Draft genome sequence of *Ruminococcus* sp. KGMB03662 isolated from healthy Korean human feces

Kook-Il Han\(^1\), Se Won Kang\(^1\), Mi Kyung Eom\(^1\), Ji-Sun Kim\(^1\), Keun Chul Lee\(^1\), Min Kuk Suh\(^1\), Han Sol Kim\(^1\), Seung-Hwan Park\(^1\), Ju Huck Lee\(^1\), Jam-Eon Park\(^1\), Byeong Seob Oh\(^1\), Seoung Woo Ryu\(^1\), 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

(Reduced July 3, 2019; Revised July 16, 2019; Accepted July 16, 2019)

*Ruminococcus* sp. KGMB03662 was isolated from fecal samples obtained from a healthy Korean. The whole-genome sequence of *Ruminococcus* sp. KGMB03662 was analyzed using the PacBio Sequel platform. The genome comprises a 2,707,502 bp chromosome with a G + C content of 43.09\%, 2,484 total genes, 2,367 protein-coding gene, 14 rRNA genes, and 53 tRNA genes. In the draft genome, genes involved in the hydrolysis enzyme, fatty acid biosynthesis, fatty acid metabolite, antibiotic biosynthesis, and antibiotic resistance have been identified. Those genes of KGMB03662 may be related to the regulation of human health and disease.

**Keywords:** *Ruminococcus* sp. KGMB03662, antibiotic biosynthesis, fatty acid biosynthesis, hydrolytic enzyme

The gut microbiome is associated closely with health and disease (Sekirov *et al*., 2010). The short chain fatty acids are the primary end-products of fermentation of non-digestible carbohydrates that become available to the gut microbiome (Morrison and Preston, 2016). Therefore, many novel species remain to be identified and characterized from the human gut (Lau *et al*., 2016). Recently, strain KGMB03662 was isolated during the investigation of the bacterial diversity of Korean gut microbiome. On the basis of the phylogenetic, phenotypic and chemotaxonomic characteristics, strains KGMB03662\(^T\) (= KCTC 15720\(^T\) = CCUG 726234\(^T\)) was found to belong to a novel species as a member of the genus *Ruminococcus* within the family *Ruminococcaceae* of *Clostridia*.

Members of the genus *Ruminococcus* are strictly anaerobic, Gram-positive, non-motile coccobacteria (Rainey, 2009). These species belong to the phyla *Firmicutes* which are the pre-
dominant bacterial groups in the human gut microbiota. The *Ruminococcus* species have been isolated from human gut (Chassard et al., 2012), feces (Simmering et al., 2002; Domingo et al., 2008; Kim et al., 2011) and cattle rumen (Sijpsteijn, 1949). The *Ruminococcus* is less abundant in human with inflammatory bowel disease (Nagao-Kitamoto and Kamada, 2017). Also, genus of *Ruminococcus* is less abundant in Parkinson’s disease (Hill-Burns et al., 2017). Here, we describe the draft genome sequence and annotation of *Ruminococcus* sp. KGMB03662 isolated from healthy Korean feces.

The study was approved by the institutional review board (IRB) of Korea Research Institute of Bioscience and Biotechnology (KRIBB, Approval number: P01-201702-31-007). The fecal sample was collected from Seoul National University Bundang Hospital, Republic of Korea. The *Ruminococcus* sp. KGMB03662 was grown in *Ruminococcus albus* broth for 5 days at 37°C under a N₂/H₂/CO₂ (8.6:0.7:0.7, by volume) gas mixture. The composition of *Ruminococcus albus* medium was as follows: 5.0 g Tryptone, 2.0 g Yeast extract, 3.0 g Glucose, 2.0 g Cellulobiose, 1.0 mg Resazurin, 920.0 g Distilled water, 40.0 ml Mineral solution 1 (per L: 0.06 g K₂HPO₄) and 40.0 ml Mineral solution 2 [per L: 0.06 g KH₂PO₄, 0.2 g (NH₄)₂SO₄, 0.12 g NaCl, 0.025 g MgSO₄·7H₂O and 0.016 g CaCl₂·2H₂O] at pH 7.0. After, media was autoclave-sterilize. Additionally, 4.0 g Na₂CO₃, 1.0 ml fatty acid mixture (10 ml Isobutyric acid, 10 ml Isovaleric acid, 10.0 ml 2-Methylbutyric acid, and 70.0 ml DW) and 500 mg L-cysteine·HCl·H₂O were added.

The genomic DNA was obtained from the cultivated cells on *Ruminococcus albus* agar during 3 days using the Wizard genomic DNA purification kit (Promega). The purified genomic DNA sheared to a size of 10 kb using a g-TUBE™ device according to the manufacturer’s instructions (Covaris). The fragmented DNA quantity was analyzed by a Qubit 2.0 fluorometer with a Qubit dsDNA high sensitivity assay kit (Invitrogen). The DNA size was measured by the Agilent 2100 Bioanalyzer with the DNA 12000 assay kit (Agilent). The Single-Molecule Real-Time (SMRT) bell library was prepared according to the manufacturer’s instructions (Pacific Biosciences) without a non-size selection. The genome sequencing was performed using a Pacific Biosciences Sequel (Pacific Biosciences) with 2.0 sequencing chemistry and 600-min movies.

<table>
<thead>
<tr>
<th>Table 1. General features of <em>Ruminococcus</em> sp. KGMB03662</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Property</strong></td>
</tr>
<tr>
<td>Genome assembly</td>
</tr>
<tr>
<td>Assemble method</td>
</tr>
<tr>
<td>Genome coverage</td>
</tr>
<tr>
<td>Genome features</td>
</tr>
<tr>
<td>Genome size (bp)</td>
</tr>
<tr>
<td>G+C content (%)</td>
</tr>
<tr>
<td>No. of contigs</td>
</tr>
<tr>
<td>Total genes</td>
</tr>
<tr>
<td>Protein-coding genes</td>
</tr>
<tr>
<td>Pseudo genes</td>
</tr>
<tr>
<td>rRNA genes (5S, 16S, 23S)</td>
</tr>
<tr>
<td>tRNA genes</td>
</tr>
<tr>
<td>CDS assigned by COG</td>
</tr>
<tr>
<td>GenBank Accession No.</td>
</tr>
</tbody>
</table>

*De novo* assembly of the sequencing reads was performed through the Hierarchical Genome Assembly Process (HGAP4, version 4.0, Pacific Biosciences). Pipeline in the SMRT Analysis (version 4.0, Graphical User Interface) used default parameters. Potential contamination in genome assemblies was checked by the ContEst16S. The tRNAs were searched by using tRNAscan-SE. The CRISPRs were predicted using PILER-CR and CRISPR Recognition Tool (CRT). The rRNAs and other non-coding RNAs were predicted by covariance model search with the inference of Rfam 12.0 (Nawroki and Eddy, 2013). The annotation of each CDS was performed by homology search against Swiss-prot (UniProt, 2015), EggNOG 4.5 (Powell et al., 2014), SEED (Overbeek et al., 2005), and KEGG (Kanehisa et al., 2014) databases. The functional assignment of genes was performed by searching translated coding DNA sequences (CDSs) against sequences in the clusters of orthologous group (COG) databases (Tatusov et al., 2000).

The genome statistics are shown in Table 1. The draft genome of *Ruminococcus* sp. KGMB03662 was composed of a 2,707,502 bp chromosome with a G + C content of 43.09%. The genome features of *Ruminococcus* sp. KGMB03662 are summarized in Fig. 1. The genome is showed to contain 2,367 Protein-coding genes, 14 rRNAs (5S, 16S, 23S), and 53 tRNAs were annotated. A total of 2,023 genes were functionally assigned to categories based on COG assignments.

The majority of the genes are related to replication, recom-
We found that effective genes involved in hydrolysis enzymes, fatty acid biosynthesis, fatty acid metabolism, antibiotic biosynthesis, and antibiotic resistance were identified in the genome. The genome sequence contained genes for hydrolysis enzymes such as cellulose, chitinase, $\beta$-glucosidase $bglX$, chitooligosaccharide deacetylase $pdaA$, fructan $\beta$-fructosidase $fruA$, $\beta$-fructofuranosidase $INV|sacA$, mannan endo-1,4-$\beta$-mannosidase $gmuG$, $\alpha$-galactosidase $galA|rafA$, and $\beta$-galactosidase $lacZ$. The genome contained the fatty acid biosynthesis and metabolism genes such as 3-oxoacyl-[acyl-carrier-protein] reductase $fabG$, Holo-[acyl-carrier-protein] synthase $acpS$, Lysophospholipase $pldB$, Acyl carrier protein, Acetyl-CoA carboxylase $accA$, Acetyl-CoA carboxylase $accD$, Biotin carboxyl carrier protein of acetyl-CoA carboxylase, 3-oxoacyl-(acyl-carrier-protein) synthase $nodE$, Beta-ketoacyl-[acyl-carrier-protein] synthase III $fabH$, [Acyl-carrier-protein] S-malonyltransferase $fabD$. The genome contained the antibiotic biosynthesis genes such as GTP diphosphokinase $relA$. Also, the genome has several antibiotic resistance factor genes, such as Zinc D-Ala-D-Ala carboxypeptidase $vanY$, Macrolide export ATP-binding/permease protein $MacB$ $macB$, Undecaprenyl-diphosphate phosphatase $bacA$, Phosphinothricin acetyltransferase $pat$, Multidrug export protein $MepA$, Bacitracin transport ATP-binding protein $BcrA$. The draft genome sequence of *Ruminococcus* sp. KGMB03662 will contribute to understanding the physiological functions of *Ruminococcus* sp. KGMB03662 in the gut.

Based on the 16S rRNA gene sequence similarity and average nucleotide identity (ANI), the strain KGMB03662 is most closely related to *Ruminococcus albus* $^7$ with the values of 94.05% and 71.35%, respectively.

**Nucleotide sequence accession number**

*Ruminococcus* sp. KGMB03662 has been deposited in the Korean Collection for Type Cultures under accession number KCTC 15720. The GenBank/EMBL/DDBJ accession number for the genome sequence of *Ruminococcus* sp. KGMB03662 is VCGV00000000.
적 요

본 연구에서는 건강한 한국인 분변으로부터 Ruminococcus sp. KGMB03662 균주를 분리하고 유전체서열을 PacBio Sequel 플랫폼을 사용하여 분석하였다. 염색체의 크기는 2,707,502 bp 로 G + C 구성 비율은 43.09%, 총 유전자수는 2,484개, 단백질 코딩 유전자는 2,367개, rRNA는 14개 및 tRNA는 53개로 구성되었다. 본 유전체로부터 가수분해효소, 지방산생합성 및 대사와 항생제생합성 및 내성 관련 유전자를 확인하였다. 이러한 유전체의 분석은 KGMB03662 균주가 사람의 건강 및 질병에 관여할 것으로 여겨진다.

Acknowledgements

This work was supported by the Bio & Medical Technology Development program (Project No. NRF-2016M3A9F3947 962) 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


