

Full-Length 16S rRNA Amplicon Sequencing for the Simple and Simultaneous Detection of Multiple Probiotic Species in Commercial Products

Gyeong-Eun Kim¹, Yoon-Soo Gwak¹, Chan-Il Bae¹, Kwangcheol C. Jeong^{2,3},
and Mi-Ju Kim^{1*}

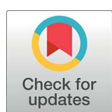
¹Institute of Life Sciences & Resources and Department of Food Science and Biotechnology, Kyung Hee University, Yongin 17104, Republic of Korea

²Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

³Department of Animal Sciences, IFAS, University of Florida, Gainesville, FL 32611, USA

To ensure efficacy and trust, probiotic strains in commercial products should be accurately identified. However, studies have consistently reported discrepancies between labeling and actual microbial composition. Therefore, we developed a full-length 16S rRNA amplicon sequencing workflow for species-level identification of probiotics. To optimize this workflow, four mock communities were analyzed using multiple pipelines and parameters; the workflow accuracy was evaluated. Optimal performance was achieved using Q10 filtering, the Emu pipeline with its default database, and a 0.1% relative abundance threshold. Subsequently, the workflow was applied to eight probiotic products, approximately 50% of which revealed labeling discrepancies with some labeled strains undetected. Furthermore, these results were compared and confirmed using PCR analysis. Despite some discrepancies, both approaches demonstrated overall consistency. Overall, the optimized sequencing workflow facilitates comprehensive species-level identification of probiotics within a single assay, providing a novel approach for quality control and labeling verification.

Keywords: Full-length 16S amplicon sequencing, Probiotic community, Species-level identification, Optimal analysis workflow, Commercial probiotic products



Received: March 5, 2026
Accepted: May 14, 2026
Published: June 4, 2026

*Corresponding author
E-mail: mijukim@khu.ac.kr

Supplementary data for this paper are available
online only at <http://jmb.or.kr>.

Conflict of Interest: The authors have no
financial conflicts of interest to declare.

eISSN 1738-8872

Copyright © 2026 by the authors. Licensee KMB.
This article is an open access article distributed
under the terms and conditions of the Creative
Commons Attribution (CC BY) license.

Introduction

Probiotics are live microorganisms that confer health benefits on the host when administered in adequate amounts [1]. Commercially available probiotic products worldwide use various probiotic strains. In Korea, the Ministry of Food and Drug Safety (MFDS) has approved 19 probiotic strains, including *Lactobacillus*, *Lactococcus*, *Enterococcus*, *Streptococcus*, and *Bifidobacterium*. Moreover, other strains that have not been approved, including *Pediococcus*, *Weissella*, and *Leuconostoc*, are utilized in probiotic products. Probiotic products contain various strains, and manufacturers provide essential information on the product labels, describing the type of microbes present in the product [2].

As the benefits and safety of microorganisms can vary by strain, accurate probiotic product labeling is crucial for product quality and consumer safety [3]. However, labeling inconsistencies in probiotic products have been repeatedly reported [4, 5]. The absence of listed strains or the presence of unlisted strains comprises these inconsistencies. Such labeling errors diminish the quality and efficacy of the products and raise safety concerns if pathogenic bacteria are included [3]. Consequently, to enhance the quality of probiotic products and provide consumer confidence, the development of effective methodologies for accurate identification and monitoring of microorganisms is essential.

Several studies have been conducted to verify probiotic information accuracy. Most of these studies mainly relied on conventional methods, including PCR and selective culture techniques. However, these methods are time-consuming and limited because they solely depend on the information of specific bacteria listed on the label, making them restrictive in application and unable to provide

detailed analyses of the products [2, 4]. Next-generation sequencing (NGS) has recently become an innovative approach for microbial community analysis, overcoming the limitations of conventional methods. The NGS technique enables culture-independent analysis and can accurately identify multiple targets simultaneously through large-scale DNA sequencing [6].

In the analysis of microbial communities using NGS, 16S rRNA amplicon sequencing is employed as the conventional approach. The 16S rRNA gene is approximately 1,550 base pairs in length and contains nine variable regions (V1–V9) that demonstrate significant sequence diversity [7]. Therefore, this gene is broadly recognized as a reliable and efficient standard for taxonomic classification [7, 8].

The 16S rRNA amplicon sequencing of probiotic products is frequently performed using a short-read based sequencing, targeting the V3–V4 region of the 16S rRNA gene [9]. This method facilitates the accurate and rapid identification of diverse bacterial taxa contained in products. However, a short-read based sequencing cannot cover the entire 16S rRNA gene; it only reads partial regions of the 16S rRNA, offering accurate probiotic bacteria identification up to the genus level [7, 10].

Third-generation sequencing has overcome these limitations, enabling full-length 16S rRNA gene sequencing using long-read based sequencers, such as Oxford Nanopore and PacBio [7, 11]. In particular, the MinION sequencer from Oxford is a compact user-friendly device with straightforward operation. Furthermore, this approach has enabled more accurate species-level analysis of microbial communities [6, 12, 13]. However, the application of Nanopore technology to probiotic products remains under-explored. In addition, a standardized analytical workflow ensures accurate and reliable taxonomic profiling has not been established.

Thus, we performed full-length 16S rRNA amplicon sequencing using Oxford Nanopore technology to establish an optimal workflow for accurate probiotic analysis. Rather than developing a new analytical method, this study aimed to establish an optimized workflow by integrating commercially available analysis pipelines with multiple analytical variables. To achieve this, we created four in-house mock communities and sequenced them to assess the accuracy of each sequencing and analysis pipeline. Nanopore sequencing targeted the full-length 16S rRNA gene, and the sequencing results were analyzed using three representative pipelines for 16S rRNA amplicon sequencing: Emu [14], Nanoclust [15], and Epi2me (v 5.1.14, Oxford Nanopore Technologies, UK; <https://epi2me.nanoporetech.com/>). Each pipeline was applied under various quality score conditions and relative abundance (RA) thresholds. The optimized workflow, confirmed through accuracy evaluation, was subsequently applied to the analysis of commercially available probiotic products, confirming product labeling accuracy.

Materials and Methods

Sample Preparation

For 16S rRNA amplicon sequencing, four in-house mock communities containing various probiotic species were utilized (Table 1). Overall, 30 probiotic strains used for constructing the mock communities were obtained from the Korean Collection for Type Cultures (KCTC, Republic of Korea), Korean Culture Center of Microorganisms (KCCM, Republic of Korea), and Korean Agricultural Culture Collection (KACC, Republic of Korea). The *Bifidobacterium* strains constituting the mock community were cultured in BL broth (MB Cell, Republic of Korea) at 37°C for 48 h

Table 1. Composition of mock communities.

Species	Strain no.	Mock 1 (%)	Mock 2 (%)	Mock 3 (%)	Mock 4 (%)
<i>Bifidobacterium animalis</i> subsp. <i>lactis</i>	KACC 16638	5.26	4	-	5
<i>Bifidobacterium bifidum</i>	KCTC 3418	5.26	4	-	-
<i>Bifidobacterium breve</i>	KACC 16639	5.26	4	-	-
<i>Bifidobacterium longum</i> subsp. <i>infantis</i>	KCTC 3249	-	-	-	5
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	KACC 20597	5.26	10	-	-
<i>Enterococcus faecalis</i>	KCTC 3206	5.26	4	-	5
<i>Enterococcus faecium</i>	KACC 11954	5.26	10	-	-
<i>Lactocaseibacillus casei</i>	KCTC 13086	5.26	4	9.1	-
<i>Lactocaseibacillus paracasei</i>	KCTC 5546	5.26	4	9.1	-
<i>Lactocaseibacillus rhamnosus</i>	KCTC 3237	5.26	4	9.1	5
<i>Lactiplantibacillus plantarum</i>	KCCM 12116	5.26	4	9.1	5
<i>Lactobacillus acidophilus</i>	KCTC 3164	5.26	4	9.1	-
<i>Lactobacillus crispatus</i>	KCTC 5054	-	-	-	5
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>	KACC 12420	5.26	4	9.1	5
<i>Lactobacillus gasseri</i>	KACC 12424	5.26	4	9.1	-
<i>Lactobacillus helveticus</i>	KACC 12418	5.26	4	9.1	-
<i>Lactobacillus johnsonii</i>	KCTC 5085	-	-	-	5
<i>Lactococcus lactis</i>	KCTC 3769	5.26	4	-	5
<i>Latilactobacillus sakei</i>	KCTC 3598	-	-	-	5
<i>Leuconostoc citreum</i>	KACC 11860	-	-	-	5
<i>Leuconostoc mesenteroides</i>	KCTC 3505	-	-	-	5
<i>Leuconostoc pseudomesenteroides</i>	KACC 12304	-	-	-	5
<i>Levilactobacillus brevis</i>	KACC 10553	-	-	-	5
<i>Ligilactobacillus salivarius</i>	KCTC 3600	5.26	10	9.1	5
<i>Limosilactobacillus fermentum</i>	KACC 11441	5.26	4	9.1	5
<i>Limosilactobacillus reuteri</i>	KCTC 3594	5.26	4	9.1	-
<i>Pediococcus acidilactici</i>	KACC 12307	-	-	-	5
<i>Pediococcus pentosaceus</i>	KCTC 3507	-	-	-	5
<i>Streptococcus thermophilus</i>	KACC 11857	5.26	10	-	5
<i>Weissella cibaria</i>	KCTC 3746	-	-	-	5

under anaerobic conditions, and the remaining bacterial strains were cultured in MRS broth (KisanBio, Republic of Korea) at 30°C for 48 h under anaerobic conditions.

Mock 1 comprised 19 MFDS-approved strains, mixed in equal proportions. Mock 2 contained the same 19 strains; however, the proportions were varied. Hypothetically, both Mocks 1 and 2 would demonstrate the detection efficiency of approved probiotics and that the RA of strains would be accurately reflected in samples containing them in either equal or varied proportions. Mock 3 encompassed 11 *Lactobacillus* genus strains from MFDS-approved probiotics, with equal proportions. This mock community was constructed for assessing the ability to distinguish closely related species within the same genus. To evaluate the ability to accurately detect a wide range of probiotic strains, Mock 4 comprised a combination of MFDS-approved probiotics and additional non-approved probiotics, also in equal proportions. The mock community, with a known taxonomic composition, was used as a standardized reference material to comparatively evaluate different analytical workflows and identify the optimal workflow [16].

Eight products were collected from local and online markets in Korea, each labeled with their probiotic composition. The type of products was either powder (n = 6) or capsule (n = 2) form, and all products were utilized for DNA extraction.

DNA Extraction

The cultured reference probiotic cells were centrifuged at 13,000 rpm for 1 min, and only the pellet was collected. Subsequently, the pellet was used for DNA extraction using the HiGene™ Genomic DNA Prep Kit (BIOFACT, Republic of Korea), with minor adjustments to the manufacturer's instructions. Furthermore, 100 mg of the probiotic products in powder and capsule forms were resuspended in 1-mL PBS and processed for DNA extraction using similar procedure. A NanoReady Touch spectrophotometer (Life Real, China) was used for measuring the purity of the extracted DNA, and Qubit dsDNA HS Assay kit on the Qubit 4.0 fluorometer (Invitrogen, USA) was used for quantifying the DNA concentration. Moreover, the DNA of the species constituting the mock communities was combined in the proportions indicated in Table 1, and the final concentration of the four mock communities was adjusted to 20 ng/μL.

Nanopore Sequencing

Full-length 16S rRNA gene amplification using two universal 16S primer sets preceded library preparation for Nanopore sequencing. Universal 16S primer sets were selected to enable detection of all probiotic bacteria present in the community through a single analysis. The primer sets included 27F (AGAGTTTGATCCTGGCTCAG) and 1492R (GGCTACCTTGTACGACTT) primer set without degenerate bases [17] and 27F (AGRGTTYGATYMTGGCTCAG) and 1492R (CGGYTACCTTGTACGACT) primer set with degenerate bases [6]. The non-degenerate 27F/1492R primer set has been reported to exhibit reduced amplification efficiency for specific taxa such as *Bifidobacterium*, which may introduce taxonomic bias in microbiome profiling [6]. Therefore, both primer sets were evaluated in this study to ensure a more accurate and representative analysis of community composition. Amplification was performed with the following conditions: 95°C for 3 min, followed by 25 denaturation cycles at 95°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 2 min, followed by a final elongation cycle at 72°C for 5 min. Subsequently, the PCR products were cleaned using AMPure XP (Beckman Coulter, USA).

The library construction process involved the end prep, native barcode ligation, and adapter ligation steps, using the Native Barcoding Kit 24 V14 (SQK-NBD114.24, Oxford Nanopore Technologies, UK). Using the Qubit dsDNA HS Assay kit on the Qubit 4.0 fluorometer (Invitrogen), the quantity of PCR products and the outcome of each step were measured; the barcoded samples were pooled in equal concentrations. Next, the pooled library was loaded into the R 10.4.1 flow cell (FLO-MIN114, Oxford Nanopore Technologies) and sequenced with MinION (Mk1B). MinKNOW software (v.24.06.16) was operated for sequencing and basecalling using the super-accuracy basecalling model.

Sequencing Data Analysis

The sequencing reads were filtered using Chopper [18] to filter the read length between 800 to 2,000 bp. Following length filtration, the quality filtering procedure was performed under two conditions as follows: quality scores of Q10 or Q20. Subsequently, the filtered reads were analyzed using pipelines designed for Nanopore 16S amplicon sequencing analysis or those providing dedicated workflows for this purpose: Emu [14], Nanoclust [15], and Epi2me (v 5.1.14, Oxford Nanopore Technologies, UK; <https://epi2me.nanoporetech.com/>). Emu and Nanoclust were run using default parameters, Emu being executed using either the Emu-developed or SILVA database, and Nanoclust using its default BLAST database. Epi2me was run using two analysis methods within the wf-metagenomic workflow, kraken2 and minimap2, using either the NCBI 16S or SILVA database. Additionally, the results of each analysis method were interpreted with RA thresholds of 0.01% and 0.1%.

Optimized Sequencing Workflow Determination

To confirm the optimized sequencing workflow, Nanopore sequencing analysis results were evaluated using accuracy evaluation metrics. These metrics were based on the number of true positives (TPs; species correctly identified in the mock), false positives (FPs; species not present in the mock but incorrectly identified), and false negatives (FNs; species not identified), and they include the following three metrics: precision, recall, and F-score. Precision denotes the proportion of claimed TPs that are truly present in the samples, and it is expressed as TP/(TP + FP). Recall refers to the percentage of expected positives that were detected by the workflow, calculated as TP/(TP + FN). F-score is the harmonic mean between the precision and the recall, which can be calculated using the formula $2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$. The method demonstrating high values across all three metrics was selected as the optimized sequencing workflow.

Optimized Sequencing Workflow Application and Validation for Probiotic Product Monitoring

Probiotic products underwent the optimized sequencing workflow, and the results were evaluated for species both labeled and identified, species identified but not labeled, and species labeled but not identified to assess probiotic labeling accuracy.

Furthermore, conventional PCR-based techniques, which are reported for labeling verification, were employed for cross-validation to verify the accuracy of the established workflow in identifying probiotic species. All eight products underwent PCR analyses using species-specific primers targeting the 19 MFDS-approved probiotic strains, based on a previous study [19, 20]. This step sought to determine whether the species listed on the label, as well as those not listed but potentially present, were consistently detected by the sequencing and PCR methods, thereby evaluating the accuracy and reliability of the amplicon sequencing approach.

Results and Discussion

Mock Community Analysis

Full-length 16S rRNA amplicon sequencing facilitates species-level analysis of bacterial communities, enabling simple and simultaneous detection of complex microbial structures. Therefore, to develop an optimized sequencing workflow for accurately profiling probiotic bacterial communities, we applied various analysis methods to this sequencing approach, following the workflow detailed in Fig. 1.

Four mock communities with known compositions, each containing various approved and non-approved probiotic strains, were prepared to accurately compare analysis methods and results. Because the taxonomic composition of each mock community was predefined, these communities were suitable for comparing workflow performance under different analytical conditions. Each mock community was amplified with two primer sets and underwent amplicon sequencing. Sequencing data were filtered on the basis of quality score and read length, and the filtered reads were analyzed using Emu, Nanoclust, and Epi2me, providing workflows for Nanopore 16S rRNA amplicon sequencing.

The overall sequencing statistics are presented in Table 2; in all cases except for Mock 1 with universal 16S primer, more than 180,000 reads were obtained following basecalling, demonstrating the high sequencing efficiency of the two primer sets. After filtering the basecalled reads with quality score and length, using Chopper under Q10 and Q20 conditions to retain reads at 800–2,000 bp, the average lengths of the filtered reads

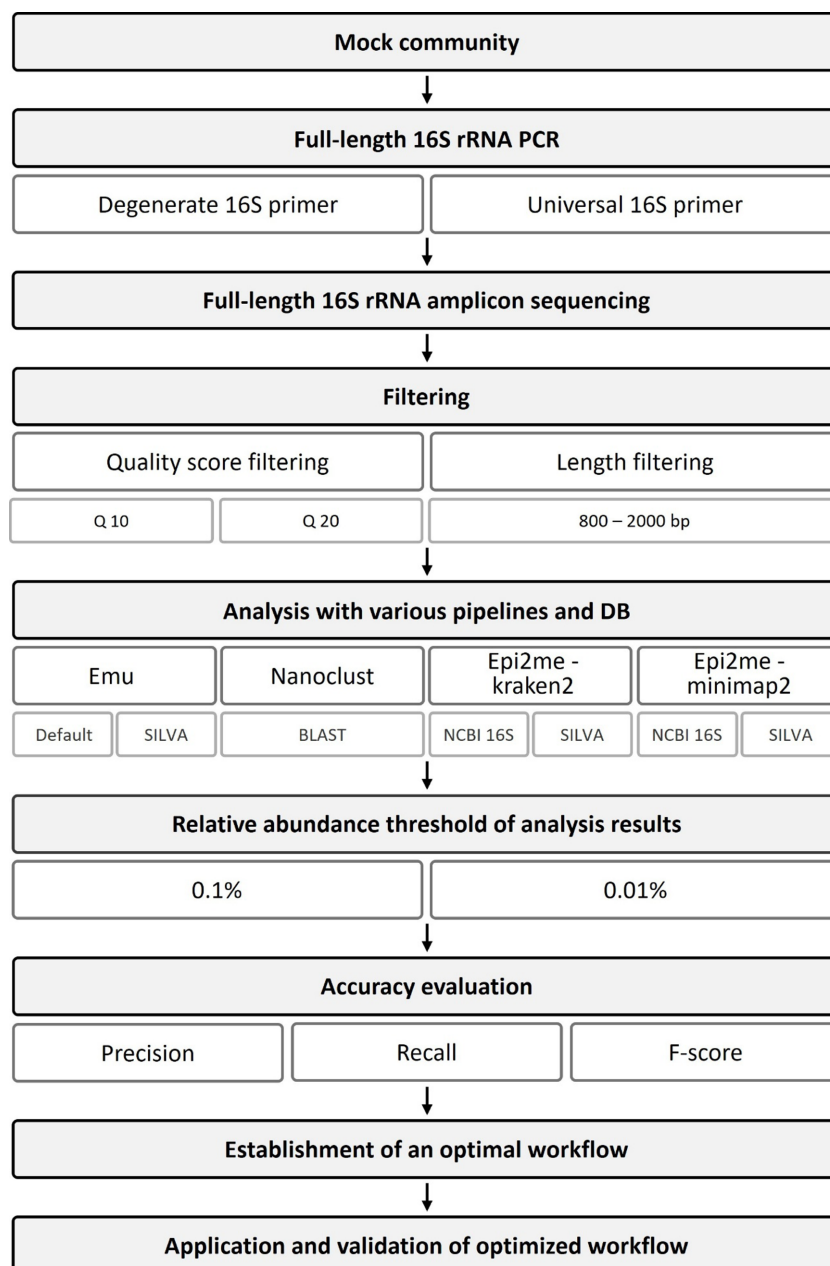


Fig. 1. Amplicon sequencing workflow used in this study.

Table 2. Statistics of full-length 16S rRNA amplicon sequencing reads.

Mock community	Primer	Basecalled reads		Q10 filtered reads		Q20 filtered reads	
		No. of reads	Avg (bp)	No. of reads	Avg (bp)	No. of reads	Avg (bp)
Mock 1	Degenerate 16S primer	251,760	1240.6	199,660	1443.9	82,188	1460.3
	Universal 16S primer	51,096	1206.4	38,411	1435.1	15,871	1458.4
Mock 2	Degenerate 16S primer	268,275	1220.7	207,978	1435.9	83,588	1459.3
	Universal 16S primer	301,025	1231.2	236,091	1445.4	98,231	1462.6
Mock 3	Degenerate 16S primer	183,790	1230.6	143,040	1449.3	60,193	1466.4
	Universal 16S primer	259,258	1217.7	199,682	1440.5	81,419	1462.5
Mock 4	Degenerate 16S primer	306,884	1239.8	242,952	1447.3	101,215	1463.0
	Universal 16S primer	279,266	1242.6	220,549	1453.9	93,735	1469.2

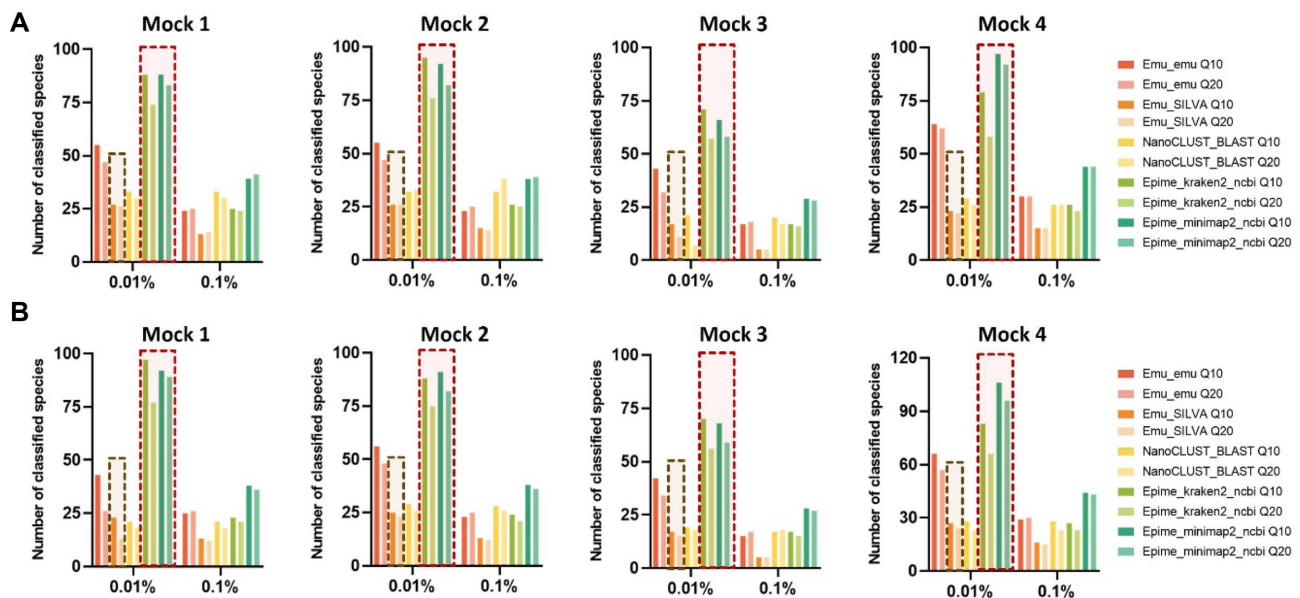


Fig. 2. Number of classified species per analysis pipelines using (A) degenerate 16S primer and (B) universal 16S primer.

were approximately 1,440 and 1,460 bp under Q10 and Q20 conditions, respectively. At this stage, when reads shorter than 800 bp were removed, the number of reads significantly decreased. However, more than 15,000 reads per sample remained after filtering, and these filtered reads were used for subsequent analysis.

Subsequently, the differences in species classification among these analysis methods were evaluated by comparing the number of detected species (Fig. 2). The highest number of species was identified with the Epi2me pipeline, which can be attributed to the tendency of Epi2me to classify a greater number of phylogenetically related species, causing the overclassification of species in the mock communities [12]. In contrast, using the SILVA database in Emu, the lowest number of species was detected. This finding may be attributed to the absence of certain species in the SILVA database provided by Emu, limiting the classification of that taxonomy to the genus level. Furthermore, the analysis performed using the SILVA database in Epi2me was performed up to the genus-level classification. This limitation can also be attributed to the identification level provided by the pipeline and database [21].

Full-Length 16S rRNA Amplicon Sequencing Performance Summary

The performance summary, which lists the TPs, FPs, and FNs for each analysis workflow, was organized for accuracy evaluation. For this evaluation, species-level detection was assumed to be accurate for all included strains, considering that mock communities were fully defined and contained only a single subspecies per species. The summaries in Tables 3 and 4, where degenerate and universal primers are used, respectively, are generally comparable; however, slight differences are noted depending on the analysis condition.

When employing the degenerate 16S primer, most analysis conditions at the species and genus levels demonstrated no FNs, except when Emu was used with the SILVA database. In contrast, using the universal 16S primer caused increased FNs in Mocks 1, 2, and 4. Importantly, the FNs were predominantly concentrated in conditions utilizing the universal 16S primer in the presence of *Bifidobacterium* taxa, indicating primer-dependent detection bias. This observation is consistent with previous reports examining sequence heterogeneities at the non-degenerate 27F primer annealing site of 16S rRNA genes. In those analyses, *Bifidobacterium* showed three base mismatches with the non-degenerate 27F forward primer, whereas most other probiotic taxa did not. Therefore, these mismatches induced biased amplification, consequently resulting in inaccurate detection and underrepresentation of *Bifidobacterium* during microbiome profiling [6].

In addition to primer selection, performance summaries also varied depending on the analytical pipeline and database. When Emu was used with the SILVA database, FN counts were markedly increased at the species level, suggesting a limited coverage of the database. However, all genera

Table 3. Performance summary of each pipeline using degenerate 16S primer.

Taxonomy level	RA Threshold (%)	Quality score	Performance metrics	Analysis pipeline						
				Emu Emu DB	Emu SILVA DB	Nanoclust	Epi2me kraken2 NCBI DB	Epi2me kraken2 SILVA DB	Epi2me minimap2 NCBI DB	Epi2me minimap2 SILVA DB
Mock 1										
Species	0.01	Q10	TP (19)	19	12	19	19	-	19	-
			FP	36	15	14	69	-	69	-
			FN	0	7	0	0	-	0	-
		Q20	TP (19)	19	12	19	19	-	19	-
			FP	28	14	11	55	-	64	-
			FN	0	7	0	0	-	0	-
	0.1	Q10	TP (19)	19	11	19	18	-	19	-
			FP	5	2	14	7	-	20	-
			FN	0	8	0	1	-	0	-
		Q20	TP (19)	19	11	19	18	-	19	-
			FP	6	3	11	6	-	22	-
			FN	0	8	0	1	-	0	-
Genus	0.01	Q10	TP (9)	9	9	9	9	9	9	9
			FP	6	10	1	7	22	8	12
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	4	8	1	7	22	7	11
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (9)	9	9	9	9	9	9	9
			FP	1	5	1	2	9	2	8
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	1	5	1	2	8	2	8
			FN	0	0	0	0	0	0	0
Mock 2										
Species	0.01	Q10	TP (19)	19	12	19	19	-	19	-
			FP	36	14	13	76	-	73	-
			FN	0	7	0	0	-	0	-
		Q20	TP (19)	19	12	19	19	-	19	-
			FP	28	14	14	57	-	63	-
			FN	0	7	0	0	-	0	-
	0.1	Q10	TP (19)	19	12	19	18	-	19	-
			FP	4	3	13	8	-	19	-
			FN	0	7	0	1	-	0	-
		Q20	TP (19)	19	11	19	18	-	19	-
			FP	6	3	13	7	-	20	-
			FN	0	8	0	1	-	0	-
Genus	0.01	Q10	TP (9)	9	9	9	9	9	9	9
			FP	7	10	2	7	25	12	11
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	5	7	1	7	25	8	12
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (9)	9	9	9	9	9	9	9
			FP	1	5	2	2	9	2	8
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	1	7	1	2	7	2	8
			FN	0	0	0	0	0	0	0
Mock 3										

Table 3. Continued

Taxonomy level	RA Threshold (%)	Quality score	Performance metrics	Analysis pipeline						
				Emu Emu DB	Emu SILVA DB	Nanoclust	Epi2me kraken2 NCBI DB	Epi2me kraken2 SILVA DB	Epi2me minimap2 NCBI DB	Epi2me minimap2 SILVA DB
Species	0.01	Q10	TP (11)	11	4	11	11	-	11	-
			FP	32	13	10	60	-	55	-
			FN	0	7	0	0	-	0	-
		Q20	TP (11)	11	4	1	11	-	11	-
			FP	21	7	6	46	-	47	-
			FN	0	7	0	0	-	0	-
	0.1	Q10	TP (11)	11	4	11	11	-	11	-
			FP	6	1	9	6	-	18	-
			FN	0	7	0	0	-	0	-
		Q20	TP (11)	11	4	11	11	-	11	-
			FP	7	1	6	5	-	17	-
			FN	0	7	0	0	-	0	-
Genus	0.01	Q10	TP (5)	5	5	5	5	5	5	5
			FP	11	14	2	13	24	13	15
			FN	0	0	0	0	0	0	0
		Q20	TP (5)	5	5	5	5	5	5	5
			FP	5	9	1	13	24	12	14
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (5)	5	5	5	5	5	5	5
			FP	1	4	1	2	11	2	7
			FN	0	0	0	0	0	0	0
		Q20	TP (5)	5	5	5	5	5	5	5
			FP	1	4	1	2	10	1	6
			FN	0	0	0	0	0	0	0
Mock 4										
Species	0.01	Q10	TP (20)	20	13	17	19	-	19	-
			FP	44	10	12	60	-	78	-
			FN	0	7	0	1	-	1	-
		Q20	TP (20)	20	13	17	19	-	19	-
			FP	42	9	9	39	-	73	-
			FN	0	7	0	2	-	1	-
	0.1	Q10	TP (20)	20	12	17	18	-	19	-
			FP	10	3	9	8	-	25	-
			FN	0	8	0	2	-	1	-
		Q20	TP (20)	19	12	17	18	-	19	-
			FP	11	3	9	5	-	25	-
			FN	1	8	0	2	-	1	-
Genus	0.01	Q10	TP (14)	14	14	14	14	14	14	14
			FP	3	5	0	5	25	6	7
			FN	0	0	0	0	0	0	0
		Q20	TP (14)	14	14	14	14	14	14	14
			FP	0	3	1	3	28	4	6
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (14)	14	14	14	14	14	14	14
			FP	0	2	0	1	8	1	4
			FN	0	0	0	0	0	0	0
		Q20	TP (14)	14	14	14	14	14	14	14
			FP	0	2	1	1	7	1	4
			FN	0	0	0	0	0	0	0

were effectively detected under similar analytical conditions, confirming the suitability of this approach for genus-level analysis. When Epi2me was used with the SILVA database, species-level identification was also not achieved, and a greater number of FPs was observed at the genus level than that using the NCBI 16S database with the same analytical method. This is because the SILVA database provided in Epi2me does not exceed the genus level, and therefore the lowest taxonomic rank available for analysis is genus. Furthermore, while the SILVA database is optimized for efficient alignment of short sequences, the NCBI 16S database is highly curated for species-level taxonomy and contains a boarder range of

Table 4. Performance summary of each pipeline using universal 16S primer.

Taxonomy level	RA Threshold (%)	Quality score	Performance metrics	Analysis pipeline						
				Emu Emu DB	Emu SILVA DB	Nanoclust	Epi2me kraken2 NCBI DB	Epi2me kraken2 SILVA DB	Epi2me minimap2 NCBI DB	Epi2me minimap2 SILVA DB
Mock 1										
Species	0.01	Q10	TP (19)	19	12	16	19	-	19	-
			FP	24	11	5	78	-	73	-
			FN	0	7	3	0	-	0	-
		Q20	TP (19)	17	9	15	19	-	19	-
			FP	9	4	3	58	-	70	-
			FN	0	10	4	0	-	0	-
	0.1	Q10	TP (19)	17	10	16	16	-	17	-
			FP	8	3	5	7	-	21	-
			FN	2	9	3	3	-	2	-
		Q20	TP (19)	17	9	15	15	-	16	-
			FP	9	3	3	6	-	20	-
			FN	2	10	4	4	-	3	-
Genus	0.01	Q10	TP (9)	9	9	9	9	9	9	9
			FP	5	9	1	9	25	9	15
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	2	6	1	9	25	9	14
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (9)	9	9	9	9	9	9	9
			FP	2	6	1	2	11	3	9
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	2	5	1	2	10	1	7
			FN	0	0	0	0	0	0	0
Mock 2										
Species	0.01	Q10	TP (19)	19	12	17	19	-	19	-
			FP	37	13	12	69	-	72	-
			FN	0	7	2	0	-	0	-
		Q20	TP (19)	19	12	16	19	-	19	-
			FP	29	11	10	56	-	63	-
			FN	0	7	3	0	-	0	-
	0.1	Q10	TP (19)	17	10	16	16	-	17	-
			FP	6	3	12	8	-	21	-
			FN	2	9	3	3	-	2	-
		Q20	TP (19)	16	9	16	15	-	16	-
			FP	9	3	10	6	-	20	-
			FN	3	10	3	4	-	3	-
Genus	0.01	Q10	TP (9)	9	9	9	9	9	9	9
			FP	8	10	1	6	24	10	14
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	4	7	1	6	25	9	14
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (9)	9	9	9	9	9	9	9
			FP	8	5	1	2	10	2	8
			FN	0	0	0	0	0	0	0
		Q20	TP (9)	9	9	9	9	9	9	9
			FP	1	5	1	2	7	1	7
			FN	0	0	0	0	0	0	0
Mock 3										

Table 4. Continued

Taxonomy level	RA Threshold (%)	Quality score	Performance metrics	Analysis pipeline						
				Emu Emu DB	Emu SILVA DB	Nanoclust	Epi2me kraken2 NCBI DB	Epi2me kraken2 SILVA DB	Epi2me minimap2 NCBI DB	Epi2me minimap2 SILVA DB
Species	0.01	Q10	TP (11)	11	4	11	11	-	11	-
			FP	31	13	8	59	-	57	-
			FN	0	7	0	0	-	0	-
		Q20	TP (11)	11	4	11	11	-	11	-
			FP	23	11	7	45	-	48	-
			FN	0	7	0	0	-	0	-
	0.1	Q10	TP (11)	11	4	11	11	-	11	-
			FP	4	1	6	6	-	17	-
			FN	0	7	0	0	-	0	-
		Q20	TP (11)	11	4	11	11	-	11	-
			FP	6	1	7	4	-	16	-
			FN	0	7	0	0	-	0	-
Genus	0.01	Q10	TP (5)	5	5	5	5	5	5	5
			FP	11	13	2	10	25	13	14
			FN	0	0	0	0	0	0	0
		Q20	TP (5)	5	5	5	5	5	5	5
			FP	7	11	1	10	24	12	13
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (5)	5	5	5	5	5	5	5
			FP	1	4	1	2	10	2	7
			FN	0	0	0	0	0	0	0
		Q20	TP (5)	5	5	5	5	5	5	5
			FP	1	4	1	2	9	1	6
			FN	0	0	0	0	0	0	0
Mock 4										
Species	0.01	Q10	TP (20)	20	13	17	19	-	19	-
			FP	46	14	11	64	-	87	-
			FN	0	7	3	1	-	1	-
		Q20	TP (20)	20	13	17	19	-	19	-
			FP	37	11	6	47	-	77	-
			FN	0	7	3	1	-	1	-
	0.1	Q10	TP (20)	18	13	17	18	-	19	-
			FP	11	3	11	9	-	25	-
			FN	2	7	3	2	-	1	-
		Q20	TP (20)	18	12	17	18	-	19	-
			FP	12	3	6	5	-	24	-
			FN	2	8	3	2	-	1	-
Genus	0.01	Q10	TP (14)	14	14	14	14	14	14	14
			FP	3	6	0	2	22	5	7
			FN	0	0	0	0	0	0	0
		Q20	TP (14)	14	14	14	14	14	14	14
			FP	0	3	0	2	26	3	6
			FN	0	0	0	0	0	0	0
	0.1	Q10	TP (14)	14	14	14	14	14	14	14
			FP	0	2	0	1	9	1	4
			FN	0	0	0	0	0	0	0
		Q20	TP (14)	14	14	14	14	14	14	14
			FP	0	2	0	1	8	0	3
			FN	0	0	0	0	0	0	0

sequences with different lengths and annotation quality [22, 23]. This finding indicates that the choice of taxonomy database can lead to different outcomes at the same taxonomic resolution, despite using the same pipeline.

Along with the pipeline settings, the RA threshold also significantly affected the performance summary. The analysis revealed a substantially reduced number of FPs at the species level when the threshold was set to 0.1%, with Q10 and Q20 conditions generating comparable results. However, Q20 demonstrated a markedly lower number of FPs than Q10 under the 0.01% threshold.

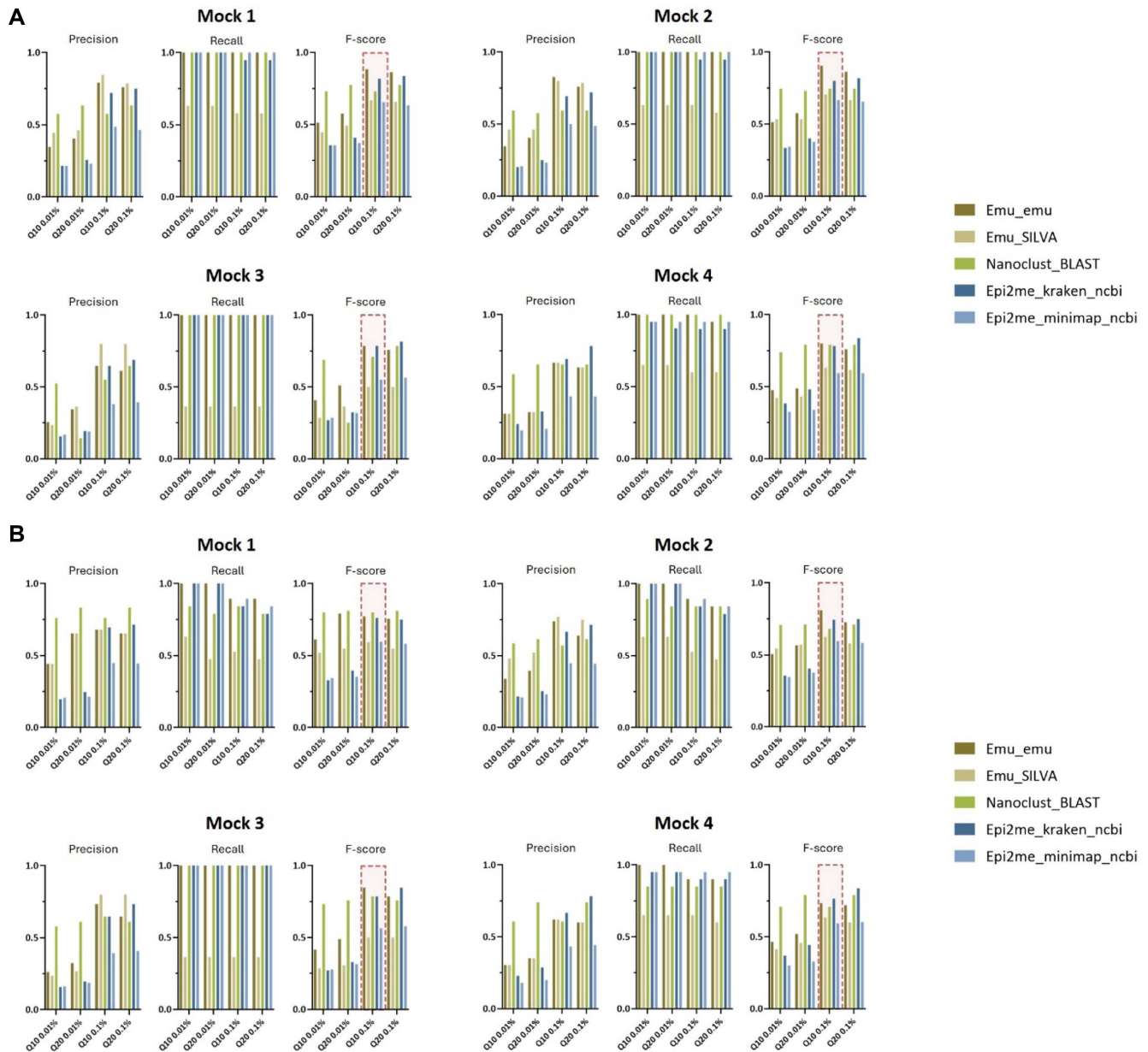


Fig. 3. Species-level accuracy evaluation of each pipeline using precision, recall, and F-score with (A) degenerate 16S primer and (B) universal 16S primer.

Accuracy Evaluation and Establishment of the Optimal Analysis Condition

The accuracy metrics, including precision, recall, and F-score, were calculated on the basis of the performance summary, and the results were used for establishing the optimal analysis condition. The values of the three metrics under different analysis conditions for all mock communities, with higher values indicating greater accuracy, are depicted in Fig. 3. The precision revealed that most pipelines exhibited higher accuracy under the 0.1% RA threshold. Notably, Emu and Epi2me with kraken2 methods displayed particularly high values. Under the 0.01% condition, recall values were slightly higher, with Emu using the SILVA database consistently yielding the lowest recall across all conditions. Moreover, recall values were further reduced when using the universal 16S primer with the 0.1% threshold, which can be attributed to *Bifidobacterium* being classified as FNs. These results demonstrate a clear trade-off between precision and recall depending on the applied RA threshold. Under the 0.1% threshold, the stringent filtering criteria reduced the number of detected taxa, effectively minimizing FPs and enhancing precision. However, this also increased FNs by omitting low-abundance true taxa, thereby lowering recall values. In contrast, the 0.01% threshold improved recall by facilitating the recovery of rare taxa and reducing FNs, while the relaxed threshold also increased FPs and reduced precision. Given the inverse relationship between these two metrics across the RA thresholds, it was difficult to determine the optimal analytical condition using either metric alone. Therefore, the F-score, which represents the harmonic mean of precision and recall and balances the trade-off between the two metrics, was used as the primary criterion for selecting the optimal workflow [24].

Table 5. Application test results of the optimized amplicon sequencing workflow to commercial probiotic products.

Sample	Labelled species	Detected species	Non-detected species	Non-labelled species
Probiotic 1	<i>L. acidophilus</i> , <i>L. gasseri</i> , <i>L. delbrueckii</i> subsp. <i>bulgaricus</i> , <i>L. helveticus</i> , <i>L. rhamnosus</i> , <i>L. casei</i> , <i>L. paracasei</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>L. plantarum</i> , <i>L. salivarius</i> , <i>Lc. lactis</i> , <i>S. thermophilus</i> , <i>B. bifidum</i> , <i>B. breve</i> , <i>B. longum</i> , <i>B. animalis</i> subsp. <i>lactis</i> , <i>Leu. citreum</i> , <i>B. infatis</i>	<i>Leu. citreum</i> , <i>L. plantarum</i> , <i>L. reuteri</i> , <i>B. longum</i> , <i>Lc. lactis</i> , <i>S. thermophilus</i> , <i>B. animalis</i> , <i>L. acidophilus</i> , <i>B. breve</i> , <i>L. fermentum</i> , <i>L. paracasei</i> , <i>L. rhamnosus</i> , <i>L. delbrueckii</i> , <i>L. gasseri</i> , <i>B. bifidum</i>	<i>B. longum</i> subsp. <i>infatis</i> , <i>L. salivarius</i> , <i>L. casei</i> , <i>L. helveticus</i>	<i>L. pentosus</i> , <i>E. faecium</i> , <i>S. salivarius</i>
Probiotic 2	<i>L. acidophilus</i> , <i>L. gasseri</i> , <i>L. delbrueckii</i> subsp. <i>bulgaricus</i> , <i>L. helveticus</i> , <i>L. rhamnosus</i> , <i>L. casei</i> , <i>L. paracasei</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>L. plantarum</i> , <i>L. salivarius</i> , <i>Lc. lactis</i> , <i>E. faecium</i> , <i>E. faecalis</i> , <i>S. thermophilus</i> , <i>B. bifidum</i> , <i>B. breve</i> , <i>B. longum</i> , <i>B. animalis</i> subsp. <i>lactis</i>	<i>L. plantarum</i> , <i>Lc. lactis</i> , <i>L. reuteri</i> , <i>B. longum</i> , <i>E. faecium</i> , <i>B. breve</i> , <i>S. thermophilus</i> , <i>B. animalis</i> , <i>L. salivarius</i> , <i>L. casei</i> , <i>L. gasseri</i> , <i>L. fermentum</i> , <i>L. delbrueckii</i> , <i>L. rhamnosus</i>	<i>L. helveticus</i> , <i>L. acidophilus</i> , <i>B. bifidum</i> , <i>L. paracasei</i> , <i>E. faecalis</i>	<i>L. pentosus</i>
Probiotic 3	<i>E. faecium</i> , <i>S. thermophilus</i> , <i>L. plantarum</i> , <i>B. bifidum</i> , <i>L. rhamnosus</i> , <i>L. fermentum</i> , <i>L. paracasei</i> , <i>Lc. lactis</i> , <i>L. gasseri</i> , <i>L. acidophilus</i> , <i>L. salivarius</i> , <i>B. animalis</i> subsp. <i>lactis</i>	<i>E. faecium</i> , <i>B. bifidum</i> , <i>L. plantarum</i> , <i>S. thermophilus</i> , <i>L. fermentum</i> , <i>L. rhamnosus</i> , <i>Lc. lactis</i> , <i>L. salivarius</i> , <i>L. gasseri</i> , <i>L. acidophilus</i> , <i>L. paracasei</i> , <i>B. animalis</i>	-	<i>E. durans</i> , <i>S. salivarius</i> , <i>E. hirae</i> , <i>L. pentosus</i>
Probiotic 4	<i>L. acidophilus</i> , <i>L. gasseri</i> , <i>L. delbrueckii</i> subsp. <i>bulgaricus</i> , <i>L. helveticus</i> , <i>L. rhamnosus</i> , <i>L. casei</i> , <i>L. paracasei</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>L. plantarum</i> , <i>L. salivarius</i> , <i>Lc. lactis</i> , <i>E. faecium</i> , <i>E. faecalis</i> , <i>S. thermophilus</i> , <i>B. bifidum</i> , <i>B. breve</i> , <i>B. longum</i> , <i>B. animalis</i> subsp. <i>lactis</i>	<i>Lc. lactis</i> , <i>L. acidophilus</i> , <i>S. thermophilus</i> , <i>L. plantarum</i> , <i>L. reuteri</i> , <i>E. faecium</i> , <i>B. animalis</i> , <i>L. salivarius</i> , <i>B. bifidum</i> , <i>L. casei</i> , <i>B. breve</i> , <i>E. faecalis</i> , <i>L. rhamnosus</i> , <i>L. fermentum</i> , <i>L. paracasei</i> , <i>L. delbrueckii</i> , <i>L. helveticus</i> , <i>L. gasseri</i>	<i>B. longum</i>	<i>E. durans</i> , <i>L. pentosus</i> , <i>S. salivarius</i> , <i>L. sakei</i> , <i>L. amylovorus</i> , <i>Lc. sp. KACC 19320</i> , <i>L. crispatus</i>
Probiotic 5	<i>L. plantarum</i> , <i>L. rhamnosus</i> , <i>L. casei</i> , <i>B. longum</i> , <i>B. breve</i> , <i>L. paracasei</i> , <i>B. bifidum</i> , <i>S. thermophilus</i> , <i>B. animalis</i> subsp. <i>lactis</i> , <i>L. fermentum</i> , <i>Bacillus coagulans</i>	<i>L. plantarum</i> , <i>L. rhamnosus</i> , <i>B. longum</i> , <i>B. breve</i> , <i>L. paracasei</i> , <i>S. thermophilus</i> , <i>L. fermentum</i> , <i>B. animalis</i> , <i>B. bifidum</i> , <i>L. casei</i> , <i>B. coagulans</i>	-	<i>B. agri</i> , <i>B. brevis</i> , <i>S. salivarius</i> , <i>L. pentosus</i> , <i>Lc. lactis</i>
Probiotic 6	<i>L. sporogenes</i> (<i>Bacillus coagulans</i>), <i>Clostridium butyricum</i> , <i>Bacillus subtilis</i> , <i>L. johnsonii</i> , <i>L. plantarum</i> , <i>B. lactis</i> , <i>L. rhamnosus</i> , <i>B. breve</i>	<i>B. coagulans</i> , <i>L. plantarum</i> , <i>B. animalis</i> , <i>L. johnsonii</i> , <i>B. subtilis</i> , <i>L. rhamnosus</i> , <i>C. butyricum</i> , <i>B. breve</i>	-	<i>L. hominis</i> , <i>L. paracasei</i> , <i>Lc. lactis</i> , <i>L. casei</i> , <i>B. velezensis</i>
Probiotic 7	<i>L. acidophilus</i> , <i>L. gasseri</i> , <i>L. delbrueckii</i> subsp. <i>bulgaricus</i> , <i>L. helveticus</i> , <i>L. rhamnosus</i> , <i>L. casei</i> , <i>L. paracasei</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>L. plantarum</i> , <i>L. salivarius</i> , <i>Lc. lactis</i> , <i>S. thermophilus</i> , <i>B. bifidum</i> , <i>B. breve</i> , <i>B. longum</i> , <i>B. animalis</i> subsp. <i>lactis</i> , <i>Bacillus subtilis</i>	<i>Lc. lactis</i> , <i>B. animalis</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>S. thermophilus</i> , <i>L. plantarum</i> , <i>L. salivarius</i> , <i>L. delbrueckii</i> , <i>B. longum</i> , <i>L. paracasei</i> , <i>L. helveticus</i> , <i>B. breve</i> , <i>L. gasseri</i> , <i>B. subtilis</i>	<i>L. rhamnosus</i> , <i>L. acidophilus</i> , <i>B. bifidum</i> , <i>L. casei</i>	-
Probiotic 8	<i>L. plantarum</i> , <i>E. faecium</i> , <i>L. reuteri</i> , <i>L. fermentum</i> , <i>B. animalis</i> ssp. <i>lactis</i> , <i>S. thermophilus</i> , <i>L. casei</i> , <i>Lc. lactis</i> , <i>L. gasseri</i> , <i>B. longum</i>	<i>E. faecium</i> , <i>L. plantarum</i> , <i>S. thermophilus</i> , <i>B. animalis</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>Lc. lactis</i> , <i>L. casei</i> , <i>B. longum</i>	<i>L. gasseri</i>	<i>E. durans</i> , <i>S. salivarius</i>

Analysis of the F-score revealed that Nanoclust achieved particularly high values under the 0.01% threshold, whereas Emu, Nanoclust, and Epi2me with the kraken2 method exhibited similarly high values under the 0.1% condition. Notably, most workflows yielded higher F-scores at the 0.1% threshold compared with the 0.01% condition. The highest overall F-score was observed under the Q10 and 0.1% RA threshold condition using Emu with its default database and a degenerate 16S primer. Consequently, based on the F-score as a balanced metric for both precision and recall, this analysis workflow was confirmed as the optimal setting and was identified as the optimal sequencing workflow of this study. In addition, this condition was found to be more suitable for species-level verification and quality control of commercial probiotic products, as it effectively reduced FPs of low abundance taxa. The detailed analysis results of the four mock communities, obtained using optimized workflow, are presented in Supplementary Table 1.

Application of the Established Amplicon Sequencing Workflow

To assess potential labeling discrepancies, the optimized workflow established through mock community analysis was applied to eight probiotic products. Therefore, PCR was performed using degenerate 16S primer, followed by full-length 16S rRNA amplicon sequencing utilizing the Oxford Nanopore platform. Subsequently, the resulting sequencing data were filtered with a quality score of Q10. Taxonomic classification was then conducted using the Emu pipeline with its default database, and the results were filtered at a 0.1% RA threshold. Next, the detected taxa were compared against the species listed on each product's label. Moreover, in this study, strains specified at the subspecies level, including *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Bifidobacterium animalis* subsp. *lactis*, were identified as successfully detected even when the result resolution was limited to the species level. Based on this comparison, the resulting data were categorized as labeled and detected species (TPs), not labeled but detected species (FPs), and labeled but not detected species (FNs) (Table 5).

Owing to amplicon sequencing, all labeled species were successfully identified in products 3, 5, and 6, indicating the absence of FNs. Conversely, products 1, 2, 4, 7, and 8 showed labeling discrepancies, with between one and five species not detected. These findings suggest that approximately 63% of the tested probiotic products contain labeling inaccuracies, wherein one or more declared species were not detected in the sequencing results. The undetected labeled species showed variation across products, which may be attributed to several factors, including low abundance, PCR amplification bias, or the actual absence of the species. Furthermore, an examination of FPs revealed that, except for product 7, all the other products demonstrated at least one unlabeled species detected with an RA of >0.1%, suggesting contamination or unintended inclusion during



Fig. 4. Comparison of the results from the optimized amplicon sequencing workflow and PCR analysis of probiotic products.

production. However, mock community analysis consistently noted FP occurrence, suggesting that such results cannot solely reflect the true composition of the products. This finding may result from sequencing errors, taxonomic misassignments by the analysis pipelines, or technical artifacts, including index swapping during sequencing [13, 25]. Therefore, to determine that the FPs from amplicon sequencing truly reflect components of the products, additional validation is warranted.

Verification of the Established Amplicon Sequencing Workflow

We performed PCR analyses, a method that is frequently used as a conventional approach for verifying product labeling, to validate the optimized sequencing workflow and evaluate the labeling discrepancies arising from amplicon sequencing. This study focused on 19 MFDS-approved probiotic strains, based on previous studies [19, 20] and applied across all products. The outcomes of the PCR assays were subsequently compared with the amplicon sequencing results (Fig. 4).

The strains initially classified as FPs in the amplicon analysis, including *E. faecium* in product 1, *L. lactis* in product 5, and *L. casei*, *L. paracasei*, and *L. lactis* in product 6, were not amplified using PCR, indicating that FPs were not present in the products and were consequence of sequencing artifacts rather than actual labeling discrepancies. Several strains listed on the product labeling but identified as FNs using amplicon sequencing, including *L. helveticus* and *L. salivarius* in product 1; *L. acidophilus*, *L. helveticus*, *L. paracasei*, *E. faecalis*, and *B. bifidum* in product 2; and *L. acidophilus*, *L. rhamnosus*, and *B. bifidum* in product 7, were all successfully detected using PCR. Based on these results, FNs were observed in only three products, and product 2 had the highest number of FNs, reflecting uneven strain abundance among complex multi-species formulation. These PCR results highlighted a major limitation of the amplicon sequencing method: the utilization of strict analysis condition and the impact of amplification bias can hinder the detection of strains that are present in small amounts, significantly impacting detection results. In addition to these observations, several strains demonstrated consistent results between amplicon sequencing and PCR, as in the case of *L. casei*, *B. longum*, *L. casei*, and *L. gasseri* in products 1, 4, 7, and 8, respectively, which were listed but not detected using either method. Accordingly, amplicon sequencing revealed both concordance and discordance with conventional PCR-based detection methods.

Furthermore, as summarized in Table 5, products 5, 6, and 7 contain strains not included in the MFDS-approved list. These non-listed strains exhibited RA values of >0.1% and clearly detected through amplicon sequencing. Although these non-approved strains did not undergo PCR verification, we assumed their presence on the basis of the finding that other strains detected with high RA by amplicon sequencing were consistently amplified using PCR. Therefore, approximately 50% of the tested products (products 2, 3, 5, and 6) contained all the labeled strains, whereas the remaining 50% (products 1, 4, 7, and 8) lacked one or more strains.

However, a discrepancy between amplicon sequencing and PCR results was observed in the case of *L. delbrueckii subsp. bulgaricus*. This strain was identified at the species level using amplicon sequencing and presumed to represent the labeled *L. delbrueckii subsp. bulgaricus* in products 1 and 7, but not amplified using PCR targeting subspecies-specific markers. These results suggest that a different *L. delbrueckii* subspecies can be present in the product, likely reflecting mislabeling, indicating a limitation of 16S rRNA amplicon sequencing in accurately resolving probiotic strains at the subspecies level. This limitation may arise because genetically closely related subspecies often share highly similar 16S rRNA gene sequences, which may prevent full-length 16S rRNA gene sequencing from providing sufficient resolution to distinguish them [13]. Moreover, the reference database utilized in optimized workflow was predominantly annotated at the species level, causing inadequate subspecies-level identification. Therefore, expanding the reference database to incorporate more detailed taxonomic annotations may enhance the accuracy of full-length 16S rRNA amplicon sequencing in future studies.

Based on these taxonomic constraints, the proposed workflow enables species level identification for probiotic product testing, rather than the subspecies level. However, this still represents an improvement over conventional partial 16S rRNA amplicon sequencing, which generally enables only genus-level identification. Therefore, this enhanced resolution makes the workflow suitable for quality control, especially for verifying product labels and detecting contaminants.

Shotgun metagenomic sequencing can be used as a complementary approach to 16S rRNA amplicon sequencing, enabling more comprehensive microbial profiling. It also allows detection of non-bacterial probiotics such as yeast and assessment of functional features such as antimicrobial resistance (AMR) genes. However, full-length 16S rRNA sequencing offers higher sensitivity for detecting bacterial taxa at low abundance that are

often missed by WGS and requires lower sequencing costs and computational resources [3]. Consequently, whereas WGS is valuable for in-depth functional assessment, the full-length 16S rRNA amplicon sequencing workflow provides a more practical, efficient, and user-friendly solution for quality control screening in the probiotic industry.

Conclusion

This study developed an optimized full-length 16S rRNA amplicon sequencing workflow for accurately and efficiently profiling probiotic microbial communities. Despite limitations in detecting low-abundance strains and resolving subspecies owing to taxonomic resolution constraints, the approach significantly decreased analysis time and complexity by facilitating comprehensive community profiling in a single sequencing run. Moreover, the approach offered superior species-level taxonomic resolution, enabling more accurate identification than conventional amplicon sequencing approaches.

Importantly, this study systematically combined multiple analytical components commonly used in 16S rRNA amplicon sequencing. These include widely applied pipelines such as Emu, Nanoclust, and Epi2me, and analytical parameters such as primer selection and analysis conditions. This approach led to the development of a workflow that provides a valuable alternative to existing methods, offering a simple and accurate approach that can be readily implemented using widely available tools. Therefore, this approach holds considerable potential for broader application as an advanced monitoring system, contributing to the safety assurance of probiotic products and the enhancement of quality control in the industry.

Author Contributions

Gyeong-Eun Kim: Methodology, Writing-Original Draft. **Yoon-Soo Gwak:** Data Curation, Writing-Original Draft. **Chan-Il Bae:** Methodology. **Kwangcheol C. Jeong:** Writing-review & editing. **Mi-Ju Kim:** Writing-Original Draft, Writing-review & editing, Conceptualization.

References

- Morelli L, Capurso L. 2012. FAO/WHO guidelines on probiotics: 10 years later. *J. Clin. Gastroenterol.* **46**: S1-S2. <https://doi.org/10.1097/mcg.0b013e318269fdd5>.
- Patro JN, Ramachandran P, Barnaba T, Mammel MK, Lewis JL, Elkins CA. 2016. Culture-independent metagenomic surveillance of commercially available probiotics with high-throughput next-generation sequencing. *MSphere* **1**: 10.1128/msphere.00057-00016. <https://doi.org/10.1128/msphere.00057-16>.
- Kruasuwat W, Jenjaroenpun P, Arigul T, Chokesajjawatee N, Leekitcharoenphon P, Poongladda S, et al. 2023. Nanopore sequencing discloses compositional quality of commercial probiotic feed supplements. *Sci. Rep.* **13**: 4540. <https://doi.org/10.1038/s41598-023-31626-4>.
- Chen T, Wu Q, Zhou H, Deng K, Wang X, Meng F, et al. 2017. Assessment of commercial probiotic products in China for labelling accuracy and probiotic characterisation of selected isolates. *Int. J. Dairy Technol.* **70**: 119-126. <https://doi.org/10.1111/1471-0307.12331>.
- Syromyatnikov M, Nesterova E, Gladikh M, Popov V. 2022. Probiotics analysis by high-throughput sequencing revealed multiple mismatches at bacteria genus level with the declared and actual composition. *LWT* **156**: 113055. <https://doi.org/10.1016/j.lwt.2021.113055>.
- Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, et al. 2021. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiol.* **21**: 35. <https://doi.org/10.1186/s12866-021-02094-5>.
- Buetas E, Jordán-López M, López-Roldán A, D'Auria G, Martínez-Priego L, De Marco G, et al. 2024. Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples. *BMC Genomics* **25**: 310. <https://doi.org/10.1186/s12864-024-10213-5>.
- Stackebrandt E, Goebel BM. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **44**: 846-849. <https://doi.org/10.1099/00207173-44-4-846>.
- Lugli GA, Mangifesta M, Mancabelli L, Milani C, Turroni F, Viappiani A, et al. 2019. Compositional assessment of bacterial communities in probiotic supplements by means of metagenomic techniques. *Int. J. Food Microbiol.* **294**: 1-9. <https://doi.org/10.1016/j.jfoodmicro.2019.01.011>.
- Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, et al. 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* **38**: e200-e200. <https://doi.org/10.1093/nar/gkq873>.
- Fujiyoshi S, Muto-Fujita A, Maruyama F. 2020. Evaluation of PCR conditions for characterizing bacterial communities with full-length 16S rRNA genes using a portable nanopore sequencer. *Sci. Rep.* **10**: 12580. <https://doi.org/10.1038/s41598-020-69450-9>.
- Lao H-Y, Wong LL-Y, Hui Y, Ng TT-L, Chan CT-M, Lo HW-H, et al. 2024. The clinical utility of Nanopore 16S rRNA gene sequencing for direct bacterial identification in normally sterile body fluids. *Front. Microbiol.* **14**: 1324494. <https://doi.org/10.3389/fmicb.2023.1324494>.
- Zhang H, Wang X, Chen A, Li S, Tao R, Chen K, et al. 2024. Comparison of the full-length sequence and sub-regions of 16S rRNA gene for skin microbiome profiling. *mSystems* **9**: e00399-00324. <https://doi.org/10.1128/msystems.00399-24>.
- Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, et al. 2022. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* **19**: 845-853. <https://doi.org/10.1038/s41592-022-01520-4>.
- Rodríguez-Pérez H, Ciuffreda L, Flores C. 2021. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics* **37**: 1600-1601. <https://doi.org/10.1093/bioinformatics/btaa900>.
- Bokulich NA, Ziemski M, Robeson II MS, Kaehler BD. 2020. Measuring the microbiome: best practices for developing and benchmarking microbiomics methods. *Comput. Struct. Biotechnol. J.* **18**: 4048-4062. <https://doi.org/10.1016/j.csbj.2020.11.049>.
- Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* **74**: 2461-2470. <https://doi.org/10.1128/AEM.02272-07>.
- De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**: btad311. <https://doi.org/10.1093/bioinformatics/btad311>.
- Kim E, Kim D, Yang S-M, Kim H-Y. 2022. Validation of probiotic species or subspecies identity in commercial probiotic products using high-resolution PCR method based on large-scale genomic analysis. *Food Res. Intern.* **154**: 111011. <https://doi.org/10.1016/j.foodres.2022.111011>.
- Kim E, Yang S-M, Lim B, Park SH, Rackerby B, Kim H-Y. 2020. Design of PCR assays to specifically detect and identify 37 *Lactobacillus* species in a single 96 well plate. *BMC Microbiol.* **20**: 96. <https://doi.org/10.1186/s12866-020-01781-z>.
- Balvočiūtė M, Huson DH. 2017. SILVA, RDP, Greengenes, NCBI and OTT—how do these taxonomies compare? *BMC Genomics* **18**: 114. <https://doi.org/10.1186/s12864-017-3501-4>.
- Hoffbeck C, Middleton DM, Nelson NJ, Taylor MW. 2025. Benchmarking 16S rRNA Gene-based approaches to bacterial taxonomy assignment based on amplicon sequencing with illumina and oxford nanopore. *Int. J. Microbiol.* **2025**: 7563096. <https://doi.org/10.1155/ijm/7563096>.
- Gupta A, Cooper V, Zemke A. 2025. Evaluation of V3-V4 and FL-16S rRNA amplicon sequencing approach for microbiota community analysis of tracheostomy aspirates. *MSphere* **10**: e00388-00325. <https://doi.org/10.1128/msphere.00388-25>.
- Sujon KM, Hassan R, Choi K, Samad MA. 2025. Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. *J. Big Data* **12**: 268. <https://doi.org/10.1186/s40537-025-01313-4>.
- Schwenker JA, Friedrichsen M, Waschina S, Bang C, Franke A, Mayer R, et al. 2022. Bovine milk microbiota: evaluation of different DNA extraction protocols for challenging samples. *Microbiologyopen* **11**: e1275. <https://doi.org/10.1002/mbo3.1275>.