Jurnal Biologi Tropis

Original Research Paper

NGS-Based Metagenomic Approaches for Bacterial Community Analysis: A Review

Rifda Hayati*, Marlina, Anthoni Agustien
Biotechnology Program, Graduate School, Universitas Andalas, Padang, West Sumatra, Indonesia

Article History
Received : June 20th, 2026
Revised : July 19th, 2026
Accepted : July 28th, 2026

*Corresponding Author: **Rifda Hayati**, Biotechnology Program, Graduate School Universitas Andalas, Padang, West Sumatera, Indonesia; Email: rifdahayati742@gmail.com

Abstract: Microorganisms play a crucial role in biological, ecological, and biogeochemical processes; however, most environmental microorganisms cannot be readily grown under standard laboratory conditions, driving the development of Next-Generation Sequencing (NGS)-based metagenomic approaches that enable culture-independent characterization of microbial communities directly from environmental DNA. This review synthesizes the principles of NGS-based metagenomics, contrasts amplicon and shotgun approaches, compares major sequencing platforms and platform-selection criteria, and evaluates the applications and limitations of bacterial community analysis across the health, agricultural, and food sectors—an integration not commonly addressed together in previous reviews. Based on literature from 2018–2026 in PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar, amplicon and shotgun metagenomics were found to offer complementary taxonomic and functional resolution, while platform choice depends on accuracy, read length, and budget. Across sectors, metagenomics supports pathogen surveillance and gut microbiome profiling in health, soil microbiome-based sustainability indicators in agriculture, and fermentation microorganism and antimicrobial resistance gene detection in food safety. Nevertheless, challenges remain regarding the accuracy of function prediction from amplicon-based data, bioinformatics pipeline standardization, and the validation of predicted microbial functions through complementary multi-omics approaches.

Keywords: Amplicon Sequencing; Bacterial Community; Metagenomics; Next-Generation Sequencing; Shotgun Metagenomics.

Introduction

Microorganisms are an essential component of various ecosystems and play a key role in biological, ecological, and biogeochemical processes. Bacterial communities contribute to nutrient cycling, the decomposition of organic matter, the transformation of nutrients, and biological interactions with other organisms. Beyond these general functions, recent studies have demonstrated that the composition and structure of bacterial communities can serve as sensitive indicators of both environmental condition and host health status. In natural ecosystems, shifts in soil bacterial community composition and diversity have been closely associated with key ecosystem functions such as nutrient cycling, organic matter decomposition, and resilience to environmental disturbance (Yang et al., 2024). In clinical settings, researchers have pinpointed a core gut microbiome signature defined by consistent co-abundance patterns among bacterial taxa that reliably tracks host health status across a range of diseases (Wu et al., 2024). These findings suggest that community-level bacterial analysis, rather than the presence of any single taxon, provides a more robust framework for assessing both ecosystem and host functional conditions. Consequently, characterizing bacterial communities has become an essential undertaking across environmental microbiology, public health,

agricultural science, and biotechnology (Kim et al., 2024).

Common methods used to determine microbial community diversity include metagenomic analysis, molecular techniques, and culture-based techniques. However, a large proportion of environmental microorganisms remain difficult to culture under laboratory conditions, with estimates varying considerably depending on habitat type and detection method; some taxa also grow so slowly that cultivation may take months or even years (Vollmers et al., 2017; Schultz et al., 2023). Limitations in microbial cultivation can also be influenced by dependence on other organisms in nature. These limitations prevent information regarding microbial diversity from being fully revealed. This situation has driven the development of molecular approaches that allow for the identification of microorganisms without the need for cultivation, one of which is metagenomic analysis (Yang et al., 2021).

Advances in metagenomic approaches address these limitations through the analysis of total DNA extracted directly from the environment. This approach enables the identification of microorganisms without the need for isolation or culture. In addition to providing taxonomic information, metagenomics can also reveal the potential genetic functions of microorganisms within an ecosystem (Moges & Megistu, 2024). The principle of metagenomic analysis is based on DNA extracted directly from communities within an ecosystem, which is then analyzed and identified using phylogenetic marker genes. In particular, 16S rRNA gene sequencing is commonly used to profile bacterial and archaeal community composition, allowing researchers to identify taxa present and estimate their relative abundance within a sample (Zhang et al., 2023).

Advances in Next-Generation Sequencing (NGS) technology have brought about significant changes in metagenomic research. This technology enables sequencing processes that can generate large amounts of data in a relatively short time, making it known as a high-throughput sequencing platform (Kim et al., 2024). At the genomic level, NGS provides broad analytical capabilities such as gene annotation, genome mapping, and sequence-based identification of homologous genes and gene interaction networks (Purwoko et al., 2018). Within the specific context of metagenomics, however, NGS applications are more directly oriented toward the analysis of bacterial communities as a whole rather than individual genomes. These applications include taxonomic profiling to characterize community composition (Zhang et al., 2023), functional profiling to infer metabolic potential and biogeochemical roles, metagenome-assembled genome (MAG) reconstruction to recover genome-resolved information from uncultured taxa (Mirete et al., 2025), resistome analysis to detect antimicrobial resistance genes within microbial communities, and pathogen surveillance for the culture-independent detection of pathogenic microorganisms (Luo et al., 2025). Collectively, these metagenomic applications have been widely adopted across health, agriculture, environmental, and biotechnological sectors to characterize bacterial community structures, diversity, and functional potential in various ecosystems.

Nevertheless, the implementation of NGS in metagenomic studies still faces various challenges, particularly in terms of data processing and result interpretation. Moreover, although numerous reviews have addressed metagenomic sequencing in general, few have specifically integrated the fundamental principles of NGS with a systematic comparison of sequencing platforms and cross-sectoral applications in the context of bacterial community analysis. Therefore, this literature review aims to (1) explain the principles of NGS technology in metagenomic analysis, (2) compare the major sequencing platforms and the criteria for platform selection relevant to bacterial community analysis, (3) review the applications of NGS in microbial community studies across the health, agricultural, and food sectors, and (4) identify the current challenges and future prospects for its development.

Materials and Methods

Four major scientific databases were consulted for this review: PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar.

Records were retrieved using a Boolean search string combining terms for the technology ("Next Generation Sequencing" OR "NGS"), the analytical approach ("Metagenomics" OR "Metagenomic Analysis"), and the biological target ("Bacterial Communities" OR "Microbiome" OR "Microbial Diversity"), with results restricted to publications from 2018 through 2026 to keep the evidence base current.

Articles were included if they (1) were original research articles, reviews, or book chapters published in peer-reviewed sources; (2) discussed the application of NGS-based metagenomic approaches in the context of bacterial community analysis; and (3) were available in full text in English or Indonesian. Articles were excluded if they (1) were conference abstracts, editorials, or non-peer-reviewed preprints; (2) focused exclusively on non-bacterial taxa (e.g., viral or fungal metagenomics only); or (3) were duplicate records retrieved from multiple databases.

The screening process began with an initial relevance check of each record's title and abstract; records that passed this stage then underwent full-text evaluation against the eligibility criteria. The screening and selection process were conducted by the author(s) based on the eligibility criteria described above. Relevant information from each selected article—including publication year, sequencing platform, analytical approach, application domain, and key findings—was extracted and organized thematically according to the objectives of this review: (1) principles of NGS in metagenomic analysis, (2) sequencing platforms and analytical pipelines, (3) applications across health, agriculture, environmental, and biotechnological sectors, and (4) current challenges and future prospects. Given the heterogeneity of study designs and application areas among the included literature, the extracted data were synthesized narratively rather than through quantitative meta-analysis.

Results and Discussion

NGS-based metagenomic approaches to bacterial community analysis rely on a set of core principles, sequencing platforms, and cross-sectoral applications that have been extensively documented across the reviewed literature. A summary of key studies discussed throughout this review, including the sequencing platform, sample or ecosystem type, application focus, and key findings of each, is presented in Table 1.

Table 1. Summary of Key Studies on NGS-Based Metagenomic Approaches in Bacterial Community Analysis

Reference (year)	Approach & Sample/Focus	Key Findings Related to Bacterial Community Analysis
Yang B et al., 2024	Amplicon (16S rRNA); forest soil ecosystems	Variation in soil bacterial taxonomic profiles and the architecture of their co-occurrence networks tracks closely with underlying soil physicochemical conditions and overall ecosystem functioning
Wu G et al., 2024	Shotgun metagenomics; human gut microbiome, multi-disease cohorts	Identified a stable "core microbiome signature" (co-abundance guilds) that serves as a reliable indicator of host health status across 15 diseases
Schultz J et al., 2023	Review of cultivation-independent methods; extreme environments	Reports that 80–85% of microbial taxa in extreme environments remain uncultured, supporting molecular/metagenomic approaches over culture-based methods
Zhang W et al., 2023	Amplicon (16S rRNA) benchmarking; multiple habitats	Confirms 16S rRNA sequencing as an effective marker for profiling community composition, though resolution is generally limited to the genus level
Mirete S et al., 2025	Shotgun metagenomics; genome-resolved metagenomics review	Metagenome-assembled genomes (MAGs) enable genome-level characterization of uncultured taxa, revealing novel biogeochemical functions
Luo H et al., 2025	Review of sequencing-based surveillance; clinical/environmental pathogens	Highlights metagenomics and metatranscriptomics as key tools for culture-independent pathogen detection and resistome profiling

Reference (year)	Approach & Sample/Focus	Key Findings Related to Bacterial Community Analysis
Terrón-Camero LC et al., 2022	Comparative review; amplicon, shotgun, and metatranscriptomic tools	Provides a bioinformatics tool comparison guide for selecting appropriate metagenomic/metatranscriptomic pipelines based on research goals
Usyk M et al., 2023	Amplicon vs. shotgun comparison; human stool (n = 1,772)	Amplicon and shotgun sequencing yield comparable genus-level taxonomic accuracy; datasets can be harmonized for pooled analysis
Breitwieser FP et al., 2019	Methodological review; metagenomic classification and assembly	Clarifies terminology distinguishing metataxonomics (amplicon-based) from shotgun metagenomics, and reviews classification/assembly tools
Mishra AK et al., 2025	Amplicon (16S/ITS); organic vs. conventional soil management	Organically managed soils show higher bacterial and fungal diversity, underscoring management-driven shifts in community composition

Basic Principles of Metagenomics

Metagenomics is defined as the direct analysis of genetic material recovered from environmental or clinical specimens, bypassing the need to isolate and culture the microorganisms involved. For the purposes of this review, metagenomic studies are organized under a single overarching framework distinguished by the type of sequencing data they generate, namely amplicon metagenomics and shotgun metagenomics.

Table 2. Comparison of Metagenomic and Related Omics Approaches

Approach	Target Molecule	Method Basis	Key Advantages	Key Limitations	Typical Output
Amplicon metagenomics (metataxonomics)	DNA (marker genes: 16S rRNA for bacteria/archaea, 18S rRNA for eukaryotes, ITS for fungi; de Vries et al., 2023)	PCR amplification of specific marker genes followed by sequencing	Low cost, high sample throughput, and well-established reference databases (Liu et al., 2021; Kinoshita et al., 2021)	Subject to PCR amplification bias and limited taxonomic resolution, typically restricted to the genus level (Breitwieser et al., 2019)	Taxonomic composition and relative abundance
Shotgun metagenomics	Total DNA	Untargeted sequencing of all genomic DNA in a sample	Provides comprehensive taxonomic and functional information without PCR bias, and enables metagenome-assembled genome (MAG) reconstruction (Terrón-Camero et al., 2022)	Requires higher cost, greater sequencing depth, and more intensive computational resources (Usyk et al., 2023)	Taxonomic profiles, functional gene content, metagenome-assembled genomes
Functional metagenomics	Total DNA (shotgun-derived)	Screening of shotgun metagenomic data for novel genes	Enables discovery of novel enzymes and metabolic pathways (Zhang et al., 2021)	Resource-intensive and dependent on adequate sequencing	Novel gene or enzyme candidates, biosynthetic gene clusters

Approach	Target Molecule	Method Basis	Key Advantages	Key Limitations	Typical Output
		or bioactive compounds		depth and annotation database coverage	
Metatranscriptomics	Total RNA (mRNA)	Sequencing of expressed transcripts from a microbial community	Reflects actively expressed functions rather than genetic potential alone (Terrón-Camero et al., 2022)	Limited by RNA instability, careful sample handling requirements, and higher cost and complexity	Active functional and metabolic profile of the community

Amplicon metagenomics relies on the sequencing of specific marker genes that are first amplified via PCR, such as the 16S rRNA, 18S rRNA, 26S rRNA, or Internal Transcribed Spacer (ITS) genes; this targeted approach has made it a common choice for profiling which taxa are present in a microbial community and how diverse that community is (Liu et al., 2021; Kinoshita et al., 2021). This approach is also referred to as "metataxonomics" by some authors, a term used specifically to distinguish amplicon-based community profiling from shotgun-based analysis (Breitwieser et al., 2019); in this review, the two terms are treated as synonymous. Shotgun metagenomics takes a different route, sequencing all DNA present in a sample rather than targeting specific genes; this untargeted strategy yields a fuller picture that spans not just which taxa are present but also what functional capabilities those microorganisms carry (Terrón-Camero et al., 2022). Empirical comparisons have shown that amplicon and shotgun sequencing yield comparable taxonomic accuracy at the genus level, although shotgun sequencing offers additional resolution for functional gene content (Usyk et al., 2023).

Within the shotgun-based framework, several application-specific subcategories have also been proposed rather than representing entirely separate data types. The term "functional metagenomics" denotes a specific application of shotgun sequencing data aimed at uncovering novel genes and bioactive compounds, whereas "sequencing metagenomics" is applied more loosely to any general exploration of microbial community diversity, whether based on amplicon or shotgun data (Zhang et al., 2021). To avoid terminological ambiguity, this review adopts amplicon and shotgun metagenomics as the two primary categories, and treats metataxonomics, functional metagenomics, and sequencing metagenomics as descriptive or application-specific terms nested within this framework rather than as independent classifications. A comparison of these approaches, along with the related technique of metatranscriptomics, is presented in Table 2.

Broadly speaking, 16S rRNA amplicon sequencing serves to reveal which taxa make up a microbial community and how diverse that community is. Advances in bioinformatics have since extended this further, making it possible to infer the likely functional capabilities of those microorganisms from the taxonomic data that 16S rRNA sequencing produces. These predictions are made by linking phylogenetic information or the taxonomic composition of microbes to existing genome and functional databases. Among the tools built for this purpose, PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (Langille et al., 2013) draws on Greengenes-derived OTU tables to approximate a community's metagenomic function (McDonald et al., 2011; Zheng et al., 2019), mapping these onto metabolic pathways catalogued in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000). Its successor, PICRUSt2, extended this functionality to Amplicon Sequence Variant (ASV) data and drew on a considerably expanded, updated reference genome set, sharpening both the accuracy and versatility of its predictions relative to the original tool (Douglas et al., 2020). A separate R-based tool, Tax4Fun, offers a comparable function by inferring microbial community capabilities from the

SILVA reference database (Asshauer et al., 2015; Quast et al., 2013). For more specific analyses, FAPROTAX provides ecological and metabolic functional annotations for prokaryotic microorganisms, such as nitrate respiration, iron respiration, plant pathogens, and symbiosis in animals (Louca et al., 2016). Meanwhile, BugBase is used to predict microbial phenotypic characteristics, including oxygen tolerance, Gram-positive or Gram-negative properties, and pathogenic potential (Ward et al., 2017). These software tools have been widely utilized in environmental research (Louca et al., 2016), agriculture (Zhang et al., 2019), animal husbandry (Ross et al., 2018), and health (Mahnert et al., 2019) to expand the biological interpretation of amplicon sequencing data.

A key caveat is that these tools output inferred, not empirically measured, functional profiles: gene content is extrapolated from existing reference genome databases rather than derived from direct sequencing of the functional genes in question. How reliable these predictions are depends heavily on the environment being studied: accuracy tends to be highest for well-studied systems like the human microbiome, but drops considerably for non-human or environmental samples, largely because such taxa are still poorly represented in existing reference databases (Sun et al., 2020). Given these limitations, functional inferences drawn from 16S rRNA data warrant a cautious reading and, wherever practical, cross-checking against complementary methods-shotgun metagenomics, metatranscriptomics, or metabolomics-that measure actual gene content, expression, or metabolic activity instead of relying on inference alone.

Table 3. Comparative Technical and Analytical Characteristics of Amplicon and Shotgun Metagenomic Approaches

Characteristic	Amplicon (16S rRNA) Sequencing	Shallow Shotgun Metagenomics	Deep Shotgun Metagenomics
Typical read depth	A few thousand reads per sample (Zhang et al., 2023)	On the order of 500,000 reads for each sample (Hillmann et al., 2018)	Ranging from several million up to tens of millions of reads for each sample (Truong et al., 2015; Wood et al., 2014; Menzel et al., 2016)
Relative cost	Low	Moderate	High
Taxonomic resolution	Generally genus-level	Species-level	Species- to strain-level
Functional potential information	Not directly available (requires prediction tools, e.g., PICRUSt2)	Available, though with reduced depth/coverage	Comprehensive and directly measured
Computational requirements	Relatively low	Moderate	High (substantial storage and processing capacity)

Shotgun metagenomics works by deeply sequencing the total DNA recovered from a sample, which allows the various microbial genomes present within that community to be identified. This sets it apart from amplicon-based methods, which target only select marker genes: shotgun sequencing instead yields a fuller picture spanning both the taxonomic makeup and the genetic functional potential of the microbiota. However, the application of this method still faces several challenges, particularly the high cost of sequencing and the need for more complex technical procedures, both during sample preparation and bioinformatics analysis. While 16S rRNA-based analysis generally yields representative data with just a few thousand reads per sample, the shotgun metagenomics, by comparison, generally needs anywhere from several million to tens of millions of high-quality reads to yield optimal results (Truong et al., 2015; Wood et al., 2014; Menzel et al., 2016). Additionally, the processes of DNA fragmentation, library preparation, and data processing require greater laboratory and computational resources. To address cost constraints, the shallow shotgun metagenomics approach has been developed-a shotgun method with lower sequencing depth that remains capable of providing information on potential metagenomic functions and identifying

microorganisms down to the species level with approximately 500,000 reads per sample (Hillmann et al., 2018). Table 3 lays out a side-by-side comparison of the technical demands and analytical capabilities associated with amplicon sequencing, shallow shotgun metagenomics, and deep shotgun metagenomics.

Next-Generation Sequencing Platform

Unlike first-generation platforms such as Sanger sequencing, Next-Generation Sequencing (NGS) relies on massively parallel processing of DNA fragments, a high-throughput design that defines the technology as a whole (Satam et al., 2023). NGS itself encompasses several platform generations with distinct underlying principles: second-generation platforms, such as Illumina and Ion Torrent, generate large volumes of relatively short reads (approximately 36–400 base pairs) with high per-base accuracy, whereas third-generation, long-read platforms, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), are capable of generating reads that stretch across several kilobases, a length advantage that makes them better equipped to resolve complex or repetitive genomic stretches, even though this has historically come at the expense of a higher raw error rate (Satam et al., 2023). Rather than sharing a single uniform capability, these platforms therefore differ considerably in sequencing principle, error profile, read length, and output, and are often selected based on the specific analytical goals of a given study. Depending on the research objective, NGS can be applied to sequence an organism's complete genome or to target a narrower set of genetic regions-ranging from the full complement of roughly 22,000 protein-coding genes captured by whole exome sequencing, down to just a handful of genes relevant to a particular study (Behjati et al., 2013).

The emergence of Next-Generation Sequencing (NGS) has reshaped molecular biology and genomics as fields of study, largely because it allows researchers to generate vast volumes of sequence data faster, more accurately, and at a lower per-base cost than older methods like Sanger sequencing. It should be noted, however, that the overall cost of an NGS-based project extends beyond per-base sequencing cost alone, as it also includes expenses related to library preparation, sample handling, and the computational infrastructure required for downstream bioinformatics analysis, which can be substantial depending on sequencing depth and data volume. Consequently, NGS is widely utilized in various research areas, including genomic, transcriptomic, and metagenomic analyses (Metzker, 2010; Goodwin et al., 2016).

A single NGS instrument can process anywhere from thousands to millions of DNA molecules at once. How well an NGS analysis turns out generally hinges on four stages working together: setting clear research objectives and an appropriate experimental design, preparing the sequencing library, running the sequencing itself, and finally analyzing the resulting data (Vincent et al., 2016).

The progress made in NGS technology-thoroughly documented in the comprehensive review by Hu et al. (2021)-underpins much of what has advanced modern genomic research. The evolution of NGS platforms has opened up increasingly broad opportunities for bacterial genome characterization, particularly through two main approaches: short-read sequencing (second generation) and long-read sequencing (third generation). Of the platforms currently in use, Illumina has earned a reputation as the benchmark for short-read accuracy specifically within bacterial genomic research; its sequencing-by-synthesis chemistry, built on reversible fluorescent terminators and paired-end reads, delivers a base error rate of roughly 0.1% (Satam et al., 2023). These strengths explain why Illumina remains the platform of choice for detecting single nucleotide polymorphisms (SNPs), studying bacterial populations, and conducting large-scale genomic surveillance. However, Illumina's short reads (typically under 300 bp) are generally insufficient to fully resolve repetitive genomic regions, such as ribosomal RNA operons, insertion sequences, or plasmids, and therefore do not usually enable complete, closed genome assembly on their own. In contrast, long-read platforms such as PacBio and Oxford Nanopore Technologies (ONT) are better suited for genome closure and structural resolution, and are increasingly used in combination with Illumina data through hybrid assembly approaches to achieve both high

accuracy and full genome completeness (Wick et al., 2023). The upshot is that selecting a sequencing platform ought to hinge on the particular analytical goal at hand-be it short-read accuracy, achieving a complete genome assembly, or enabling real-time, portable sequencing-rather than assuming any one platform outperforms the rest across every use case.

In addition to Illumina, the Ion Torrent platform offers a different approach through semiconductor-based pH change detection technology that does not require an optical imaging system. This technology enables faster sequencing times—approximately 2.5–4 hours—though it has the limitation of a higher error rate in homopolymer regions. Meanwhile, the development of third-generation technology, represented by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), has revolutionized genomic analysis capabilities by generating reads larger than 10 kb. This capability is crucial for resolving complex and repetitive genomic regions, identifying structural variations, and generating complete bacterial genomes (closed genomes).

Raw read accuracy for ONT sequencing started out around 87–92% in its early chemistries; newer pore designs (the shift from R9.4.1 to R10.4/R10.4.1) paired with improved basecalling algorithms have since pushed modal accuracy up to nearly 99% (Sereika et al., 2022). The extent of this improvement depends on several interacting factors, including flow cell and chemistry version, the basecaller model used, read quality filtering thresholds, and whether additional consensus polishing is applied; consequently, accuracy figures reported in earlier literature may no longer reflect the performance of current ONT platforms. Several bacterial isolate genome studies have demonstrated that merging short-read data (Illumina) with long-read data (ONT or PacBio) in a hybrid assembly offsets the weaknesses inherent to each method on its own: the resulting assembly retains the base-level precision of short reads while gaining the long reads' capacity to bridge repetitive or structurally complicated genomic stretches (Wick et al., 2023). Using the newer ONT chemistries, researchers have also managed to assemble near-complete bacterial genomes straight from long-read-only data-no short-read polishing required-provided the sequencing depth is adequate (Sereika et al., 2022). That said, hybrid or long-read-only assembly does not succeed uniformly across all cases-outcomes hinge on sequencing coverage, the quality and integrity of the input DNA, the assembly algorithm chosen, and, especially in metagenomic settings, how complex and evenly distributed the microbial community is; results obtained for a single bacterial isolate genome cannot simply be extrapolated to metagenome-assembled genomes (MAGs) or to more intricate, multi-species metagenome assemblies (Sereika et al., 2022). Thus, the integration of various NGS platforms not only improves the quality of genomic data but also accelerates research in microbiology, molecular epidemiology, and the deeper exploration of the genetic potential of microorganisms, provided that platform selection and assembly strategy are matched to the complexity of the sample and the goals of the study.

Stages of NGS-Based Metagenomic Analysis

Carrying out NGS-based metagenomic analysis typically involves working through a sequence of stages that span both bench work (wet-lab procedures) and computational analysis, as Figure 1 summarizes. These stages run from sample collection through DNA extraction, library preparation, sequencing, quality control, taxonomic and functional annotation, and finally biological interpretation, with each one examined more closely in the subsections that follow.

Metagenomics enables the analysis of the composition of functional genes within a microbial community, thereby providing more comprehensive information compared to phylogenetic approaches, which generally focus only on the diversity of specific genes, such as the 16S rRNA gene. In metagenomic research, the sampling process is a crucial initial step because it determines the quality of the data to be obtained. The next stage is DNA extraction, which aims to obtain DNA representative of all microorganisms present in the sample. The extracted DNA must be of sufficient quality and quantity to be optimally utilized in the sequencing process. For this reason, the DNA extraction method chosen needs to match the specific characteristics of the sample type in

question, so that the DNA obtained is both representative and of high quality (Knight et al., 2012).

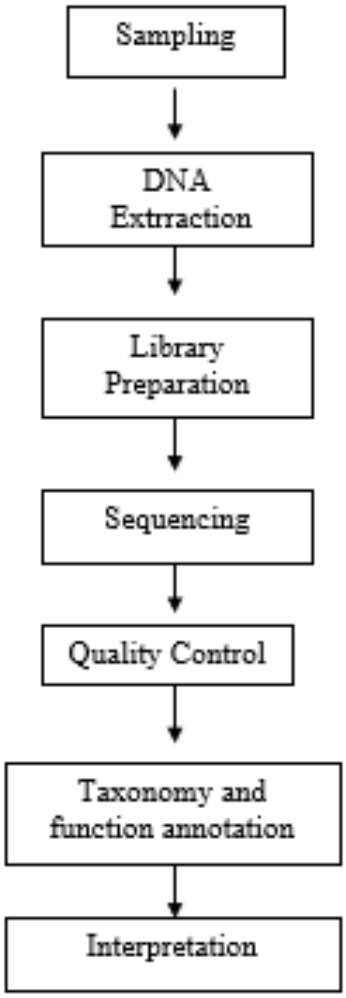

Figure 1. The NGS-based metagenomic analysis workflow, beginning with wet-laboratory steps (sampling, DNA extraction, library preparation, and sequencing) and continuing through bioinformatics analysis (quality control, taxonomic and functional annotation, and interpretation).

Various sample types can be used in metagenomic analysis, including environmental, human, animal, industrial process (e.g., composting, fermentation), and biofilm samples, each requiring an appropriate DNA extraction technique to obtain high-quality genomic DNA. The choice of extraction method, however, is not the only—or even the primary—source of variation in metagenomic data: evidence from comparative studies indicates that how deep a sample is sequenced matters more for detecting bacterial richness than which extraction method was used (Sanchez-Cid et al., 2022). In addition, DNA extraction reagents and kits can introduce their own background microbial contamination, with contamination profiles varying not only between reagent brands but also between manufacturing lots of the same brand-a batch effect that can confound result interpretation, particularly in low-biomass samples (Lai et al., 2025). Addressing these potential sources of bias-by incorporating negative controls, keeping extraction protocols consistent throughout a study, and ensuring sequencing depth is both adequate and standardized-is thus essential if metagenomic data are to be accurate and reproducible.

In microbiome research, amplicon sequencing and shotgun metagenomics stand as the two dominant approaches, each pursuing different objectives and each carrying its own degree of analytical complexity. The amplicon sequencing workflow starts by isolating DNA from the sample, after which polymerase chain reaction (PCR) is used to amplify specific marker genes-for bacteria, typically the 16S rRNA gene. This stage aims to obtain taxonomic information and to reveal which microbial communities are present within it. Once generated, the PCR amplicons undergo purification and are converted into a sequencing library, after which reading takes place on a high-throughput platform such as Illumina. The resulting sequence output is subsequently processed through a range of bioinformatics pipelines that allow researchers to identify, classify, and characterize the microbial community structure captured by the marker gene data.

In contrast to the gene-targeted nature of amplicon sequencing, shotgun metagenomics captures the full genetic content present within a sample. The workflow starts with a thorough DNA isolation step designed to recover genomic material from every member of the microbial community. This extracted DNA is fragmented into smaller pieces and assembled into a sequencing library. The library is then read on high-throughput NGS instruments, such as those from Illumina or comparable platforms. Bioinformatics analysis of the resulting reads is considerably more involved, encompassing gene identification, functional prediction, and genome reconstruction for the organisms detected in the

sample. Because it captures the whole genetic pool rather than a single marker region, shotgun metagenomics offers a substantially richer picture of both taxonomic composition and functional potential within a microbial community. The fundamental distinction between shotgun and amplicon metagenomics, along with their respective marker genes and target organism groups, has been described in the "Basic Principles of Metagenomics" section and summarized in Table 2.

Criteria for Sequencing Platform Selection

Deciding between sequencing platforms and approaches hinges on several practical factors: what the study aims to achieve, how many samples must be processed, the budget at hand, and the computational resources realistically available (Van Reckem et al., 2021; Suminda et al., 2022). For large-scale surveys of microbiota composition, amplicon sequencing tends to be the more economical choice (Weinroth et al., 2022). Shotgun metagenomics, on the other hand, becomes preferable whenever functional genetic data are needed-for instance, when the goal is to pinpoint bacteriocin genes, screen for antimicrobial resistance genes (ARGs), or characterize the metabolic profile of communities tied to product quality and health outcomes (De Ibeke et al., 2021). This added depth of information on microbial structure and function comes at the price of higher sequencing costs and greater computational demand relative to amplicon-based methods (Joseph et al., 2021; Wensel et al., 2022). These platform-selection principles are not confined to a single domain; they resurface throughout later sections addressing bacterial community analysis in health, agriculture, and food-related contexts.

Applications of NGS in Bacterial Community Analysis

NGS-based metagenomic approaches have been applied across a wide range of scientific and industrial domains; however, this review focuses specifically on the health, agriculture, and food sectors, as these three fields represent the most extensively documented and rapidly growing areas of application for bacterial community analysis using NGS-based metagenomics (Kim et al., 2024). Health-related applications largely concern the characterization of the human microbiome and pathogen surveillance; agricultural applications focus on soil and plant-associated bacterial communities relevant to crop productivity and sustainability; and food sector applications address microbial safety, fermentation processes, and quality control. Each of these applications is discussed in turn in the following subsections.

Health Sector

The emergence of Next-Generation Sequencing has transformed numerous areas of scientific inquiry and clinical genomics, largely owing to its capacity for high-throughput sequencing. Its role in advancing translational medicine stems not only from faster, more efficient sequencing workflows but also from parallel progress in bioinformatics software capable of handling, analyzing, and interpreting genomic datasets at scale. Several large-scale microbiome sequencing initiatives have generated substantial reference data that support human microbiome and metagenomic research, among them the Human Microbiome Project, which mapped microbial communities at several clinically important body sites in healthy individuals (Human Microbiome Project Consortium, 2012), healthy human cohort (Human Microbiome Project Consortium, 2012), while the MetaHIT consortium built a reference gene catalog encompassing millions of non-redundant genes drawn from the human gut microbiome (Qin et al., 2010). Metagenomic sequencing has not been limited to cataloging human-associated microbiota; it has also increasingly been directed toward pathogen surveillance, offering a culture-independent means of detecting both established and newly emerging microbial pathogens in near real time directly from clinical or environmental samples (Ko et al., 2022). The accumulation of such extensive microbiome and pathogen-surveillance datasets has, in turn, spurred the creation of reference databases and analytical frameworks that underpin both clinical research and the wider effort to characterize bacterial communities of relevance to human health.

Virtually every ecosystem harbors microorganisms, which form intricate relationships with their hosts-ranging from symbiosis and commensalism to pathogenicity. Collectively, these interactions constitute a microbiome that contributes substantially to physiological balance and overall human health

(Qin et al., 2010). Gaining a thorough, ongoing understanding of how the host and its microbiome interact-and of how members of the microbial community interact with one another-has consequently become a central concern in contemporary health research. Metagenomics offers the tools needed for such comprehensive characterization, supporting work on pathogen surveillance, dysbiosis, and the therapeutic possibilities the microbiome may hold.

A substantial and expanding body of literature links shifts in gut microbiota composition to various health outcomes, though the strength and reproducibility of these links differ markedly depending on the study design and the disease under investigation. For instance, some meta-analyses have found significant compositional differences in gut microbiota among individuals with depressive disorder, while others report inconsistent or non-significant differences, highlighting substantial heterogeneity across studies (Gao et al., 2023). Similarly, specific gut microbial taxa have shown associations with rheumatoid arthritis risk in recent Mendelian randomization analyses, although most associations did not reach strict statistical significance thresholds and causal mechanisms remain to be fully established (Wei et al., 2025). A comparable pattern emerges in metabolic disease research: meta-analyses and Mendelian randomization studies have connected particular gut microbial taxa to type 2 diabetes risk, while cautioning that correlation observed in these studies does not, on its own, demonstrate a causal relationship (Liu et al., 2024).

Metagenomic approaches further contribute by flagging pathogenic genes and antimicrobial-resistance genes, thereby aiding efforts to track the spread of drug resistance. Out of this line of research has emerged the field of pharmacomicrobiomics, which examines how the resident microbiota shapes an individual's response to medication and, conversely, how medications reshape the host's microbial community-work that may eventually pave the way toward more personalized and effective treatment strategies (Satam et al., 2023)

Agriculture

Agricultural productivity depends heavily on soil quality and fertility, both of which are conventionally gauged through a combination of physical, chemical, and biological indicators. Among these, the biological dimension-specifically, the composition of soil microbial communities-has drawn particular attention because of the central role these communities play in sustaining soil ecosystem function (Villafuerte et al., 2024; Malone et al., 2023). The discussion that follows draws together how NGS-based metagenomics has informed agricultural research along four interconnected lines: biomarkers of soil health, genes governing nutrient cycling, microbiomes that promote plant growth, and the effects of land-use change on sustainable farming practices.

Soil health biomarkers

Soil microorganisms respond quickly to environmental disturbance, a property that has made them a favored biological proxy for assessing soil quality and fertility. Urra et al. (2019), for instance, found that changes in the structure and enzymatic activity of the soil microbial community closely mirror shifts in how the soil is managed, and Kabiri et al. (2016) reported that microbial biomass and respiration rates shift rapidly in response to contamination often ahead of any detectable change in physicochemical measurements (see also Richter et al., 2018; Ghosh et al., 2020; Michael et al., 2018).

Nutrient cycling genes.

NGS-based metagenomic approaches enable the the diversity of the soil microbiome, while simultaneously profiling the functional genes underlying key biogeochemical cycles. Through metagenomic analysis, genes tied to carbon, nitrogen, and phosphorus cycling have been successfully pinpointed and put forward as molecular markers of soil health (Jia et al., 2025)—findings that align with previously documented links among microbial biomass, microbial activity, and soil organic carbon levels (Nunes et al., 2012; Kumar et al., 2014; see also Singh et al., 2016; Virmal et al., 2017). These microbially driven processes are reinforced by soil enzymes involved in nitrogen cycling, humus formation, the breakdown of foreign compounds, and nitrification-denitrification reactions (Daunoras et al., 2024).

Plant-growth-promoting microbiomes.

A range of metagenomic surveys has profiled soil bacterial groups known to directly benefit plant growth. Azotobacter and Azospirillum species, for example, fix atmospheric nitrogen and release vitamins, siderophores, and phytohormones that improve nutrient availability to plants (Aasfar et al., 2021; Mariotti et al., 2021; Souza et al., 2017). Actinobacteria, meanwhile, break down complex organic matter, contribute to humus formation, and produce enzymes, antibiotics, and secondary metabolites that enhance soil fertility (Bhatti et al., 2017; Adenan & Ting, 2022). Soil fungi add a further layer of support by helping plants withstand abiotic stress from drought, salinity, and xenobiotic contamination (Yuvaraj & Ramasamy, 2020). Taken together, this body of work frames the soil microbiome-through the metagenomic identification of its plant-growth-promoting members as a valuable biotechnological asset for sustainable crop management (Vincze et al., 2024).

Land-use change and sustainable agriculture

Beyond biomarker discovery, metagenomic data have proven useful for pinpointing microorganisms capable of boosting soil carbon storage and for tracking how different land-use practices reconfigure microbial community structure and the ecological functions it performs (Beattie et al., 2025; Jin et al., 2025). Pairing metagenomics with metabolomics is opening up a fuller picture of how microorganisms, metabolites, and ecological processes interrelate within soil systems (Freire-Zapata et al., 2024). Metabarcoding based on 16S rRNA and ITS genes remains the more affordable and widely adopted option, yet newer bioinformatics tools are now able to infer a community's metabolic functions directly from taxonomic data, without additional functional sequencing (Belcour et al., 2025). This progress has translated into tangible applications for sustainable agriculture comparing how organic versus inorganic fertilizers affect microbial diversity, using machine learning to flag microbial biomarkers, and guiding ecosystem-restoration planning grounded in microbiome data (Mishra et al., 2025; Mo et al., 2024; Peddle et al., 2025). A number of countries have even begun exploring national-scale microbiome monitoring programs as a strategy for protecting soil health and securing the long-term sustainability of their agricultural systems (Neale et al., 2024).

Food Sector

The growing sophistication of Next-Generation Sequencing has fueled a broader shift toward metagenomic approaches in fermented-food research. Improvements in accuracy and sequencing throughput, coupled with falling analysis costs, have prompted many players in the food industry and regulatory bodies alike to move away from conventional culturing and classical molecular techniques in favor of NGS-based methods (Jagadeesan et al., 2019). Third-generation platforms such as those from Oxford Nanopore Technologies add a further advantage: the ability to identify microorganisms in real time, which in turn speeds up decision-making when responding to foodborne outbreaks or bacteriophage contamination on production lines (Juul et al., 2015; Bao et al., 2021). A related innovation, Read Until technology, allows sequencing to be targeted selectively, which boosts analytical efficiency, prolongs flow-cell usability, and cuts down on overall sequencing expense (Edwards et al., 2019; Payne et al., 2021; Marquet et al., 2022).

Metagenomics also makes it possible to flag genes linked to both desirable and undesirable traits within the microbial communities found in fermented foods. Detecting antibiotic-resistance genes (ARGs) is a particular priority, given the growing food-safety concerns they raise (EFSA BIOHAZ panel et al., 2021; Leech et al., 2020). To this end, databases such as CARD and ResFinder are commonly used to screen sequencing data for the presence of ARGs (McArthur et al., 2013; Florensa et al., 2022). Metagenomics is not limited to flagging harmful genes, however; it can equally reveal genes tied to flavor development, texture formation, and the health-promoting qualities of fermented products. Genes governing the synthesis of organic acids and ethanol, along with those involved in amino acid, sugar, lipid, and protein metabolism, are routinely examined to assess how fermentation is progressing and what characteristics the final product will display (Walsh et al., 2020; Walsh et al., 2016; Obafemi et al., 2022). Genes related to lactose metabolism

warrant particular attention as well, since they determine a microorganism's capacity to reduce lactose levels during fermentation-an outcome that can make dairy products more accessible to individuals with lactose intolerance (Dimidi et al., 2019).

A body of research has documented several ways in which microorganisms present in fermented foods can benefit health: by modulating immune activity, strengthening intestinal barrier integrity, keeping pathogens from colonizing the gut, neutralizing microbial toxins, and generating antimicrobial substances such as bacteriocins (Obafemi et al., 2022; Walsh et al., 2022; Leech et al., 2020; Chen et al., 2017; De Filippis et al., 2020; Kok et al., 2018). Metagenomic approaches have likewise proven effective at picking out genes tied to prebiotic activity and to the synthesis of beneficial metabolites like short-chain fatty acids (SCFAs) (Aslam et al., 2020). A microorganism's probiotic potential, meanwhile, is frequently linked to genes controlling exopolysaccharide production, urease activity, bile salt hydrolase, and mucin-binding proteins-all of which help the organism survive passage through the digestive tract (Harper et al., 2018; Dimidi et al., 2019; De Filippis et al., 2020).

Pairing NGS with metagenomics allows for a fuller analysis of the microbial communities driving fermentation, capturing the intricate cross-species interactions that shape both the sensory qualities and the health benefits of the resulting food products (Blasche et al., 2021). Such an approach yields richer insight than examining isolates one at a time, since many metabolic pathways only manifest through interaction among community members within a given food matrix (Wolfe et al., 2014). Metagenomics further extends into virome research within fermented foods, opening a window onto the bacteriophages and foodborne viruses present-organisms that can bear directly on product quality and safety (Roux et al., 2021). Low viral abundance and comparatively small genome sizes remain persistent obstacles to viral analysis, though emerging direct RNA sequencing technologies are beginning to open up more thorough characterization of the food virome (Tamang et al., 2022). Consistent with the platform-selection considerations discussed earlier, deciding between amplicon and shotgun metagenomics for fermented-food research ultimately comes down to the specific analytical aim in view-whether that aim is taxonomic surveying or a deeper functional profile of the microbial community.

Conclusion

NGS-based metagenomics stands out as a compelling alternative to conventional culture-dependent methods, which fall short in capturing the vast majority of environmental microorganisms. Within this field, two complementary strategies have taken shape: amplicon metagenomics, built around marker genes such as 16S rRNA, and shotgun metagenomics, which sequences a sample's entire genetic content-each carrying its own set of trade-offs depending on what the research sets out to achieve. Complementing these strategies is a growing array of NGS platforms, from second-generation systems like Illumina and Ion Torrent to third-generation options such as PacBio and ONT, giving researchers room to balance accuracy, read length, and budget as needed; hybrid assembly strategies add a further layer of refinement to the genomic data produced. Across the health, agricultural, and food sectors, NGS-based metagenomic applications have made tangible contributions, supporting pathogen surveillance, microbiome dysbiosis analysis, and the development of pharmacomicrobiomics concepts in healthcare; the characterization of soil microbial communities as indicators of soil quality and fertility in agriculture; and the food sector, it has aided in identifying fermentation microorganisms, detecting antibiotic-resistance genes, and uncovering genes tied to flavor and health benefits. Several obstacles nonetheless persist: bioinformatics pipelines are still inconsistent across studies, data sharing and interoperability remain limited, and functional predictions-especially those derived from amplicon-based data continue to depend heavily on experimental confirmation rather than direct measurement. Looking ahead, the field stands to benefit from several parallel efforts: standardizing bioinformatics pipelines more broadly, encouraging open and interoperable data-sharing practices, drawing metagenomics into closer dialogue with complementary multi-omics approaches such as metatranscriptomics

and metabolomics, and placing greater emphasis on experimentally validating the microbial genes and pathways that current methods predict. Progress on these fronts will be key to producing insights into bacterial community structure and function that are more reproducible, more comparable across studies, and more biologically meaningful.

Acknowledgments

The authors wish to express their gratitude to all those whose support and contributions made this review possible, with particular thanks extended to their academic advisors and affiliated institutions for the guidance, resources, and opportunities extended throughout the writing process.

References

Aasfar, A., Bargaz, A., Yaakoubi, K., Hilali, A., Bennis, I., Zeroual, Y., et al. (2021). Nitrogen fixing *Azotobacter* species as potential soil biological enhancers for crop nutrition and yield stability. *Frontiers in Microbiology, 12*, 628379. https://doi.org/10.3389/fmicb.2021.628379

Adenan, N. H., & Ting, A. S.-Y. (2022). Actinobacteria from soils and their applications in environmental bioremediation. In P. Chowdhary, S. Mani, & P. Chaturvedi (Eds.), *Microbial Biotechnology*. Wiley. https://doi.org/10.1002/9781119834489.ch16

Aslam, H., Green, J., Jacka, F. N., Collier, F., Berk, M., Pasco, J., & Dawson, S. L. (2020). Fermented foods, the gut and mental health: A mechanistic overview with implications for depression and anxiety. *Nutritional Neuroscience, 23*(9), 659–671. https://doi.org/10.1080/1028415X.2018.1544332

Asshauer, K. P., Wemheuer, B., Daniel, R., & Meinicke, P. (2015). Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics, 31*(17), 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Bao, Y., Wadden, J., Erb-Downward, J. R., Ranjan, P., Zhou, W., McDonald, T. L., et al. (2021). SquiggleNet: Real-time, direct classification of nanopore signals. *Genome Biology, 22*(1), 298. https://doi.org/10.1186/s13059-021-02511-y

Beattie, G. A., Edlund, A., Esiobu, N., Gilbert, J., Nicolaisen, M. H., Jansson, J. K., et al. (2025). Soil microbiome interventions for carbon sequestration and climate mitigation. *mSystems, 10*, e01129-24. https://doi.org/10.1128/msystems.01129-24

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition, 98*(6), 236–238.

Belcour, A., Megy, L., Stephant, S., Michel, C., Rad, S., Bombach, P., et al. (2025). Predicting coarse-grained representations of biogeochemical cycles from metabarcoding data. *Bioinformatics, 41*(Supplement_1), i49–i57. https://doi.org/10.1093/bioinformatics/btaf230

Bhatti, A. A., Haq, S., & Bhat, R. A. (2017). Actinomycetes benefaction role in soil and plant health. *Microbial Pathogenesis, 111*, 458–467. https://doi.org/10.1016/j.micpath.2017.09.036

Blasche, S., Kim, Y., Mars, R. A. T., Machado, D., Maansson, M., Kafkia, E., et al. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nature Microbiology, 6*(2), 196–208.

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics, 20*(4), 1125–1136.

Daunoras, J., Kačergius, A., & Gudiukaitė, R. (2024). Role of soil microbiota enzymes in soil health and activity changes depending on climate change and the type of soil ecosystem. *Biology, 13*(2), 855. https://doi.org/10.3390/biology13020085

De Filippis, F., Valentino, V., Alvarez-Ordóñez, A., Cotter, P. D., & Ercolini, D. (2021). Environmental microbiome mapping as a strategy to improve quality and safety in the food industry. *Current Opinion in Food Science, 38*, 168–176. https://doi.org/10.1016/j.cofs.2020.11.012

Devaraj, S., Hemarajata, P., & Versalovic, J. (2013). The human gut microbiome and body metabolism: Implications for obesity and diabetes. *Clinical Chemistry, 59*(4), 617–628. https://doi.org/10.1373/clinchem.2012.187617

Dimidi, E., Cox, S. R., Rossi, M., & Whelan, K. (2019). Fermented foods: Definitions and characteristics, impact on the gut microbiota and effects on gastrointestinal health and disease. *Nutrients, 11*(8), 1806. https://doi.org/10.3390/nu11081806

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology, 38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Edwards, H. S., Krishnakumar, R., Sinha, A., Bird, S. W., Patel, K. D., & Bartsch, M. S. (2019). Real-time selective sequencing with RUBRIC: Read until with basecall and reference-informed criteria. *Scientific Reports, 9*(1), 11475. https://doi.org/10.1038/s41598-019-47857-3

Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D., & Aarestrup, F. M. (2022). ResFinder—An open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics, 8*(1), Article 000748. https://doi.org/10.1099/mgen.0.000748

Foster, J. A., & McVey Neufeld, K.-A. (2013). Gut–brain axis: How the microbiome influences anxiety and depression. *Trends in Neurosciences, 36*(5), 305–312. https://doi.org/10.1016/j.tins.2013.01.005

Freire-Zapata, V., Holland-Moritz, H., Cronin, D. R., Aroney, S., Smith, D. A., Wilson, R. M., et al. (2024). Microbiome–metabolite linkages drive greenhouse gas dynamics over a permafrost thaw gradient. *Nature Microbiology, 9*, 2892–2908. https://doi.org/10.1038/s41564-024-01800-z

Gao, M., Wang, J., Liu, P., Tu, H., Zhang, R., Zhang, Y., Sun, N., & Zhang, K. (2023). Gut microbiota composition in depressive disorder: A systematic review, meta-analysis, and meta-regression. *Translational Psychiatry, 13*(1), 379. https://doi.org/10.1038/s41398-023-02670-5

Ghosh, A., Singh, A. B., Kumar, R. V., Manna, M. C., Bhattacharyya, R., Rahman, M. M., et al. (2020). Soil enzymes and microbial elemental stoichiometry as bio-indicators of soil quality in diverse cropping systems and nutrient management practices of Indian vertisols. *Applied Soil Ecology, 145*, 103304. https://doi.org/10.1016/j.apsoil.2019.06.007

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*(6), 333–351.

Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., Knight, R., & Knights, D. (2018). Evaluating the information content of shallow shotgun metagenomics. *mSystems, 3*(6), e00069-18. https://doi.org/10.1128/mSystems.00069-18

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology, 82*(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature, 486*(7402), 207–214. https://doi.org/10.1038/nature11234

Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., et al. (2019). The use of next generation sequencing for improving food safety: Translation into practice. *Food*

Microbiology, 79, 96–115. https://doi.org/10.1016/j.fm.2018.11.005

Jain, A., Bhoyar, R. C., Pandhare, K., Mishra, A., Sharma, D., Imran, M., et al. (2021). IndiGenomes: A comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Research, 49*(D1), D1225–D1232. https://doi.org/10.1093/nar/gkaa923

Jia, J., de Goede, R., Li, Y., Zhang, J., Wang, G., Zhang, J., et al. (2025). Unlocking soil health: Are microbial functional genes effective indicators? *Soil Biology and Biochemistry, 204*, 109768. https://doi.org/10.1016/j.soilbio.2025.109768

Jin, W., Zhang, Y., Su, X., Xie, Z., Wang, R., Wang, Y., et al. (2025). Effects of different land use on functional genes of soil microbial carbon and phosphorus cycles in the desert steppe zone of the Loess Plateau. *BMC Microbiology, 25*, 607. https://doi.org/10.1186/s12866-025-04305-9

Juul, S., Izquierdo, F., Hurst, A., Dai, X., Wright, A., Kulesha, E., et al. (2015). What's in my pot? Real-time species identification on the MinION™. *bioRxiv*. https://doi.org/10.1101/030742

Kabiri, V., Raiesi, F., & Ghazavi, M. A. (2016). Tillage effects on soil microbial biomass, SOM mineralization and enzyme activity in a semi-arid Calcixerepts. *Agriculture, Ecosystems & Environment, 232*, 73–84. https://doi.org/10.1016/j.agee.2016.07.022

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research, 28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kim, C., Pongpanich, M., & Porntaveetus, T. (2024). Unraveling metagenomics through long-read sequencing: A comprehensive review. *Journal of Translational Medicine, 22*(1), 111. https://doi.org/10.1186/s12967-024-05006-4

Kim, N., Ma, J., Kim, W., Kim, J., Belenky, P., & Lee, I. (2024). Genome-resolved metagenomics: A game changer for microbiome medicine. *Experimental & Molecular Medicine, 56*(7), 1501–1512. https://doi.org/10.1038/s12276-024-01268-z

Kinoshita, Y., Niwa, H., Uchida-Fujii, E., & Nukada, T. (2021). Establishment and assessment of an amplicon sequencing method targeting the 16S-ITS-23S rRNA operon for analysis of the equine gut microbiome. *Scientific Reports, 11*, 11884. https://doi.org/10.1038/s41598-021-91387-8

Ko, K. K. K., Chng, K. R., & Nagarajan, N. (2022). Metagenomics-enabled microbial surveillance. *Nature Microbiology, 7*(4), 486–496. https://doi.org/10.1038/s41564-022-01089-w

Kok, C. R., & Hutkins, R. (2018). Yogurt and other fermented foods as sources of health-promoting bacteria. *Nutrition Reviews, 76*(Suppl. 1), 4–15. https://doi.org/10.1093/nutrit/nuy056

Kumar, A., Maurya, B. R., & Raghuwanshi, R. (2014). Isolation and characterization of PGPR and their effect on growth, yield and nutrient content in wheat (*Triticum aestivum* L.). *Biocatalysis and Agricultural Biotechnology, 3*(2), 121–128. https://doi.org/10.1016/j.bcab.2014.08.003

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology, 31*(9), 814–821. https://doi.org/10.1038/nbt.2676

Liu, T., Cao, Y., Liang, N., Ma, X., Fang, J., & Zhang, X. (2024). Investigating the causal association between gut microbiota and type 2 diabetes: A meta-analysis and Mendelian randomization. *Frontiers in Public Health, 12*, 1342313. https://doi.org/10.3389/fpubh.2024.1342313

Liu, Y.-X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., & Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein & Cell, 12*(5), 315–330. https://doi.org/10.1007/s13238-020-00724-8

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science, 353*(6305), 1272–1277. https://doi.org/10.1126/science.aaf4507

Lu, Y., Chan, Y.-T., Tan, H.-Y., Li, S., Wang, N., & Feng, Y. (2020). Epigenetic regulation in human cancer: The potential role of epi-drug in cancer therapy. *Molecular Cancer, 19*, 79. https://doi.org/10.1186/s12943-020-01197-3

Luo, H., Wang, Y., Hou, H., Yang, J., & Liu, Y.-X. (2025). Advances and applications in sequencing-based pathogen surveillance. *Advanced Biotechnology, 3*, 100004. https://doi.org/10.1016/j.abiote.2025.100004

Mahnert, A., Moissl-Eichinger, C., Zojer, M., Bogumil, D., Mizrahi, I., Rattei, T., Martinez, J. L., & Berg, G. (2019). Man-made microbial resistances in built environments. *Nature Communications, 10*, 968. https://doi.org/10.1038/s41467-019-08864-0

Malone, Z., Berhe, A. A., & Ryals, R. (2023). Impacts of organic matter amendments on urban soil carbon and soil quality: A meta-analysis. *Journal of Cleaner Production, 419*, 138148. https://doi.org/10.1016/j.jclepro.2023.138148

Mariotti, L., Scartazza, A., Curadi, M., Picciarelli, P., & Toffanin, A. (2021). *Azospirillum baldaniorum* Sp245 induces physiological responses to alleviate the adverse effects of drought stress in purple basil. *Plants, 10*(6), 1141. https://doi.org/10.3390/plants10061141

Marquet, M., Zöllkau, J., Pastuschek, J., Viehweger, A., Schleußner, E., Makarewicz, O., et al. (2022). Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. *Scientific Reports, 12*(1), 4000. https://doi.org/10.1038/s41598-022-08003-8

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R., & Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal, 6*(3), 610–618. https://doi.org/10.1038/ismej.2011.139

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications, 7*, 11257. https://doi.org/10.1038/ncomms11257

Metzker, M. L. (2010). Sequencing technologies—The next generation. *Nature Reviews Genetics, 11*(1), 31–46. https://doi.org/10.1038/nrg2626

Michael, S., Nannipieri, P., Sørensen, S. J., & van Elsas, J. D. (2018). Microbial indicators for soil quality. *Biology and Fertility of Soils, 54*(1), 1–10. https://doi.org/10.1007/s00374-017-1248-3

Mirete, S., Sánchez-Costa, M., Díaz-Rullo, J., González de Figueras, C., Martínez-Rodríguez, P., & González-Pastor, J. E. (2025). Metagenome-assembled genomes (MAGs): Advances, challenges, and ecological insights. *Microorganisms, 13*(5), 985. https://doi.org/10.3390/microorganisms13050985

Mishra, A. K., Yadav, P., Sharma, S., & Maurya, P. (2025). Comparison of microbial diversity and community structure in soils managed with organic and chemical fertilization strategies using amplicon sequencing of 16S and ITS regions. *Frontiers in Microbiology, 15*, 1444903. https://doi.org/10.3389/fmicb.2024.1444903

Mo, Y., Bier, R., Li, X., Daniels, M., Smith, A., Yu, L., et al. (2024). Agricultural practices influence soil microbiome assembly and interactions at different depths identified by machine learning. *Communications Biology, 7*, 1349. https://doi.org/10.1038/s42003-024-07059-8

Moges, B., & Mengistu, D. Y. (2024). Metagenomics approaches for studying the human microbiome. *All Life, 17*(1), 2350166. https://doi.org/10.1080/26895293.2024.2350166

Neale, D., Cullen, L., & Ranout, A. S. (2024). Improving soil health in the UK: Why a microbial approach is indispensable in attaining sustainable soils. *Sustainable Microbiology, 1*, qvae026. https://doi.org/10.1093/sumbio/qvae026

Nunes, J. S., Araujo, A. S. F., Nunes, L. A. P. L., Lima, L. M., Carneiro, R. F. V., Salviano, A. A. C., et al. (2012). Impact of land degradation on soil microbial biomass and activity in northeast Brazil. *Pedosphere, 22*(1), 88–95. https://doi.org/10.1016/S1002-0160(11)60194-X

Obafemi, Y. D., Oranusi, S. U., Ajanaku, K. O., Akinduti, P. A., Leech, J., & Cotter, P. D. (2022). African fermented foods: Overview, emerging benefits, and novel approaches to microbiome profiling. *npj Science of Food, 6*(1), 15. https://doi.org/10.1038/s41538-022-00130-w

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., & Loose, M. (2021). Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nature Biotechnology, 39*(4), 442–450. https://doi.org/10.1038/s41587-020-00746-x

Peddle, S. D., Hodgson, R. J., Borrett, R. J., Brachmann, S., Davies, T. C., Erickson, T. E., et al. (2025). Practical applications of soil microbiota to improve ecosystem restoration: Current knowledge and future directions. *Biological Reviews, 100*, 1–18. https://doi.org/10.1111/brv.13124

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59–65. https://doi.org/10.1038/nature08821

Richter, A., Ó Huallacháin, D., Doyle, E., Clipson, N., van Leeuwen, J. P., Heuvelink, G. B., et al. (2018). Linking diagnostic features to soil microbial biomass and respiration in agricultural grassland soil: A large-scale study in Ireland. *European Journal of Soil Science, 69*(3), 414–428. https://doi.org/10.1111/ejss.12551

Ross, A. A., Müller, K. M., Weese, J. S., & Neufeld, J. D. (2018). Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class *Mammalia*. *Proceedings of the National Academy of Sciences of the United States of America, 115*(25), E5786–E5795. https://doi.org/10.1073/pnas.1801302115

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., Malonia, S. K., et al. (2023). Next-generation sequencing technology: Current trends and advancements. *Biology, 12*(7), 997. https://doi.org/10.3390/biology12070997

Scher, J. U., & Abramson, S. B. (2011). The microbiome and rheumatoid arthritis. *Nature Reviews Rheumatology, 7*(10), 569–578. https://doi.org/10.1038/nrrheum.2011.121

Schultz, J., Modolon, F., Peixoto, R. S., & Rosado, A. S. (2023). Shedding light on the composition of extreme microbial dark matter: Alternative approaches for culturing extremophiles. *Frontiers in Microbiology, 14*, 1167718. https://doi.org/10.3389/fmicb.2023.1167718

Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., & Albertsen, M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nature Methods, 19*(7), 823–826. https://doi.org/10.1038/s41592-022-01539-7

Singh, J. S., Kumar, A., Rai, A. N., & Singh, D. P. (2016). Cyanobacteria: A precious bio-resource in agriculture, ecosystem, and environmental sustainability. *Frontiers in Microbiology, 7*, 529. https://doi.org/10.3389/fmicb.2016.00529

Souza, M. S. T., de Baura, V. A., Santos, S. A., Fernandes-Júnior, P. I., Reis Junior, F. B., Marques, M. R., et al. (2017). *Azospirillum* spp. from native forage grasses in Brazilian Pantanal floodplain: Biodiversity and plant growth promotion potential. *World Journal of Microbiology and Biotechnology, 33*, 81. https://doi.org/10.1007/s11274-017-2251-4

Suminda, G. G. D., Bhandari, S., Won, Y., Goutam, U., Pulicherla, K. K., Son, Y. O., & Ghosh, M. (2022). High-throughput sequencing technologies in the detection of livestock pathogens, diagnosis, and zoonotic surveillance. *Computational and Structural Biotechnology Journal, 20*, 5378–5392. https://doi.org/10.1016/j.csbj.2022.09.028

Sun, S., Jones, R. B., & Fodor, A. A. (2020). Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome, 8*(1), 46. https://doi.org/10.1186/s40168-020-00815-y

Tamang, J. P., Das, S., Kharnaior, P., Pariyar, P., Thapa, N., Jo, S. W., et al. (2022). Shotgun metagenomics of Cheonggukjang, a fermented soybean food of Korea: Community structure, predictive functionalities and amino acids profile. *Food Research International, 151*, 110904. https://doi.org/10.1016/j.foodres.2021.110904

Terrón-Camero, L. C., Gordillo-González, F., Salas-Espejo, E., & Andrés-León, E. (2022). Comparison of metagenomics and metatranscriptomics tools: A guide to making the right choice. *Genes, 13*(12), 2280. https://doi.org/10.3390/genes13122280

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., & Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods, 12*(10), 902–903. https://doi.org/10.1038/nmeth.3589

Urra, J., Alkorta, I., Lanzén, A., Mijangos, I., & Garbisu, C. (2019). The application of fresh and composted horse and chicken manure affects soil quality, microbial composition and antibiotic resistance. *Applied Soil Ecology, 135*, 73–84. https://doi.org/10.1016/j.apsoil.2018.11.005

Usyk, M., Peters, B. A., Karthikeyan, S., McDonald, D., Sollecito, C. C., Vazquez-Baeza, Y., Shaffer, J. P., Gellman, M. D., Talavera, G. A., Daviglus, M. L., Thyagarajan, B., Knight, R., Qi, Q., Kaplan, R., & Burk, R. D. (2023). Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. *Cell Reports Methods, 3*(1), 100391. https://doi.org/10.1016/j.crmeth.2022.100391

Villafuerte, A. B., Soria, R., Rodríguez-Berbel, N., Zema, D. A., Lucas-Borja, M. E., Ortega, R., et al. (2024). Short-term evaluation of soil physical, chemical and biochemical properties in an abandoned cropland treated with different soil organic amendments under semiarid conditions. *Journal of Environmental Management, 349*, 119372. https://doi.org/10.1016/j.jenvman.2023.119372

Vimal, S. R., Singh, J. S., Arora, N. K., & Singh, S. (2017). Soil–plant–microbe interactions in stressed agriculture management: A review. *Pedosphere, 27*(2), 177–192. https://doi.org/10.1016/S1002-0160(17)60309-6

Vollmers, J., Wiegand, S., & Kaster, A.-K. (2017). Comparing and evaluating metagenome assembly tools from a microbiologist's perspective—Not only size matters! *PLoS ONE, 12*(1), e0169662. https://doi.org/10.1371/journal.pone.0169662

Walsh, A. M., Macori, G., Kilcawley, K. N., & Cotter, P. D. (2020). Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. *Nature Food, 1*(8), 500–510. https://doi.org/10.1038/s43016-020-0129-3

Ward, T., Larson, J., Meulemans, J., Hillmann, B., Lynch, J., Sidiropoulos, D., Spear, J. R., Caporaso, J. G., Blekhman, R., Knight,

R., Fink, R., & Knights, D. (2017). BugBase predicts organism-level microbiome phenotypes. *bioRxiv*. https://doi.org/10.1101/133462

Wei, F., Zhao, M., Sun, X., Ma, H., Yin, H., & Shen, X. (2025). Causal associations between gut microbiota and rheumatoid arthritis: A two-sample Mendelian randomization study. *Medicine, 104*(22), e42596. https://doi.org/10.1097/MD.0000000000042596

Weinroth, M. D., Belk, A. D., Dean, C., Noyes, N., Dittoe, D. K., Rothrock, M. J., Jr., et al. (2022). Considerations and best practices in animal science 16S ribosomal RNA gene sequencing microbiome studies. *Journal of Animal Science, 100*(2), skab346. https://doi.org/10.1093/jas/skab346

Wensel, C. R., Pluznick, J. L., Salzberg, S. L., & Sears, C. L. (2022). Next-generation sequencing: Insights to advance clinical investigations of the microbiome. *The Journal of Clinical Investigation, 132*(7), e154944. https://doi.org/10.1172/JCI154944

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology, 15*(3), R46. https://doi.org/10.1186/gb-2014-15-3-r46

Wu, G., Xu, T., Zhao, N., Lam, Y. Y., Ding, X., Wei, D., Fan, J., Shi, Y., Li, X., Li, M., Ji, S., Wang, X., Fu, H., Zhang, F., Shi, Y., Zhang, C., Peng, Y., & Zhao, L. (2024). A core microbiome signature as an indicator of health. *Cell, 187*(23), 6550–6565.e11. https://doi.org/10.1016/j.cell.2024.09.019

Yang, B., Feng, W., Zhou, W., He, K., & Yang, Z. (2024). Association between soil physicochemical properties and bacterial community structure in diverse forest ecosystems. *Microorganisms, 12*(4), 728. https://doi.org/10.3390/microorganisms12040728

Yuvaraj, M., & Ramasamy, M. (2020). Role of fungi in agriculture. In S. M. Mirmajlessi & R. Radhakrishnan (Eds.), *Biostimulants in Plant Science*. IntechOpen. https://doi.org/10.5772/intechopen.89718

Zhang, J., Liu, Y.-X., Zhang, N., Hu, B., Jin, T., Xu, H., Qin, Y., Yan, P., Zhang, X., Guo, X., et al. (2019). NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nature Biotechnology, 37*(6), 676–684. https://doi.org/10.1038/s41587-019-0104-4

Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., et al. (2021). Advances in metagenomics and its application in environmental microorganisms. *Frontiers in Microbiology, 12*, 766364. https://doi.org/10.3389/fmicb.2021.766364

Zheng, M., Zhou, N., Liu, S., Dang, C., Liu, Y.-X., He, S., Zhao, Y., Liu, W., & Wang, X. (2019). N_2O and NO emission from a biological aerated filter treating coking wastewater: Main source and microbial community. *Journal of Cleaner Production, 213*, 365–374. https://doi.org/10.1016/j.jclepro.2018.12.164.