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

National Instrumentation Center for Environmental Management at Seoul National University. The details of the sequence data are provided in Supplementary Table S1 and Fig. S1.

Raw read data (sff) were modified for de Bruijn assembly (SOAPdenovo [20], Allpaths-LG [13]). Sff file format reads were converted into fastq using sff2fasta (http:// github.com/indraniel/sff2fastq). Linker sequences were removed from 454 8 kb mate pair reads and converted into a paired-end library using in-house software. Quality control and trimming of 454 shotgun reads was performed using FastQC [2] and FastX-toolkit [15]. Then the reads were converted into overlapped paired-end fastq for Allpaths-LG input.

The overall process of *de novo* assembly used in this paper is described in Fig. 1. Newbler assembly software [8] (gsAssembler 2.8) was used to perform *de novo* genome assembly using 454-FLX sequence data. Independent *de novo* assembly was performed three times to generate contigs using SOAPdenovo and Allpaths-LG. First, Illumina reads were error-corrected using the Allpaths-LG error-correction module, and assembled using SOAPdenovo. Second, converted 454 reads were edited for SOAPdenovo assembly using the Allpath-LG error-correction module. Last, modified 454 reads were converted into paired-end reads for assembly using Allpaths-LG. Gap filling was conducted by Gapcloser [20] and Gapfiller [25]; Illumina and 454 reads were error-corrected using the Allpaths-LG module, and the combined contigs from the de Bruijn assembly were used for Gapcloser. For Gapfiller, error-corrected Illumina reads were used as input data. Gapcloser and Gapfiller were reiterated until no change in N base number was seen.

BLAST was used to identify scaffolds that could be connected to each other. For each scaffold, 200 bp were cut from both ends, and the resulting collection of 200-bp sequences was used as query sequences. Contigs from SOAPdenovo, Allpaths-LG, and Newbler were used as the BLAST database. Before connecting two scaffolds, the matches of base pairs in the connection end of each scaffold were manually checked using Bioedit [16] (Supplementary Fig. S2).

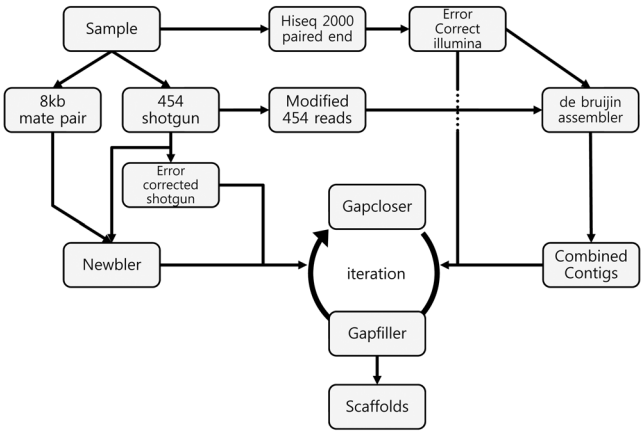

Fig. 1. *De novo* assembly pipeline used for *E. faecalis* KACC 91532 genome assembly.

Enterococcus faecalis reference genome sequences from four strains (EF62, D32, V583, OG1RF) were obtained from the NCBI database. Genome sequences were annotated using the RAST [3] server pipeline. Assembly of *E. faecalis* KACC 91532 was compared against reference sequences using a BLAST dotplot with the RAST SEED viewer. The sequences and functions of genes mentioned in this paper were manually confirmed using BLASTp; Cn3D [40] was used to identify the location of variants in the 3D protein structure.

dn/dS Analysis

To conduct dN/dS analysis using the branch model, we gathered nucleotide and amino acid sequences with the same FIGfam IDs as our RAST annotation results. The orthologous gene sets were aligned by PRANK [18] using the default settings; poorly aligned sites were eliminated using Gblocks [7]. To build a standard phylogenetic tree, we performed bootstrap analysis on the combined data set sequences using PHYLIP (Seqboot) [31]. We used TREE-PUZZLE [38] to estimate the Ts/Tv ratio and calculated the distance of each strain using a Kimura 2-parameter model. A consensus tree was built using the neighbor-joining method. The maximum likelihood method (codeml of PAML 4) [41] was used to estimate the dN (the rate of non-synonymous substitution), dS (the rate of synonymous substitution), and ω (the ratio of non-synonymous substitutions to the rate of synonymous substitutions)

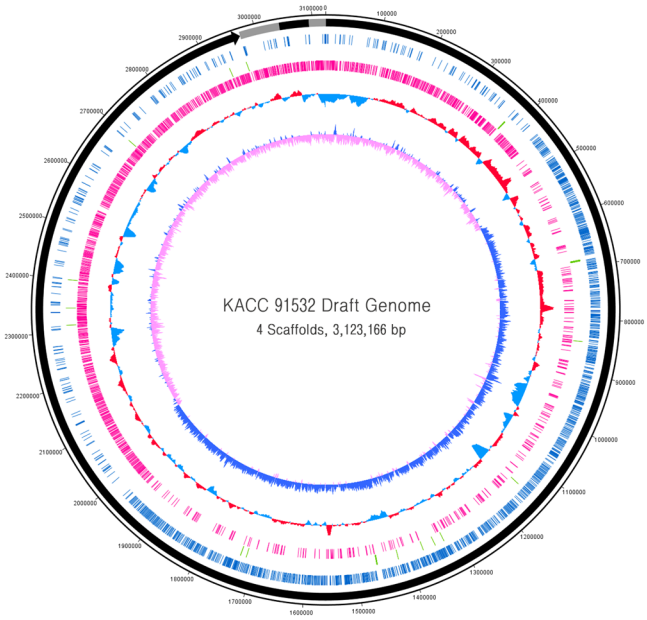

Fig. 2. *E. faecalis* KACC 91532 genome assembly map generated using DNAPlotter [5].
From outside to inside, tracks describe (i) four scaffolds, (ii) CDS on forward strand in blue, (iii) RNA genes in green, (iv) CDS on reverse strands in pink, (v) GC content, and (vi) GC skew.

with F3X4 codon frequencies under the branch model (model = 2, NS sites = 0) and basic model (model = 0, NS sites = 0). Orthologs with dS > 3 or ω > 5 were filtered [6, 29]. The overall process of dN/dS analysis is provided in Supplementary Fig. S3.

Results

Physiological Features of *E. faecalis* KACC 91532

The details of test results pertaining to acid resistance, heat resistance, and antibiotic resistance are provided in Supplementary Table S2. *E. faecalis* KACC 91532 was able to survive at pH 4.8, and was resistant to gentamicin. It was also able to withstand 2 min heat treatment at 95°C.

De Novo Assembly and Gene Contents of *E. faecalis* KACC 91532

The first assembly using Newbler (gsAssembler 2.8) produced five scaffolds with an N50 of 2,966,033 and a total length of 3,120,175 bp, with 63,577 N bases (2.037%). Newbler provided hybrid assembly using illumina reads or contigs. However, the results of hybrid assembly had more fragmented scaffolds with more N bases than the assembled genome, which used 454 raw reads only. For example, when we added contigs from SOAPdenovo and Allpaths-LG to Newbler, the assembly result had six scaffolds with 63,785 N bases. After gap filling using the GapCloser process based on the first assembly using 454 raw data only, the numbers of N bases were reduced from 63,577 to 1,046. Additional gap filling using GapCloser did not close the gap anymore, and the continued gap-filling process using Gapfiller increased the number of N bases to 1,117. However, Gapfiller helped GapCloser to close more gaps and we could reduce the number of N bases to 71. A BLAST search using 200-bp ends from each of the five scaffolds revealed the longest and shortest scaffolds to be reverse complementary to each other; these two scaffolds were subsequently joined to form a single scaffold. After all iterated gap filling and bridging, the final assembly consisted of four scaffolds (2.97 Mb, 70,526 bp, 51,709 bp, 28,528 bp) with a total length of 3,123,166 bp and 71 N bases. A comparison of the *E. faecalis* KACC 91532 assembly sequence with four reference genome sequences using a BLAST dotplot is provided in Supplementary Fig. S4. The BLAST dotplot shows a bidirectional comparison of two genome sequences, and the closer genome sequence shows a more diagonal line. The *E. faecalis* KACC 91532 assembly showed the least disconnection in a diagonal line and the highest query cover rate (94%) with the EF62 strain isolated from Norwegian infants. The *E. faecalis* KACC 91532 assembly was the most similar to the EF62 strain isolated from Norwegian infants [39].

A comparison of four *E. faecalis* reference genomes and *E. faecalis* KACC 91532 using RAST is provided in Table 1. The *E. faecalis* KACC 91532 draft genome had 346 subsystems, 3,061 CDS, and 67 structural RNAs. After removing hypothetical proteins, a total of 2,286 CDSs remained. The four reference genomes (EF62, D32, V583, OG1RF) had 2,887, 2,919, 3,172, and 2,548 CDSs each. We found numerous genes associated with resistance to antibiotics and toxic compounds, including the bile hydrolysis gene for survival in the gastrointestinal tract, the tetracycline-resistance gene, beta-lactamase (which confers resistance to beta-lactam antibiotics such as penicillin), topoisomerase IV genes (which confer resistance to fluoroquinolone antibiotics), and multidrug resistance efflux pump genes, all within the *E. faecalis* KACC 91532 assembly. There are 11 KACC 91532-specific genes compared with the EF62 strain, 37 genes compared with the D32 strain, 25 genes compared with the V583 strain, and 39 genes compared with the OG1RF strain. There are nine KACC 91532-specific genes compared with the four references, and two of them (exodeoxynuclease V beta and cadmium-transporting ATPase) could be assigned with an enzyme commission (EC) number. All the KACC 91532-specific genes are provided in Supplementary Table S3.

dN/dS Analysis

The dN/dS analysis revealed 18 evolutionarily accelerated genes with p-value < 0.05 and FDR < 0.2; the gene list is provided in Table 2. Among 18 evolutionarily accelerated

Table 1. RAST annotation summary for four *E. faecalis* reference genomes and KACC 91532.

Genome	Genome size	Subsystem	CDS	Structural RNAs	Hypothetical CDS removed
E62	2,988,673 bp	349	2,887	67	2,214
D32	2,987,450 bp	344	2,919	74	2,187
V583	3,218,031 bp	355	3,172	80	2,327
OG1RF	2,739,625 bp	332	2,548	71	2,021
KACC 91532	3,123,166 bp	346	3,061	67	2,286

Table 2. Evolutionarily accelerated genes in *E. faecalis* KACC 91532 ($p < 0.05$, FDR < 0.2).

Figfam ID	Function	*p*-Value
FIG00000165	Translation elongation factor LepA	0.03784
FIG00000184	Putative deoxyribonuclease YcfH	0.02807
FIG00000289	tRNA nucleotidyltransferase (E.C. 2.7.7.21) (E.C. 2.7.7.25)	0.04455
FIG00000298	GTP-binding protein EngB	0.00505
FIG00000860	*O*-Succinylbenzoic acid-CoA ligase (E.C. 6.2.1.26)	0.02428
FIG00001104	Arginine deiminase (E.C. 3.5.3.6)	0.03786
FIG00001384	Catalyzes the cleavage of *p*-aminobenzoyl-glutamate to p-aminobenzoate and glutamate, subunit A	0.02479
FIG00001663	GTP-sensing transcriptional pleiotropic repressor codY	0.04811
FIG00002133	Two-component sensor histidine kinase, malate (E.C. 2.7.3.-)	0.03056
FIG00002968	Undecaprenyl-phosphate galactosephosphotransferase (E.C. 2.7.8.6)	0
FIG00004373	Putative membrane protein YeiH	0.01028
FIG00018224	Substrate-specific component MtsA of methionine-regulated ECF transporter	0.02596
FIG00051666	Lon-like protease with PDZ domain	0.0133
FIG00134695	Predicted PTS system, galactosamine-specific IIA component (E.C. 2.7.1.69)	0.04318
FIG00139589	Probable L-ascorbate-6-phosphate lactonase UlaG (E.C. 3.1.1.-) (L-ascorbate utilization protein G)	0.03166
FIG01115339	ABC transporter substrate-binding protein	0.03338
FIG01321817	Quaternary ammonium compound-resistance protein sugE	0.01516
FIG01333098	NtrC family transcriptional regulator, ATPase domain	0.0176

Table 3. Number of substitution sites and non-synonymous sites in evolutionarily accelerated enzymes.

Figfam ID	Function	S.site	K-S.site	NS.site	K-NS.site
FIG00000289	tRNA nucleotidyltransferase	28	10	2	2
FIG00000860	*O*-Succinylbenzoic acid-CoA ligase	21	1	7	1
FIG00001104	Arginine deiminase	18	3	1	1
FIG00002133	Two-component sensor histidine kinase, malate	14	2	5	1
FIG00002968	Undecaprenyl-phosphate galactosephosphotransferase	51	32	11	11
FIG00134695	Predicted PTS system, galactosamine-specific IIA component	10	3	3	2

S.site: Substitution site; K-S.site: KACC91532 Substitution site; NS.site: Non-synonymous substitution site; K-NS.site: KACC91532 Non-synonymous site

genes, seven (arginine deiminase (E.C. 3.5.3.6), *O*-succinylbenzoic acid-CoA ligase (E.C. 6.2.1.26), probable L-ascorbate-6-phosphate lactonase UlaG (E.C. 3.1.1.-), two-component sensor histidine kinase, malate (E.C. 2.7.3.-), tRNA nucleotidyltransferase (E.C. 2.7.7.21,25), predicted PTS system, galactosamine-specific IIA component (E.C.2.7.1.69), and UDP-galactosephosphotransferase (E.C. 2.7.8.6)) had been assigned with an enzyme commission number. With the exception of probable L-ascorbate-6-phosphate lactonase UlaG (E.C. 3.1.1.-), the sequences and annotations of the remaining six genes were confirmed using BLASTp. A summary comparing the nucleotide and amino acid sequences of these six genes is provided in Table 3. There are three genes (tRNA nucleotidyltransferase, arginine deiminase, UDP-galactosephosphotransferase) whose non-synonymous sites were found only in *E. faecalis* KACC 91532. In tRNA nucleotidyltransferase, glutamic acid and serine had been changed to lysine and asparagines; for arginine deiminase, arginine had been changed to histidine. UDP-galactosephosphotransferase had many variable positions near the end of its amino acid sequence region. The amino acid sequence changes in these three genes are provided in Fig. 3A. Arginine deiminase had an available protein 3D structure in NCBI and we identified the location of the variant in the 3D protein structure using Cn3D. The variant in arginine deiminase was estimated to be located in a chain between an α-helix and β-strand sheet of 1LXYA (Fig. 3B).

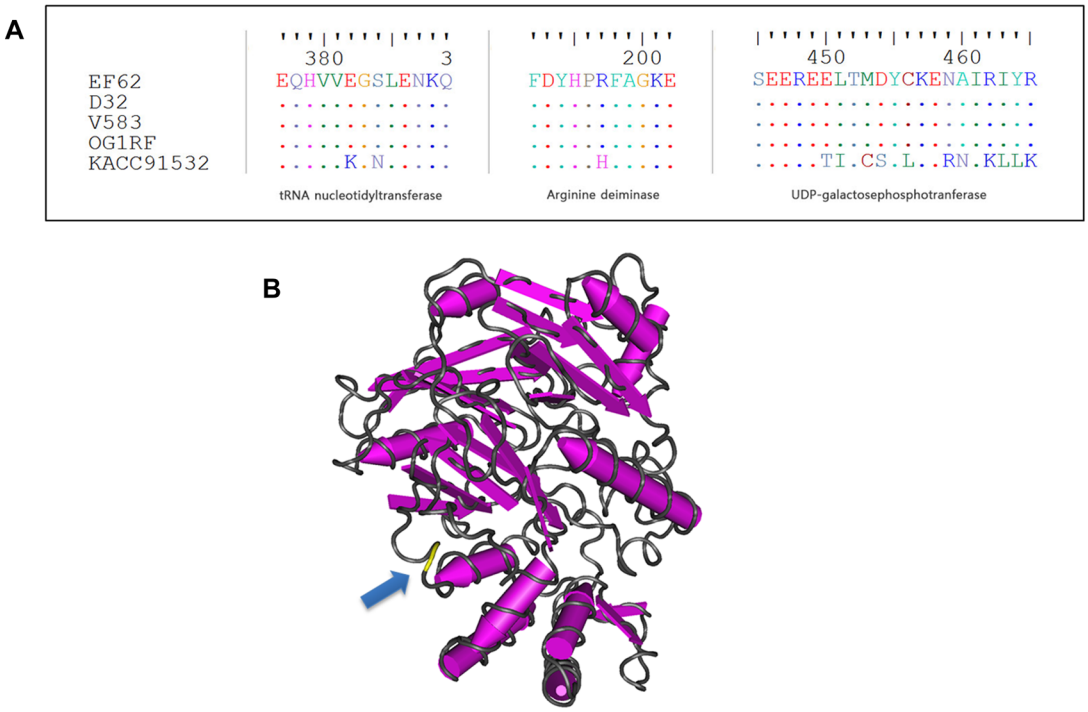

Fig. 3. *E. faecalis* KACC 91532-specific amino acid variants in evolutionarily accelerated genes.
(**A**) Amino acid variant in tRNA-nucleotidyltransferase, arginine deiminase, and UDP-galactosephosphotransferase. (**B**) Location of the variant in arginine deiminase.

Discussion

454 GS FLX+ (Roche) using pyrosequencing chemistry is routinely used for microorganism genome sequencing owing to its long read length. However, the high error rate in homopolymers and high reagent costs for pyrosequencing are significant obstacles to whole genome sequencing [21]. The addition of Illumina sequencing reads (sequencing by synthesis) to pyrosequencing data provides a useful and cost-effective way to get more complete genome assemblies. Accordingly, many studies have employed a combination of these two different methods for whole genome sequencing. Through testing various combinations of read data sets and various *de novo* assembly programs employing different algorithms, we have developed a newly constructed *de novo* assembly pipeline. Using Newbler under default conditions and only 454 reads showed better performance than using both 454 reads and Illumina reads at the same time. We think that this result stems from the characteristics of the *E. faecalis* genome, including its small genome size and low repeat content. Allpaths-LG is a *de novo* assembler that uses Illumina reads, and was not designed to perform error correction for raw read data. However, it has been used to generate input data for other *de novo* assembly programs [34]. An independent error correction module (Errorcorrectread.pl) from Allpaths-LG (http://www.broadinstitute.org/software/allpaths-lg/blog/) can be used for error correction of raw read data more easily. The result of gap filling showed that the combinational use of GapCloser and Gapfiller effectively closed gaps that existed in the scaffolds. We think that this result comes from the algorithm of Gapfiller, which removes the low quality edge and extends gap edges with k-mers. The *de novo* assembly pipeline described in this paper is an effective pipeline that uses independent and suitable programs for each step. This *de novo* assembly pipeline can reduce costs for iterated Sanger sequencing to generate complete genomes and provide more accurate information about microorganisms.

No known gentamicin resistance genes were identified in *E. faecalis* KACC 91532 to account for the observed gentamicin-resistant phenotype. However, *E. faecalis* KACC 91532 does possess multidrug resistance (MDR) efflux pump genes. As drug efflux systems are a well-established mechanism of antibiotic resistance [22] [1], we believe that the gentamicin resistance of *E. faecalis* KACC 91532 is related to this system. *E. faecalis* KACC 91532 exhibited thermal resistance in a heat resistance assay. Chaperone protein DnaK, one of the heat-shock proteins, was found in

E. faecalis KACC 91532 and may be related to its thermotolerance. In terms of virulence factors, *E. faecalis* KACC 91532 had sex pheromone-related genes (aggregation substance Asa1/PrgB) present in only one of four reference strains (V583). Aggregation substance Asa1/ProgB encodes for cell-surface protein Asc10. This protein is involved in cell aggregation, and can lead to the horizontal transfer of antibiotics resistance genes, such as pheromone-inducible tetracycline-resistance plasmid pCF10 [10], to other bacteria. Hence, *E. faecalis* KACC 91532 should only be used as a probiotic strain for Korean infants after careful consideration. This analysis demonstrates that by understanding gene contents using *de novo* assembly, we can build upon existing probiotics selection criteria to produce safer probiotics and health supplements.

Evolutionarily Accelerated Genes and KACC 91532-Specific Variants

tRNA nucleotidyltransferase performs 3′-terminal-CCA tRNA sequence repair in conjunction with poly(A) polymerase I and polynucleotide phosphorylase; this process is essential for the growth of the bacteria [32]. *O*-Succynylbenzoic acid-CoA ligase (E.C. 6.2.1.26) is used in vitamin K production, which can in turn stimulate microbial growth [26]. As *E. faecalis* KACC 91532 is the fastest-growing strain among all enterococci isolates examined, the presence of these two evolutionarily accelerated genes may account for the elevated growth rate of this strain.

Arginine deiminase (ADI) is the first enzyme in the arginine deiminase pathway, and is commonly found in acid-resistant LAB. This pathway produces ammonia by converting L-arginine into L-citrulline; the resulting ammonia helps to buffer the organism under acidic conditions. The acid resistance of *E. faecalis* KACC 91532 was above average compared with other *E. faecalis* isolates. More research is needed to establish a direct correlation between the variation in arginine deiminase and acid resistance in *E. faecalis* KACC 91532; gene-specific population analysis can be used to determine the effect of these variants.

Acknowledgments

This study was supported by funding of the RDA projects PJ0085852013 and PJ008487.

References

1. Alekshun MN, Levy SB. 2007. Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128:** 1037-1050.
2. Andrews S. 2012. FastQC. A quality control tool for high throughput sequence data. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc.
3. Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, *et al.* 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9:** 75.
4. Baronets NG. 2003. Vitamin K as a stimulator of microbial growth. *Zh. Mikrobiol. Epidemiol. Immunobiol.* **4:** 104-105 [In Russian].
5. Cariolato D, Andrighetto C, Lombardi A. 2008. Occurrence of virulence factors and antibiotic resistances in *Enterococcus faecalis* and *Enterococcus faecium* collected from dairy and human samples in North Italy. *Food Control* **19:** 886-892.
6. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25:** 119-120.
7. Castillo-Davis C, Kondrashov F, Hartl D, Kulathinal R. 2004. The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. *Genome Res.* **14:** 802-811.
8. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17:** 540.
9. Chaisson MJ, Pevzner PA. 2008. Short read fragment assembly of bacterial genomes. *Genome Res.* **18:** 324-330.
10. Chow J, Thal L, Perri M, Vazquez J, Donabedian S, Clewell D, *et al.* 1993. Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. *Antimicrob. Agents Chemother.* **37:** 2474-2477.
11. Chung JW, Bensing BA, Dunny GM. 1995. Genetic analysis of a region of the *Enterococcus faecalis* plasmid pCF10 involved in positive regulation of conjugative transfer functions. *J. Bacteriol.* **177:** 2107-2117.
12. Eaton TJ, Gasson MJ. 2001. Molecular screening of *Enterococcus* virulence determinants and potential for genetic exchange between food and medical isolates. *Appl. Environ. Microbiol.* **67:** 1628-1635.
13. Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A, *et al.* 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. tomato DC3000. *Proc. Natl. Acad. Sci. USA* **102:** 11064-11069.
14. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, *et al.* 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA* **108:** 1513-1518.
15. Gordon A, Hannon G. 2010. Fastx-toolkit : FASTQ/A short-reads pre-processing tools http://hannonlab.cshl.edu/fastx_toolkit.
16. Hall T. 2005. Bioedit version 7.0. 4. Department of Microbiology, North Carolina State University
17. Hancock LE, Gilmore MS. 2006. Pathogenicity of enterococci, pp. 299-311. *In* Fischetti V, Novick R, Ferretti J, Portnoy D,

Rood J (eds.). *Gram-positive Organisms.* ASM Press, Washington, DC.
18. Kim MK, Choi A, Han GS, Jeong SG, Chae HS, Jang A, *et al.* 2011. Development of probiotic dairy product for the normalization of microbial flora in Korean infants. *Korean J. Food Sci. Anim. Resour.* **31:** 290-295.
19. Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. USA* **102:** 10557.
20. Lempiäinen H, Kinnunen K, Mertanen A, Wright A. 2005. Occurrence of virulence factors among human intestinal enterococcal isolates. *Lett. Appl. Microbiol.* **41:** 341-344.
21. Li R. 2009. Short oligonucleotide analysis package: SOAPdenovo 1.03. Beijing Genomics Institute, Beijing.
22. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, *et al*. 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol*. **10:** 599-606.
23. Lomovskaya O, Watkins W. 2001. Inhibition of efflux pumps as a novel approach to combat drug resistance in bacteria. *J. Mol. Microbiol. Biotechnol.* **3:** 225-236.
24. Martín R, Jiménez E, Olivares M, Marín M, Fernández L, Xaus J, *et al*. 2006. *Lactobacillus salivarius* ECT 5713, a potential probiotic strain isolated from infant feces and breast milk of a mother–child pair. *Int. J. Food Microbiol.* **112:** 35-43.
25. Mercenier A, Pavan S, Pot B. 2003. Probiotics as biotherapeutic agents: present knowledge and future prospects. *Curr. Pharmaceut. Des.* **9:** 175-191.
26. Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* **13:** S8.
27. Ochman H, Moran NA. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292:** 1096-1099.
28. Pavlova S, Kilic A, Kilic S, So JS, Nader-Macias M, Simoes J, *et al*. 2002. Genetic diversity of vaginal lactobacilli from women in different countries based on 16S rRNA gene sequences. *J. Appl. Microbiol.* **92:** 451-459.
29. Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, *et al.* 2007. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat. Genet.* **39:** 839-847.
30. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, *et al*. 2006. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118:** 511-521.
31. PLOTREE D, D PLOTGRAM. 1989. PHYLIP-phylogeny inference package (version 3.2).
32. Reuven NB, Zhou Z, Deutscher MP. 1997. Functional overlap of tRNA nucleotidyltransferase, poly (A) polymerase I, and polynucleotide phosphorylase. *J. Biol. Chem.* **272:** 33255-33259.
33. Ryan KJ, Ray CG. 2010. *Sherris Medical Microbiology*. McGraw Hill Medical.
34. Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, *et al.* 2012. GAGE: a critical evaluation of genome assemblies and assembly algorithms. *Genome Res.* **22:** 557-567.
35. Salzberg SL, Sommer DD, Puiu D, Lee VT. 2008. Gene-boosted assembly of a novel bacterial genome from very short reads. *PLoS Comput. Biol.* **4:** e1000186.
36. Sarkar S. 2008. Effect of probiotics on biotechnological characteristics of yoghurt: a review. *Br. Food J.* **110:** 717-740.
37. Schlievert PM, Gahr PJ, Assimacopoulos AP, Dinges MM, Stoehr JA, Harmala JW, *et al*. 1998. Aggregation and binding substances enhance pathogenicity in rabbit models of *Enterococcus faecalis* endocarditis. *Infect. Immun.* **66:** 218-223.
38. Schmidt HA, Strimmer K, Vingron M, Von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18:** 502-504.
39. Solheim M, Aakra Å, Snipen LG, Brede DA, Nes IF. 2009. Comparative genomics of *Enterococcus faecalis* from healthy Norwegian infants. *BMC Genomics* **10:** 194.
40. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. 2000. Cn3D: sequence and structure views for Entrez. *Trends Biochem. Sci.* **25:** 300.
41. Wikler MA. 2006. Performance standards for antimicrobial susceptibility testing: sixteenth informational supplement. Clinical and Laboratory Standards Institute.
42. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24:** 1586-1591.