Korean Journal of Microbiology (2019) Vol. 55, No. 3, pp. 289-292
DOI https://doi.org/10.7845/kjm.2019.9086

pISSN 0440-2413
eISSN 2383-9902

건강한 한국인 분변으로부터 분리된 *Lachnospiraceae* bacterium KGMB03038 (=KCTC 15821) 균주의 유전체 염기서열 초안

김지선[1] · 강세원[1] · 한국일[1] · 이근철[1] · 엄미경[1] · 서민국[1] · 김한솔[1] · 이주혁[1] · 박승환[1] · 박잠언[1] · 오병섭[1] · 유승엽[1] · 최승현[1] · 유승우[1] · 이동호[2] · 윤혁[2] · 김병용[3] · 이제희[3] · 이정숙[1,4*]

[1]한국생명공학연구원 생물자원센터, [2]분당서울대학교병원, [3]천랩, [4]과학기술연합대학원대학교

Complete genome sequence of *Lachnospiraceae* bacterium KGMB03038 (=KCTC 15821) isolated from healthy Korean feces

Ji-Sun Kim[1], Se Won Kang[1], Kook-Il Han[1], Keun Chul Lee[1], Mi Kyung Eom[1], Min Kuk Suh[1], Han Sol Kim[1], Ju Huck Lee[1], Seung-Hwan Park[1], Jam-Eon Park[1], Byeong Seob Oh[1], Seung Yeob Yu[1], Seung-Hyeon Choi[1], Seoung Woo Ryu[1], Dong Ho Lee[2], Hyuk Yoon[2], Byung-Yong Kim[3], Je Hee Lee[3], and Jung-Sook Lee[1,4*]

[1]*Korean Collection for Type Cultures, Korea Research Institute of Bioscience and Biotechnology, Jeongeup 56212, Republic of Korea*
[2]*Seoul National University Bundang Hospital, Seongnam 13620, Republic of Korea*
[3]*ChunLab, Inc., Seoul 06725, Republic of Korea*
[4]*University of Science and Technology (UST), Daejeon 34113, Republic of Korea*

(Received August 7, 2019; Revised August 21, 2019; Accepted September 1, 2019)

Lachnospiraceae bacterium KGMB03038 (=KCTC 15821) belonging to the class *Clostridia* in phylum *Firmicutes*, was isolated from a stool sample of a healthy Korean. Herein, we report the complete genome sequence of strain KGMB03038 analyzed using the PacBio Sequel platform. The genome comprises of 3,334,474 bp with G + C content of 47.8%, which includes 3,099 predicted protein-coding genes, 12 ribosomal RNAs, 54 transfer RNAs, and 4 ncRNAs. Genome analysis revealed that strain KGMB03038 possesses a number of genes involved in hydrolysis of carbohydrates, including mono-, di-, and oligo-saccharides, and biosynthesis of various amino acids.

Keywords: *Lachnospiraceae*, KGMB03038, carbohydrate, human feces

***For correspondence.** E-mail: jslee@kribb.re.kr;
Tel.: +82-63-570-5618; Fax: +82-63-570-5609

The human gastrointestinal tract contains a diverse microbial community such as bacteria, virus, archaea and eukarya. This microbial community plays a pivotal role in human health and disease. In particular, the family *Lachnospiraceae* has been known as major players in the human gut (Manson *et al*., 2008) because of their ability to produce secondary metabolites such as short chain fatty acids (SCFAs). The designation of the family *Lachnospiraceae* was first proposed by Rainey (2009). The family *Lachnospiraceae* is a phylogenetically and morphologically heterogeneous taxon of the class *Clostridia* in phylum *Firmicutes*. The members of the family are anaerobic, fermentative, chemoorganotrophic, and some have potential hydrolytic activity such as pectin methylesterase, pectate lyase, xylanase, galactosidase, and glucosidase. The digestive tract of humans or animals is a major habitat for most members and some have been isolated from the oral cavity and soil. On the

basis of phylogenetic analysis of 16S rRNA gene sequence, *Lachnospiraceae* bacterium KGMB03038 was closely related to *Merdimonas faecis* BR31^T (94.3% 16S rRNA gene similarity) and formed a distinct genus-level lineage within the family *Lachnospiraceae*. This result indicated that strain KGMB03038 represents a novel species within the novel genus belonging to family *Lachnospiraceae* of the class *Clostridia*.

Lachnospiraceae bacterium KGMB03038 was isolated from a healthy Korean feces. The fresh stool sample was collected in anaerobic pouch from Seoul National University Bundang Hospital. The isolation and cultivation of bacteria were performed in the anaerobic chamber (Coy Laboratory Products Inc.) filled with 86% N_2, 7% CO_2, and 7% H_2. The fecal sample suspended in 0.85% saline solution was serially diluted and spread onto Eggerth-Gagnon (EG) agar [2.4 g lab-lemco meat extract, 10 g proteose peptone No. 3, 5 g yeast extract, 4 g Na_2HPO_4, 1.5 g D-(+)-glucose, 0.5 g soluble starch, 0.2 g L-cystine, 0.5 g L-cysteine, 50 ml horse blood per 1 L; pH 7.6].

Genomic DNA was extracted from cells grown on EG medium as described previously (Chun and Goodfellow, 1995). The genomic DNA of strain KGMB03038 was sequenced using the Pacific Biosciences Sequel platform using a 10 kb Single-Molecule Real-Time (SMRT) bell library by Chun Lab, Inc. *De novo* genome assembly was performed with the Hierarchical Genome Assembly Process (HGAP4) pipeline in the SMRT Analysis version 4.0 (GUI) using default parameters. Potential contamination in genome assembles was inspected by the Contamination Estimator by 16S (ContEst 16S) and CheckM tools (Lee *et al*., 2017). The protein coding sequences (CDS) were predicted by prodigal (Hyatt *et al*., 2010) and the CRISPRs were searched using PILER-CR (Edgar, 2007) and CRISPR Recognition Tool (CRT) (Bland *et al*., 2007). The annotation of each CDS was performed using the National Center for Biotechnology Information (NCBI)'s Prokaryotic Genome Annotation Pipeline 2.0 (PGAP) (Tatusova *et al*., 2016).

Whole genome sequencing by the PacBio platform produced a total of 46,321 reads with an average length of 3,288 bp and genome coverage depth, about 383. As described in Table 1, the complete genome of strain KGMB03038 consists of a single circular 3,334,474 bp chromosome with G + C content of 47.8%. No functional plasmid was detected in genome. The genome is predicted to contain 3,099 coding sequences (CDSs), 12 rRNAs (5S, 16S, 23S), 54 tRNAs, and 4 ncRNA genes. A total of 2,924 genes were functionally assigned to categories based on clusters of orthologous group (COG) assignments and the categorized genes were presented in the circular representation with color codes (Fig. 1). Among them, major categories were following as: "unknown function" (S, 37.3%), "amino acid transport and metabolism" (E, 8.7%), "transcription" (K, 8.2%), "replication, recombination and repair" (L, 6.9%), and "carbohydrate transport and metabolism" (G, 6.4%). Genome analysis revealed that strain KGMB03038 possesses putative enzymes for a variety of carbohydrates metabolism. The genome sequence contains key enzymes in glycerol metabolism (glycerol kinase and glycerol dehydrogenase) and D-sorbitol metabolism (aldose reductase and sorbitol dehydrogenase). This strain also possesses α-/β-galactosidase, galactokinase, α-glucosidase, xylanase, α-amylase, and amylosucrase, which hydrolyze mono-, di-, and oligosaccharides such as galactose, maltose, sucrose, xylan, and starch. In addition, genome contains enzymes related with biosynthesis of various amino acid such as alanine, glycine, serine, threonine, cysteine, valine, lysine, arginine, phenylalanine, tyrosine, tryptophan and selenocompound. Overall, the

Table 1. General genomic features of *Lachnospiraceae* bacterium KGMB 03038

Attribute	Value
Genome assembly	
Assembly method	SMRT Analysis version 4.0 (HGAP4)
Genome coverage	382.92X
Genome features	
Genome size (bp)	3,334,474
G+C content (%)	47.8
No. of contig	1
Total genes	3,275
Protein coding CDS	3,099
Genes assigned to COGs	2,924
rRNA genes (5S, 16S, 23S)	12 (4, 4, 4)
tRNA genes	54
ncRNA genes	4
Pseudogenes	106
GenBank accession	NZ_CP041667

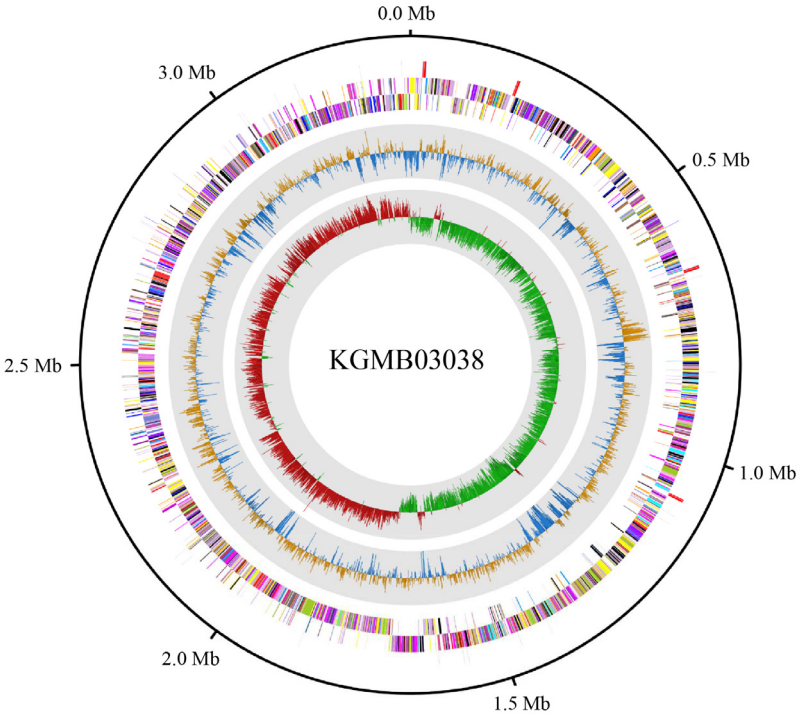

Fig. 1. Complete genome map of *Lachnospiraceae* bacterium KGMB03038. From the center to the outside: GC skew (red and green), G + C content (yellow and blue), CDS on the forward strand (colored by COG categories), CDSs on the reverse strand (colored by COG categories), and RNA genes (rRNAs-red and tRNAs-blue).

genome analysis suggests that strain KGMB03038 is able to break down various kinds of mono-, di-, and oligo-saccharides, and biosynthesizes amino acids in human gut, indicating that this novel *Lachnospiraceae* bacterium KGMB03038 will contribute to human health promotion.

Nucleotide sequence accession number

Lachnospiraceae bacterium KGMB03038 has been deposited in the Korean Collection for Type Cultures under accession number, KCTC 15821. The GenBank/EMBL/DDBJ accession number for the genome sequence of *Lachnospiraceae* bacterium KGMB03038 is NZ_CP041667.

적 요

Lachnospiraceae bacterium KGMB03038 (=KCTC 15821)은 건강한 한국인의 대변 시료로부터 분리된 *Lachnospiraceae* 과, *Clostridia* 강, *Firmicutes* 문에 속하는 신속 균주이다. 본 연구에서는 PacBio Sequel 플랫폼을 이용하여 KGMB03038 균주의 유전체를 해독하고 분석하였으며 그 결과, 47.8% G + C 함량을 가진 3,334,474 bp 길이의 완전한 하나의 유전체 컨티그를 얻었다. 이 유전체는 3,099개의 단백질 암호화 유전자와 12개의 rRNA 유전자, 54개의 rRNA 유전자, 그리고 4개의 ncRNA 유전자를 포함하고 있다. 이 유전체 분석 결과, KGMB03038 균주가 탄수화물의 가수화와 아미노산의 생합성에 관련되어 있는 중요 유전자들을 가지고 있음을 확인하였다.

Acknowledgements

This work was supported by the Bio & Medical Technology Development program (Project No. NRF-2016M3A9F3947962) of the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) of the Republic of Korea and a grant from the Korea Research Institute of Bioscience & Biotechnology (KRIBB) Research initiative program.

References

Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, and Hugenholtz P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Biofinformatics* **18**, 209.

Chun J and Goodfellow M. 1995. A phylogenetic analysis of the genus *Nocardia* with 16S rRNA gene sequences. *Int. J. Syst. Bacteriol.* **45**, 240–245.

Edgar RC. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, and Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Biofinformatics* **11**, 119.

Lee I, Chalita M, Ha SM, Na SI, Yoon SH, and Chun J. 2017. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* **67**, 2053–2057.

Manson JM, Rauch M, and Gilmore MS. 2008. The commensal microbiology of the gastrointestinal tract. *Adv. Exp. Med. Biol.* **635**, 15–28.

Rainey FA. 2009. Family V. *Lachnospiraceae* fam. nov., pp. 921. *In* De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, *et al.* (eds.), Bergey's Manual of Systematic Bacteriology, 2nd edn, Springer, New York, USA.

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, and Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624.