Original Article

DOI: 10.5582/bst.2024.01393

Comparative analysis of human gut bacterial microbiota between shallow shotgun metagenomic sequencing and full-length 16S rDNA amplicon sequencing

Suwalak Chitcharoen[1,2], Vorthon Sawaswong[3], Pavit Klomkliew[4], Prangwalai Chanchaem[4], Sunchai Payungporn[4,*]

[1] Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand;
[2] Research and Diagnostic Center for Emerging Infectious Diseases, Khon Kaen University, Khon Kaen, Thailand;
[3] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand;
[4] Center of Excellence in Systems Microbiology, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.

SUMMARY: The human gut microbiome is increasingly recognized as important to health and disease, influencing immune function, metabolism, mental health, and chronic illnesses. Two widely used, cost-effective, and fast approaches for analyzing gut microbial communities are shallow shotgun metagenomic sequencing (SSMS) and full-length 16S rDNA sequencing. This study compares these methods across 43 stool samples, revealing notable differences in taxonomic and species-level detection. At the genus level, *Bacteroides* was most abundant in both methods, with *Faecalibacterium* showing similar trends but *Prevotella* was more abundant in full-length 16S rDNA. Genera such as *Alistipes* and *Akkermansia* were more frequently detected by full-length 16S rDNA, whereas *Eubacterium* and *Roseburia* were more prevalent in SSMS. At the species level, *Faecalibacterium prausnitzii*, a key indicator of gut health, was abundant across both datasets, while *Bacteroides vulgatus* was more frequently detected by SSMS. Species within *Parabacteroides* and *Bacteroides* were primarily detected by 16S rDNA, contrasting with higher SSMS detection of *Prevotella copri* and *Oscillibacter valericigenes*. LEfSe analysis identified 18 species (9 species in each method) with significantly different detection between methods, underscoring the impact of methodological choice on microbial diversity and abundance. Differences in classification databases, such as Ribosomal Database Project (RDP) for 16S rDNA and Kraken2 for SSMS, further highlight the influence of database selection on outcomes. These findings emphasize the importance of carefully selecting sequencing methods and bioinformatics tools in microbiome research, as each approach demonstrates unique strengths and limitations in capturing microbial diversity and relative abundances.

Keywords: bacterial profile, microbiome, oxford nanopore technologies (ONT), ion torrent sequencing

1. Introduction

The gut microbiome is a complex ecosystem consisting of a diverse community of microorganisms residing in the human gastrointestinal tract. Numerous studies have demonstrated that a diverse and balanced population of gut microbiota is crucial for maintaining gut and overall health by facilitating digestion, nutrient absorption, and supporting the immune system. Furthermore, mounting evidence has suggested that dysbiosis, an imbalance in the composition and functionality of gut microbiota, is directly or indirectly associated with the pathogenesis of various diseases, including obesity (*1*), diabetes, chronic kidney disease (CKD) (*2*), liver diseases, colorectal cancer (CRC) or adenoma (*3*), and even mental and neurodegenerative disease (*4-6*). Therefore, modulation of the gut microbiome composition has been proposed as a potential therapeutic target, and dietary interventions have been suggested as a means to achieve this goal (*7*). In particular, patients with autoimmune diseases display reduced levels of beneficial bacteria, such as *Bifidobacterium* spp., *Faecalibacterium* spp., *Roseburia* spp., and *Coprococcus eutactus*, alongside increased levels of pathogenic bacteria like *Escherichia coli*, *Staphylococcus aureus*, and *Clostridioides difficile*, accompanied by microbial-driven TH17/TH1 activation and reduced Regulatory T cells, worsening inflammation (*8-10*). Consequently, modulating gut microbiota composition is proposed as a therapeutic target, and dietary interventions are suggested as a viable approach

for achieving this modulation (*11*).

Identifying microorganisms at the species level offers precise insights that can aid clinicians in designing targeted treatments to promote gut health and mitigate disease risk. For instance, within the genus *Bifidobacterium*, *Bifidobacterium lactis* has been shown to reduce the risk of diarrhea and fever in children and infants (*11*), whereas *Bifidobacterium bifidum* is known to enhance the immune system and combat pathogens (*12*). Similarly, *Faecalibacterium prausnitzii*, a member of the Firmicutes phylum, is positively correlated with gut health and plays a role in reducing inflammation and colorectal cancer risk. However, health effects are not uniform across all species within the Firmicutes phylum (*13,14*). Current methods for microbiota analysis primarily include 16S rDNA amplicon sequencing and shotgun metagenomics. The 16S rDNA amplicon sequencing method is based on amplifying a specific region of the 16S rRNA gene, allowing the identification of distinct taxa through variations in the less-conserved regions (*15*). This approach is relatively cost-effective and straightforward. However, taxa assignments are based on a single genomic region, which can introduce amplification biases and affect taxonomic representation due to primer choice and amplification error (*16,17*). The 16S rDNA contains nine hypervariable regions (V1–V9) surrounded by conserved regions, with species-specific variants that enable community-level identification down to the genus level. Full-length 16S rDNA sequencing facilitates species-level identification(*18,19*). In contrast, shotgun metagenomics sequences the entire microbial community's DNA, necessitating greater sequencing depth, thus increasing costs, analytical complexity, and potential host DNA contamination (*20*). The downstream analysis in shotgun metagenomics relies on reference databases for genome assembly, which can result in false positives (*21*). Despite these limitations, shotgun metagenomics provides comprehensive microbial genomic information, including gene function analysis and insights into other microbiome components, such as fungi and viruses. In clinical settings, balancing accuracy, cost, and processing time is vital for achieving species-level microbial profiling.

Oxford Nanopore Technologies (ONT) offers ultra-long nucleic acid sequencing, with read lengths exceeding 2 million base pairs, which enables full-length 16S rDNA sequencing can improves taxonomic resolution by providing a comprehensive sequence of informative sites. Additionally, ONT's devices offer real-time data acquisition, allowing for immediate insights during sequencing runs. This sequencing approach facilitates faster (sequencing time 1-2 hours) and more accurate microbial community analysis (*22-24*). For metagenomic approaches, studies have shown that ~7 Gb of paired-end sequencing data are necessary to achieve > 20X coverage for microbes at > 1% relative abundance, indicating that shallow shotgun metagenomics sequencing (SSMS) is viable for preliminary screening (*25,26*). Ion Torrent offers short-read sequencing platforms that leverage semiconductor-based technology to deliver high-throughput data with rapid turnaround times. Notably, the GeneStudio™ S5 System enhances efficiency with automated library preparation, enabling high-throughput sequencing data in approximately 2 to 4 hours while maintaining ease of use (*27*). These platforms are cost-effective, making them ideal for time-sensitive applications such as SSMS.

However, comparative studies between SSMS and full-length 16S rDNA sequencing remain limited. Most prior research has focused on comparing hypervariable regions of the 16S rDNA gene, such as V4 or V3-V4, using short-read sequencing. These studies have reported biases in taxonomic detection due to the targeted nature of hypervariable region sequencing, which, despite being cost-effective, can lead to incomplete microbial profiles (*15,28*). While some studies have explored full-length 16S rDNA sequencing, they have primarily compared it with deep shotgun metagenomics rather than shallow shotgun sequencing. These studies indicate that full-length 16S rDNA sequencing provides a more comprehensive representation of dominant microorganisms and offers enhanced taxonomic resolution for low-abundance taxa in food-related matrices (*29*).

This research gap underscores the need for a direct comparison between SSMS and full-length 16S rDNA sequencing, particularly in terms of cost, time efficiency, and taxonomic resolution across diverse microbial communities. To address this, our study employs ONT-based full-length 16S rDNA sequencing alongside SSMS using the Ion GeneStudio S5 System to analyze the gut microbiota of healthy individuals. We compare alpha and beta diversity, identify taxa unique to each method, and discuss the broader implications of these sequencing strategies for gut microbiome research and future applications.

2. Methods

2.1. Sample collection and DNA extraction

In this study, forty-three stool samples were collected from consenting participants using collection tubes that incorporated DNA/RNA Shield (Zymo Research, USA) to preserve microbial specimens. These samples were subsequently stored at -20 °C until DNA extraction was performed. For DNA extraction, each fecal sample was thawed on ice, and 20 mg of material was processed using the ZymoBIOMICS DNA Miniprep kit (Zymo Research, USA) according to the manufacturer's protocol. The extracted DNA was preserved at -20 °C until further processing for sequencing.

2.2. Library construction and sequencing

2.2.1. Full length 16S rDNA nanopore sequencing

The full-length of bacterial 16S rDNA was polymerase chain reaction (PCR) amplified for 20 cycles using primers for targeting regions V1-V9 of the 16S rDNA. Primers were described Forword: 5'-TTTCTGTTGGT GCTGATATTGCAGRGTTYGATYMTGGCTCAG-3' and Reverse: 5'-ACTTGCCTG TCGCTCTATCTTCC GGYTACCTTGTTACGACTT-3' (*30*). The 20 µL PCR reaction contained 1 µg of DNA template, 0.2 µM of each primer, 0.2 mM of dNTPs, and 0.4 U of Phusion DNA Polymerase (Thermo Scientific, USA). The 1 µg of DNA template were used in the total volume (20 µL) of PCR reaction. The barcode sequences were added to PCR products using the PCR Barcoding Expansion Kit (Oxford Nanopore Technologies, UK). The products were checked by 1% agarose gel electrophoresis and purified using the QIAquick® PCR Purification Kit (QIAGEN, Germany). The samples were pooled at equal concentration and purified using AMPure XP beads (Beckman Coulter, USA). The final products of full-length V1-V9 region of 16S rDNA, were sequenced using Ligation Sequencing Kit (Oxford Nanopore Technologies, UK) and flow cell version R10.4 (Oxford Nanopore Technologies, UK).

2.2.2. Shallow shotgun metagenomic sequencing (SSMS)

Library preparation was performed using the Ion Xpress™ Fragment Library Kit (Thermo Fisher Scientific) with 100 ng of DNA as input. Adapter ligation, size selection, nick repair, and amplification followed the manufacturer's protocol. Sequencing was conducted on the Ion GeneStudio S5 System (Thermo Fisher Scientific, USA).

2.3. Bioinformatics analysis

2.3.1. 16S rDNA nanopore sequencing (16S rDNA)

The FAST5 files were base called by using Guppy basecaller software v6.0.7 (*31*) (Oxford Nanopore Technologies, UK) with a super-accuracy model to generate pass reads (FASTQ format) with a minimum acceptable quality score (Q > 10). The quality of reads was examined by MinIONQC (*32*). Then, FASTQ sequences were demultiplexed and adaptor-trimmed using Porechop v0.2.4 (*https://github.com/rrwick/Porechop*). The filtered reads were then clustered, polished, and taxonomically classified by NanoCLUST (*33*) based on the size sequences for the V1-V9 region of 16S rDNA sequences from Ribosomal Database Project (RDP) database (*34*). The abundance taxonomic assignment data were converted into QIIME2 software v2021.2 (*35*) data format for illustrating the richness and evenness of bacterial species based on their taxa abundances.

2.3.2. Shallow shotgun metagenomic sequencing (SSMS)

The taxonomic classification and abundance estimation of the shallow shotgun metagenomic sequencing data obtained from the gut microbiome samples were performed using Kraken2 (*36*) and Bracken2 (*37*), respectively. The raw reads were aligned against the PlusPF database (version 9/19/2020 available at *https://benlangmead.github.io/aws-indexes/k2*), which includes both the NCBI and RefSeq microbial genomes and has been demonstrated to have higher accuracy than other databases. Bracken2 was then used to estimate the taxonomic abundances at different levels of the classification hierarchy by adjusting the classification counts based on the distribution of read lengths. The resulting output was a table of taxonomic abundances at various levels of the classification hierarchy, which provided insight into the composition of the gut microbiome. An overview method used in this study is illustrated in Figure 1.

2.4. Statistical analysis

Statistical analyses were conducted on both 16S rDNA and SSMS datasets. The data for bacteria with a relative abundance greater than 1% were visualized using

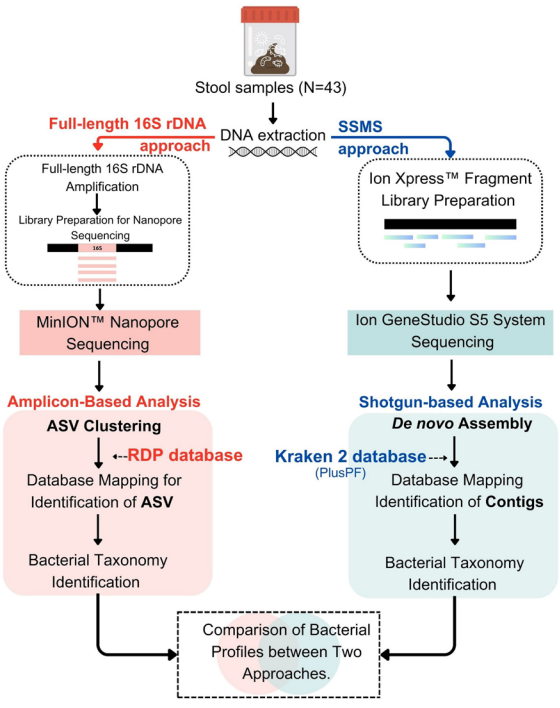

Figure 1. Overview of analytical plan for bacterial taxonomy identification from stool samples using Full-Length 16S rDNA Amplicon Sequencing (Full-length 16S rDNA) and Shallow Shotgun Metagenome sequencing (SSMS).

threshold cut-off values (*38*). Alpha diversity measures, including Observed Species and Chao1, were utilized to assess species richness, while the Shannon and Simpson indices were employed to evaluate both richness and evenness. Each alpha diversity measure was calculated using the R software (version 3.5.0) with the vegan package, aiming to examine microbiota diversity across the datasets. The relative abundance was compared between the two approaches using the Wilcoxon signed-rank test, which was performed in Python using the Pandas and SciPy libraries. Beta-diversity was analyzed using PERMANOVA tests based on Bray-Curtis and Jaccard distances, as implemented in the MicrobiomeAnalyst tools (*39*). Statistical significance was attributed to *P*-values less than 0.01 for ensuring robust statistical interpretation.

2.5. Data availability

The raw sequence reads generated during this study have been submitted to the NCBI Sequence Read Archive database under the BioProject accession number PRJNA1089554. The raw reads for full-length 16S rDNA sequencing and SSMS are available under BioSample accessions SAMN40544624 and SAMN40544783, respectively.

2.6. Ethics statement

The experiments were conducted after obtaining the approval of Ethical Committee of the Khon Kaen University Ethics Committee for Human Research on HE681056. This Research was conducted in accordance with the Declaration of Helsinki.

3. Results

3.1. Sequencing data

In this microbiome sequencing study, we compared two different sequencing methods: 16S rDNA full-length by Oxford Nanopore Technology sequencing (16S rDNA) and shallow shotgun metagenomic sequencing by Ion Torrent System (SSMS). The dataset comprised 43 samples, yielding a range of 3,622 to 89,831 raw reads for the 16s rDNA (mean: 11,677 ± 2,38) and 1,590,861 to 3,200,974 raw reads for the SSMS 1,590,861 to 3,200,974 (mean: 2,449,982 ± 44,489). The mapped reads for 16S rDNA ranged from 1,453 to 62,973 with an average of 8,489 ± 1,705. For SSMS, the mapped reads from 501,452 to 1,846,203, with a mean of 1,167,404 ± 52,511. Percentages of mapped reads were 46.96% (95% CI: 4.51-19.02%) and 71.91% (95% CI: 40.12-87.87%) for 16S rDNA full-length and shallow shotgun metagenomic sequencing, respectively (Table 1). Our findings provide important insights into the performance of these two sequencing methods and their potential application in microbiome studies.

3.2. Diversity of bacterial composition between shallow shotgun metagenomic sequencing and full-length 16S rDNA amplicon sequencing

To investigate gut microbial patterns associated with technical methods, we compared available microbiome data generated by two different approaches (shallow shotgun metagenomic sequencing (SSMS) and full-length 16S rDNA sequencing (full-length 16S rDNA)). Initial analysis without data cut-off parameter revealed a core microbiome of 7 phyla and 81 bacterial species common to both protocols. However, SSMS demonstrated greater sensitivity, identifying an additional 31 phyla, 1,235 genera, and 2,613 bacterial species. The full-length 16S rDNA approach also detected unique microbes, with 181 species belonging to 109 genera not found in the SSMS dataset. Applying a 1% abundance threshold narrowed the focus to a diverse bacterial community composed of 7 distinct phyla, 83 genera, and 205 species. There was significant overlap between the methods 47 genera (37.93%) and 113 species (54.00%) detected by both approach), while each method also demonstrated unique detection capabilities (13 genera (10.13%) and 38 species (16.50%) unique to SSMS, 23 genera (51.90%) and 54 species (29.50%) unique to full-length 16S rDNA sequencing (Figure 2). Furthermore, we observed discrepancies in bacterial nomenclature

Table 1. Sequencing statistic

Types	Statistic value	Sequencing approach	
		full-length 16S rDNA	SSMS
No. of Reads	Minimum	3,622	6,363,443
	Maximum	89,831	12,803,895
	Mean±Std.	11,677 ± 2,387	9,799,926 ± 177,956
	95% CI of median (0.9685)	6,737 - 9,001	9,151,579 - 10,276,515
No. of Mapped Reads	Minimum	1,453	501,452
	Maximum	62,973	1,846,203
	Mean ± Std.	8,489 ± 1,705	1,167,404 ± 52,511
	95% CI of median (0.9685)	4,436	1,000,816
%mapped reads	Mean ± Std.	71.91 ± 1.562	11.99 ± 0.5205
	95% CI of median (0.9685)	40.12 - 87.87	4.51 - 19.02

across databases. For example, *Bacteroides vulgatus* was designated as *Phocaeicola vulgatus, Eubacterium eligens* as *Lachnospira eligens*, and *Clostridium bolteae* as *Enterocloster bolteae* (Supplemental Table S1, *https://www.biosciencetrends.com/supplementaldata/249*). After consolidating taxa names, our analysis identified 200 bacterial species with 79 genera. The full-length 16S rDNA sequencing method detected a higher number of species (161 species) compared to SSMS (96 species) (Figure 2) and will therefore be used for further analysis.

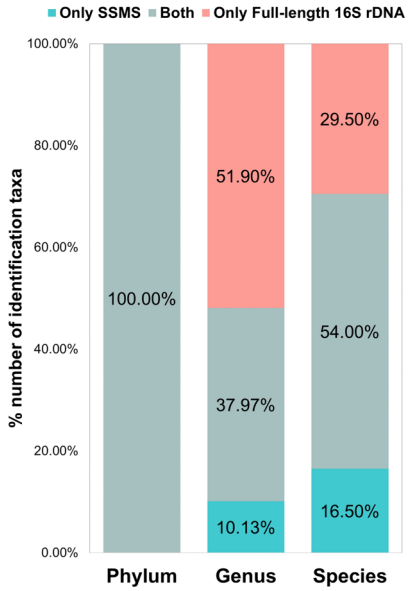

Figure 2. Bacterial taxonomic identification counts by sequencing approach (phyla, genera, and species).

Alpha diversity was quantified by observed richness (Figure 3A), Chao1 index (Figure 2B), Shannon's diversity (Figure 3C), and Simpson's diversity (Figure 3D), to evaluate bacterial richness and evenness across the two identification approaches. Analysis of full-length 16S rDNA sequencing revealed significantly higher bacterial diversity compared to the SSMS method, as demonstrated by all four diversity indices (Wilcoxon test, $p < 0.01$). Beta diversity analysis of the gut microbiome, assessed using Bray-Curtis and Jaccard dissimilarity indices, revealed significant separation between datasets generated by the two sequencing approaches ($p < 0.001$, Figure 3E and 3F).

3.3. Relative abundance and core species of gut microbiome from shallow shotgun metagenomic sequencing and full-Length 16S rDNA amplicon sequencing

At the phylum level, Bacteroidetes predominated in the SSMS method, accounting for 57.60% of the total abundance, whereas it was the second most abundant phylum in the full-length 16S rDNA sequencing method, constituting 30.93%. Conversely, Firmicutes was the most abundant phylum detected by the full-length 16S rDNA method, representing 57.40% of the observed microbiota, and was observed as the second most abundant phylum in the SSMS method, with a relative abundance of 28.93%. Additionally, Proteobacteria exhibited a higher prevalence in the

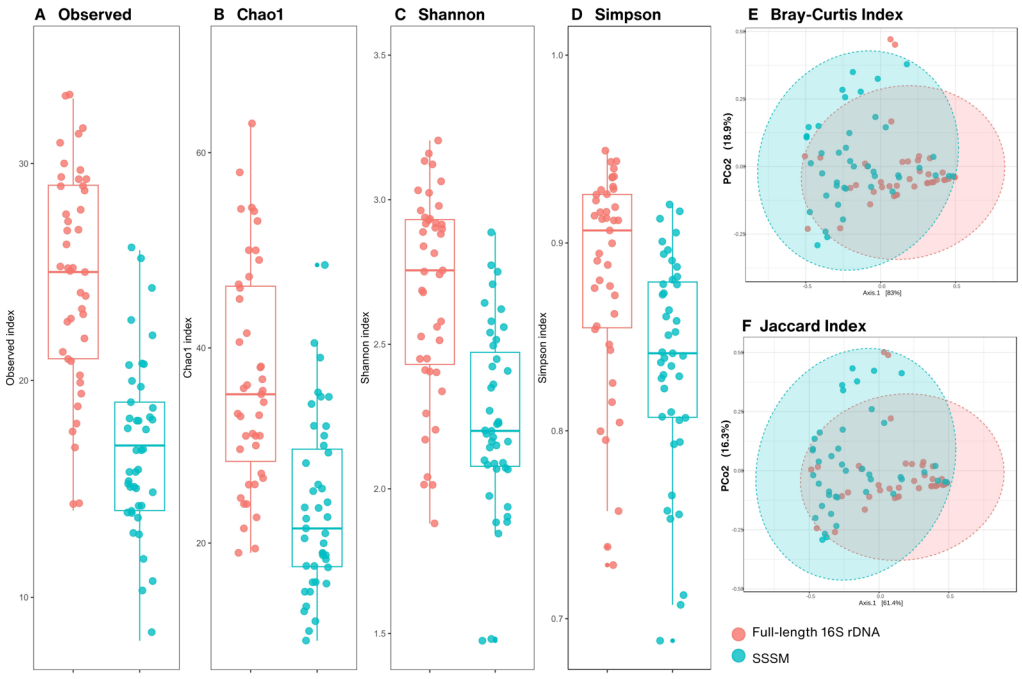

Figure 3. Comparison of gut microbiome diversity measures between sequencing approaches. Alpha diversity is represented by Observed species **(A)**, Chao1 **(B)**, Shannon index **(C)**, and Simpson index **(D)**, with significant differences determined by the Wilcoxon rank-sum test ($p < 0.001$). Beta diversity is visualized using Principal Coordinate Analysis (PCoA) with Bray-Curtis **(E)** and Jaccard **(F)** dissimilarity indices, with statistical significance determined by the PERMANOVA test ($p < 0.001$).

SSMS dataset, with a relative abundance of 7.17%, compared to 5.62% in the full-length 16S rDNA dataset (Supplemental Table S2, *https://www.biosciencetrends.com/supplementaldata/249*). The phyla Actinobacteria, Fusobacteria, Lentisphaerae, and Verrucomicrobia displayed low abundance in both methodologies (Figure 4A). However, the statistical analysis using the Wilcoxon signed-rank test showed that the total abundance at the phylum level is not significantly different between the two approaches ($p = 0.974$).

At the genus level, *Bacteroides* emerged as the most abundant genus within both datasets, although its presence was significantly greater in the SSMS dataset (47.18%) compared to the full-length 16S rDNA method (Figure 4C). Conversely, *Faecalibacterium* ranked as the second most abundant genus in the SSMS dataset (10.10%) but demonstrated markedly lower abundance in the full-length 16S rDNA dataset (1.11%). In contrast, *Prevotella* exhibited a high relative abundance of 8.36% in the full-length 16S rDNA dataset, significantly exceeding its presence in the SSMS dataset (1.83%). Other genera, including *Alistipes*, *Escherichia*, *Parabacteroides,* and *Akkermansia*, also showed higher relative abundances in the full-length 16S rDNA dataset compared to SSMS. Meanwhile, *Eubacterium*, *Roseburia*, *Bifidobacterium*, *Prevotella*, *Oscillibacter, Clostridium*, *Blautia*, and *Ruminococcus* were more prevalent in the full-length 16S dataset than in SSMS. Despite these variances at the genus level, there was no significant difference in the overall abundance of bacterial communities between the two methods ($P = 0.443$), as shown in Figure 4B and

Figure 4. The relative abundance of gut microbiota between sequencing approaches is shown across different taxonomic levels, including phylum (A), genus (B), and species (C). The percentages of relative abundance at each level are displayed for individual samples and group averages for both full-length 16S rRNA and SSMS sequencing approaches.

Supplemental Table S2 (*https://www.biosciencetrends.com/supplementaldata/249*).

Notably, at the species level, *Faecalibacterium prausnitzii* maintained a consistent dominance in both the SSMS (11.07%) and full-length 16S rDNA methodologies (9.94%). In stark contrast, *Bacteroides vulgatus* was significantly more dominant in the SSMS dataset, with 15.71%, compared to a considerably lower prevalence of 4.66% in the full-length 16S rDNA dataset. Additionally, a suite of species within the *Parabacteroides* and *Bacteroides* genera exhibited higher abundances solely in the full-length 16S rDNA dataset, including *Parabacteroides distasonis, Bacteroides distasonis, Bacteroides dorei, Bacteroides uniformis, Bacteroides fragilis, Bacteroides thetaiotaomicron, Bacteroides xylanisolvens, Bacteroides caccae, and Bacteroides ovatus. Conversely, Prevotella copri and Oscillibacter valericigenes* showed a notably higher prevalence in the SSMS dataset, with relative abundances of 6.90% and 5.16%, respectively. These data reveal disparities at the species level, indicating a statistically significant difference in the total relative abundance of species between the two datasets, with an extremely low *P*-value ($P = 2.27e{-}13$) as shown in Figure 4C and Supplemental Table S2 (*https://www.biosciencetrends.com/supplementaldata/249*).

3.4. Differential detection of microbes using SSMS and full-length 16S rDNA sequencing

The LEfSE analysis revealed marked significant differences ($p < 0.01$, LDA > 5, LDA < -5) across 18 bacterial species, as depicted in Figure 4A. The nine species—*Escherichia coli, Bacteroides ovatus, Bacteroides caccae, Parabacteroides distasonis, Bacteroides thetaiotaomicron, Bacteroides fragilis, Bacteroides dorei, Bacteroides uniformis,* and *Bacteroides vulgatus* exhibited elevated detection rates when analyzed using SSMS, as shown in Figure 5B. In contrast, an equivalent number of species, including *Oscillibacter valericigenes, Bacteroides plebeius, Lachnospira pectinoschiza, Blautia obeum, Gemmiger formicilis, Ruminococcus torques, Bacteroides massiliensis, Bacteroides stercoris,* and *Megamonas rupellensis*, demonstrated 50 to 0 percent relative abundance when sequenced using the full-length 16S rDNA approach but were not detectable *via* SSMS, as illustrated in Figure 5C. These findings underscore significant disparities in the detection of relative abundances of gut microbiota attributable to the two sequencing methodologies employed.

4. Discussion

To comprehensively evaluate the bacteria taxa detection from pair samples using two methods: 16S full-length rDNA sequencing from oxford nanopore technology with classification by the RDP database and shallow shotgun metagenomic sequencing by ion torrent with classification by Kraken2. Although the SSMS was applied in this study, the number of bacterial reads was identified as 1,167,404 reads (the average of 43 samples), consistent with previous reports that SSMS data were assigned accuracy taxonomic in species levels (*40*). Despite the similar time and cost requirements of the two approaches, there is a substantial difference in their data output, with one method yielding approximately 839 times more data than the other. This divergence is primarily attributed to the specific gene amplification and random sequencing of all nucleotides, which results in a low mapping ratio for the SSMS approach, recorded at only 11.91%. Consequently, the SSMS method required a significantly higher number of reads compared to the full-length 16S rDNA sequencing approach, which achieved a much higher mapping percentage of approximately 71.91%. These findings align with previous studies using V4 region 16S rDNA sequencing, which reported mapping percentages of 94.4% (*41*).

Numerous studies have compared V4 or V3-V4 16S rDNA sequencing with shotgun and shallow shotgun metagenomic sequencing, with most findings suggesting that SSMS is more effective for identifying bacterial species than V4 or V3-V4 16S rDNA sequencing (*25,28,42*). However, our results using full-length 16S rDNA sequencing indicate the opposite. Our findings show that 29.5% of the identified species were detected exclusively by the full-length 16S rDNA method, while only 16.5% were identified solely by the SSMS method. Although without applying a cutoff to exclude species with a relative abundance lower than 1%, SSMS identified a larger number of species (2,613 species). Many of these identifications were at very low abundance, suggesting that they may be artifact reads. After applying the 1% cutoff, only 94 species remained. This discrepancy may be due to the different methodological sensitivities and biases inherent in each approach. Full-length 16S rDNA sequencing provides more comprehensive coverage of the rRNA gene, which may lead to more accurate species identification, particularly for low-abundance or rare taxa. In contrast, SSMS, while effective at capturing a broad range of species, may include a higher number of false positives, especially when low-abundance thresholds are not applied. We observed significant differences in the gut microbial profiles between the two approaches, from alpha diversity (richness) to beta diversity, even though the Shannon index showed no significant difference. This suggests that the methods differ in their ability to capture species diversity and community composition. Despite these differences, both methods consistently identified Bacteroidetes and Firmicutes as the predominant phyla, which aligns with previous studies of the gut microbiome in healthy individuals (*43,44*).

The comparative analysis of shallow shotgun

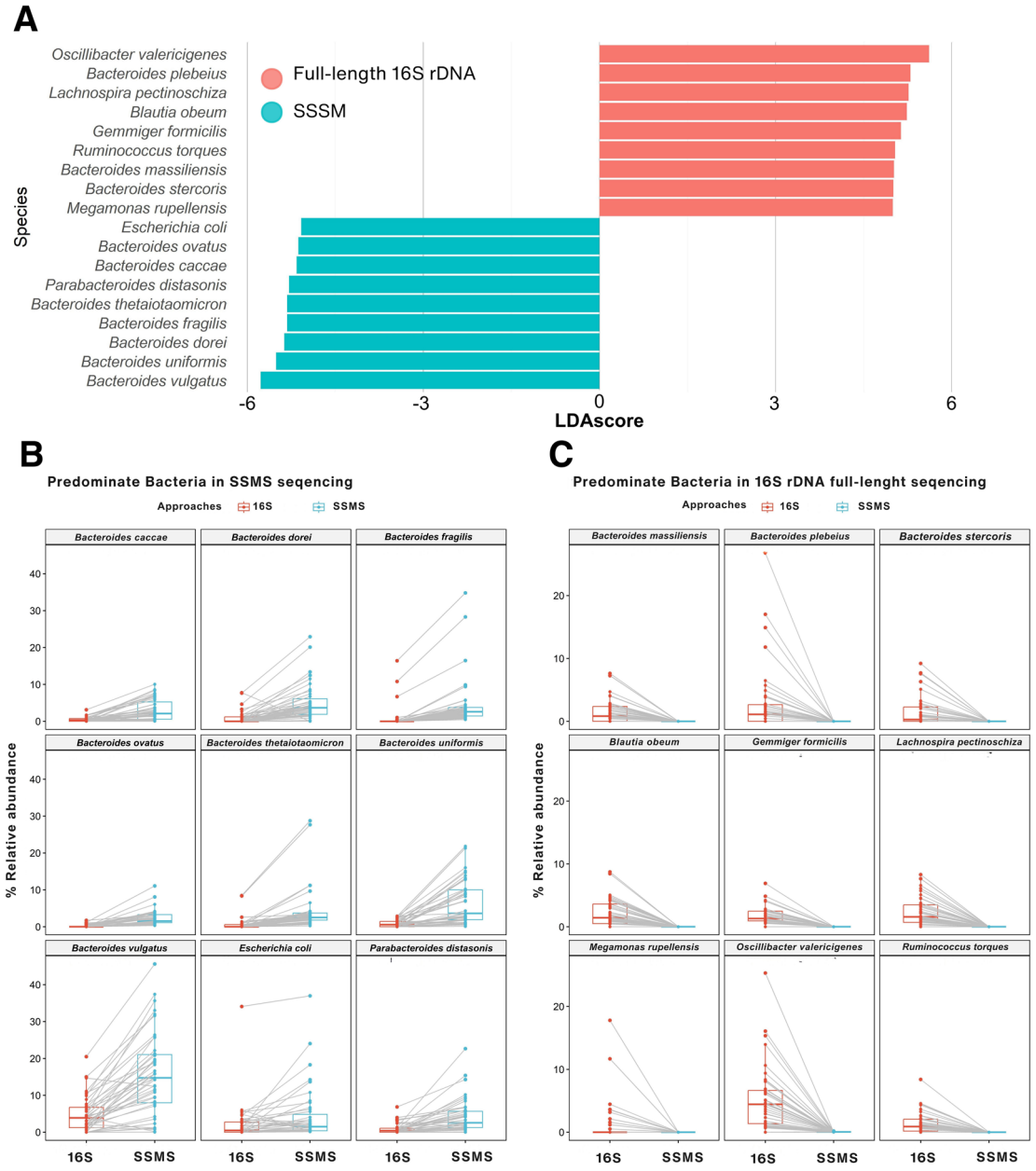

Figure 5. Identification of differentially abundant bacterial species. (A) Linear discriminant analysis (LDA) scores from LEfSe analysis reveal species with differential abundance ($p < 0.01$, LDA > 5 or LDA < -5). SSMS **(B)** and full-length 16S rDNA sequencing **(C)** each show species with significantly higher abundance as detected by their respective approaches, with significant differences determined by the Wilcoxon rank-sum test ($p < 0.001$).

metagenomic sequencing (SSMS) and full-length 16S rDNA sequencing highlights distinct discrepancies in the relative abundances of bacterial genera and species within the gut microbiome. Both methods efficiently capture major microbial groups; however, they demonstrate significant variation in detecting less abundant taxa. At the phylum level, Bacteroides and Faecalibacterium were more prevalent in the SSMS dataset, potentially reflecting the method's increased sensitivity to these groups due to broader genomic coverage and a more extensive database (*45*). This observation aligns with prior studies that indicate SSMS method efficacy in detecting a wide range of taxa, particularly those with greater genomic diversity.

At the species level, *Faecalibacterium prausnitzii* exhibited stable abundance across both methods, underscoring its role as a resilient and central component of the gut microbiome. *F. prausnitzii*, recognized for its high prevalence within the human gut, has been consistently linked to beneficial gut health effects, with decreased levels associated with inflammatory diseases such as Crohn's disease and ulcerative colitis (*13,46-48*). In contrast, notable differences in the abundance of *Bacteroides vulgatus* and other *Bacteroides* species between the SSMS and full-length 16S rDNA datasets suggest that SSMS may either overestimate or capture

strain-level variations not detected by full-length 16S rDNA sequencing (*49*). Additionally, the prominence of *Prevotella copri* and *Oscillibacter valericigenes* in the SSMS dataset suggests that SSMS may better capture specific species; however, this observation could be influenced by low-abundance artifacts(*50*).

The SSMS method identified *P. vulgatus*, a bacterium associated with gastrointestinal diseases such as inflammatory bowel disease (IBD), colorectal cancer, and obesity (*50*). Interestingly, *Prevotella copri*, frequently associated with both beneficial and detrimental health effects, was predominantly detected in the SSMS dataset, while *Oscillibacter valericigenes*, a challenging bacterium to culture linked to bacteremia, showed low abundance in SSMS, suggesting a potential advantage of the full-length 16S rDNA method for profiling low-abundance taxa (*51*). Moreover, *Lachnospira eligens*, also referred to by its basionym *Eubacterium eligens*, was detected by SSMS, while *E. eligens* was primarily identified through full-length 16S rDNA sequencing, illustrating taxonomic discrepancies between the two methods due to database differences. The consistent detection of *Bacteroides dorei* by SSMS, known for promoting the proliferation of gut probiotics, highlights SSMS's potential utility in identifying functionally significant species (*52*).

The taxonomic naming discrepancies observed between databases underscore the critical role of database choice in microbiome research. The RDP database, commonly used for full-length 16S rDNA classification, contrasts with Kraken2, which efficiently processes large datasets from high-throughput sequencing platforms like Illumina and Ion Torrent Torrent (*53,54*). Previous studies affirm that database selection significantly affects the detection and classification of microbiota, further complicating comparisons across sequencing techniques (*45*).

Overall, both full-length 16S rDNA sequencing and shallow shotgun metagenomic sequencing (SSMS) demonstrated time- and cost-efficiency, making them suitable for clinical applications. However, method and database selection significantly impact the detection of low-abundance gut microbiome species, emphasizing the need for careful evaluation. The findings highlight the need for critical evaluation of these methodologies, as each offers unique benefits and limitations regarding microbial diversity and relative abundance resolution. A major strength of this study is the first comparative analysis of full-length 16S rDNA sequencing and SSMS within the same sample set, minimizing inter-sample variability while providing a cost-effective, species-level microbiome characterization. 16S rDNA sequencing offers higher taxonomic resolution, particularly for dominant bacterial species, whereas SSMS captures some broader genomic insights detection the functional genes as antibiotic resistance and virulence factors, making it valuable for infection control. However, SSMS requires higher data and cost compared to 16S rDNA sequencing, which is approximately two times more cost-effective, making it more practical for routine clinical microbiome profiling. Despite these advantages, certain limitations must be acknowledged. The small sample size (n = 43) may impact generalizability, and database-dependent taxonomic biases could influence microbial classification. The observed method-dependent differences suggest that an integrative approach combining SSMS and full-length 16S rDNA sequencing could provide a more comprehensive microbiome profile. To advance microbiome research, standardized classification pipelines are needed to reduce inter-study variability. Expanding sample sizes and diversifying study populations will enhance the robustness and clinical relevance of findings. This approach will enhance considerations for selecting gut microbiome detection methods, facilitating its integration into clinical diagnostics.

5. Conclusion

The comparative study of SSMS and full-length 16S rDNA sequencing highlights the impact of sequencing method and database choice on gut microbiome analysis. Despite comparable time and cost requirements, SSMS yielded significantly more data, primarily due to its broad genomic coverage. However, the full-length 16S rDNA approach offered higher mapping accuracy and identified unique bacterial taxa, particularly at low abundances. Differences in taxonomic classification between RDP and Kraken2 further emphasize the influence of database selection on identification accuracy. Notably, *Bacteroides vulgatus*, *Prevotella copri* and *Oscillibacter valericigenes* exhibited method-dependent detection patterns, underscoring the critical role of methodological choice in microbial analysis. Given these differences, integrating SSMS and full-length 16S rDNA sequencing may provide a more comprehensive relevant representation of gut microbiota. To advance microbiome research and its clinical applications, the development of standardized classification pipelines and expansion of study cohorts with diverse populations are essential. These efforts will enhance the accuracy, consistency, and clinical relevance of microbial community profiling, ultimately deepening our understanding of the gut microbiome's role in health and disease.

Funding: This study was supported by a grant from the by the Fundamental Fund which has received funding support from the National Science Research and Innovation Fund, Thailand (project number 67A103000113).

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

References

1. Spyrou N, Vallianou N, Kadillari J, Dalamaga M. The interplay of obesity, gut microbiome and diet in the immune check point inhibitors therapy era. Seminars in Cancer Biology. 2021; 73:356-376.
2. Hu X, Ouyang S, Xie Y, Gong Z, Du J. Characterizing the gut microbiota in patients with chronic kidney disease. Postgraduate Medicine. 2020; 132:495-505.
3. Chen F, Dai X, Zhou CC, Li KX, Zhang YJ, Lou XY, Zhu YM, Sun YL, Peng BX, Cui W. Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. Gut. 2022; 71:1315-1325.
4. Onywera H, Meiring TL. Comparative analyses of Ion Torrent V4 and Illumina V3-V4 16S rRNA gene metabarcoding methods for characterization of cervical microbiota: taxonomic and functional profiling. Scientific African. 2020; 7:e00278.
5. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. Bmj. 2018; 361.
6. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell research. 2020; 30:492-506.
7. Cani PD. Human gut microbiome: hopes, threats and promises. Gut. 2018; 67:1716-1725.
8. Sadeghpour Heravi F. Gut Microbiota and Autoimmune Diseases: Mechanisms, Treatment, Challenges, and Future Recommendations. Current Clinical Microbiology Reports. 2024; 11:18-33.
9. Wang X, Yuan W, Yang C, Wang Z, Zhang J, Xu D, Sun X, Sun W. Emerging role of gut microbiota in autoimmune diseases. Frontiers in Immunology. 2024; 15.
10. Miyauchi E, Shimokawa C, Steimle A, Desai MS, Ohno H. The impact of the gut microbiome on extra-intestinal autoimmune diseases. Nature Reviews Immunology. 2023; 23:9-23.
11. Shaheen WA, Quraishi MN, Iqbal TH. Gut microbiome and autoimmune disorders. Clinical and Experimental Immunology. 2022; 209:161-174.
12. Turroni F, Taverniti V, Ruas-Madiedo P, Duranti S, Guglielmetti S, Lugli GA, Gioiosa L, Palanza P, Margolles A, Sinderen Dv, Ventura M. Bifidobacterium bifidum PRL2010 Modulates the Host Innate Immune Response. Applied and Environmental Microbiology. 2014; 80:730-740.
13. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. The ISME Journal. 2017; 11:841-852.
14. Huang Y, Wang Y, Huang X, Yu X. Unveiling the overlooked fungi: the vital of gut fungi in inflammatory bowel disease and colorectal cancer. Gut Pathogens. 2024; 16:59.
15. Bars-Cortina D, Ramon E, Rius-Sansalvador B, Guinó E, Garcia-Serrano A, Mach N, Khannous-Lleiffe O, Saus E, Gabaldón T, Ibáñez-Sanz G. Comparison between 16S rRNA and shotgun sequencing in colorectal cancer, advanced colorectal lesions, and healthy human gut microbiota. BMC genomics. 2024; 25:730.
16. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, Segovia BT, Lansac-Toha FA, Lemke M, DeSalle R. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Scientific reports. 2017; 7:6589.
17. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, List M, Neuhaus K. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. Msphere. 2021; 6:10.1128/msphere. 01202-01220.
18. Olivier SA, Bull MK, Strube ML, Murphy R, Ross T, Bowman JP, Chapman B. Long-read MinION™ sequencing of 16S and 16S-ITS-23S rRNA genes provides species-level resolution of Lactobacillaceae in mixed communities. Frontiers in Microbiology. 2023; 14:1290756.
19. Qing W, Shi Y, Chen R, Zou Ya, Qi C, Zhang Y, Zhou Z, Li S, Hou Y, Zhou H. Species-level resolution for the vaginal microbiota with short amplicons. Msystems. 2024; 9:e01039-01023.
20. Zhou Y, Liu M, Yang J. Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: methods, applications, challenges, and opportunities. Microbiological research. 2022; 260:127023.
21. Garrido-Sanz L, Àngel Senar M, Piñol J. Drastic reduction of false positive species in samples of insects by intersecting the default output of two popular metagenomic classifiers. PLoS One. 2022; 17:e0275790.
22. Dommann J, Kerbl-Knapp J, Albertos Torres D, Egli A, Keiser J, Schneeberger PH. A novel barcoded nanopore sequencing workflow of high-quality, full-length bacterial 16S amplicons for taxonomic annotation of bacterial isolates and complex microbial communities. Msystems. 2024; e00859-00824.
23. Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, Wu Q, Graeber E, Finzer P, Mendling W. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. Nature methods. 2022; 19:845-853.
24. Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, Chen W. The newest Oxford Nanopore R10. 4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. Applied and Environmental Microbiology. 2023; 89:e00605-00623.
25. La Reau AJ, Strom NB, Filvaroff E, Mavrommatis K, Ward TL, Knights D. Shallow shotgun sequencing reduces technical variation in microbiome analysis. Scientific Reports. 2023; 13:7668.
26. Hillmann B, Al-Ghalith Gabriel A, Shields-Cutler Robin R, Zhu Q, Gohl Daryl M, Beckman Kenneth B, Knight R, Knights D. Evaluating the Information Content of Shallow Shotgun Metagenomics. mSystems. 2018; 3:10.1128/msystems.00069-00018.
27. Marine RL, Magaña LC, Castro CJ, Zhao K, Montmayeur AM, Schmidt A, Diez-Valcarce M, Ng TFF, Vinjé J, Burns CC. Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses. Journal of virological methods. 2020; 280:113865.
28. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Scientific Reports. 2021; 11:3030.
29. Rubiola S, Macori G, Civera T, Fanning S, Mitchell M, Chiesa F. Comparison Between Full-Length 16S rRNA Metabarcoding and Whole Metagenome Sequencing Suggests the Use of Either Is Suitable for Large-Scale

Microbiome Studies. Foodborne Pathogens and Disease. 2022; 19:495-504.
30. Jitvaropas R, Mayuramart O, Sawaswong V, Kaewsapsak P, Payungporn S. Classification of salivary bacteriome in asymptomatic COVID-19 cases based on long-read nanopore sequencing. Exp Biol Med (Maywood). 2022; 247:1937-1946.
31. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biology. 2019; 20:129.
32. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics. 2018; 35:523-525.
33. Rodríguez-Pérez H, Ciuffreda L, Flores C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. Bioinformatics. 2020; 37:1600-1601.
34. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Research. 2003; 31:442-443.
35. Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology. 2019; 37:852-857.
36. Garcia BJ, Simha R, Garvin M, Furches A, Jones P, Gazolla JGFM, Hyatt PD, Schadt CW, Pelletier D, Jacobson D. A k-mer based approach for classifying viruses without taxonomy identifies viral associations in human autism and plant microbiomes. Computational and Structural Biotechnology Journal. 2021; 19:5911-5919.
37. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science. 2017; 3:e104.
38. de Cena JA, Zhang J, Deng D, Damé-Teixeira N, Do T. Low-Abundant Microorganisms: The Human Microbiome's Dark Matter, a Scoping Review. Front Cell Infect Microbiol. 2021; 11:689197.
39. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nature Protocols. 2020; 15:799-821.
40. Xu W, Chen T, Pei Y, Guo H, Li Z, Yang Y, Zhang F, Yu J, Li X, Yang Y. Characterization of shallow whole-metagenome shotgun sequencing as a high-accuracy and low-cost method by complicated mock microbiomes. Frontiers in Microbiology. 2021; 12:678319.
41. Usyk M, Peters BA, Karthikeyan S, *et al*. Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. Cell Reports Methods. 2023; 3.
42. Lamoureux C, Surgers L, Fihman V, Gricourt G, Demontant V, Trawinski E, N'debi M, Gomart C, Royer G, Launay N. Prospective comparison between shotgun metagenomics and sanger sequencing of the 16S rRNA gene for the etiological diagnosis of infections. Frontiers in Microbiology. 2022; 13:761873.
43. Vaiserman A, Romanenko M, Piven L, Moseiko V, Lushchak O, Kryzhanovska N, Guryanov V, Koliada A. Differences in the gut Firmicutes to Bacteroidetes ratio across age groups in healthy Ukrainian population. BMC microbiology. 2020; 20:1-8.
44. Singh P, Rawat A, Alwakeel M, Sharif E, Al Khodor S. The potential role of vitamin D supplementation as a gut microbiota modifier in healthy individuals. Scientific reports. 2020; 10:1-14.
45. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, Jones CMA, Wright RJ, Dhanani AS, Comeau AM, Langille MGI. Microbiome differential abundance methods produce different results across 38 datasets. Nature Communications. 2022; 13:342.
46. Maioli TU, Borras-Nogues E, Torres L, Barbosa SC, Martins VD, Langella P, Azevedo VA, Chatel J-M. Possible Benefits of Faecalibacterium prausnitzii for Obesity-Associated Gut Disorders. Frontiers in Pharmacology. 2021; 12.
47. Fujimoto T, Imaeda H, Takahashi K, Kasumi E, Bamba S, Fujiyama Y, Andoh A. Decreased abundance of F aecalibacterium prausnitzii in the gut microbiota of C rohn's disease. Journal of gastroenterology and hepatology. 2013; 28:613-619.
48. Lopez-Siles M, Enrich-Capó N, Aldeguer X, Sabat-Mir M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. Alterations in the abundance and co-occurrence of Akkermansia muciniphila and Faecalibacterium prausnitzii in the colonic mucosa of inflammatory bowel disease subjects. Frontiers in cellular and infection microbiology. 2018; 8:281.
49. Bank NC, Singh V, McCourt B, Burberry A, Roberts KD, Grubb B, Rodriguez-Palacios A. Antigenic operon fragmentation and diversification mechanism in Bacteroidota impacts gut metagenomics and pathobionts in Crohn's disease microlesions. Gut Microbes. 2024; 16:2350150.
50. Verbrugghe P, Van Aken O, Hållenius F, Nilsson A. Development of a real-time quantitative PCR method for detection and quantification of Prevotella copri. BMC Microbiology. 2021; 21:23.
51. Sydenham TV, Arpi M, Klein K, Justesen US. Four Cases of Bacteremia Caused by Oscillibacter ruminantium, a Newly Described Species. Journal of Clinical Microbiology. 2014; 52:1304-1307.
52. Tufail MA, Schmitz RA. Exploring the Probiotic Potential of Bacteroides spp. Within One Health Paradigm. Probiotics and Antimicrobial Proteins. 2024.
53. Lu J, Salzberg SL. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. Microbiome. 2020; 8:1-11.
54. Terrón-Camero LC, Gordillo-González F, Salas-Espejo E, Andrés-León E. Comparison of Metagenomics and Metatranscriptomics Tools: A Guide to Making the Right Choice. Genes. 2022; 13:2280.

Received December 11, 2024; Revised March 24, 2025; Accepted March 28, 2025.

**Address correspondence to:*
Sunchai Payungporn, Center of Excellence in Systems Microbiology, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.
E-mail: sp.medbiochemcu@gmail.com

Released online in J-STAGE as advance publication April 4, 2025.