Identification and Monitoring of *Lactobacillus delbrueckii* Subspecies Using Pangenomic-Based Novel Genetic Markers

Eiseul Kim, Eun-Ji Cho, Seung-Min Yang, and Hae-Yeong Kim*

Institute of Life Sciences and Resources and Department of Food Science and Biotechnology, Kyung Hee University, Yongin 17104, Republic of Korea

Genetic markers currently used for the discrimination of *Lactobacillus delbrueckii* subspecies have low efficiency for identification at subspecies level. Therefore, our objective in this study was to select novel genetic markers for accurate identification and discrimination of six *L. delbrueckii* subspecies based on pangenome analysis. We evaluated *L. delbrueckii* genomes to avoid making incorrect conclusions in the process of selecting genetic markers due to mislabeled genomes. Genome analysis showed that two genomes of *L. delbrueckii* subspecies deposited at NCBI were misidentified. Based on these results, subspecies-specific genetic markers were selected by comparing the core and pangenomes. Genetic markers were confirmed to be specific for 59,196,562 genome sequences via in silico analysis. They were found in all strains of the same subspecies, but not in other subspecies or bacterial strains. These genetic markers also could be used to accurately identify genomes at the subspecies level for genomes known at the species level. A real-time PCR method for detecting three main subspecies (*L. delbrueckii* subsp. *delbrueckii*, *lactis*, and *bulgaricus*) was developed to cost-effectively identify them using genetic markers. Results showed 100% specificity for each subspecies. These genetic markers could differentiate each subspecies from 44 other lactic acid bacteria. This real-time PCR method was then applied to monitor 26 probiotics and dairy products. It was also used to identify 64 unknown strains isolated from raw milk samples and dairy products. Results confirmed that unknown isolates and subspecies contained in the product could be accurately identified using this real-time PCR method.

Keywords: *Lactobacillus delbrueckii* subspecies, pangenome, genetic marker, identification, real-time PCR, probiotic product

Introduction

Lactobacillus delbrueckii comprises six subspecies, namely *delbrueckii*, *lactis*, *bulgaricus*, *indicus*, *jakobsenii*, and *sunkii*, all of which have historically been differentiated based on their ability to metabolize different carbohydrates [1]. Among these subspecies, *L. delbrueckii* subsp. *lactis* and *bulgaricus* are usually associated with the manufacture of dairy products such as cheeses and yogurt [2]. *L. delbrueckii* subsp. *bulgaricus* is one of the starter culture components for the production of yogurt [1, 3]. This subspecies displays probiotic properties [4]. On the other hand, *L. delbrueckii* subsp. *lactis* is traditionally used in cheese production and can be distinguished from *L. delbrueckii* subsp. *bulgaricus* by its extensive carbohydrate metabolizing capabilities [1, 5]. *L. delbrueckii* subsp. *delbrueckii* cannot ferment lactose. It is typically associated with fermented vegetables [2]. *L. delbrueckii* subsp. *indicus*, *jakobsenii*, and *sunkii* are relatively minor subspecies isolated from Indian dairy products, fermented alcoholic beverages, and non-salted pickle as a traditional Japanese food, respectively [6–9].

Accurate identification of *L. delbrueckii* subspecies in food samples is an important issue to confirm probiotic properties and perform product quality assessment [4]. Genetic markers and molecular-based methods have been used to efficiently identify and detect lactic acid bacteria commonly used in commercial probiotic and dairy products. Molecular-based methods for the identification and typing of lactic acid bacteria have been reported, including amplified fragment length polymorphism (AFLP), DNA–DNA hybridization (DDH), multi-locus sequence analysis (MLST), and restriction fragment length polymorphism (RFLP) [6, 10–12]. However, these techniques are labor-intensive, expensive, and time-consuming with low reproducibility whereas PCR-based methods are rapid, sensitive, and reliable for identifying lactic acid bacteria [4]. Of these methods, genetic markers such as the 16S rRNA gene and 16S–23S rRNA intergenic spacer region have been used to distinguish *L. delbrueckii* used in PCR-based methods [13]. Although genetic markers described above are useful for identifying *L. delbrueckii* at the species level, they cannot be applied to distinguish *L. delbrueckii* at the subspecies level [4].

Received: September 21, 2020
Accepted: October 19, 2020

First published online:
October 22, 2020

*Corresponding author
Phone: +82-31-201-2123
Fax: +82-31-204-8116
E-mail: hykim@khu.ac.kr

pISSN 1017-7825
eISSN 1738-8872

Copyright© 2021 by
The Korean Society for
Microbiology and
Biotechnology

Recently, the development of whole-genome sequencing (WGS) and the increase in genome sequences have made it possible to rapidly and freely process large-scale sequence data on microorganisms in public repositories [14]. Pangenome analysis based on WGS has a wide range of applications, including prediction of lifestyles of microorganisms, pathogenicities, resistome, and taxonomy [15]. Pangenome analysis also allows reclassification of bacterial species and/or subspecies, improving and clarifying criteria previously presented [16]. In the present study, we selected six *L. delbrueckii* subspecies-specific genetic markers by pangenome analysis to develop a real-time PCR method for rapid identification of bacterial strains. The real-time PCR method we developed was then applied to bacterial strains isolated from raw milk, probiotic products, and dairy products to identify and differentiate three *L. delbrueckii* subspecies.

Materials and Methods

Pangenome Analysis and Selection of Genetic Markers

The *in silico* scheme for selecting the genetic markers of six *L. delbrueckii* subspecies is illustrated in Fig. 1. A total of 41 genomes belonging to the subspecies *L. delbrueckii* subsp. *delbrueckii*, *lactis*, *bulgaricus*, *indicus*, *jakobsenii*, and *sunkii* were obtained from the National Center for Biotechnology Information (NCBI) (Table 1). Phylogenetic analysis based on the pangenome was performed using microbial pangenomics in Anvi'o v6 software [17]. According to the developer's recommendation, a genome database for pangenome analysis was constructed using Anvi'o genome storage. The pangenome was then analyzed using the NCBI BLASTp and MCL algorithm. Subsequently, a phylogenetic tree was constructed based on pangenome cluster frequencies.

The pangenome of *L. delbrueckii* subspecies was calculated using Bacterial Pan Genome Analysis (BPGA) pipeline ver. 1.3 (identity cut off = 50%) [18]. The pangenome was formatted into two local databases: a pangenome database and a core-genome database for each subspecies. Candidate genetic markers were selected by comparing the pangenome database composed of protein-coding genes, present in all genomes except for the target subspecies, and the core-genome database composed of protein-coding genes present in all genomes of target subspecies. Candidate genetic markers were then aligned with 59,196,562 sequences using BLASTN. Genetic markers only present in target subspecies but not present in other bacterial genomes were selected.

In Silico Specificity Confirmation and Development of Subspecies-Specific Primer

A total of four genome sequences registered at the species level were obtained from the NCBI. Genetic marker specificity was aligned with 45 genome sequences, including genomes known to the species level using USEARCH ver. 9.0 [19]. Alignment results are presented as a heatmap using Seaborn python library in Matplotlib. Subspecies-specific primer pairs for *L. delbrueckii* subsp. *lactis*, *bulgaricus*, and *delbrueckii* were designed from their genetic markers using Primer Designer (Scientific and Education Software, USA).

DNA Extractions from *L. delbrueckii* Subspecies and Lactic Acid Bacteria

For specificity testing of primer pairs developed in this study, the reference strains of lactic acid bacteria mainly isolated from probiotic and dairy products were used. A total of 54 strains of lactic acid bacteria including *L. delbrueckii* subspecies were obtained from the Korean Agricultural Culture Collection (KACC, Korea), the Korean Collection for Type Cultures (KCTC, Korea), the Korean Culture Center of Microorganisms (KCCM, Korea), the NITE Biological Resource Center (NBRC, Japan), and the Laboratory Isolates (LI, Korea) (Table 2). All reference strains were grown in MRS broth (Difco, Becton & Dickinson, USA) for extraction of genomic DNA. *L. delbrueckii* and other bacterial strains were cultured for 48 h at 42°C and 37°C under anaerobic condition, respectively. Bacterial cells were centrifuged at 13,600 ×g for 5 min and the supernatant was removed. Genomic DNA of reference strains was extracted using a DNeasy Blood & Tissue Kit (Qiagen, Germany) following the protocol described previously [13, 20]. DNA concentration and purity were confirmed using a MaestroNano spectrophotometer (MaestroGen, USA).

Specificity and Accuracy of Specific Primer Pairs

Real-time PCR assay was conducted to determine the specificity and accuracy of primer pairs using a 7500 Real-time PCR system. Each reaction contained 20 ng of genomic DNA, 10 µl of 2X Thunderbird SYBR qPCR Mix (Toyobo, Japan), 500 nM of primer pairs, and distilled water up to 20 µl total volume. Real-time PCR conditions consisted of initiation at 95°C for 2 min followed by 30 amplification cycles of 95°C for 5 s and 60°C for 30 s. Melting curves were obtained at 95°C for 15 s, 60°C for 1 min, 95°C for 30 s, and 60°C for 15 s. Specificity of primer

Fig. 1. The *in silico* approach for genetic marker selection of six *L. delbrueckii* subspecies.

Table 1. Summary in genome features of 41 *L. delbrueckii* subspecies.

Organism name	Strain	Size (Mb)	GC%	CDS	Assembly	Accession No.
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	ATCC BAA-365	1.85695	49.7	1579	Complete	CP000412.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	ATCC 11842	1.865	49.7	1561	Complete	CR954253.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	2038	1.87292	49.7	1562	Complete	CP000156.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	CNCM I-1519	1.79654	49.9	1630	Contig	AGHW01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	INRA-MIG	1.85324	49.8	1692	Scaffold	CCET01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	DSM 20081	1.75853	49.9	1533	Scaffold	JQAV01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	MN-BM-F01	1.87507	49.7	1585	Complete	CP013610.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	CFL1	1.75792	49.8	1539	Contig	CZPS01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	LBB.B5	1.77788	49.8	1558	Contig	LUGK01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	DSM 20080	1.86818	49.8	1564	Complete	CP019120.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	ND04	1.86175	49.6	1538	Complete	CP016393.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	MBT 92059	1.83117	49.8	1648	Scaffold	QOV01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	L99	1.84811	49.7	1575	Complete	CP017235.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	KLDS1.0207	1.86918	49.8	1620	Complete	CP032451.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	FAM 21277	2.01984	49.2	1830	Contig	VBSR01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	NBRC 13953	1.72582	50.0	1519	Contig	BJMY01
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	KLDS1.1011	1.88749	49.8	1634	Complete	CP041280.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	LJJ	1.89109	49.5	1604	Complete	CP049052.1
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	ACA-DC 87	1.856	49.8	1579	Complete	LT899687.1
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	KACC 13439	1.76619	50.0	1485	Contig	LHPL01
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	KCTC 13731	1.91051	50.0	1600	Complete	CP018216.1
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	DSM 20074	1.95372	49.6	1577	Complete	CP018615.1
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	TUA4408L	2.01244	49.9	1718	Complete	CP021136.1
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	NBRC 3534	1.78742	50.3	1588	Contig	BJLM01
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	NBRC 3202	1.91031	50.1	1653	Complete	AP019750.1
<i>L. delbrueckii</i> subsp. <i>indicus</i>	JCM 15610	1.87741	49.5	1627	Contig	LGAS01
<i>L. delbrueckii</i> subsp. <i>indicus</i>	DSM 15996	1.86357	49.6	1621	Scaffold	AZFL01
<i>L. delbrueckii</i> subsp. <i>indicus</i>	JCM 15610	2.02186	49.4	1694	Complete	CP018614.1
<i>L. delbrueckii</i> subsp. <i>jakobsenii</i>	ZN7a-9	1.73081	50.2	1552	Contig	ALPY01
<i>L. delbrueckii</i> subsp. <i>jakobsenii</i>	DSM 26046	1.74924	50.3	1568	Scaffold	JQCG01
<i>L. delbrueckii</i> subsp. <i>jakobsenii</i>	DSM 26046	1.8918	50.1	1614	Complete	CP018218.1
<i>L. delbrueckii</i> subsp. <i>jakobsenii</i>	DSM 26046	1.78119	50.1	1585	Scaffold	PUGF01
<i>L. delbrueckii</i> subsp. <i>lactis</i>	CRL581	2.13682	49.6	1639	Scaffold	ATBQ01
<i>L. delbrueckii</i> subsp. <i>lactis</i>	KCCM 34717	2.26338	49.1	1905	Complete	CP018215.1
<i>L. delbrueckii</i> subsp. <i>lactis</i>	DSM 20072	2.16598	49.0	1793	Complete	CP022988.1
<i>L. delbrueckii</i> subsp. <i>lactis</i>	KCTC 3034	2.23761	49.0	1889	Complete	CP023139.1
<i>L. delbrueckii</i> subsp. <i>lactis</i>	NBRC 3734	1.81291	50.2	1653	Contig	BJLO01
<i>L. delbrueckii</i> subsp. <i>lactis</i>	lactis1	2.05032	49.6	1694	Complete	LS991409.1
<i>L. delbrueckii</i> subsp. <i>lactis</i>	NWC_2_2	2.269179	48.7	1934	Complete	CP031023.1
<i>L. delbrueckii</i> subsp. <i>sunkii</i>	JCM 17838	1.94526	50.1	1713	Contig	LGHR01
<i>L. delbrueckii</i> subsp. <i>sunkii</i>	JCM 17838	2.00434	50.1	1726	Complete	CP018217.1

pairs was tested against a total of 10 strains of *L. delbrueckii* subspecies and 44 other lactic acid bacteria. For the accuracy test, genomic DNA from each reference strain was serially diluted and used for real-time PCR.

Application of Real-Time PCR Method

To test the developed real-time PCR method, 64 isolates, 15 probiotic products, and 11 dairy products were used. *L. delbrueckii* subspecies were isolated from three raw milk samples and three dairy products. Serially diluted samples were spread onto MRS agar plates (Difco, Becton & Dickinson, USA) and incubated at 42°C for 48 h under anaerobic conditions. Probiotic and dairy products were obtained from markets around the world. Genomic DNAs were extracted from isolates and products under the same conditions as described in section 2.3. DNA Extraction of *L. delbrueckii* Subspecies and Lactic Acid Bacteria. For the application test, genomic DNAs of isolates or products were added to wells of 96-well plates containing 2X qPCR mix (Toyobo) and subspecies-specific primer pairs. The real-time PCR condition was the same as that described in section 2.4. Specificity and Accuracy for Specific Primer Pairs.

Results and Discussion

Pangenome Analysis

Many studies have previously reported a mislabeling issue regarding subspecies or closely related species in the NCBI genome database [20, 21]. In these studies, the majority of the mislabeled genomes were closely related species [20, 22, 23]. Such genomes should therefore be evaluated to avoid reaching incorrect conclusions in a comparative genomic analysis. In the present study, for the first time, we evaluated the genomes of *L. delbrueckii*

Table 2. List of reference strains used in this study.

Species	Strain no.
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	KACC ^a 12420
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	LI ^b 00010
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	LI 00011
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	LI 00012
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	LI 00013
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	LI 00014
<i>Lactobacillus delbrueckii</i> subsp. <i>lactis</i>	KACC 12417
<i>Lactobacillus delbrueckii</i> subsp. <i>lactis</i>	LI 00015
<i>Lactobacillus delbrueckii</i> subsp. <i>delbrueckii</i>	KACC 13439
<i>Lactobacillus delbrueckii</i> subsp. <i>delbrueckii</i>	KCTC 13730
<i>Lactobacillus acidipiscis</i>	KACC 12394
<i>Lactobacillus acidophilus</i>	KACC 12419
<i>Lactobacillus agilis</i>	KACC 12433
<i>Lactobacillus amylolyticus</i>	KACC 12374
<i>Lactobacillus amylophilus</i>	KACC 11430
<i>Lactobacillus amylovorus</i>	KACC 12435
<i>Lactobacillus brevis</i>	KCTC ^c 3498
<i>Lactobacillus buchneri</i>	KACC 12416
<i>Lactobacillus casei</i>	KACC 12413
<i>Lactobacillus chiayiensis</i>	NBRC ^d 112906
<i>Lactobacillus coryniformis</i>	KACC 12411
<i>Lactobacillus crustorum</i>	KACC 16344
<i>Lactobacillus curvatus</i>	KACC 12415
<i>Lactobacillus farciminis</i>	KACC 12423
<i>Lactobacillus fermentum</i>	KACC 11441
<i>Lactobacillus gallinarum</i>	KACC 12370
<i>Lactobacillus gasseri</i>	KCTC 3163
<i>Lactobacillus heilongjiangensis</i>	KACC 18741
<i>Lactobacillus helveticus</i>	KACC 12418
<i>Lactobacillus jensenii</i>	KCTC 5194
<i>Lactobacillus johnsonii</i>	KCTC 3801
<i>Lactobacillus kunkeei</i>	KACC 19371
<i>Lactobacillus lindneri</i>	KACC 12445
<i>Lactobacillus mucosae</i>	KACC 12381
<i>Lactobacillus parabuchneri</i>	KACC 12363
<i>Lactobacillus paracasei</i>	KCTC 3165
<i>Lactobacillus paraplantarum</i>	KACC 12373
<i>Lactobacillus paraplantarum</i>	KCTC 5045
<i>Lactobacillus pentosus</i>	KACC 12428
<i>Lactobacillus pentosus</i>	KCCM ^e 40997
<i>Lactobacillus plantarum</i> subsp. <i>argenteratensis</i>	KACC 12404
<i>Lactobacillus plantarum</i> subsp. <i>plantarum</i>	KACC 11451
<i>Lactobacillus reuteri</i>	KCTC 3594
<i>Lactobacillus rhamnosus</i>	KCTC 3237
<i>Lactobacillus ruminis</i>	KACC 12429
<i>Lactobacillus sakei</i>	KCTC 3603
<i>Lactobacillus salivarius</i>	KCTC 3600
<i>Lactobacillus sanfranciscensis</i>	KACC 12431
<i>Lactobacillus zymae</i>	KACC 16349
<i>Bifidobacterium animalis</i> subsp. <i>lactis</i>	KACC 16638
<i>Bifidobacterium bifidum</i>	KCTC 3418
<i>Bifidobacterium breve</i>	KACC 16639
<i>Bifidobacterium longum</i> subsp. <i>infantis</i>	KCTC 3249
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	KCCM 11953

^aKACC, the Korean Agricultural Culture Collection^bLI, the Laboratory Isolate^cKCTC, the Korean Collection for Type Cultures^dNBRC, the NITE Biological Resource Center^eKCCM, the Korean Culture Center of Microorganisms

subspecies by phylogenetic analysis based on the pangenome before specific genetic markers were selected. Phylogenetic analysis results based on pangenome frequencies were displayed along with the distribution of subspecies' specific regions. Each bar represents *L. delbrueckii* subspecies genomes and each layer presents

Fig. 2. Pangenome distribution of the 41 *L. delbrueckii* subspecies genomes. The color bar of black, yellow, red, blue, green, and green represents *L. delbrueckii* subsp. *bulgaricus*, *jakobsenii*, *delbrueckii*, *sunkii*, *indicus*, and *lactis* genomes, respectively. The dark color and tinted bright of the bar indicate core-genome presence and absence, respectively. The phylogenetic tree on the right is based on gene cluster frequencies.

pangenome distribution (Fig. 2). Most genomes clustered according to the subspecies. However, some genomes of *L. delbrueckii* subsp. *bulgaricus* and *delbrueckii* clustered with different subspecies. Genomes of *L. delbrueckii* subsp. *bulgaricus* FAM 21277 and *delbrueckii* TUA4408L clustered with *L. delbrueckii* subsp. *lactis* and *sunkii*, respectively. Based on these results, *L. delbrueckii* subsp. *bulgaricus* FAM 21277 and *delbrueckii* TUA4408L were determined as *L. delbrueckii* subsp. *lactis* and *sunkii*, respectively. These genomes should be indicated correctly in the genome database to avoid further misidentification. We also suggest implementing measures to prevent and correct taxonomic errors in the NCBI database to avoid confusion in future research.

Closely related strains in phylogenetic analysis can be distinguished using efficient and customized mining methods for genome sequences [20, 24, 25]. Conventional methods can be used to successfully distinguish pathogenic bacteria that are difficult to differentiate, although these methods only focus on pathogenic bacteria. Studies on lactic acid bacteria are still lacking. Here, we employed a pangenome approach to identify novel genetic markers for specific identification and detection of *L. delbrueckii* subspecies.

As a result of pangenome analysis, a total of 67,178 genes from 41 *L. delbrueckii* subspecies yielded a pangenome size of 3,456 genes. The core-genome, accessory-genome, and unique-genome had 749, 2,071, and 636 genes, respectively. Six subspecies-specific genetic markers were then obtained by pangenome analysis. Genetic markers were found to be protein-coding genes present in the same subspecies but absent in other subspecies or bacterial strains. By comparing genomes of the same subspecies, 995 to 1,628 protein-coding genes were found in common in the genomes of each subspecies and considered as the core-genome for each subspecies. After comparing each core-genome with pangenome for protein-coding genes present in all genomes except for target genomes, 5 to 50 protein-coding genes were selected as candidate genetic markers for each subspecies. These candidate genetic markers were aligned with 59,196,562 genome sequences. Genes not present in other bacterial strains except target subspecies were finally selected as genetic markers. Selected genetic markers of *L. delbrueckii* subsp.

Fig. 3. Heatmap shows the presence/absence of genetic markers in 45 genomes. The heatmap shows that the similarity of genetic markers present in each genome is visualized with a color bar of blue (high identity) to sky blue (low identity). The bottom of the heatmap presents the six subspecies-specific genetic markers; the right of the heatmap presents the 45 genomes of *L. delbrueckii* species or subspecies.

bulgaricus, *lactis*, *delbrueckii*, *indicus*, *jakobsenii*, and *sunkii* were identified as YcaO-like family protein (Accession No. ABJ57813.1), Ser/Thr protein kinase (Accession No. EGD27260.1), choline kinase (Accession No. KNZ37552.1), DNA methyltransferase (Accession No. KNE31255.1), RpoD family RNA polymerase sigma factor (Accession No. EOD03403.1), and hypothetical protein (Accession No. APG74821.1), respectively.

Genetic Marker Specificity Test

The specificity of genetic markers was tested using 45 genomes including genomes registered at species level by *in silico* analysis. The heatmap for identities of genetic markers in genomes is shown with color codes, ranging from blue (region with high identity) to sky blue (region with low identity) (Fig. 3). Each genetic marker shared more than 95% sequence identity with genomes of most corresponding subspecies. In contrast, a genetic marker for *L. delbrueckii* subsp. *bulgaricus* was present in 19 *bulgaricus* genomes (95–100% identity), but one genome had the genetic marker for *L. delbrueckii* subsp. *lactis* instead of *bulgaricus* (99% identity). A genetic marker for *L. delbrueckii* subsp. *delbrueckii* was present in six *delbrueckii* genomes (99–100% identity), but one genome had the genetic marker for *L. delbrueckii* subsp. *sunkii* instead of *delbrueckii* (100% identity). These results were the same as those of pangenome analysis. Genetic markers were aligned with their genomes to determine the subspecies of genomes registered at the species level. *L. delbrueckii* AVK, TJA31, and 328M contained the genetic marker for *L. delbrueckii* subsp. *bulgaricus* (96–97% sequence identity). *L. delbrueckii* LDEL18P1 contained the genetic marker for *L. delbrueckii* subsp. *lactis* (100% sequence identity).

In previous studies, genes such as 16S rRNA, 16S–23S rRNA intergenic spacer region, and the elongation factor Tu (*tuf*) gene have been used to distinguish microorganisms at the species or subspecies level [4, 9, 26–29].

Table 3. Primer pairs designed in this study.

Target species	Primer name	Sequence (5'-3')	Size (bp)	Target gene	Accession No.
<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	Bulgaricus_F	TAC CGC TGT TCT GTC TCA AGG	102	YcaO-like family protein	ABJ57813.1
	Bulgaricus_R	TAT GCC TCC GTG AGC GAT CT			
<i>L. delbrueckii</i> subsp. <i>lactis</i>	Lactis_F	TTG TGC AAG AGC CAG CTG AA	113	Ser/Thr protein kinase	EGD27206.1
	Lactis_R	GCC GCC ATT ACT GAA GTG GA			
<i>L. delbrueckii</i> subsp. <i>delbrueckii</i>	Delbrueckii_F	CAT GGA ACT TCT GCG AAG GT	110	Choline kinase	KNZ37552.1
	Delbrueckii_R	TAG ATC CGG AGC TGT TCC AC			

However, some studies have reported that these genes share high sequence similarities without showing sufficient variabilities to allow for the differentiation between *L. delbrueckii* subspecies [4, 13]. In contrast, we selected genetic markers specific to the genomes of each subspecies using pangenome analysis. The markers selected in this study were specific to *L. delbrueckii* subspecies and other bacterial strains. They were able to accurately identify subspecies level for unknown genomes.

Specificity and Accuracy for Specific Primer Pairs

The method to identify *L. delbrueckii* subspecies with genetic markers selected in this study requires WGS and bioinformatics analysis to confirm the presence of their markers. This method can accurately identify *L. delbrueckii* subspecies. However, the cost associated with WGS and its informational capacities must be considered. In addition, specialized researchers are needed to handle bioinformatics analysis [14, 30]. Therefore, we developed a real-time PCR method to cost-effectively identify many *L. delbrueckii* isolates using relatively simple procedures. This real-time PCR method is designed to identify three main subspecies, *L. delbrueckii* subsp. *delbrueckii*, *lactis*, and *bulgaricus* [2] that are mainly isolated from food or used for fermenting dairy products. Subspecies-specific primer pairs were designed from selected genetic markers. Information for primer pairs is shown in Table 3. The specificity test for these designed subspecies-specific primer pairs was performed using 54

Fig. 4. The specificity of subspecies-specific primer pairs against 54 lactic acid bacteria. (A) Specificity of *L. delbrueckii* subsp. *bulgaricus* primer pair, amplification curve: *L. delbrueckii* subsp. *bulgaricus* KACC 12420, LI 00010, LI 00011, LI 00012, LI 00013, and LI 00014; (B) Specificity of *L. delbrueckii* subsp. *lactis* primer pair, amplification curve: *L. delbrueckii* subsp. *lactis* KACC 12417 and LI 00015; (C) Specificity of *L. delbrueckii* subsp. *delbrueckii* primer pair, amplification curve: *L. delbrueckii* subsp. *delbrueckii* KACC 13439 and KCTC 13730.

Fig. 5. Real-time PCR amplification plot, melt curve, and standard curve. (A) *L. delbrueckii* subsp. *bulgaricus* amplification plot (left), melt curve (middle), and standard curve (right); (B) *L. delbrueckii* subsp. *lactis* amplification plot (left), melt curve (middle), and standard curve (right); (C) *L. delbrueckii* subsp. *delbrueckii* amplification plot (left), melt curve (middle), and standard curve (right).

reference strains of lactic acid bacteria. The genomic DNA of each subspecies generated a positive signal for corresponding primer pairs, whereas genomic DNAs from other *L. delbrueckii* subspecies and lactic acid bacteria did not generate any signal (Fig. 4). The Ct value ranged from 12.72 to 16.94 for each subspecies-specific primer pair. Genomic DNAs of three subspecies were used to confirm the accuracy of primer pairs. Standard curves were generated using serial diluted genomic DNA at an amount ranging from 0.002 ng to 20 ng. Slopes for standard curves of *L. delbrueckii* subsp. *bulgaricus*, *lactis*, and *delbrueckii* were -3.44 , -3.46 , and -3.34 , respectively. All correlation coefficient values (R^2) were greater than 0.998 and all amplification efficiencies were more than 94% (Fig. 5). All of these values met real-time PCR conditions indicating a high efficiency [31]. Thus, the method we developed in the present study shows high accuracies. Our real-time PCR method targeting specific genetic markers enables rapid and accurate identification of three *L. delbrueckii* subspecies.

Table 4. Application test of real-time PCR method to probiotic and dairy products.

Products	Type	Label claims	Detected subspecies
Products monitoring			
A1	Probiotic product (powder, Korea)	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A2	Probiotic product (powder, Korea)	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A3	Probiotic product (capsules, Canada)	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A4	Probiotic product (capsules, Canada)	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A5	Probiotic product (powder, Korea)	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A6	Probiotic product (capsules, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A7	Probiotic product (capsules, USA)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A8	Probiotic product (capsules, USA)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A9	Probiotic product (powder, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A10	Probiotic product (powder, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A11	Probiotic product (capsules, Canada)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A12	Probiotic product (powder, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A13	Probiotic product (capsules, Canada)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A14	Probiotic product (powder, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
A15	Probiotic product (powder, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B1	Dairy product (yogurt, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B2	Dairy product (yogurt, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B3	Dairy product (yogurt, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B4	Dairy product (yogurt, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B5	Dairy product (yogurt, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B6	Dairy product (yogurt, Korea)	<i>L. bulgaricus</i>	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B7	Dairy product (yogurt, Korea)	Lactic acid bacteria	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B8	Dairy product (yogurt, Korea)	Lactic acid bacteria	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B9	Dairy product (yogurt, Korea)	Lactic acid bacteria	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B10	Dairy product (yogurt, Korea)	Lactic acid bacteria	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
B11	Dairy product (yogurt, Korea)	Lactic acid bacteria	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
Identification of isolates			
I1~I4	Raw milk (cow's milk, Korea)	Unknown isolates	<i>L. delbrueckii</i> subsp. <i>lactis</i>
I5~I8	Raw milk (cow's milk, Korea)	Unknown isolates	<i>L. delbrueckii</i> subsp. <i>lactis</i>
I9~I17	Raw milk (cow's milk, Korea)	Unknown isolates	<i>L. delbrueckii</i> subsp. <i>lactis</i>
I18~I21	Dairy product (powder, Korea)	Unknown isolates	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
I22~I57	Dairy product (yogurt, Korea)	Unknown isolates	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>
I58~I64	Dairy product (yogurt, Korea)	Unknown isolates	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i>

Application of Real-Time PCR Method

Fifty-one isolates, 15 probiotic products, and 11 dairy products were used to perform the application test of the developed real-time PCR method. Results of its application to probiotic and dairy products were compared with their label claims. A total of 26 products were detected with the same subspecies as their label claims (Table 4). However, for 16 products, subspecies was incorrectly claimed on the label. According to the nomenclature of subspecies, these should be labeled as “*L. delbrueckii* subsp. *bulgaricus*,” not “*L. bulgaricus*” [12]. Dairy products labeled only with lactic acid bacteria were confirmed to contain *L. delbrueckii* subsp. *bulgaricus* by real-time PCR. As a result of the application of our method to different isolates, a total of 64 strains isolated from raw milk and dairy products were identified as *L. delbrueckii* subsp. *lactis* ($n = 17$) and *bulgaricus* ($n = 47$) (Table 4). These results confirmed that the real-time PCR developed in this study could accurately identify strains present in probiotic and dairy products and bacterial isolates to the subspecies level.

Conclusion

In conclusion, pangenome analysis was performed to select genetic markers for six *L. delbrueckii* subspecies. These genetic markers were present in all genomes of the same subspecies but absent in genomes of other subspecies and bacterial strains. To rapidly and cost-effectively identify *L. delbrueckii* subspecies, subspecies-specific primer pairs for three subspecies mainly isolated from food samples were designed. The real-time PCR method using these genes could accurately identify *L. delbrueckii* subspecies and other lactic acid bacteria with high specificity. The developed real-time PCR method was able to successfully monitor probiotic and dairy products and identify various isolates. Thus, our method can be used to accurately identify *L. delbrueckii* subspecies and determine the nomenclature of these subspecies. Furthermore, it can contribute to safety in the food industry by ensuring products are labeled to show the correct strain.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant no. 2020R1A6A3A01100168).

References

1. El Kafsi H, Binesse J, Loux V, Buratti J, Boudebouze S, Dervyn R, et al. 2014. *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: A chronicle of evolution in action. *BMC. Genomics* **15**: 407.
2. Zanni E, Schifano E, Motta S, Sciubba F, Palleschi C, Mauri P, et al. 2017. Combination of metabolomic and proteomic analysis revealed different features among *Lactobacillus delbrueckii* subspecies *bulgaricus* and *lactis* strains while in vivo testing in the model organism *Caenorhabditis elegans* highlighted probiotic properties. *Front. Microbiol.* **8**: 1206.
3. van De Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, et al. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl. Acad. Sci. USA* **103**: 9274–9279.
4. Nikolau A, Saxami G, Kourkoutas Y, Galanis A. 2011. A new methodology for rapid detection of *Lactobacillus delbrueckii* subsp. *bulgaricus* based on multiplex PCR. *J. Microbiol. Methods* **84**: 362–364.
5. Weiss N, Schillinger U, Kandler O. 1983. *Lactobacillus lactis*, *Lactobacillus leichmannii* and *Lactobacillus bulgaricus*, subjective synonyms of *Lactobacillus delbrueckii*, and description of *Lactobacillus delbrueckii* subsp. *lactis* comb. nov. and *Lactobacillus delbrueckii* subsp. *bulgaricus* comb. nov. *Syst. Appl. Microbiol.* **4**: 552–557.
6. Dellaglio F, Felis GE, Castioni A, Torriani S, Germond JE. 2005. *Lactobacillus delbrueckii* subsp. *indicus* subsp. nov., isolated from Indian dairy products. *Int. J. Syst. Evol. Microbiol.* **55**: 401–404.
7. Kudo Y, Oki K, Watanabe K. 2012. *Lactobacillus delbrueckii* subsp. *sunkii* subsp. nov., isolated from sunki, a traditional Japanese pickle. *Int. J. Syst. Evol. Microbiol.* **62**: 2643–2649.
8. Adimpong DB, Nielsen DS, Sørensen KI, Vogensen FK, Sawadogo-Lingani H, Derx PMF, et al. 2013. *Lactobacillus delbrueckii* subsp. *jakobsenii* subsp. nov., isolated from dolo wort, an alcoholic fermented beverage in Burkina Faso. *Int. J. Syst. Evol. Microbiol.* **63**: 3720–3726.
9. Song Y, Sun Z, Guo C, Wu Y, Liu W, Yu J, et al. 2016. Genetic diversity and population structure of *Lactobacillus delbrueckii* subspecies *bulgaricus* isolated from naturally fermented dairy foods. *Sci. Rep.* **6**: 22704.
10. Giraffa G, Lazzi C, Gatti M, Rossetti L, Mora D, Neviani E. 2003. Molecular typing of *Lactobacillus delbrueckii* of dairy origin by PCR-RFLP of protein-coding genes. *Int. J. Food Microbiol.* **82**: 163–172.
11. Cebeci A, Gürakan GC. 2011. Comparative typing of *L. delbrueckii* subsp. *bulgaricus* strains using multilocus sequence typing and RAPD-PCR. *Eur. Food Res. Technol.* **233**: 377–385.
12. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, et al. 2020. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int. J. Syst. Evol. Microbiol.* **70**: 2782–2858.
13. Kim E, Yang SM, Lim B, Park SH, Rackerby B, Kim HY. 2020. Design of PCR assays to specifically detect and identify 37 *Lactobacillus* species in a single 96 well plate. *BMC Microbiol.* **20**: 96.
14. Ibrahim GM, Morin PM. 2018. *Salmonella* serotyping using whole genome sequencing. *Front. Microbiol.* **9**: 2993.
15. Caputo A, Fournier PE, Raoult D. 2019. Genome and pan-genome analysis to classify emerging bacteria. *Biol. Direct.* **14**: 5.
16. Caputo A, Merhej V, Georgiades K, Fournier PE, Croce O, Robert C, et al. 2015. Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: the *Klebsiella* paradigm. *Biol. Direct.* **10**: 55.
17. Eren AM, Esen Ö, Quince C, Vineis J, Sogin M, Delmont T. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**: e1319.
18. Chaudhari NM, Gupta VK, Dutta C. 2016. BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* **6**: 24373.
19. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
20. Kim E, Yang SM, Cho EJ, Kim HY. 2020. Novel real-time PCR assay for *Lactobacillus casei* group species using comparative genomics. *Food Microbiol.* **90**: 103485.
21. Wuyts S, Wittouck S, De Boeck I, Allonsius CN, Pasolli E, Segata N, et al. 2017. Large-scale phylogenomics of the *Lactobacillus casei* group highlights taxonomic inconsistencies and reveals novel clade-associated features. *mSystems* **2**: e00061–17.

22. Beaz-Hidalgo R, Hossain MJ, Liles MR, Figueras MJ. 2015. Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for *Aeromonas* genomes in the genbank database. *PLoS One* **10**: e0115813.
23. Tran PN, Savka MA, Gan HM. 2017. *In-silico* taxonomic classification of 373 genomes reveals species misidentification and new genospecies within the genus *Pseudomonas*. *Front. Microbiol.* **8**: 1296.
24. Laing CR, Whiteside MD, Gannon VPJ. 2017. Pan-genome analyses of the species *Salmonella enterica*, and identification of genomic markers predictive for species, subspecies, and serovar. *Front. Microbiol.* **8**: 1345.
25. Velsko IM, Perez MS, Richards VP. 2019. Resolving phylogenetic relationships for *Streptococcus mitis* and *Streptococcus oralis* through core- and pan-genome analyses. *Genome Biol. Evol.* **11**: 1077-1087.
26. Lu W, Kong W, Yang P, Kong J. 2015. A one-step PCR-based method for specific identification of 10 common lactic acid bacteria and *Bifidobacterium* in fermented milk. *Int. Dairy J.* **41**: 7-12.
27. Stachelska MA. 2017. Identification of *Lactobacillus delbrueckii* and *Streptococcus thermophilus* strains present in artisanal raw cow milk cheese using real-Time PCR and classic plate count methods. *Polish. J. Microbiol.* **66**: 491-499.
28. Kim E, Kim HJ, Yang SM, Kim CG, Choo DW, Kim HY. 2019. Rapid identification of *Staphylococcus* species isolated from food samples by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Microbiol. Biotechnol.* **29**: 548-557.
29. Ha M, Jo HJ, Choi EK, Kim Y, Kim J, Cho HJ. 2019. Reliable identification of *Bacillus cereus* group species using low mass biomarkers by MALDI-TOF MS. *J. Microbiol. Biotechnol.* **29**: 887-896.
30. Huang CH, Huang L. 2018. Rapid species- and subspecies-specific level classification and identification of *Lactobacillus casei* group members using MALDI Biotyper combined with ClinProTools. *J. Dairy Sci.* **101**: 979-991.
31. Broeders S, Huber I, Grohmann L, Berben G, Taverniers I, Mazzara M, et al. 2014. Guidelines for validation of qualitative real-time PCR methods. *Trends Food Sci. Technol.* **37**: 115-126.