in carbohydrate metabolism and the L-lysine biosynthetic pathway in *C. glutamicum* ATCC 21300 strains.

Our analyses identified 11 genes, involved in carbohydrate metabolism and the L-lysine biosynthetic pathway, containing genetic variations in the *C. glutamicum* ATCC 21300 strain compared to the wild type ATCC 13032 strain (Table 1). Four of these genes, including *gnd* (cg1643, encoding 6-phosphogluconate dehydrogenase), *icd* (cg0766, encoding isocitrate dehydrogenase), *pyc* (cg0791, encoding pyruvate carboxylase), and *zwf* (cg1778, encoding glucose-6-phosphate 1-dehydrogenase), have been previously reported as potential targets for mutations that are useful for increasing L-lysine production (Lee, 2010). However, those genes have mutations with 1 guanine or 1 thymine insertion instead of point mutations (SNPs) in the *C. glutamicum* ATCC 21300 strain (Ohnishi *et al.*, 2005; Lee, 2010). None of these specific mutations have been reported previously. In addition, we did not observe any significant differences in their expression levels through RT-PCR between *C. glutamicum* ATCC 21300 and *C. glutamicum* ATCC 13032, despite the presence of mutations (InDels or SNPs) in the Open Reading Frames (ORFs) in the former strain (Fig. 2). We do not know yet, however, these results suggest that some type of modulation of these transcripts may occur at the posttranscriptional or translational level. To assess this, it will be necessary to analyze whether these transcripts are translated into proteins.

Corynebacterium glutamicum has a long history of classical breeding, which has resulted in a huge variety of industrially useful mutants. However, strains developed through undirected mutagenesis have several disadvantages. Those strains have acquired a large number of mutations that are not beneficial for a stable process. For example, these strains are usually more sensitive to higher temperature or unfavorable pH and are often strongly affected by certain limitations of vitamins and micronutrients (Kelle *et al.*, 2005). To evaluate the significance of the large number of mutations and the effect on the L-lysine process, methods for multidimensional optimization have to be developed. In the next decade conventional production strains generated by random mutagenesis and selection will be replaced more and more by carefully designed strains with a defined set of mutations. In the present study, a HiSeq2000 sequencing system (Illumina, Inc.) was used to identify genome-wide mutations within the *C. glutamicum* ATCC 21300 strain. Advances in genomics such as next-generation sequencing technologies have facilitated the identification of beneficial mutations from each genome strain. Moreover, defining the genome-wide genetic variations will provide new and valuable insights for the development of future strategies of metabolic engineering for industrial high-level production. The combination of modern metabolic engineering technologies with genetic informations derived from high-throughput analyses based on NGS will lead to the exploitation of new valuable resources in a cost-efficient and time-saving manner.

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

References

Bentley, D.R. 2006. Whole-genome re-sequencing. *Curr. Opin.*

Genet. Dev. **16**, 545–552.

Eikmanns, B.J., Thum-Schmitz, N., Eggeling, L., Lüdtke, K.U., and Sahm, H. 1994. Nucleotide sequence, expression and transcriptional analysis of the *Corynebacterium glutamicum gltA* gene encoding citrate synthase. *Microbiology* **140**, 1817–1828.

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., and Mitsuhashi, S. 2005. Genome breeding of an amino acid-producing *Corynebacterium glutamicum* mutant, pp. 179–189. *In* Barredo, J.L.S. (ed.), Microbial Processes and Products. Humana Press, Totowa, USA.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B.J., Gaigalat, L., and *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.

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.

Kinoshita, S., Udaka, S., and Shimono, M. 1957. Studies on amino acid fermentation. Part I. Production of L-glutamic acid by various microorganisms. *J. Gen. Appl. Microbiol.* **3**, 193–205.

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

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and *et al.* 2010. The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

Nakagawa, S. 2002. The complete genome sequencing of *Corynebacterium glutamicum* ATCC 13032. Abstr. 21. Proceedings of the 9th International Symposium on the Genetics of Industrial Microorganisms.

Ohnishi, J., Katahira, R., Mitsuhashi, S., Kakita, S., and Ikeda, M. 2005. A novel *gnd* mutation leading to increased L-lysine production in *Corynebacterium glutamicum*. *FEMS Microbiol. Lett.* **242**, 265–274.

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.

Udaka, S. 1960. Screening method for microorganisms accumulating metabolites and its use in the isolation of *Micrococcus glutamicus*. *J. Bacteriol.* **79**, 754–755.

von Bubnoff, A. 2008. Next-generation sequencing: The race is on. *Cell* **132**, 721–723.