Draft Genome Sequences of Enterococcus faecalis Strains Isolated from Healthy Japanese Individuals

Kazuki Tanaka,a,b,c Tsubasa Watabe,a,b Kumiko Kato,d Tomoya Tsukimi,a,b Mitsuhiko P. Sato,e Toshitaka Odamaki,d Masaru Tomita,a,b Shinji Fukuda,a,b,c,f,g

4Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan
5Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa, Japan
6Intestinal Microbiota Project, Kanagawa Institute of Industrial Science and Technology, Kawasaki, Kanagawa, Japan
7Next Generation Science Institute, Morinaga Milk Industry Co., Ltd., Zama, Kanagawa, Japan
8Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, Higashi, Fukuoka, Japan
9Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki, Japan
10PRESTO, Japan Science and Technology Agency, Saitama, Japan

ABSTRACT Enterococcus faecalis is a common commensal of the intestines of humans and other mammals but is also a frequent cause of serious ailments. Here, we report 14 draft genome sequences of strains of Enterococcus faecalis, a normal inhabitant and Gram-positive bacterium that was isolated from 7 healthy Japanese volunteers.

Enterococcus faecalis is a common commensal of the intestines of humans and other mammals. There has been an increasing number of multidrug-resistant strains and nosocomial infections of the bacterium due to the frequent use of antibiotics (1), which drives the need to understand the bacterial infection mechanism of the species.

E. faecalis was isolated from 7 healthy Japanese volunteers on a transoligosaccharide (TOS)-propionate agar medium plate (Yakult Pharmaceutical Industry) and anaerobically cultured for 48 h at 37°C. The isolated colonies from each subject were further incubated with de Man, Rogosa, and Sharpe (MRS) liquid medium (Wako) and cultured anaerobically for 24 h at 37°C. Genomic DNA was isolated using DNeasy blood and tissue kits (Qiagen). Sequencing libraries were prepared using the NEBNext Ultra II FS DNA library prep (New England BioLabs) and sequenced with the HiSeq 2500 platform (Illumina) using 150-bp paired-end read technology. The reads were filtered and trimmed using Platanus_trim v1.0.7 (http://platanus.bio.titech.ac.jp/platanus_trim), and 940,782 to 1,393,444 reads per sample remained. Trimmed reads were assembled with Platanus v1.2.4 (2), with default parameters, and contigs equal to or shorter than 300 bp were discarded with an in-house script (https://github.com/MitsuhikoP/cut_short_fasta). Genome completeness was analyzed using BUSCO v1 (3) on the gVolante Web server (4), and gene annotation was analyzed using DFAST (2), as previously described (5). Identification of E. faecalis was performed with BLAST+ v2.4.0 (6, 7) and pyani v0.2.7 (https://github.com/widdowquinn/pyani). The draft genome sequences of E. faecalis have 2,794,898 to 2,930,781 bp in a total of 13 to 33 contigs, with an N50 value of 218,789 to 563,424 bp and GC content of approximately 37.5%. The genomes were predicted to contain 2,627 to 2,793 putative coding sequences.

For comparative analysis, GenBank files of 25 E. faecalis strains were downloaded from the NCBI FTP site on 25 April 2019. Nucleotide sequence alignments for core genes were produced using Roary v3.12.0 (8), with a minimum blastp percentage identity of 95, and MAFFT v7.407 (9). A phylogenetic tree was constructed using FastTree v2.1.3 (10) with the general time-reversible (GTR) plus category (CAT) model.


Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine

Copyright © 2019 Tanaka et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Shinji Fukuda, sfukuda@fc.keio.ac.jp.

Received 17 July 2019
Accepted 31 August 2019
Published 3 October 2019
The phylogenetic tree was rooted using the phangorn package v2.4.0 and drawn using the ape package v5.3 in R v3.3.3.

The core-genome phylogeny indicated that the clade consisting of the strains isolated from the Japanese volunteers was divided into a clade close to the L12 strain and a clade close to the D32 strain (Fig. 1), which are pig-derived bacteria (GenBank accession numbers CP018102 and CP003726, respectively). Therefore, these strains

![Midpoint rooting phylogenetic tree obtained from a concatenated nucleotide sequence alignment of the 1,849 core genes of the 14 Enterococcus faecalis strains isolated from the 7 volunteers and reference genomes of 25 E. faecalis strains.](image)

The phylogenetic tree was rooted using the phangorn package v2.4.0 and drawn using the ape package v5.3 in R v3.3.3.

The core-genome phylogeny indicated that the clade consisting of the strains isolated from the Japanese volunteers was divided into a clade close to the L12 strain and a clade close to the D32 strain (Fig. 1), which are pig-derived bacteria (GenBank accession numbers CP018102 and CP003726, respectively). Therefore, these strains

### TABLE 1 Genomic features of strains used in this study

<table>
<thead>
<tr>
<th>Sample</th>
<th>GenBank accession no.</th>
<th>DRA accession no.</th>
<th>GC content (%)</th>
<th>No. of contigs</th>
<th>Total contig size (bp)</th>
<th>Largest contig size (bp)</th>
<th>N_50 (bp)</th>
<th>No. of CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>BJTH0000000000</td>
<td>DRR179066</td>
<td>37.5</td>
<td>32</td>
<td>2,928,338</td>
<td>659,735</td>
<td>218,789</td>
<td>2,793</td>
</tr>
<tr>
<td>B1</td>
<td>BJTI0000000000</td>
<td>DRR179067</td>
<td>37.3</td>
<td>31</td>
<td>2,850,529</td>
<td>890,826</td>
<td>301,644</td>
<td>2,688</td>
</tr>
<tr>
<td>B2</td>
<td>BJTI0000000000</td>
<td>DRR179068</td>
<td>37.5</td>
<td>32</td>
<td>2,930,781</td>
<td>659,737</td>
<td>249,285</td>
<td>2,791</td>
</tr>
<tr>
<td>B3</td>
<td>BJTK0000000000</td>
<td>DRR179069</td>
<td>37.4</td>
<td>29</td>
<td>2,794,898</td>
<td>835,194</td>
<td>318,734</td>
<td>2,627</td>
</tr>
<tr>
<td>C1</td>
<td>BJTL0000000000</td>
<td>DRR179070</td>
<td>37.5</td>
<td>29</td>
<td>2,871,953</td>
<td>889,827</td>
<td>261,062</td>
<td>2,714</td>
</tr>
<tr>
<td>C2</td>
<td>BJTM0000000000</td>
<td>DRR179071</td>
<td>37.5</td>
<td>13</td>
<td>2,873,694</td>
<td>1,435,291</td>
<td>563,268</td>
<td>2,271</td>
</tr>
<tr>
<td>D1</td>
<td>BJTN0000000000</td>
<td>DRR179072</td>
<td>37.5</td>
<td>22</td>
<td>2,872,519</td>
<td>881,255</td>
<td>514,512</td>
<td>2,717</td>
</tr>
<tr>
<td>D2</td>
<td>BJTN0000000000</td>
<td>DRR179073</td>
<td>37.5</td>
<td>17</td>
<td>2,874,611</td>
<td>889,949</td>
<td>480,305</td>
<td>2,716</td>
</tr>
<tr>
<td>E1</td>
<td>BJTP0000000000</td>
<td>DRR179074</td>
<td>37.5</td>
<td>18</td>
<td>2,873,047</td>
<td>881,086</td>
<td>563,424</td>
<td>2,716</td>
</tr>
<tr>
<td>E2</td>
<td>BJTP0000000000</td>
<td>DRR179075</td>
<td>37.5</td>
<td>19</td>
<td>2,872,883</td>
<td>881,086</td>
<td>562,779</td>
<td>2,713</td>
</tr>
<tr>
<td>F1</td>
<td>BJTR0000000000</td>
<td>DRR179076</td>
<td>37.5</td>
<td>27</td>
<td>2,873,489</td>
<td>881,203</td>
<td>361,800</td>
<td>2,713</td>
</tr>
<tr>
<td>F2</td>
<td>BJTS0000000000</td>
<td>DRR179077</td>
<td>37.5</td>
<td>33</td>
<td>2,928,825</td>
<td>659,737</td>
<td>249,285</td>
<td>2,790</td>
</tr>
<tr>
<td>G1</td>
<td>BJTT0000000000</td>
<td>DRR179078</td>
<td>37.5</td>
<td>25</td>
<td>2,873,406</td>
<td>733,617</td>
<td>424,659</td>
<td>2,712</td>
</tr>
<tr>
<td>G2</td>
<td>BJTU0000000000</td>
<td>DRR179079</td>
<td>37.5</td>
<td>27</td>
<td>2,874,408</td>
<td>881,085</td>
<td>480,088</td>
<td>2,716</td>
</tr>
</tbody>
</table>

*The sample name represents the subject and the colony number.

*GC content (%) is the relative frequency (percentage) of guanine and cytosine (G+C)/(A+T+G+C).

*CDSs, coding sequences.
could be colonized both in humans and in livestock. Furthermore, since a large number of phylogenetically similar strains were also detected from multiple subjects, the strains may be transmitted between humans frequently.

This study was approved by the ethics committee of Keio University Shonan Fujisawa Campus and the Kenshokai under approval numbers 195 and 20170915-4, respectively. All subjects were informed of the purpose of this study, and written consent was obtained from all subjects. Data availability. The GenBank and DDBJ Sequence Read Archive (DRA) accession numbers are listed in Table 1.

ACKNOWLEDGMENTS

We thank Yasuyuki Ohkawa, Tetsuya Hayashi, Yasuhiro Gotoh, and Keiji Nakamura for genome sequencing, Yoshitoshi Ogura for data analysis, and Haruo Suzuki for critical discussions. This work was supported in part by JSPS KAKENHI (grants 16H04901, 17H05654, and 18H04805 to S.F.), JST PRESTO (grant JPMJPR1537 to S.F.), AMED-CREST (grant JP19gm1010009 to S.F.), the Takeda Science Foundation (to S.F.), the Food Science Institute Foundation (to S.F.), the Yamagata Prefectural Government and the city of Tsuruoka (to M.T.), and the Taikichiro Mori Memorial Research Fund (to K.T.). Sequence analysis in this study was supported by JSPS KAKENHI (grant 16H06279 to PAGS).

REFERENCES


