J. Microbiol. Biotechnol. (2013), 23(7), 966–973
http://dx.doi.org/10.4014/jmb.1303.03045

jmb

De Novo Assembly and Comparative Analysis of the *Enterococcus faecalis* Genome (KACC 91532) from a Korean Neonate S

Jun Sang Ham[1†], Woori Kwak[2,3†], Oun Ki Chang[1], Gi Sung Han[1], Seok Geun Jeong[1], Kuk Hwan Seol[1], Hyoun Wook Kim[1], Geun Ho Kang[1], Beom Young Park[1], Hyun-Jeong Lee[4], Jong Geun Kim[6], Kyu-won Kim[2], Samsun Sung[3], Taeheon Lee[4], Seoae Cho[3*], and Heebal Kim[2,3,5*]

[1]*Animal Products Research and Development Division, National Institute of Animal Science, Rural Development Administration, Suwon 441-706, Republic of Korea*
[2]*Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-747, Republic of Korea*
[3]*C&K Genomics, Seoul 151-919, Republic of Korea*
[4]*Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Suwon 441-706, Republic of Korea*
[5]*Department of Agricultural Biotechnology, Animal Biotechnology Major, and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea*
[6]*Forage and Grassland Division, National Institutue of Animal Science, Rural Development Adminstration, Cheonan 331-801, Republic of Korea*

Received: March 13, 2013
Revised: May 14, 2013
Accepted: May 15, 2013

First published online
May 27, 2013

**Corresponding authors*
H.K.
Phone: +82-2-880-4803;
Fax: + 82-2-883-8812;
E-mail: heebal@snu.ac.kr
S.C.
Phone: +82-2-876-8830;
Fax: +82-2-876-8827;
E-mail: seoae@cnkgenomics.com

†These authors contributed equally to this work.

Supplementary data for this paper are available on-line only at http://jmb.or.kr.

pISSN 1017-7825, eISSN 1738-8872

Using a newly constructed *de novo* assembly pipeline, finished genome level assembly had been conducted for the probiotic candidate strain *E. faecalis* KACC 91532 isolated from a stool samples of Korean neonates. Our gene prediction identified 3,061 genes in the assembled genome of the strain. Among these, nine genes were specific only for the *E. faecalis* KACC 91532, compared with all of the four known reference genomes (EF62, D32, V583, OG1RF). We identified genes related to phenotypic characters and detected *E. faecalis* KACC 91532-specific evolutionarily accelerated genes using dN/dS analysis. From these results, we found the potential risk of KACC 91532 as a useful probiotic strain and identified some candidate genetic variations that could affect the function of enzymes.

Keywords: *Enterococcus faecalis*, KACC 91532, *de novo* assembly, dN/dS, arginine deiminase

Introduction

Ever since Ilya Mechnikov first proposed lactic acid from fermented milk as a secret to longevity, lactic acid bacteria (LAB) have been of considerable interest [24]. LAB have been shown to provide numerous health benefits, including in the prevention and treatment of diarrhea, immunological activation against pathogens and cancer, prevention of allergies, improved gastrointestinal function, improved lactose tolerance, and treatment of high blood

pressure [36]. Because of these wide-ranging benefits, significant research has been aimed at isolating LAB from feces for food and medicinal uses [23]. Based on existing probiotics selection criteria, we hypothesized that the most numerous LAB in feces would exhibit superior proliferation and adaptability in the intestine of neonates [30], and could therefore be used as a probiotic strain for infants and children. Accordingly, we collected feces samples from newborns from all areas of South Korea. From these samples, *Enterococcus faecalis* isolate KACC 91532 exhibited the fastest growth rate and was chosen as a potential new probiotic bacterial strain [17].

E. faecalis is the first LAB to colonize the intestines of infants, and is generally considered a nonpathogenic commensal of the mammalian gastrointestinal tract [33]. However, recent studies have established *E. faecalis* as the major cause of enterococcal infection [17]; other virulence factors in *E. faecalis* and their role in pathogenicity have also been established [9, 37]. Many studies have shown that the presence or absence of specific virulence factors is important for enterococcal infections, and that these virulence factors can enhance the ability of *E. faecalis* to cause disease [4, 11, 19]. As bacteria undergo significant horizontal gene transfer and gene loss compared with eukaryotic species, whole genome *de novo* assembly is essential to fully understand the genetic composition of newly isolated bacteria [27]. For example, a previous study [12] demonstrated large differences in gene contents between fully sequenced pathovars of *Pseudomonas syringae*. Such divergence between known reference genomes and newly isolated bacteria makes it difficult to obtain complete genome sequences for newly isolated bacteria using resequencing or reference-based assembly [35]. Therefore, we sought to characterize the gene content and virulence factors of *E. faecalis* KACC 91532 using whole genome *de novo* assembly.

In this study, we identified acid, heat, and antibiotic resistance in *E. faecalis* isolate KACC 91532, and conducted whole genome *de novo* assembly using a newly constructed *de novo* assembly pipeline. Based on this assembled genome sequence, we identified the gene contents and virulence factors of *E. faecalis* KACC 91532, and compared them with four available *E. faecalis* reference genomes. We were also able to identify evolutionarily accelerated genes and variation in *E. faecalis* KACC 91532 using dN/dS analysis. We establish the potential risk of *E. faecalis* KACC 91532 as a probiotic strain, and present a newly constructed *de novo* assembly pipeline that can be used for performing *de novo* assembly of other microorganisms.

Materials and Methods

LAB Isolation

Fecal samples were collected from 25 neonates (16 males, 9 females) born over the course of 5 days across 6 regions of South Korea (Seoul, Incheon, Gang-won, Chungcheong, Jeolla, and Gyeongsang). Samples were stored under anaerobic conditions at 4°C.

Fecal samples were plated on BCP (Bromocresol Purple) agar (Eiken, Japan) and incubated at 37°C for 48 h. LAB were quantified by manually counting all yellow colonies, which then were subcultured in MRS broth (Difco, USA) and screened on TOS agar (Eiken, Japan). Then, isolated colonies were cultured on MRS agar (Difco, USA) under anaerobic incubation, and preserved in cryovials (Key Scientific Products, USA) at -70°C for further study.

To identify the LAB isolate, we performed 16S rRNA gene sequence analysis according to the method of Pavlova *et al*. [28]. To amplify 16S rDNA, we used universal primers corresponding to six conserved regions of the *Escherichia coli* numbering system. Chromosomal DNA was isolated using a genomic DNA extraction kit (Qiagen, Germany). PCR was performed in a 50 µl reaction mixture containing primers (50 pmol), template DNA (50 ng), 5 µl of 10× *Taq* DNA polymerase buffer, 4 µl of dNTP at 2.5 mM, and 1 U of *Taq* DNA polymerase (Takara, Japan). The PCR amplification product was purified using a QIAquick gel extraction kit (Qiagen), ligated into a pSTBlue-1 vector (Novagen, USA), and transformed into *E. coli* DH5α competent cells. The recombinant plasmids were purified using a DNA purification kit (Qiagen) and digested with *Eco*RI to confirm the insert. The nucleotide sequence of the insert was determined using a BigDye-terminator sequencing kit and ABI PRISM 377 sequencer (Perkin-Elmer, USA), according to the manufacturer's instructions. The 16S rDNA sequences were subjected to a similarity search of the GenBank database. The strain exhibiting the highest growth rate in MRS broth from 26 XR7 strain *E. faecalis* was donated to the Rural Development Administration (RDA)-Genebank Information Center, Republic of Korea.

Testing Physiological Features

Antimicrobial susceptibility to erythromycin, gentamicin, oxacillin, tylosin, and vancomycin was performed by disk diffusion in accordance with Clinical Laboratory Standard Institute guidelines [41]. For heat challenge and survival measurement, growth phase cultures were heat treated at 95°C for 30 sec, 1 min, and 2 min, respectively. Heat-treated culture (100 µl) was spread on MRS agar (Difco, USA) and incubated at 37°C for 24 h. Growth phase culture (100 µl) was spread on MRS agar (Difco, USA) adjusted to pH 4.8, 5.0, and 5.5 with 1.0 N HCl, and incubated at 37°C for 24 h for the acid tolerance test.

Genomic Sequencing, Assembly, and Annotation

Roche 454 pyrosequencing reads (shotgun and 8 kb mate pair) and Illumina Hiseq 2000 sequencing reads were generated by the