Korean Journal of Microbiology (2019) Vol. 55, No. 2, pp. 160-163
DOI https://doi.org/10.7845/kjm.2019.9019

pISSN 0440-2413
eISSN 2383-9902

Draft genome sequence of *Senegalimassilia* sp. KGMB 04484 isolated from healthy Korean human feces

Kook-Il Han[1], Se Won Kang[1], Ji-Sun Kim[1], Keun Chul Lee[1], Mi Kyung Eom[1], Min Kuk Suh[1], Han Sol Kim[1], Seung-Hwan Park[1], Ju Huck Lee[1], Jam-Eon Park[1], Byeong Seob Oh[1], Seung Yeob Yu[1], Seung-Hyeon Choi[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*

건강한 한국인 분변으로부터 분리된 *Senegalimassilia* sp. KGMB 04484 균주의 유전체 염기서열 초안

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

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

(Received February 14, 2019; Revised March 13, 2019; Accepted March 20, 2019)

Senegalimassilia sp. KGMB 04484 was isolated from fecal samples obtained from a healthy Korean. The whole-genome sequence of *Senegalimassilia* sp. KGMB 04484 was analyzed using the PacBio Sequel platform. The genome comprises a 2,748,041 bp chromosome with a G+C content of 61.18%, 2,300 total genes, 2,139 protein-coding gene, 21 rRNA genes, and 51 tRNA genes. Also, we found that strain KGMB 04484 had some genes for hydrolysis enzyme, fatty acid biosynthesis and metabolism in its genome based on the result of genome analysis. Those genes of KGMB 04484 may be related to regulation of human health and digest.

Keywords: *Senegalimassilia* sp. KGMB 04484, cellulase, fatty acid, feces

Human gut microbiome is the complex community of micro-organisms that live in the digestive tracts. The majority of the bacteria reside in the colon, with estimates of about $10^{11\text{-}14}$ cells/g bacteria (Sender *et al*., 2016). The four major bacterial phyla in the human gut are *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, and *Firmicutes* (Khanna and Tosh, 2014). The gut microbiome products, such as short-chain fatty acids (SCFAs) and membrane proteins, may affect host metabolism by regulating digestion, inflammation, and other functions (Cuevas-Sierra *et al*., 2019). Strain KGMB 04484 was isolated during the investigation of the bacterial diversity of Korean gut microbiome. On the basis of the phylogenetic, phenotypic, and chemotaxonomic characteristics, strain KGMB 04484^T (= KCTC 15721^T = CCUG 72347^T) was found to belong to a novel species as a member of the genus *Senegalimassilia* within the family *Coriobacteriaceae* of *Actinobacteria*.

The genus *Senegalimassilia* was first proposed by Lagier *et al*. (2013). *Senegalimassilia* species have been isolated from

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

the fecal flora of a healthy Senegalese subject [*Senegalimassilia anaerobia* (Lagier *et al*., 2013)]. The genus *Senegalimassilia* are Gram-positive, anaerobic, indole-negative, and motile coccobacillus bacteria (Lagier *et al*., 2013). Here, we describe the draft genome sequence and annotation of *Senegalimassilia* sp. KGMB 04484 isolated from healthy Korean human feces.

The fecal sample was collected from Seoul National University Bundang Hospital, Republic of Korea. The *Senegalimassilia* sp. KGMB 04484 was grown in Columbia blood agar supplemented with 5% (v/v) horse blood for 3 days at 37°C under a $N_2/H_2/CO_2$ (86:7:7, by vol) gas mixture. The genomic DNA was extracted using a Wizard genomic DNA purification kit (Promega). Purified genomic DNA sheared to a size of 10 kb using a g-TUBE^{TM} device according to the manufacturer's instructions (Covaris). Fragmented DNA quantity was analyzed by a Qubit 2.0 fluorometer with a Qubit dsDNA HS Assay Kit (Invitrogen). DNA size was measured by the Agilent 2100 Bioanalyzer with the DNA 12000 assay kit (Agilent). Single-Molecule Real-Time (SMRT) bell library was prepared according to the manufacturer's instructions (Pacific Biosciences) without a non-size selection. Genome sequencing was performed using a Pacific Biosciences Sequel (Pacific Biosciences) with 2.0 sequencing chemistry and 600 min movies.

The *de novo* genome assembly was performed with the Hierarchical Genome Assembly Process (HGAP version 4.0, Pacific Biosciences). Pipeline in the SMRT Analysis (version 4.0, Graphical User Interface) using default parameters. Potential contamination in genome assembles were checked by the ContEst16S (Lee *et al*., 2017). The tRNA was predicted by using tRNAscan-SE (Lowe and Chan, 2016). The CRISPRs were detected using PILER-CR and CRISPR Recognition Tool. The rRNAs and other non-coding RNAs were searched by covariance model search with inference of Rfam 12.0 (Nawrocki *et al*., 2015). The gene annotation of each CDSs were performed with homology searches against Swiss-prot (Bairoch and Apweiler, 2000), EggNOG 4.5 (Huerta-Cepas *et al*., 2016), SEED (Aziz *et al*., 2012), and KEGG databases.

Table 1. General features of *Senegalimassilia* sp. KGMB 04484

Property	Value
Genome assembly	
Assemble method	SMRT Analysis version 4.0
Genome coverage	153×
Genome features	
Genome size (bp)	2,748,041
G+C content (%)	61.18
No. of contigs	2
Total genes	2,300
Protein-coding genes	2,139
Pseudo genes	85
rRNA genes (5S, 16S, 23S)	21 (7, 7, 7)
tRNA genes	51
CDS assigned by COG	2,010
GenBank accession no.	SDPW00000000

The genome statistics are showed in Table 1. The draft genome of *Senegalimassilia* sp. KGMB 04484 was composed of a 2,748,041 bp chromosome with a G+C content of 61.18%. The genome features of *Senegalimassilia* sp. KGMB 04484 are summarized in Fig. 1. The genome is showed to contain 2,139 protein-coding genes, 21 rRNAs (5S, 16S, 23S), and 51 tRNAs were annotated. A total of 2,010 genes were functionally assigned to categories based on clusters of orthologous group (COG) assignments. The majority of the genes are related to energy production and conversion [175 genes (8.7%)] and amino acid transport and metabolism [174 genes (8.6%)].

We found that various genes involved in hydrolytic enzymes and fatty acid were identified in the genome. The genome revealed the presence of cellulase, which involved in the degradation of cellulose. The genome sequence contained genes for fatty acid biosynthesis and metabolism such as [acyl-carrier-protein] S-malonyltransferase *accA*, [acyl-carrier-protein] S-malonyltransferase *fabD*, holo-[acyl-carrier-protein] synthase *acpS*, β-ketoacyl-[acyl-carrier-protein] synthase III *fabH*, oxaloacetate decarboxylase *oadA*, acetyl-CoA carboxylase *accC*, acyl carrier protein, 3-oxoacyl-[acyl-carrier-protein] reductase *fabG*, enoyl-CoA hydratase, and 3-hydroxybutyryl-CoA dehydrogenase *paaH|hbd|fadB|mmgB* gene. The draft genome sequence of *Senegalimassilia* sp. KGMB 04484 will contribute to understanding the physiological functions of *Senegalimassilia* sp. KGMB 04484 in the gut.

Based on the 16S rRNA gene sequence similarity and average nucleotide identity, the strain KGMB 04484 is most closely related to *Senegalimassilia anaerobia* JC110^{T} with the values of 96.49% and 82.74%, respectively.

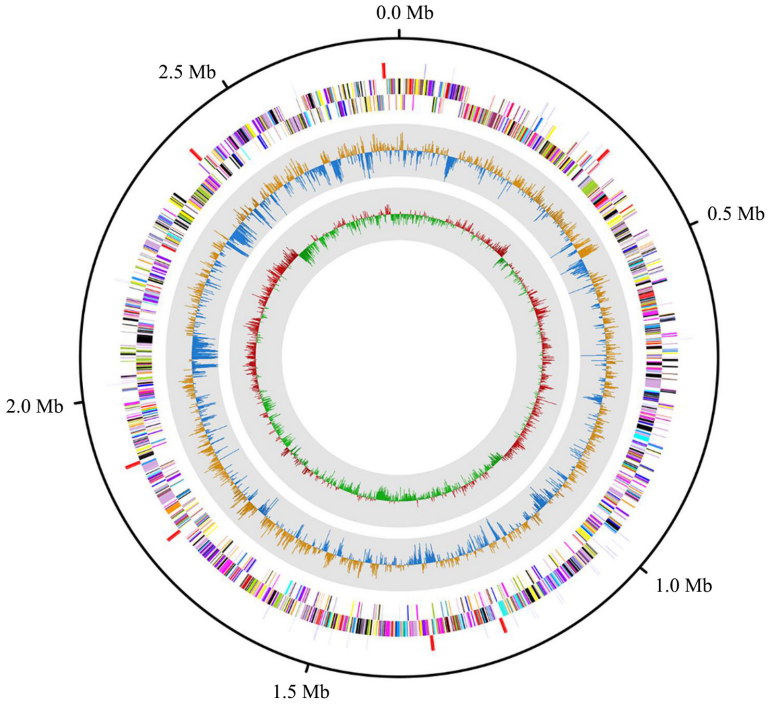

Fig. 1. Graphical circular map of *Senegalimassilia* sp. KGMB 04484. Marked characteristics are shown from outside to the center; coding DNA sequences (CDS) on forward strand, CDS on reverse strand, tRNA, rRNA, GC content, and GC skew.

Nucleotide sequence accession number

Senegalimassilia sp. KGMB 04484 has been deposited in the Korean Collection for Type Cultures under accession number KCTC 15721. The GenBank/EMBL/DDBJ accession number for the genome sequence of *Senegalimassilia* sp. KGMB 04484 is SDPW00000000.

적 요

본 연구에서는 건강한 한국인 분변으로부터 *Senegalimassilia* sp. KGMB 04484 균주를 분리하였으며 PacBio Sequel 플랫폼을 이용하여 유전체서열을 분석하였다. 유전체는 G+C 구성 비율이 61.18%이며, 2,300개의 유전자와 2,139개의 단백질 코딩 유전자, 21개의 rRNA 및 51개 tRNA로 구성되었으며, 염색체의 크기는 2,748,041 bp였다. 유전체의 주요 특징은 가수분해효소와 지방산생합성 및 대사에 관련된 유전자를 포함한다. 이러한 유전체의 분석은 KGMB 04484 균주가 사람의 건강 및 소화에 관여할 것으로 여겨진다.

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

Aziz RK, Devoid S, Disz T, Edwards RA, Henry CS, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, *et al.* 2012. SEED servers: high-performance access to the SEED genomes, annotations, and metabolic models. *PLoS One* **7**, e48053.

Bairoch A and Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **1**, 45-48.

Cuevas-Sierra A, Ramos-Lopez O, Riezu-Boj JI, Milagro FI, and Martinez JA. 2019. Diet, gut microbiota, and obesity: Links with host genetics and epigenetics and potential applications. *Adv.*

Nutr. **10**, S17–S30.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, *et al.* 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic, and viral sequences. *Nucleic Acids Res*. **4**, D286–D293.

Khanna S and Tosh PK. 2014. A clinician's primer on the role of the microbiome in human health and disease. *Mayo Clin. Proc*. **89**, 107–114.

Lagier JC, Elkarkouri K, Rivet R, Couderc C, Raoult D, and Fournier PE. 2013. Noncontiguous-finished genome sequence and description of *Senegalemassilia anaerobia* gen. nov., sp. nov. *Stand. Genomic Sci.* **7**, 343–356.

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.

Lowe TM and Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. **8**, W54–W57.

Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, *et al.* 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43**, D130–D137.

Sender R, Fuchs S, and Milo R. 2016. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol*. **14**, e1002533.