

RESEARCH ARTICLE

Genomic Features and Niche-Adaptation of *Enterococcus faecium* Strains from Korean Soybean-Fermented Foods

Eun Bae Kim^{1,2*}, Gwi-Deuk Jin¹, Jun-Yeong Lee³, Yun-Jaie Choi^{3,4}

1 Department of Animal Life Science, College of Animal Life Sciences, Kangwon National University, Chuncheon, Kangwon-do, Republic of Korea, **2** Division of Applied Animal Science, College of Animal Life Sciences, Kangwon National University, Chuncheon, Kangwon-do, Republic of Korea, **3** Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea, **4** Research Institute for Agriculture and Life Science, Seoul National University, Seoul, Republic of Korea

* itanimal@kangwon.ac.kr



CrossMark
click for updates

OPEN ACCESS

Citation: Kim EB, Jin G-D, Lee J-Y, Choi Y-J (2016) Genomic Features and Niche-Adaptation of *Enterococcus faecium* Strains from Korean Soybean-Fermented Foods. PLoS ONE 11(4): e0153279. doi:10.1371/journal.pone.0153279

Editor: Willem van Schaik, University Medical Center Utrecht, NETHERLANDS

Received: November 13, 2015

Accepted: March 25, 2016

Published: April 12, 2016

Copyright: © 2016 Kim et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All draft genome sequences are available from NCBI GenBank database (accession numbers LBIT00000000, LDNE00000000, LDNF00000000, LDNC00000000, LDND00000000, LDNG00000000, LDNH00000000, LDNI00000000, LDNJ00000000, LDNK00000000).

Funding: This study was supported by a grant from the Strategic Initiative for Microbiomes in Agriculture and Food, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea as part of the multi-ministerial Genome Technology to Business Translation Program (Grant ID: 914005-04). The funders had no role in study design, data collection

Abstract

Certain strains of *Enterococcus faecium* contribute beneficially to human health and food fermentation. However, other *E. faecium* strains are opportunistic pathogens due to the acquisition of virulence factors and antibiotic resistance determinants. To characterize *E. faecium* from soybean fermentation, we sequenced the genomes of 10 *E. faecium* strains from Korean soybean-fermented foods and analyzed their genomes by comparing them with 51 clinical and 52 non-clinical strains of different origins. Hierarchical clustering based on 13,820 orthologous genes from all *E. faecium* genomes showed that the 10 strains are distinguished from most of the clinical strains. Like non-clinical strains, their genomes are significantly smaller than clinical strains due to fewer accessory genes associated with antibiotic resistance, virulence, and mobile genetic elements. Moreover, we identified niche-associated gene gain and loss from the soybean strains. Thus, we conclude that soybean *E. faecium* strains might have evolved to have distinctive genomic features that may contribute to its ability to thrive during soybean fermentation.

Introduction

Enterococcus faecium is a Gram-positive bacterium found in the gastro-intestinal tracts (GIT) of animals [1], fermented foods [2,3], and a variety of other environments [4]. *E. faecium* is a lactic acid bacterium, which produces lactic acid as a final product of carbohydrate fermentation [5], and it may have hypothetical roles in the early stages of soybean fermentation [3,6]. Some strains of *E. faecium* have beneficial roles in GIT health as commensal or probiotic bacteria [7]; others are associated with nosocomial infections such as bacteremia and endocarditis in humans [1]. Recently, certain *E. faecium* strains have emerged as critical hospital pathogens due to their resistance to vancomycin treatments [1].

Although the pathogenicity of *E. faecium* has not been yet fully defined, many antibiotic resistance (AR) genes and virulence factors (VF) have been identified [1]. Mobile genetic

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

elements (ME) contribute to horizontal transfer of co-localized AR and VF genes [8]. AR, VF, and ME genes are more frequently found in clinically isolated *E. faecium* strains than community-associated non-clinical strains [8]. Moreover, comparative genomic analysis of *E. faecium* strains revealed different lineages that represent clinical and non-clinical strains [8–10].

Many Korean foods have long been prepared based on fermentation of soybeans. Fermented soybean blocks (Meju), fermented soybean paste (Doenjang), and soy sauce (Ganjang) are representative fermented soybean foods in Korea. A recent study showed that *Enterococcus* spp were found at considerable levels in fermented soybeans at the early stage of fermentation [6], and *E. faecium* strains were frequently isolated from the fermented products [3]. Despite the important roles of *E. faecium* in soybean fermentation, genomic features and contents of *E. faecium* have never been evaluated by using genomic information. Here, we sequenced genomes of 10 *E. faecium* strains isolated from fermented soybeans to characterize their genomic features.

Materials and Methods

E. faecium strains

The 10 *E. faecium* strains used in this study were obtained from a Microorganism Collection, Korean Agricultural Culture Collection (KACC). They were all isolated from fermented soybean products as listed in Table 1. To avoid clonality or geographical relatedness among the 10 strains, we selected them from the products of seven independent commercial companies.

Preparation of Genomic DNA Library and Sequencing

Bacterial cells were harvested from overnight BHI (Brain-Heart Infusion) broth culture. Harvested cells were washed two times with 1X PBS buffer. The cells were further processed to extract genomic DNA using G-spin Genomic DNA Extraction Kit (iNtRON Biotechnology, Cat #17121, South Korea). Genomic DNA was fragmented using NEBNext dsDNA Fragmentase (NEB, Cat #M0348S, MA, USA). The fragmented DNA was further processed to construct a genomic DNA library using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, Cat #E7370S, MA, USA). Genomic DNA libraries were constructed with ~350-bp inserts and sequenced by Illumina HiSeq2500 for 100 bp paired-end reads.

Genome Assembly and Annotation

Sequenced reads were quality-filtered using in-house Perl scripts [11]. In brief, when 95% of nucleotide bases in a read were given a quality score over 31 (Illumina 1.8+) and the read length was ≥ 70 bp, the read was used for genome assembly. The filtered paired and single reads were assembled using Ray 2.3.1 [12] with a k-mer size of 31 bp. The assembled draft genome sequences were uploaded to an annotation server, RAST [13] with default options for bacteria.

Genome Comparison and Strain Clustering

For the ortholog collection, *E. faecium* nucleotide sequences were downloaded on 07/03/2014 from the NCBI GenBank database. We inspected protein coding sequences (CDS) from GenBank information using in-house Perl scripts. In brief, we excluded premature stop codons, codon shifts by deletions and insertions, errors in CDS length, etc. as shown previously [8]. CDS were extracted from NCBI *E. faecium* nucleotides and our 10 annotated *E. faecium* genomes. The CDS were clustered based on the sequence similarity as suggested previously [8]. In brief, CDS were mutually aligned to highly similar CDS to generate CDS clusters using high-throughput sequence alignment software, GASSST [14] under the parameters of $\geq 90\%$

Table 1. Ten soybean *E. faecium* strains and genomes sequenced in this study.

Strains ^a	Origins ^b / Producers	Genome (bp)	Contigs	Max Contig (bp)	N50 (bp)	GC (%)	Accession ^c
15689	Doenjang / A	2,824,234	128	213,108	72,492	38.0	LBIT00000000
15960	Doenjang / B	2,692,315	93	185,163	71,626	38.1	LDNE00000000
15962	Doenjang / C	2,719,801	99	216,188	65,696	38.2	LDNF00000000
15700	Ganjang / A	2,852,054	123	213,345	81,177	38.0	LDNC00000000
15711	Meju / A	2,781,534	145	175,610	61,421	38.0	LDND00000000
16076	Meju / D	2,719,289	73	354,186	104,708	38.1	LDNG00000000
16093	Meju / E	2,585,336	96	197,388	57,918	38.1	LDNH00000000
16097	Meju / F	2,630,779	64	194,680	104,580	38.2	LDNI00000000
16100	Meju / G	2,837,155	104	159,513	83,784	38.0	LDNJ00000000
16106	Meju / B	2,673,905	74	268,871	107,917	38.0	LDNK00000000

^aStrain numbers assigned by KACC

^bDoenjang, fermented soybean paste; Ganjang, soy source; Meju, fermented soybean block

^cNCBI GenBank Accession Numbers. Both chromosome and plasmids are included.

doi:10.1371/journal.pone.0153279.t001

sequence identity and sensitivity level 5 (maximum). Each collection of clustered CDS was further assembled to make a consensus orthologous CDS. From the clusters, a total of 13,820 orthologous CDS were finally defined. Each CDS was examined to determine gene presence/absence. For strain clustering, we used genomes of 51 clinical and 52 non-clinical *E. faecium* strains (S1 Table) that were available from NCBI. Strain clustering analysis was conducted based on the presence/absence of each CDS as previously suggested [8]. Briefly, distance between strains was calculated according the Euclidean distance method. Existence of clades was statistically confirmed by 1,000 times re-sampling using the Pvcust R package [15].

Phylogenetic Analysis

To analyze clonal relatedness among soybean strains and assign their lineage among 113 *E. faecium* strains, we collected single nucleotide polymorphisms (SNPs) from the 113 strains. We used 990 core genes that are commonly present in all genomes. Due to incomplete genome assemblies, 100 core genes are excluded from SNP selection. To collect core gene sequences from each genome, ortholog sequences were aligned to each genome using GASSST under the 75–100% sequence similarity option. The aligned regions in each genome were extracted and collected for SNP selection. The collected core gene sequences from each genome were aligned among 113 strains using the alignment tool MUSCLE [16]. On the basis of the core gene alignment, we could detect SNPs. All of the SNPs in the core genes were further used to construct phylogenetic trees using the bioinformatics tool MEGA [17]. The evolutionary history and distance were inferred and computed using the Neighbor-Joining method [18] with bootstrap tests [19] and the Maximum Composite Likelihood method [20], respectively.

Detection of VF, AR, ME, and BA Genes

As previously defined [8], 34 VF genes, 748 AR orthologs, and ME genes were collected based on literature review [8], ARDB database [21], and sequence annotations [8], respectively. Bacteriocin-associated (BA) genes including both bacteriocins and their regulatory genes were collected based on known bacteriocin genes found in a bacteriocin database (BAGEL, <http://bagel.molgenrug.nl/>) [22]. Each genome was screened for the presence of these genes as described above.

Detection of Soybean Niche-Specific Genes

The frequency of each orthologous CDS was counted for *E. faecium* from all three different origins; soybean origin, non-soybean non-clinical origin, and clinical origin. Genes with higher frequency in the soybean origin than other origins were regarded as soybean-specific. Significance of different frequencies was evaluated by Fisher's Exact Test at $P < 0.05$.

Results

Relationships of Soybean *E. faecium* Strains

Before we describe the differences between 10 soybean strains and others from different origins, we need to demonstrate whether any specific strains are clonally related, which may affect subsequent comparisons. In total, 8,850 SNPs from 990 core genes, which are commonly found in 113 strains, were identified among the 10 soybean strains. We constructed a phylogenetic tree (Fig 1) that was based on these SNPs. While seven soybean strains are not closely related, three other strains (KACC 15689, KACC 15700 and KACC 15711) were very close. The three closely related strains were isolated from different products produced by the same company. However, we also identified 125 SNPs among these three strains. This observation indicates that 10 soybean strains are not highly biased and are appropriate for comparison between 10 soybean strains and others from different origins.

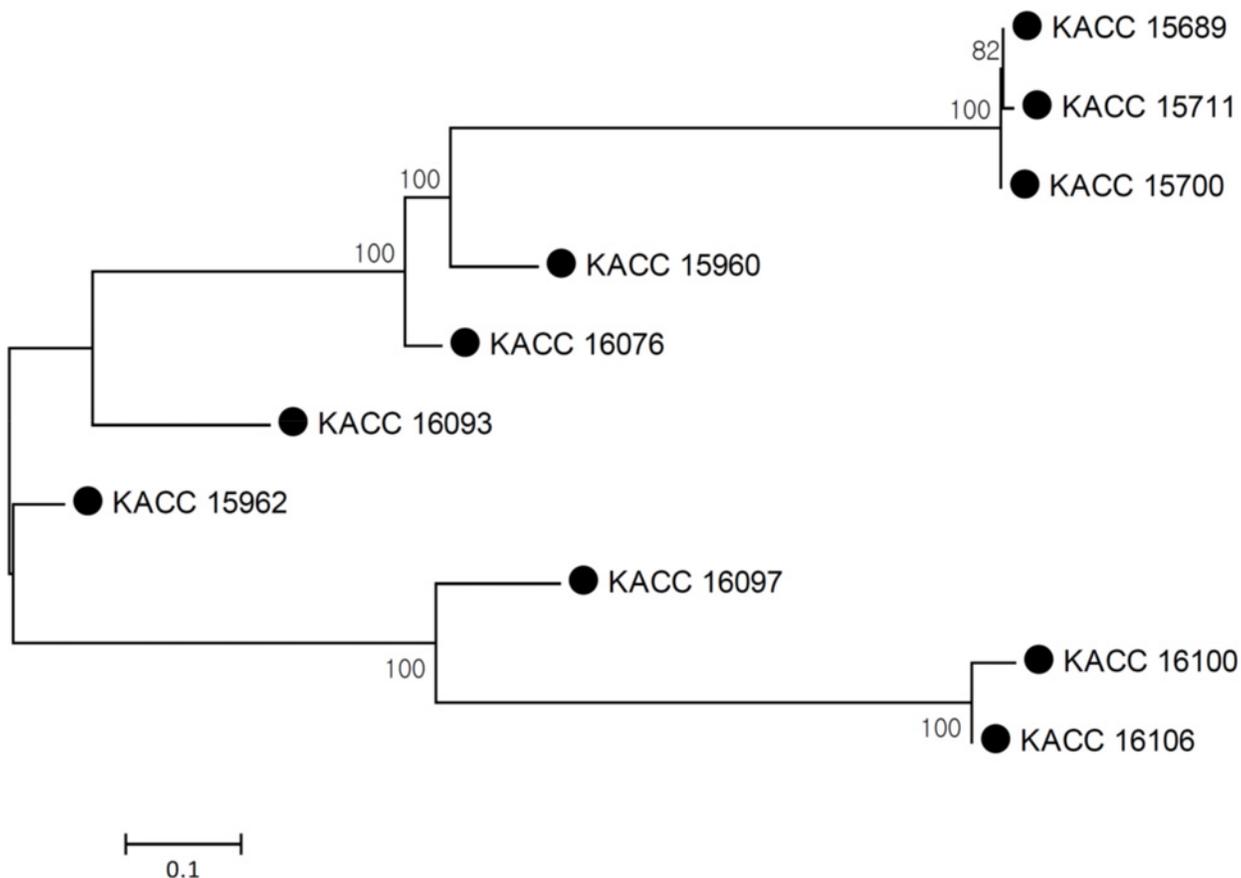


Fig 1. Evolutionary relationships of 10 soybean strains based on 8,850 SNPs from 990 core genes. The optimal tree with the sum of branch length = 2.15 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances are in the units of the number of base substitutions per site.

doi:10.1371/journal.pone.0153279.g001

Genome Sequencing and Genome Assembly

Ten *E. faecium* strains from fermented soybean were sequenced and assembled to obtain draft genomes (Table 1). The numbers of contigs ranged from 64 to 145. The average genome size of the strains was 2.73 ± 0.09 Mbp, and the GC content ranged from 37.95% to 38.16%. The average genome size of soybean-isolated strains ($n = 10$, Fig 2A) was significantly smaller than that of the clinical strains ($n = 51$, 2.88 ± 0.13 Mbp, $P = 0.006$), but not compared to that of non-clinical strains ($n = 52$, 2.78 ± 0.15 Mbp). From our pan-genome analysis, we identified 1,090 core CDS that are found in all 113 *E. faecium* genomes (10 soybean strains, 51 clinical strains, and 52 non-clinical strains). The number of accessory CDS that were found only in some strains was compared among the three groups (Fig 2B). Soybean-isolated strains have a total of $1,828 \pm 74$ accessory CDS, which are significantly fewer than in clinical strains ($2,149 \pm 197$, $P < 0.001$). Non-clinical strains ($1,985 \pm 217$) also have the higher numbers of accessory CDS compared to clinical strains ($P < 0.001$). The average GC content of soybean-isolated strains ($38.05 \pm 0.08\%$, Fig 2C) was similar to that of clinical strains ($38.01 \pm 0.19\%$), and GC content was higher in both groups compared to non-clinical strains ($37.86 \pm 0.20\%$, $P = 0.011$ and $P < 0.001$).

Lineage of *E. faecium* Strains from Soybean-Fermented Foods

According to previous studies [8,11], non-clinical strains belong to two major groups called the NC1 and NC2 clades; these are distinct from the CL clade, which is enriched with clinical strains. Here, we examined where the tested soybean strains were positioned by a hierarchical clustering analysis. Based on the presence or absence of orthologous CDS found in our pan-

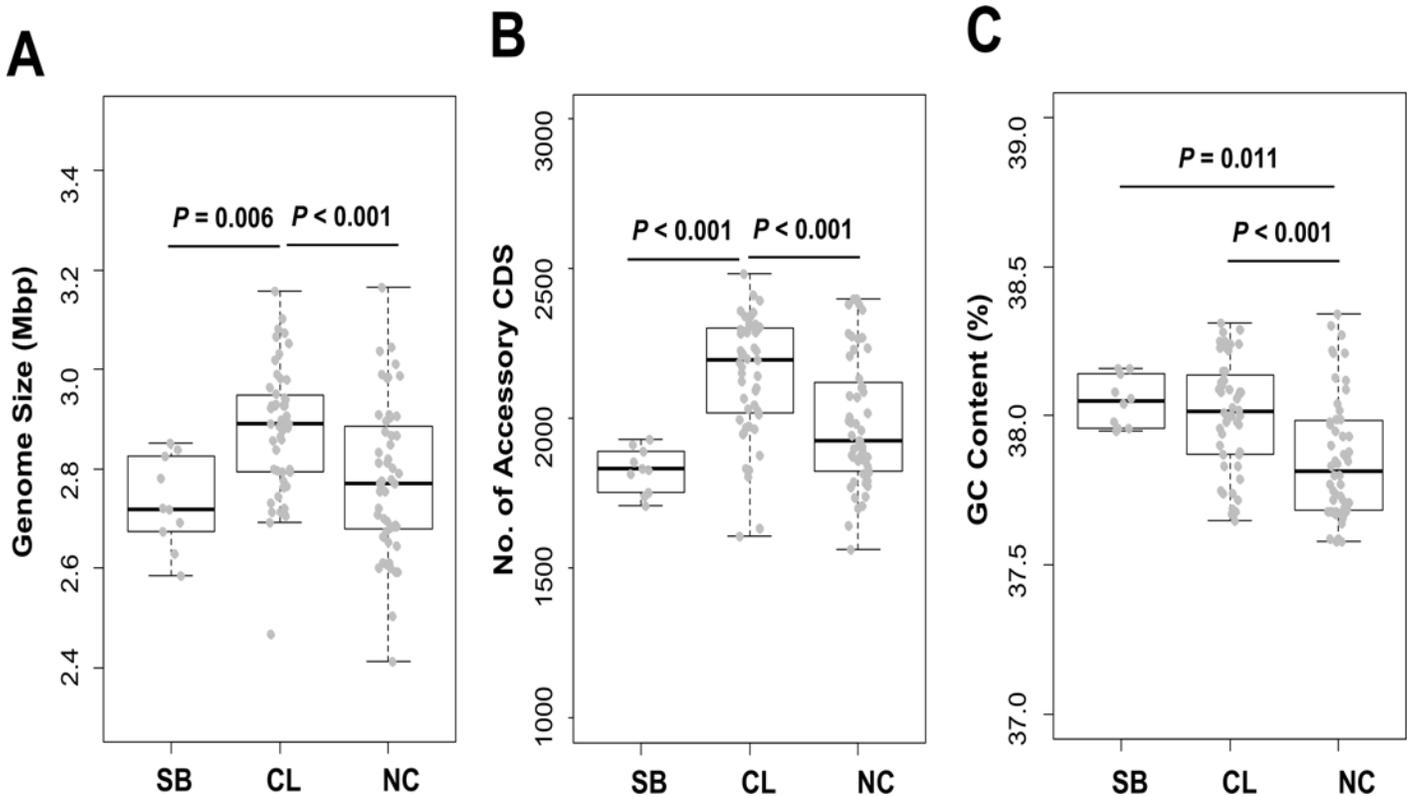


Fig 2. Genome size, number of accessory CDS, and GC content (%) of soybean strains. Genome sizes (A), accessory genes (B), and GC content (C) were compared among soybean (SB), clinical (CL) and non-clinical (NC) groups. Each gray spot indicate a single *E. faecium* strain.

doi:10.1371/journal.pone.0153279.g002

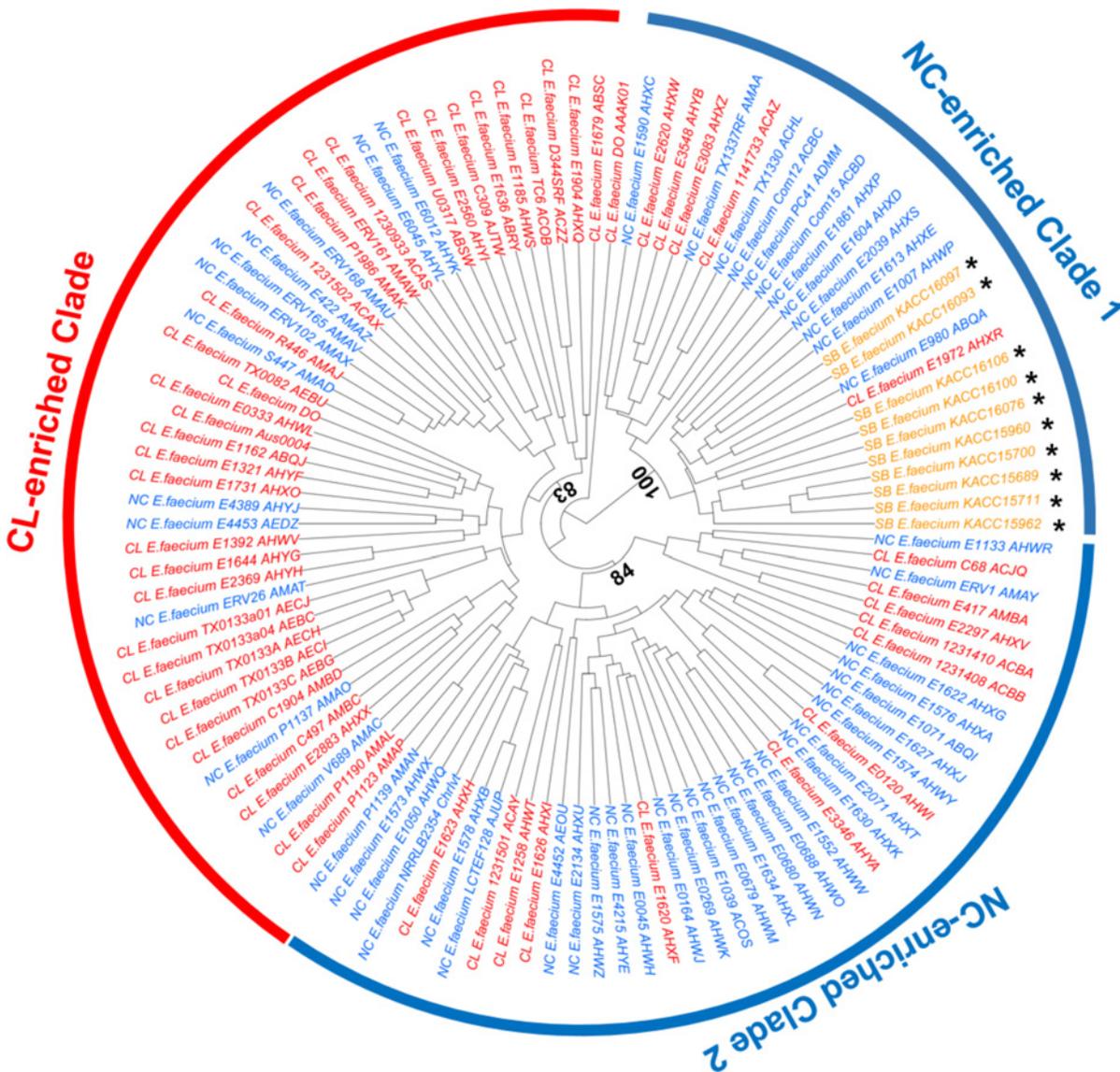


Fig 3. Hierarchical clustering of the 113 *E. faecium* strains used in this study. Based on ortholog presence/absence, clusters were obtained. Asterisks indicate the 10 soybean strains (Prefix: SB). Clinical and non-clinical strains were indicated by different prefixes, CL and NC, respectively. Approximately unbiased probabilities for the bootstrapping are shown in the center of the dendrogram.

doi:10.1371/journal.pone.0153279.g003

genome analysis, the 113 *E. faecium* genomes were positioned as shown in Fig 3. As expected, the soybean-isolated strains were positioned in NC-enriched clade 1 but not in NC-enriched clade 2, which clustered together with a major CL-enriched clade. The NC-enriched clades and CL-enriched clade were matched to previously defined clades, NC1, NC2 and CL [8,11]. We also identified that 10 soybean strains are positioned at the NC1 clade on a phylogenetic tree based on 59,739 SNPs detected from 990 core genes (Fig 4). This result indicates that soybean strains are genetically much closer to non-clinical strains than to clinical strains.

Prevalence of AR, VF, ME, and BA genes

AR, VF, ME, and BA genes contribute not only to the pathogenicity of *E. faecium* species [1] but also to environmental adaptation [8,11,23,24]. To better understand soybean *E. faecium*

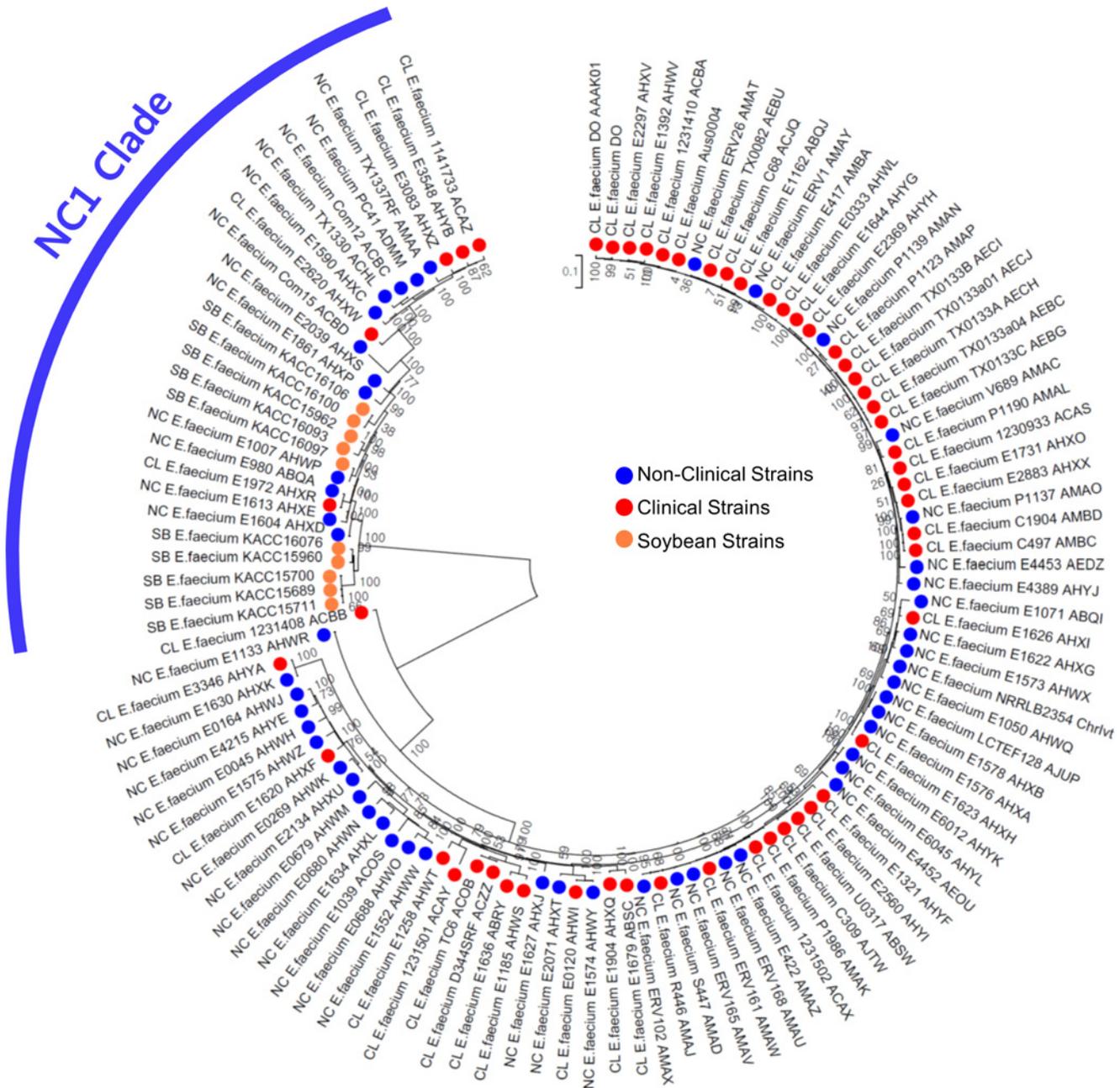


Fig 4. Evolutionary relationships of 113 strains based on 59,739 SNPs from 990 core genes. The optimal tree with the sum of branch length = 3.21 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances are in the units of the number of base substitutions per site.

doi:10.1371/journal.pone.0153279.g004

strains, we analyzed the prevalence of these genes (Fig 5). The soybean strains have fewer AR genes than both clinical and non-clinical strains. Only two AR genes, *bacA* and *tetU*, were found in certain soybean strains although a role of *tetU* as a resistance gene is under debate [25], and no vancomycin resistance genes were found. The soybean strains also have fewer VF and ME genes than the clinical strains but have a similar number of VF and ME genes as non-clinical strains [8,11]. Some VF genes (*efmA*, *acm*, *sgrA*, *cad*, and *cbh3*) were found in all

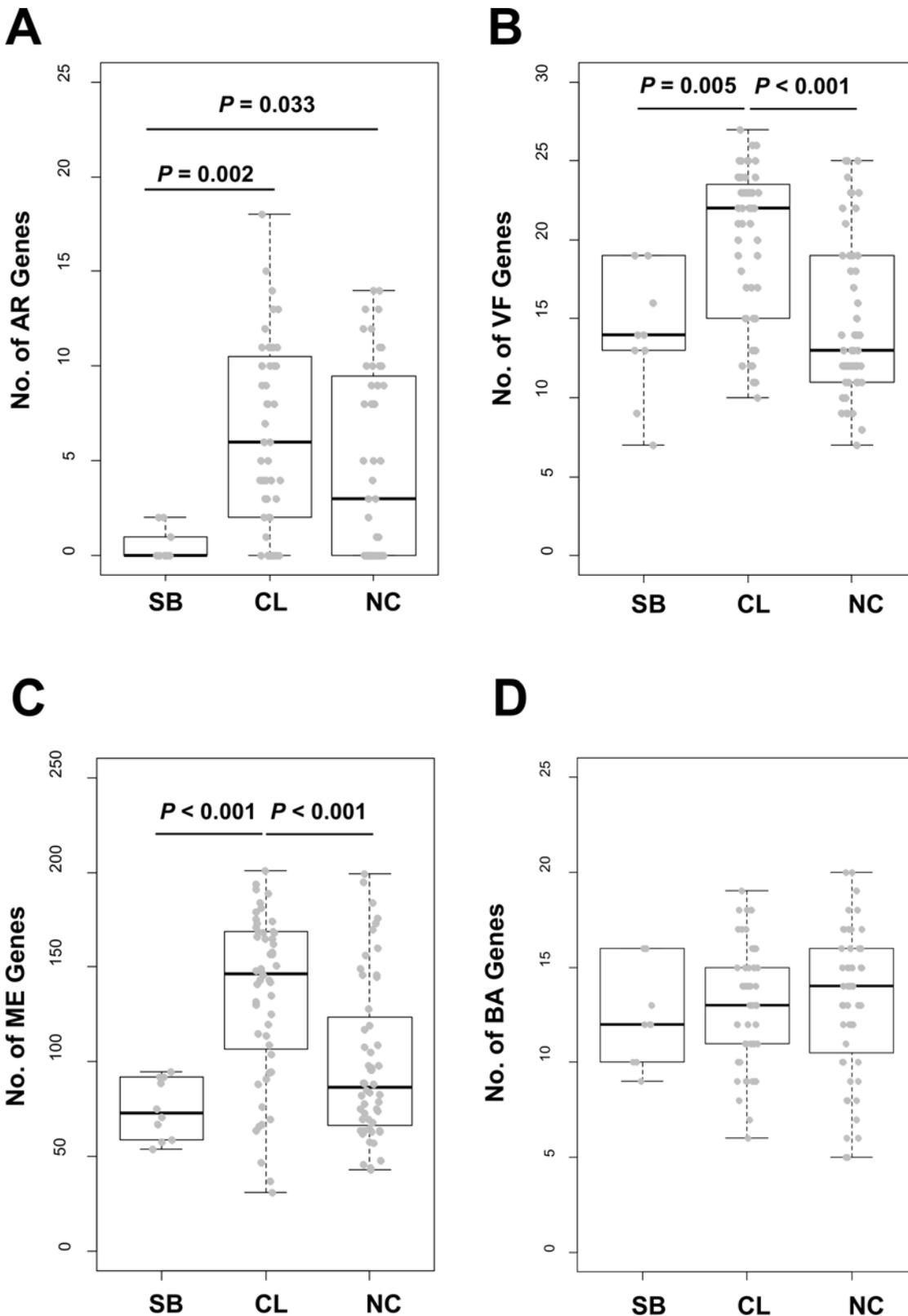


Fig 5. Frequencies of AR, VF, ME, and BA genes. Frequencies of four types of genes, AR (A), VF(B), ME (C) and BA (D), were compared among soybean (SB), clinical (CL) and non-clinical (NC) groups. Statistical significance was examined using Tukey's HSD (Honestly Significant Difference). Non-significant comparisons were omitted ($P > 0.05$). Each gray spot indicate a single *E. faecium* strain.

doi:10.1371/journal.pone.0153279.g005

soybean strains; however, the 3 major VF genes (*IS16*, *hyl_{efm}*, and *esp*) that are used for safety assessment in Europe [26] were not found in any soybean strain. The number of BA genes was not significantly different among the three groups. These results suggest that the genomic content of soybean strains of *E. faecium* is more similar to non-clinical strains than to the clinical strains.

Niche-Enriched Genes in Soybean *E. faecium* Strains

Although the genomic patterns of soybean *E. faecium* strains are similar to non-clinical strains as described above, we expected that they would have their own niche-enriched genes that are not frequently found in isolates of other origins. We identified not only such niche-enriched genes, but also missing genes in soybean *E. faecium* strains (Fig 6 and S2 Table). In total, 81 niche-specific genes were enriched in soybean strains and they were significantly fewer in the clinical ($P < 0.005$) and non-clinical strains ($P < 0.005$). Among the 81 genes, 6 gene clusters were identified (S1 Fig and S2 Table). Three clusters (Clusters 1, 3, and 5) were completely aligned to three known *E. faecium* plasmids (GenBank Accession No. CP013995, CP006030, and DQ198088), and another cluster (Cluster 4) was found not only in many *E. faecium* contigs but also in *Enterococcus faecalis* (GenBank Accession No. KJ756557) and *Enterococcus durans* (GenBank Accession No. CP012367). Clusters 1, 4, 5, and 6 are related to sugar metabolism, cell wall-associated function, the enterocin AS-48 system, and mannose/fructose transportation, respectively. We identified 17 genes that were frequently found in *E. faecium* from both clinical and other origins but were frequently missing from the soybean *E. faecium* strains. Among the 17 genes, we identified two major clusters (S1 Fig and S2 Table). Clusters 7 and 8 are related to arabino-furanose metabolism and mannose/fructose transportation, respectively. These results suggest that soybean *E. faecium* strains have niche-specific and nutrition/environment-adaptive genomic features, which distinguish them from *E. faecium* from other origins.

Discussion

Although *E. faecium* strains are frequently found in soybeans [27,28] and soybean-fermented foods [3,6], to our knowledge, the genomes of soybean-originated *E. faecium* have never been investigated. Here, we sequenced the genomes of *E. faecium* strains from fermented soybean foods and compared them to the genomes of *E. faecium* isolated from other sources.

Several genomic studies on *E. faecium* strains showed differences in *E. faecium* lineages—clinical and non-clinical clades [8,11,29,30]. *E. faecium* has been recognized as commensal and/or probiotic species. However, pathogenic phenotypes have emerged in certain clinical *E. faecium* strains due to overuse of antibiotics in hospitals [1], and these pathogenic strains show different genomic and phenotypic features because they have obtained niche-adapted genetic resources such as AR and VF genes to survive in the presence of antibiotics and to grow in hospitalized patients [8,11]. In this study, we identified that soybean *E. faecium* strains are closely clustered together with non-clinical strains (Fig 3). Unlike clinical strains, the soybean strains may not require extra AR and VF genes to survive in the soybean environments because the fermented soybean environments may be conducive for growth and are rich in nutrients that are produced due to the activity of *Bacillus* sp. [31,32].

Genomic GC contents of the soybean strains were slightly higher than those of CL and NC strains. The higher GC contents of the soybean strains were influenced by accessory genes, which are lower in GC content than core genes (Core: 39.14%, Accessory: 38.13%, $P < 0.0001$). The soybean strains (1927.5±74.4 genes) have a lower number of accessory genes than CL and NC strains (2178.5±292.0 genes), which differentially influenced the GC contents of the entire

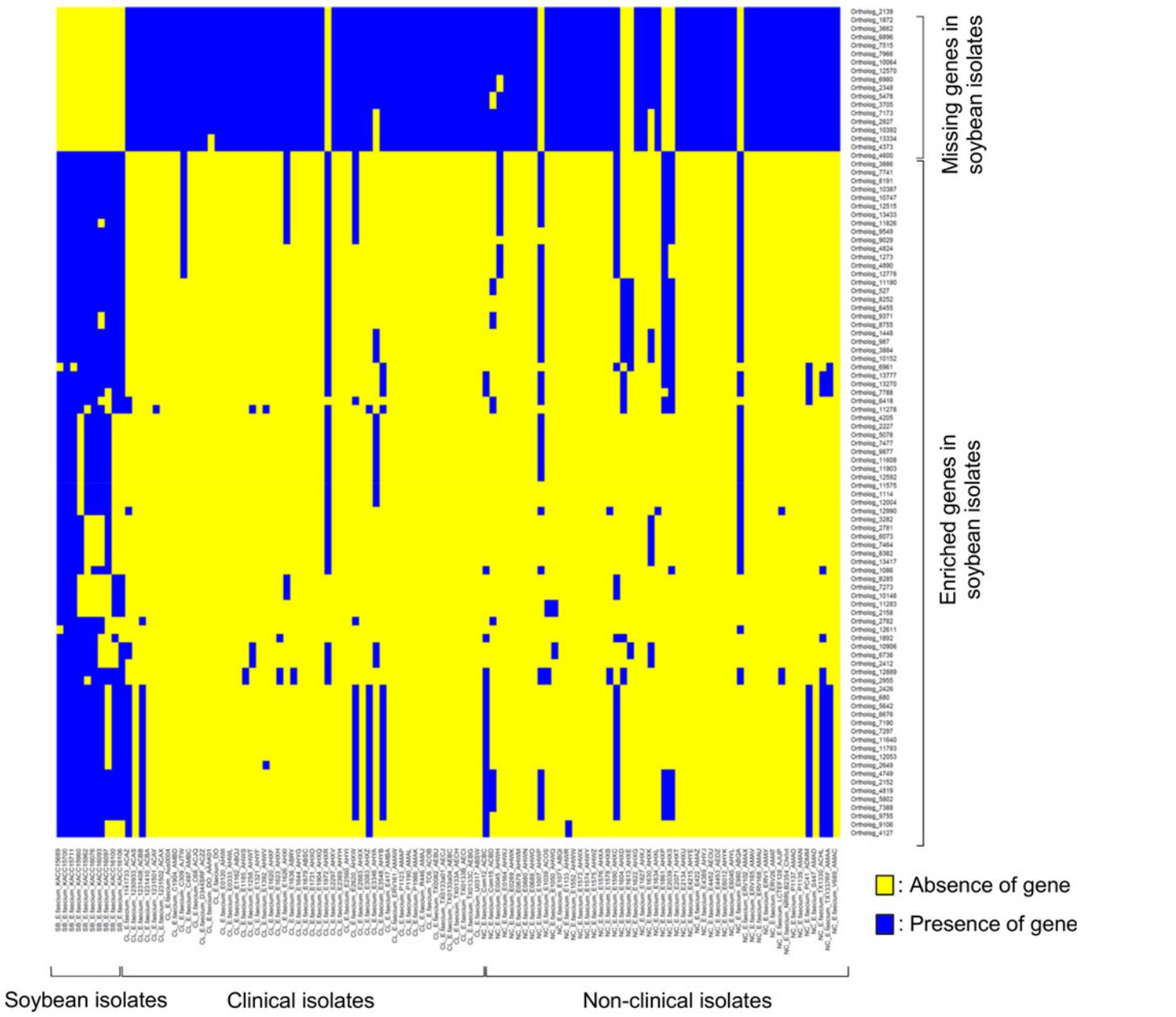


Fig 6. Niche-specific gene gain and loss in soybean *E. faecium* strains. Enriched or missing genes in soybean strains were identified and sorted according to *P* values (S2 Table). If *P* values of a gene both between soybean and clinical strains and between soybean and non-clinical strains were less than 0.005 (Fisher’s exact test), the gene was regarded as enriched or missing in the soybean strains.

doi:10.1371/journal.pone.0153279.g006

genome. That is, lower-GC accessory genes are more abundant in NC and CL strains than in soybean strains. More specifically, ME genes (37.05%) were lower in GC content than other genes (38.25%, $P < 0.001$), and there were significantly fewer such ME genes in the soybean strains (75.2 ± 15.8 genes) than in both CL and NC strains (117.4 ± 46.5 genes, $P < 0.001$). However, soybean-enriched genes may have a limited effect on the higher GC contents of the soybean strains, although they are significantly lower in GC content (34.20%) than other genes (38.23%, $P < 0.001$). For now, the reason for the slightly higher GC contents in soybean strains

is not fully understood. Other types of accessory genes or codon usage bias may play a role in such a high GC content, which must be clarified in future studies.

Recently, the safety of *E. faecium* for use in food products has been questioned, due to widespread antibiotic resistance in this species [26]. *E. faecium* used in soybean fermentation lacks virulence and antibiotic resistance genes and therefore safe for human use, as was previously shown for the dairy strain NRRL B-2354 [11].

We didn't observe any significant differences in the number of BA genes between soybean-isolated *E. faecium* and other strains from different origins (Fig 5D). These data indicate either that soybean *E. faecium* may have antimicrobial activities similar to *E. faecium* from different origins or that clinical strains might have maintained their BA genes because of unknown ecological reasons such as competitive exclusion against other bacteria in the infected patients. BA genes associated with only immunity to enterocin AS-48 [33], which is antimicrobial against *Bacillus subtilis*, a major species in fermented soybean foods [34], were identified in soybean strains (S2 Table).

Soybeans are distinctive environments that are different from other plant-derived food sources and from the animal guts because they are richer in plant proteins. We hypothesized that soybean *E. faecium* strains might have been influenced by such nutrient-rich environments. We identified niche-enriched or missing genes in soybean strains compared with other *E. faecium* strains. Genes associated with fructose and mannose metabolism were simultaneously found in both enriched and missing genes (S2 Table). The missing functions mediated by Cluster 8 (mannose/fructose metabolism) may have been complemented by other similar carbohydrate metabolic genes (Cluster 6) in the soybean environments. Gain and loss of nutrient transporting genes, which are associated with PTS systems and permeases, may have also been influenced by the distinctive nutritional environments that are formed during protein and carbohydrate degradation mediated by *B. subtilis* [35,36]. As clusters 1 and 2 have integrases, they may act as mobile genetic elements. Similarly, Cluster 4 may be transferred among *E. faecium*, *E. faecalis*, and *E. durans*, which are found in soybean-fermented foods [3]. As many niche-enriched genes encode hypothetical proteins, further studies are required to clarify their functions and roles in soybean fermentation.

In this study, we characterized the genomes of 10 soybean *E. faecium* strains to identify genomic features and contents that distinguished them from *E. faecium* from other origins. Our results suggest that soybean-isolated *E. faecium* strains are closely related to non-clinical *E. faecium* strains, have distinct genomic features from the clinical strains, and have evolved to have niche-associated genomic contents. We hope that our findings will contribute to a better understanding of the role of *E. faecium* in soybean fermentation.

Supporting Information

S1 Fig. Clusters of niche-enriched or missing genes in soybean *E. faecium* strains. Cluster sequences were blasted to NCBI nr or WGS sequence database.
(DOC)

S1 Table. Clinical and non-clinical *E. faecium* strains used in this study.
(XLS)

S2 Table. Frequency of genes in soybean (SB), clinical (CL), and non-clinical (NC) isoates.
(XLS)

Acknowledgments

This study was supported by a grant from the Strategic Initiative for Microbiomes in Agriculture and Food, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea as part of

the multi-ministerial Genome Technology to Business Translation Program (Grant ID: 914005–04). We appreciate KACC (Korean Agricultural Culture Collection) for donation of *E. faecium* strains for this study.

Author Contributions

Conceived and designed the experiments: EBK YJC. Performed the experiments: EBK GDJ. Analyzed the data: EBK GDJ JYL. Contributed reagents/materials/analysis tools: EBK GDJ. Wrote the paper: EBK GDJ JYL YJC.

References

1. Arias CA, Murray BE The rise of the *Enterococcus*: beyond vancomycin resistance. *Nat Rev Microbiol*. 2012; 10(4): 266–278. doi: [10.1038/nrmicro2761](https://doi.org/10.1038/nrmicro2761) PMID: [22421879](https://pubmed.ncbi.nlm.nih.gov/22421879/)
2. Barbosa J, Borges S, Teixeira P Selection of potential probiotic *Enterococcus faecium* isolated from Portuguese fermented food. *Int J Food Microbiol*. 2014; 191(144–148). doi: [10.1016/j.ijfoodmicro.2014.09.009](https://doi.org/10.1016/j.ijfoodmicro.2014.09.009) PMID: [25268323](https://pubmed.ncbi.nlm.nih.gov/25268323/)
3. Jeong DW, Kim HR, Jung G, Han S, Kim CT, Lee JH Bacterial community migration in the ripening of doenjang, a traditional Korean fermented soybean food. *J Microbiol Biotechnol*. 2014; 24(5): 648–660. PMID: [24548930](https://pubmed.ncbi.nlm.nih.gov/24548930/)
4. Castillo-Rojas G, Mazari-Hiriart M, Ponce de León S, Amieva-Fernández RI, Agis-Juárez RA, Huebner J, et al. Comparison of *Enterococcus faecium* and *Enterococcus faecalis* strains isolated from water and clinical samples: Antimicrobial susceptibility and genetic relationships. *PLoS One*. 2013; 8(4): e59491. doi: [10.1371/journal.pone.0059491](https://doi.org/10.1371/journal.pone.0059491) PMID: [23560050](https://pubmed.ncbi.nlm.nih.gov/23560050/)
5. Masood MI, Qadir MI, Shirazi JH, Khan IU Beneficial effects of lactic acid bacteria on human beings. *Crit Rev Microbiol*. 2011; 37(1): 91–98. doi: [10.3109/1040841X.2010.536522](https://doi.org/10.3109/1040841X.2010.536522) PMID: [21162695](https://pubmed.ncbi.nlm.nih.gov/21162695/)
6. Kim YS, Jeong DY, Hwang YT, Uhm T-B Bacterial community profiling during the manufacturing process of traditional soybean paste by pyrosequencing method. *Korean J Microbiol*. 2011; 47(275–280).
7. Willems RJL, van Schaik W Transition of *Enterococcus faecium* from commensal organism to nosocomial pathogen. *Future Microbiol*. 2009; 4(9): 1125–1135. doi: [10.2217/fmb.09.82](https://doi.org/10.2217/fmb.09.82) PMID: [19895216](https://pubmed.ncbi.nlm.nih.gov/19895216/)
8. Kim EB, Marco ML Nonclinical and clinical *Enterococcus faecium* strains, but not *Enterococcus faecalis* strains, have distinct structural and functional genomic features. *Appl Environ Microbiol*. 2014; 80(1): 154–165. doi: [10.1128/AEM.03108-13](https://doi.org/10.1128/AEM.03108-13) PMID: [24141120](https://pubmed.ncbi.nlm.nih.gov/24141120/)
9. Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, et al. Comparative genomics of enterococci: Variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. *MBio*. 2012; 3(1).
10. Lam MMC, Seemann T, Bulach DM, Gladman SL, Chen H, Haring V, et al. Comparative analysis of the first complete *Enterococcus faecium* genome. *J Bacteriol*. 2012; 194(9): 2334–2341. doi: [10.1128/JB.00259-12](https://doi.org/10.1128/JB.00259-12) PMID: [22366422](https://pubmed.ncbi.nlm.nih.gov/22366422/)
11. Kopit LM, Kim EB, Siezen RJ, Harris LJ, Marco ML Safety of the surrogate *Enterococcus faecium* NRRL B-2354 for use in thermal process validation. *Appl Environ Microbiol*. 2014; 80(6): 1899–1909. doi: [10.1128/AEM.03859-13](https://doi.org/10.1128/AEM.03859-13) PMID: [24413604](https://pubmed.ncbi.nlm.nih.gov/24413604/)
12. Boisvert S, Laviolette F, Corbeil J Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol*. 2010; 17(11): 1519–1533. doi: [10.1089/cmb.2009.0238](https://doi.org/10.1089/cmb.2009.0238) PMID: [20958248](https://pubmed.ncbi.nlm.nih.gov/20958248/)
13. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008; 9(75). doi: [10.1186/1471-2164-9-75](https://doi.org/10.1186/1471-2164-9-75) PMID: [18261238](https://pubmed.ncbi.nlm.nih.gov/18261238/)
14. Rizk G, Lavenier D GASSST: global alignment short sequence search tool. *Bioinformatics*. 2010; 26(20): 2534–2540. doi: [10.1093/bioinformatics/btq485](https://doi.org/10.1093/bioinformatics/btq485) PMID: [20739310](https://pubmed.ncbi.nlm.nih.gov/20739310/)
15. Suzuki R, Shimodaira H PvcLust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006; 22(12): 1540–1542. PMID: [16595560](https://pubmed.ncbi.nlm.nih.gov/16595560/)
16. Edgar RC MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32(5): 1792–1797. PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
17. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30(12): 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)

18. Saitou N, Nei M The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4(4): 406–425. PMID: [3447015](#)
19. Felsenstein J Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* 1985; 39(4): 783–791.
20. Tamura K, Nei M, Kumar S Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A.* 2004; 101(30): 11030–11035. PMID: [15258291](#)
21. Liu B, Pop M ARDB—Antibiotic resistance genes database. *Nucleic Acids Res.* 2009; 37(suppl 1): D443–D447.
22. van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Research.* 2013; 41(W1): W448–W453.
23. Hill C Virulence or niche factors: What's in a name? *J Bacteriol.* 2012; 194(21): 5725–5727. doi: [10.1128/JB.00980-12](#) PMID: [22821969](#)
24. Yang S-C, Lin C-H, Sung CT, Fang J-Y Antibacterial activities of bacteriocins: Application in foods and pharmaceuticals. *Front Microbiol.* 2014; 5(
25. Caryl JA, Cox G, Trimble S, O'Neill AJ “tet(U)” Is Not a Tetracycline Resistance Determinant. *Antimicrob Agents Chemother.* 2012; 56(6): 3378–3379. doi: [10.1128/AAC.05957-11](#) PMID: [22491689](#)
26. Authority EFS Guidance on the safety assessment of *Enterococcus faecium* in animal nutrition. *EFSA J.* 2012; 10(5): 10.
27. Martínez-Villaluenga C, Torino MI, Martín V, Arroyo R, García-Mora P, Estrella Pedrola I, et al. Multi-functional properties of soy milk fermented by *Enterococcus faecium* strains isolated from raw soy milk. *J Agric Food Chem.* 2012; 60(41): 10235–10244. doi: [10.1021/jf302751m](#) PMID: [22978423](#)
28. Lee Y-D, Park J-H Rapid fermentation starter *Enterococcus faecium* of Soybean for soy-sauce like product. *Korean J Food Nutr.* 2012; 25(1): 188–195.
29. de Regt MJ, van Schaik W, van Luit-Asbroek M, Dekker HA, van Duijkeren E, Koning CJ, et al. Hospital and community ampicillin-resistant *Enterococcus faecium* are evolutionarily closely linked but have diversified through niche adaptation. *PLoS One.* 2012; 7(2): e30319. doi: [10.1371/journal.pone.0030319](#) PMID: [22363425](#)
30. Lebreton F, van Schaik W, Manson McGuire A, Godfrey P, Griggs A, Mazumdar V, et al. Emergence of Epidemic Multidrug-Resistant *Enterococcus faecium* from Animal and Commensal Strains. *MBio.* 2013; 4(4): e00534–00513. doi: [10.1128/mBio.00534-13](#) PMID: [23963180](#)
31. Ochiai A, Itoh T, Kawamata A, Hashimoto W, Murata K Plant cell wall degradation by saprophytic *Bacillus subtilis* strains: gene clusters responsible for rhamnogalacturonan depolymerization. *Appl Environ Microbiol.* 2007; 73(12): 3803–3813. PMID: [17449691](#)
32. Kim M-H, Han S-Y, Ko J-M, Kim Y-S Degradation characteristics of proteins in cheonggukjang (fermented unsalted soybean paste) prepared with various soybean cultivars. *Food Sci Biotechnol.* 2012; 21(1): 9–18.
33. Diaz M, Valdivia E, Martínez-Bueno M, Fernández M, Soler-González AS, Ramírez-Rodrigo H, et al. Characterization of a New Operon, as-48EFGH, from the as-48 Gene Cluster Involved in Immunity to Enterocin AS-48. *Appl Environ Microbiol.* 2003; 69(2): 1229–1236. PMID: [12571051](#)
34. Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiol.* 2010; 10(22). doi: [10.1186/1471-2180-10-22](#) PMID: [20105292](#)
35. Jo YJ, Cho IH, Song CK, Shin HW, Kim YS Comparison of fermented soybean paste (Doenjang) prepared by different methods based on profiling of volatile compounds. *J Food Sci.* 2011; 76(3): C368–C379. doi: [10.1111/j.1750-3841.2011.02068.x](#) PMID: [21535802](#)
36. Shukla S, Choi TB, Park HK, Kim M, Lee IK, Kim JK Determination of non-volatile and volatile organic acids in Korean traditional fermented soybean paste (Doenjang). *Food Chem Toxicol.* 2010; 48(8–9): 2005–2010. doi: [10.1016/j.fct.2010.04.034](#) PMID: [20434502](#)