



Review

Bioinformatics Tools and Big Data Analytics for Understanding Agricultural Microbial Ecosystems

Madan L. Verma ^{1,*}, Meenu Thakur ², Prachi Arora ³, Prince Sharma ³ and Sarita Devi ⁴

¹ Department of Biotechnology, School of Basic Sciences, Indian Institute of Information Technology Una, Una 177209, Himachal Pradesh, India

² Department of Biotechnology, Shoolini Institute of Life Sciences and Business Management, Solan 173212, Himachal Pradesh, India

³ School of Computing, Indian Institute of Information Technology Una, Una 177209, Himachal Pradesh, India

⁴ Microbiology and Fermentation Technology Laboratory, Fermentation and Phytofarming Technology Division, CSIR—Institute of Himalayan Bioresource Technology, P.O. Box 06, Palampur 176061, Himachal Pradesh, India

* Correspondence: madanverma@gmail.com or madanverma@iiitu.ac.in

Abstract

Microbial diversity and sustainable agriculture are key players for the mutual harmony of the nature/agriculture ecosystem that align with the Sustainable Development Goals (SDGs). However, to harness their best potential, it is necessary to understand the complex interactions between microbial communities and plants. This has become feasible with the intervention of advanced multiomics techniques including genomics, metagenomics, transcriptomics, proteomics, metabolomics, etc. With the advent of next-generation technologies, production of biological data at a cost-effective expense has escalated rapidly. Thus, it is feasible for all researchers to explore these advanced technologies for a better understanding of how to harness the full potential of microbial diversity toward sustainable agriculture. The growth of biological databases, particularly microbial databases, and interphasic tools are bound to the rapid growth of bioinformatics domains. Additionally, bioinformatics provides the direction for better understanding of the interaction of the microbe–plant system. Keeping in view the expansion of escalated biological datasets, the role of big data analytics is vital to understand the interplay of the diverse datasets along with the growing biological databases. The relevant case studies employing big data analytics in sustainable agriculture have been discussed. This will help to make better decisions toward the productivity of sustainable agriculture. This review is an attempt to showcase the recent growth of bioinformatics databases and the role of big data analytics in achieving sustainable agriculture.

Keywords: microbial diversity; next-generation sequencing; biological data; bioinformatics; big data analytics

Received: 10 April 2026

Revised: 22 June 2026

Accepted: 25 June 2026

Published: 28 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](#) license.

1. Introduction

There is an increasing global food security challenge due to large population growth, uncertain climate conditions, and reduced dependence on agrochemicals faced by all agriculturists worldwide. Sustainable agricultural practices are the need of the hour for sustenance of life on our planet [1]. Crop microbiome/microbial communities associated with

various crops play a significant role in enhancing soil fertility, dealing with biotic and abiotic stresses, and thus improving the overall vitality and productivity of crops. Diverse microbial forms can be studied by trivial culture-dependent methods and advanced culture-independent methods involving molecular biology as well as omics-based technologies such as genomics, metagenomics, transcriptomics, proteomics, and metabolomics, which have revolutionized our ability to study plant-associated microorganisms [2].

The complex nature of plant–soil microorganism interactions has escalated the need for multiomics approaches to combine taxonomical and functional data [3]. Metagenomics has helped us better understand microbial life by enabling the direct analysis of environmental DNA, thereby deciphering the diverse forms of microorganisms [4]. The systematic and accurate identification of bioactive compounds, “bioprospecting”, holds excellent potential in uncultivable microorganisms [5]. So, microbiomics explores the diverse microbial forms inhabiting different environments to compute the beneficial microorganisms using high-throughput bioinformatics approaches [6].

Next-generation sequencing (NGS) approaches have become cost-effective, sophisticated, and automated and are integrated with computational biology, assessing metagenomic data for studying microbial diversity in agricultural ecosystems [7]. Several recent research studies on the isolation and characterization of soil microbiomes have been performed using various advanced methods, as depicted in Figure 1 [8]. Microorganisms can be broadly employed in three ways: (i) as biostimulants for enhancing soil fertility and plant-growth-promoting (PGP) activity and various soil microbiome interactions, (ii) for degradation of solid waste to be used for growth and development, and (iii) for screening of pathogenic microorganisms for crop losses.

Microbiome interactions in the post-genomic era, with advent of NGS approaches, have considerably impacted sustainability in agriculture [9,10]. Screening of microorganisms using marker-assisted selection of resistant genes against abiotic stresses is one of the approaches. However, the whole-genome sequencing (WGS) approach has been utilized for creating recombinants and hybrids with desirable traits in less time. Beneficial microorganisms have been studied extensively regarding agricultural sustainability, resulting in the screening of microorganisms that was otherwise not possible using trivial methods such as culture-dependent methods [7–11].

Microbial diversity for sustainable agriculture is a pressing demand that specifically aligns with the Sustainable Development Goals (SDGs) as defined by the United Nations, which go with the environmental dimension of sustainability on the planet. Agricultural sustainability is concerned with SDG-2, which refers to zero hunger by increasing soil fertility, nutrient cycling, and enhancing crop productivity through microbiome interaction, thereby reducing the dependence on chemical fertilizers [11]. It connects with SDG-12 for responsible consumption and production because soil–microbiome interactions lead to an eco-friendly approach. Further, it is aligned with SDG-13 (climate action) by improving carbon sequestration. Microbial diversity helps maintain social sustainable development goals such as good health and well-being (SDG-3) by providing chemical-residue-free crop yield, no poverty (SDG-1), economic goals such as SDG-8 (decent work) with economic growth and, finally, industry, innovation and infrastructure (SDG-9) by maintaining sustainable agriculture in biotechnological applications [12]. Thus, microbial diversity can be considered a natural solution for connecting environmental, social and economic sustainability.

To deal with the rapid pace of biological data production through the advent of NGS, the bioinformatics domain is coming into existence where computer scientists, biologists, and statisticians are the key players [13]. Bioinformatics databases, algorithms, and tools are assisting in data mining, endorsing the hypothesis of data-driven theory [14,15].

Precisely, data science and big data analytics in biology are growing domains that will assist in achieving sustainable agriculture.

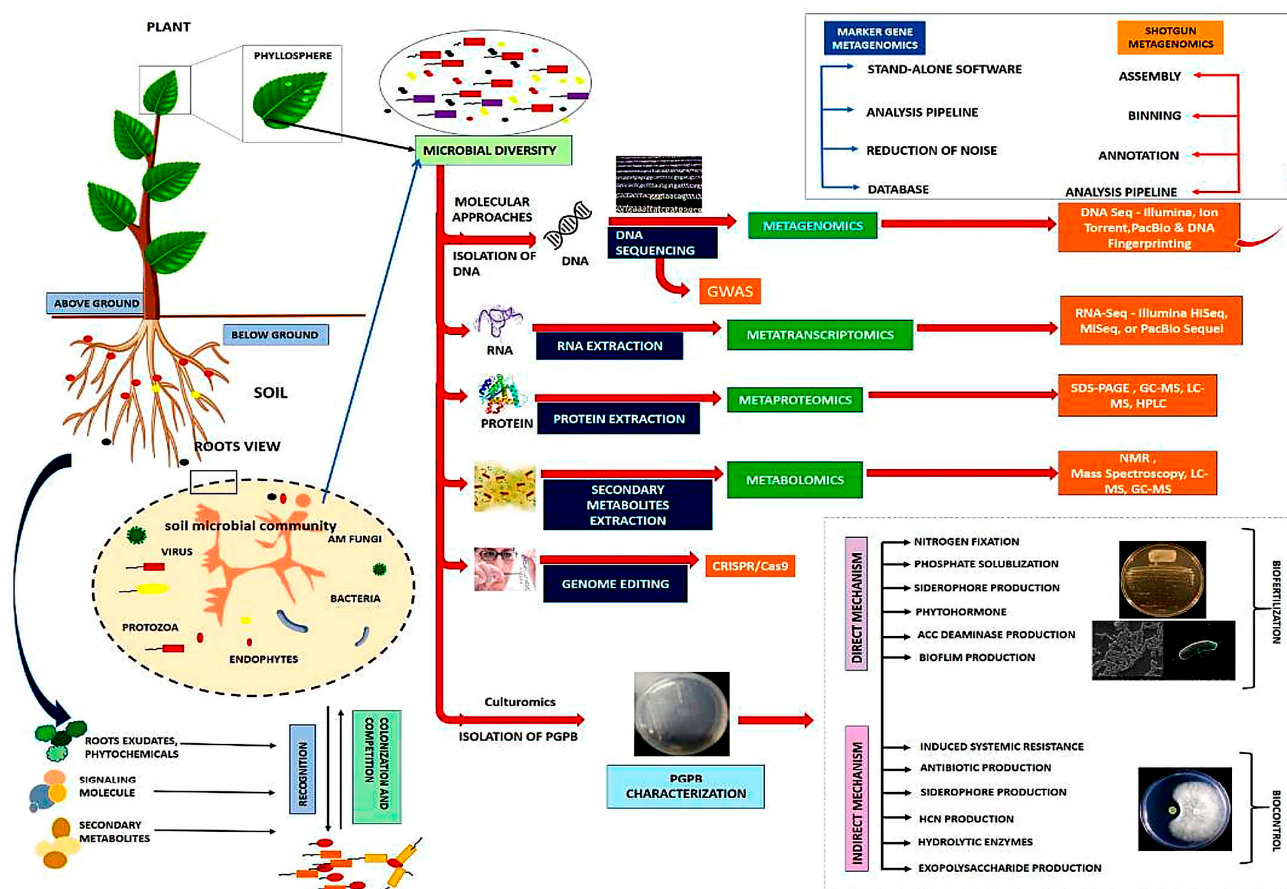


Figure 1. Application of Next-Generation Sequencing approaches in screening of isolates enhancing microbiome interactions with overall soil fertility and growth-promoting activity for sustainable agriculture. Legend: NMR—nuclear magnetic resonance; LC-MS—liquid chromatography–mass spectrometry; GC-MS—gas chromatography–mass spectrometry; RNA-Seq—RNA sequencing; DNA Seq—DNA sequencing. (Adapted from reference [8]).

The present article focusses on the role of microbial biodiversity in achieving sustainable agriculture, leveraged through powerful computation tools, where omics technologies employed for investigating microbial diversity are elucidated. Data driven via omics technologies has hastened the development of many microbial databases. And, further, the role of big data analytics is highlighted with the conclusion of future trends along with the resolution of contemporary technologies.

2. Methodology

2.1. Databases and Search Engines

The literature survey for this review was conducted using multiple scientific databases and search engines to ensure broad coverage of peer-reviewed publications and relevant book chapters. Databases used included PubMed, Scopus, Web of Science, Google Scholar, ScienceDirect, SpringerLink, Frontiers journals, MDPI journals, and IEEE Xplore. Additional references were identified through citation chaining—both backward and forward reference searching—from key articles.

2.2. Search Keywords

The search strategy combined individual terms with Boolean operators (AND, OR) to maximize retrieval of relevant studies: microbiome and omics terms, metagenomics, metaomics, multiomics, plant–microbe interaction, next-generation sequencing, bioinformatics, machine learning, artificial intelligence, big data analytics, and sustainable agriculture.

2.3. Boolean Search Strings and Literature Screening

Boolean search strings used include: “Metagenomics” AND “Sustainable Agriculture”, “Plant-microbe interaction” AND “Multi-omics” OR “Meta-omics”, “Soil microbiome” AND “Bioinformatics” OR “Machine learning”, “Next-generation sequencing” OR “Long-read sequencing”, “Artificial intelligence” OR “Big data” AND “agriculture”, “Metagenomic classifiers” OR “Microbial databases” AND “Sustainable agriculture”.

The search primarily included studies published between 2014 and 2026, with emphasis on 2020–2026 due to rapid developments in multiomics technologies and their applications in sustainable agriculture. However, seminal older studies and foundational methodological papers were also retained. Only English-language publications were included to ensure consistency in interpretation and data extraction; non-English publications and papers lacking complete English translations were excluded.

3. Omics Technologies for Deciphering Microbial Diversity Towards Sustainable Agriculture

3.1. Next-Generation Sequencing Techniques

Genome sequencing technology has rapidly evolved after first-generation sequencing and made it possible to transition from culture-dependent screening to culture-independent screening methods to characterize microbial communities at a mass level [16]. The advancement of first-generation sequencing and transition into second- and third-generation sequencing techniques allowed a deep scale of analysis, providing high throughput and generation of vast amounts of data (Table 1). First-generation sequencing is based upon Sanger sequencing that forms the base of the sequencing technologies involving the chain termination method. Automated systems such as the Applied Biosystems ABI 370 with improved efficiency of fluorescent labeling were introduced. The major limitation has been low throughput, high cost and time-consuming processes restricting it to small-scale genomic studies. The introduction of second-generation sequencing has revolutionized genomic analysis by enabling massive sequencing in one go and it is thus called next-generation sequencing (NGS). Platforms such as Roche 454 pyrosequencing, ion torrent, SOLiD and Illumina have been widely adopted due to reduced cost and fast processing time, finding applications in transcriptomics, whole-genome sequencing and targeted sequencing. However major limitations have been PCR amplification bias and difficulty in sequencing repeated sequences. The third-generation sequencing technique has emerged to solve the problem of short reads by enabling real-time sequencing of single DNA molecules without amplification using advanced platforms such as Pacific Biosciences Single Molecule Real-Time (SMRT) sequencing and Oxford Nanopore sequencing technology for portable real-time analysis [17].

In modern agriculture, soil microbiome studies using DNA sequencing technologies provide deep understanding of structural and functional aspects. Specific gene sequencing, for example, the 16S rRNA gene using Illumina, Oxford Nanopore, and Pacific Biosciences platforms, provides high resolution and thus accuracy in studying the diversity of microorganisms present in the rhizosphere. This has led to a high understanding of microbiome analysis, giving sustainable agricultural solutions for agricultural

management, better productivity, and increased crop yields. However, each sequencing platform has its own pros and cons based on cost, sensitivity, depth of coverage, and data analysis. Understanding the characteristics of each method is crucial in selecting a method for carrying out specific objective-based sequencing. New trends and novel bioinformatics tools have been used for improved metagenomics studies [7]. Recent advances in metagenomics involve metagenome-assembled genomes constructed directly from environmental DNA, which has revolutionized approaches to study unculturable organisms from the ecosystem [18].

Table 1. Next-Generation Sequencing (NGS): current technology and advancements [7,9,17–19].

Generation	Technology Name/Platform	Key Principle	Throughput and Accuracy	Advantages	Limitations
1st Generation	Sanger sequencing/ABI 3730 and Sanger Capillary Sequencer	Chain termination using dideoxynucleotide	Low with very high accuracy 99.9%	High accuracy with long reads using gold standard	Expensive, time-consuming, low throughput
2nd Generation	Sequencing by synthesis/ligation/Illumina, HiSeq, Miseq, SoLiD, IonTorrent	Massive parallel sequencing of amplified DNA	High throughput (98%) with very high accuracy (>99%)	Cost effective, high throughput and fast	Short reads, PCR bias and complex data analysis
3rd Generation	Single molecule sequencing/PacBio, SMRT, Oxford Nanopore	Real-time sequencing without amplification	Moderate with moderately high accuracy	Very long reads can detect structural variants	Higher cost with more error

NGS has the advantage of early and precise detection of crop pathogens over traditional methods such as culture-based methods, biochemical tests, molecular testing, and mass spectrometry. Trivial methods are very complex, involving many steps for detection, and time-consuming. NGS has been the cutting-edge technology in identification of various plant pathogens, which can actually prevent crop losses [19]. Although NGS has revolutionized the understanding of microbial dynamics in the rhizosphere, one of the major limitations has been the volume and nature of complex data [20]. Another limitation posed by NGS is storage, processing, and lack of computational infrastructure and expertise. Amplification biases, sequencing errors, and variations in bioinformatics necessitate stringent quality control [21]. But the integrative approach, utilizing both multiomics data and advancing bioinformatics tools, is essential for understanding the functional significance of microbial diversity in sustainable agriculture.

Bioinformatics tools for sequencing data analysis have set new trends for molecular biology. Bioinformatics tools play a crucial role in the effective application of next-generation sequencing (NGS) for studying and screening agriculturally beneficial microorganisms. These approaches can analyze diverse environmental samples such as rhizospheric soil, plant roots, plant tissues and hydroponic systems using the Qiagen Plant Mini Kit, Fast DNA SPIN Kit for soil, ZR Soil Microbe DNA kit and GeneJET soil DNA kit for DNA extraction which is the first step in the analysis followed by qualitative assessment of the DNA samples using agarose gel electrophoresis, NanoDrop, Qubit and Agilent TapeStation to ensure high-quality DNA samples.

Advanced sequencing platforms such as Illumina, PacBio and Oxford Nanopore generate massive genomic datasets for soil microbiome analysis. Amplicon sequencing enables taxonomic identification and diversity analysis through marker genes, while shotgun metagenomics provides deeper insights into microbial functions, metabolic pathways and unculturable populations. Due to the large volume of raw data, bioinformatics tools like Trimmomatic, Trim Galore and UCHIME are essential for quality control, removal of low-

quality reads, and elimination of chimeric sequences ensuring accurate microbiome analysis. Bioinformatics platforms such as QIIME 2 and MOTHUR play a key role in analyzing amplicon sequencing data by clustering OTUs, assigning taxonomy and assessing microbial diversity. Databases like the Ribosomal Database Project enable identification of beneficial soil microbial groups associated with plant growth and soil fertility, viz., *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, *Bacteroidetes*, *Chloroflexi* and *Cyanobacteria*. Shotgun metagenomics advanced assembly tools such as MEGAHIT, metaSPAdes, and Ray Meta facilitate genome reconstruction from complex soil communities while tools like SqueezeMeta and DAS Tool improve genome binning and assembly accuracy for diverse microbial populations. Functional annotation using MG-RAST helps in assembling and identifying agriculturally important genes displaying nitrogen fixation, phosphate solubilization, indole acetic acid production, stress tolerance and nutrient cycling. Phylogenetic analysis has been performed using ASTRAL, UniFrac, MEGAX, EDGAR and IMG/M to analyze and understand evolutionary relationships and compare genomes [22].

Complete shotgun sequencing using bioinformatics involves many steps, including quality control, trimming and filtering, mapping/alignment, binning and taxonomic classification [7]. A popular tool for analyzing the reads is FastQC, which evaluates the quality of sequences by graphical methods based on position, GC content, and adapter presence [23]. NanoQC and NanoComp are popular for comparison of various datasets [24]. PycoQC has been curated to explore the details of next-generation sequencing data, providing interactive graphs to filter. LongQC is a tool to assess read length using K-mer overlaps without the need for a reference genome [25]. LongReadSum is the most advanced tool using C++ for maximum performance to identify any modification and for detection [26].

3.2. Genomics, Transcriptomics and Proteomics Approaches

Transcriptomics and RNA sequencing are used for gene expression and regulatory circuits that can be modulated to achieve the optimal level of metabolite production. For example, biosynthetic gene clusters, vital for the bioactive compound synthesis, have been explored. The potential of bioactive compounds in microorganisms can be explored by advanced omics-based genome mining and metagenomics approaches [5].

Omics technology advancements have revolutionized rhizosphere dynamic studies. Metabolomics is concerned with studying the small molecules that lead to plant–microbe symbiotic interactions that can enhance resistance characteristics in plants and optimizing agricultural practices [27]. Integration of genomics, metagenomics, transcriptomics, and proteomics leads to facilitation and holistic understanding of microbial ecosystems, regulating networks and metabolic pathways critical for plant–microbe interactions [28]. Variations in root exudates attracting and harboring microorganisms in the rhizosphere depend on the signaling molecules and signaling mechanisms present in microorganisms [29].

Implications of integrated omics technologies, NGS technologies and bioinformatics tools on diverse plant–microbe interactions ranging from nitrogen fixation and mycorrhizal association to pathogen–host interaction have been reviewed [30]. Genomics and metagenomics involve 16S rRNA sequencing, whole-genome sequencing, shotgun metagenomics, ChIP-seq and ATAC-seq using platforms like Illumina, PacBio, Oxford Nanopore and Ion Torrent. Bioinformatics tools such as QIIME2, MOTHUR, MG-RAST, Kraken2, MEGAHIT, MetaSPAdes, HUMAnN, KEGG and MetaCyc have been utilized to process raw sequence data to identify diverse microorganisms followed by genome assembly and understanding the functional role of genes involved in diverse plant–microbe interactions. However, major challenges include handling the large datasets, host DNA contamination and incomplete databases [31]. Transcriptomics and metatranscriptomics

using RNA-seq, dual RNA-seq and microarrays are carried out to study active gene expressions. Tools such as STAR, HISAT2, DESeq2, edgeR and Cufflinks have been employed to identify genes involved in plant defense and response to stress, addressing the major challenge to separate plant and microbial transcripts [32]. Further, LC-MS/MS, MALDI-TOF-MS and iTRAQ have been used to identify proteins and in proteomic studies. Tools such as Mascot, MaxQuant and STRING can be used to analyze protein functions with major limitations such as complexity of proteins, which is associated with high costs [30]. To study and understand the metabolic pathways of root exudates, signaling compounds and synthesis of antimicrobial compounds in plants, metabolomics approaches use GC-MS, LC-MS, NMR and HPLC. These approaches are followed by tools and databases such as METLIN, HMDB, KEGG and MetaboAnalyst to identify the metabolites involved in nutrient exchange and defense mechanisms of plants. Major challenges involve the instability of the metabolites along with the availability of the limited databases. Advanced omics technology involves epigenomics approaches using bisulfite sequencing, ChIP-seq and ATAC-seq including tools like Bismark, MACS and Bowtie to study DNA methylation and thereby understand gene regulation in the microbial interaction [33]. Another emerging approach has been using single-cell omics and phylogenomics using tools such as Seurat, Scanpy, MEGA X, ASTRAL and UniFrac for studying cell-specific responses and microbial evolution. Thus, these integrated omics approaches helped to identify important microorganisms with potential applications as biofertilizers, biopesticides, and nutrient cyclers in sustainable agriculture but major challenges involve computational requirements and validation in field experiments.

There are still many challenges in synchronization of data at various molecular levels in the case of integration of multiomics approaches [11]. Further advances in next-generation sequencing technology have provided easy access to genomes and metagenomes. Pan-genomic analysis has started including genomics and phenotypes of mass environmental communities instead of individual organisms. Pan-genomics and metagenomics are significant in carrying out in-depth profiling of microbial communities. This strategy, tailored to genetic diversity, dynamics, and phylogeny at the multiomics level, can be used for studying genomic profiles and functions of culture-independent microbial communities. By combining both strategies, pan-genomes and metagenomes can efficiently reveal the genomics of mass communities and extend the catalogue of diverse microorganisms as compared to traditional cultural methods. The combined strategies of pan-genomics and metagenomics have applications in studies of microbial dynamics, evolution, and function [10].

3.3. Microbial Inoculants and Agricultural Stimulants for Sustainable Agriculture

Microorganisms and their diverse metabolites act as agricultural biostimulants by enhancing plant growth (Table 2). Biostimulants encompass microbial inoculants, organic acids like humic acids and fulvic acids, protein hydrolysates, and amino acids as well as seaweed extracts. The market for biostimulants was expected to reach \$2241 million by 2018, with a compound annual growth rate of 12.5% from 2013–2018 [34].

Biological soil crusts, biocrusts, harbor various microorganisms beneficial for the soil and increase soil fertility. Most of the time, cyanobacteria and microalga initiate biocrust formation, promoting soil stability and nutrient absorption and leading to the development of other microbial communities [35]. In a study, it has been demonstrated that cyanobacteria and microalgae isolated from central Portugal possess strong biostimulant potential. Genomic DNA has been isolated with the Plant/Fungi gDNA isolation kit and amplified by PCR amplification using 16S rRNA for bacteria, 23S rRNA genes for cyanobacteria and the 18S rRNA gene for microalgae followed by purification of the DNA fragments using the NZY Gelpure kit. Sequence data thus obtained has been deposited in

GenBank. Phylogenetic analysis using bioinformatics tools involved NCBI BLASTn and Clustal Omega using the maximum likelihood (ML) method. FastTree has been implemented as general time reversible model. Statistical evaluation has been done using the Akaike information criterion (AIC) and MEGA X for validation. Molecular identification has revealed filamentous cyanobacteria (*Trichocoleus*, *Nodosilinea*, *Microcoleus*), heterocystous *Nostoc* and the microalga *Klebsormidium*, all capable of producing phytohormones and siderophores. Their positive interaction with plant exudates indicates efficient plant-microbe signaling, supporting their application as eco-friendly biostimulants for sustainable agriculture [35].

In one study, the impact of two commercial biostimulants on soil microbiome composition, diversity and physico-chemical properties across five sites was assessed. High-throughput 16S rRNA and ITS amplicon sequencing revealed distinct bacterial and fungal indicator taxa associated with biostimulant treatments. Sequencing data was processed using QIIME 2 software version after removing demultiplexed raw sequences using Cutadapt v2019.7.0 and DADA 2 v2019.7.0. Taxonomic classification was carried out to identify unique amplicon sequence variants using the q2 feature classifier. Moreover, multivariate analysis (Bray–Curtis dissimilarity) has confirmed treatment-driven microbial diversity. Overall, biostimulant application has improved soil health parameters, indicating potential long-term benefits for soil sustainability, crop yield, and produce quality [36].

The intricate relationship between root exudates and microbiome interactions in the rhizosphere has been studied and emphasized the potential of SynCom for engineering the rhizosphere with known microorganisms as a tool for sustainable agriculture [11]. SynCom offers immense benefits such as nutrient cycling, enhanced nutrient uptake and absorption, greater soil fertility, and therefore sustainable agriculture [11]. In a similar study, organic acids such as oxalic acid, lactic acid, and citric acid have been identified as biostimulants in the rhizosphere. Biochemical activity of soil has been confirmed by the presence of phosphatase and dehydrogenase enzymes. Total genomic DNA was extracted from the soil samples using the DNeasy power soil DNA isolation kit. A library was prepared using V3–V4 hypervariable regions using primer pairs using an Illumina-specific platform. The libraries were purified utilizing Mag-Bind and quantified by Qubit dsDNA HS assay. Sequencing data thus obtained was processed using QIIME version 1.9.0. Raw data was demultiplexed and trimmed by Avada pt 1.3, merged by FLASH and filtered using QIIME 1.9.0. Additionally, chimeric sequences were identified and removed by the UCHIME algorithm followed by using the Greengenes database. It has been confirmed that organic acids can change the microbial community, favoring degradation and enhancing soil fertility, predominantly belonging to the *Bacillus* genus and *Micrococcus* sp. The degradation of lactic acid led to the development and induction of specific taxa such as *Sinorhizobium* and *Lysobacter* genera and the *Pseudomonaceae* family. Similarly, with degradation of citric acid, the presence of the *Clostridiaceae* family was confirmed [37].

Plant biostimulants can help crops adapt to abiotic stresses and improve tolerance. To understand the mechanism, root exudates of biostimulants based on *Ascophyllum nodosum* in the rhizosphere of *Zea mays* and other plants have been investigated and shown to influence the motility of *Pseudomonas protegens*, a plant-growth-promoting rhizobacterium. DNA from soil samples were extracted using the E.Z.N.A soil DNA kit (Omega Bio-Tek, USA) and quantified by NanoDrop and V3/V4 region 16S rRNA sequencing was performed. Data analysis was performed using STAMP v 2.1.3 and weighted unifracs of the bacterial community was assessed using Qiime2view. To confirm the action, gene expression analysis of the root exudates was studied, and upregulation of genes associated with chemotaxis (*cheV*, *cheW*), hydrogen cyanide (*hcnA*), pyoverdine (*pvdS*), and pyrrolnitrin (*prnD*) biosynthesis was compared to controls. They were found to significantly alter the abundance of genera such as *Chryseolinea*, *Novosphingobium*,

Kitasatospora, *Pseudoxanthomonas*, *Quadrisphaera*, and *Turneriella*. Concentrations of benzoxazinoids 2,4-dihydroxy-7-methoxy-1,4 benzoxazin-3-one and 6-methoxybenzoxazolinone in maize roots increased by 1.4- and 1.76-fold, respectively [38].

Another study showed that free-living nitrogen-fixing bacteria isolated from sugarcane-associated niches, characterized as *Priestia megaterium*, *Sphingobium yanoikuyae*, and *Burkholderia paludism*, function as a biostimulant consortium to enhance plant growth. The identification of the bacterial isolate was determined by DNA sequence analysis. DNA was extracted from single colonies using the DNeasy Power Food Microbial Kit by QIAGEN, Australia followed by PCR using universal bacterial primers 27F (5' AGAGTTT-GATCMTGGCTCAG 3') and 152R (5' AAGGAGGTGWTCCARCC 3'). The PCR products were sent to the Australian Genome Research Facility (AGRF) for sequencing followed by BLASTn analysis. The rRNA/ITS were searched by using MegaBLAST. Thus, it can be inferred that technology evolves, and omics-based approaches may be used to accelerate sustainable bioprospecting studies for sustainable agriculture, thereby enhancing crop yields. The collective growth-promoting activity highlights the value of microbial consortia in sustainable agriculture, while the concluding insight emphasizes that integrating omics-based approaches can significantly accelerate bioprospecting and optimize microbial applications for improved crop yields and long-term sustainability [39].

Plant-growth-promoting microorganisms exist in complex soil ecosystems and, along with environmental factors, various interactions occur between them, ultimately affecting stability, productivity and resilience. Micropores are the site of microbial growth and gas diffusion. Out of all, only 1% of microorganisms can be cultured and characterized, rendering soil ecosystems unexplored. Metagenomics approaches include 16S rRNA sequencing and assigning functional characteristics leading to phylogenetic diversity. Further multiomics studies have proved the presence of several gene families related to nitrogen and carbon fixation and resistance to environmental stress. Analysis of the structure of microbial communities in soil involves targeted methods, for example, using a diversity index as a quantitative metric representing biodiversity [40]. Three types of metrics can be used: alpha diversity for richness, gamma diversity for the overall number of species in an extensive region and beta diversity, followed by using bioinformatics tools such as Phyloseq and vegan designed to manipulate the data. Vegan is needed for ecological analysis under targeted sequencing. Whole-genome metagenomics can be employed for information related to genetic and genomic diversity and analysis of novel functioning [41].

Another novel approach is culturomics and in situ cultivation techniques enabling the study of unculturable microorganisms. Various methods like soil substrate membrane, diffusion chambers and microfluidic chips can simulate soil conditions. High-throughput techniques can be used to customize the culture media, followed by metabolic profiling using BIOLOG ecoplates integrated with MALDI-TOF, MS, GC-MS, LC-MS and NMR to improve soil health research and climate adaptation [42]. For example, researchers have engineered a culturomics platform by automated microbiome imaging and isolation for carrying out robotic strain isolation and genotyping employing an intelligent imaging algorithm [43]. This advancement has helped the generation of cultured biobanks customized to specific needs and response to climate change adaptation [41].

Functional gene and metabolic pathway annotation involves the identification of microbial genes and prediction of their metabolic functions. Metagenomic studies have revealed the prevalence of Proteobacteria important in nutrient cycling. The annotation process starts with assembling DNA sequence data into contigs using tools like SPAdes and MEGAHIT. Gene prediction tools such as Prodigal, Glimmer and MetaGeneMark identify the open reading frames (ORFs). Functional annotation is performed post gene prediction by comparing the genes with databases like GenBank, KEGG and UniProt using BLAST.

Tools such as Pfam, InterPro and COG help in identification of conserved protein domains and families. Mapper is used to map the metabolic pathways. Metabolic modeling techniques such as flux balance analysis (FBA) and genome-scale metabolic models (GEMs) are used to predict microbial activity. These predictions are validated through lab methods including gene knockouts, transcriptomics, proteomics and metabolomics. Machine learning and advancements in algorithms have improved the accuracy of gene function prediction and pathway reconstruction [44].

Table 2. Screening of various microorganisms beneficial for sustainable agriculture.

Sr. No.	Name and Type of Microorganism	Application in Agriculture	Bioinformatics Tools Used for Analysis	Reference
1.	<i>Trichocoleus</i> , <i>Nodosilinea</i> , <i>Microcoleus</i> (cyanobacteria), <i>Nostoc</i> and <i>Klebsormidium</i> (microalgae)	Biostimulants enhancing the growth of plants by producing phytohormones and siderophores	BLASTn, Clustal omega, FastTree, MegaX	[35]
2.	<i>Ellisterales</i> , <i>Propionibacterials</i> , <i>Solibacterales</i> , <i>Candidatus</i> , <i>Reyranellales</i> and <i>Sphingomonadales</i> as bacterial indicator species <i>Filobasidiales</i> and <i>Pleosporales</i> (fungal sp.)	Biostimulants enhancing soil fertility, crop yield and overall production	QIIME 2, Cutadapt V 2019.7.0, DADA 2 v 2019:7.0 q2 feature classifier	[36]
3.	<i>Bacillus</i> , <i>Micrococcus</i> , <i>Synorhizobium</i> and <i>Lysobacter</i>	Presence of oxalic acid and citric acid with growth promotion activity and adaptation of crops to abiotic stresses	Mag-Bind, Qubitd SDNA QIIME version 1.9.0, AVADA PT 1.3, FLASH, CHIME Algorithm and Greengenes database	[37]
4.	<i>Chryseolinea</i> , <i>Pseudoxanthomonas</i> , <i>Novosphingobium</i> , <i>Quadrifida</i> , <i>Turneriella</i> and <i>Kitscheltospora</i>	Increased concentration of benzoxazinoids in crop plants, increasing immunoprotective activities against pests	STAMP v 2.1.3, QIIME 2view	[38]
5.	<i>Priestia megaterium</i> , <i>Sphingobium yanoikuyae</i> and <i>Burkholderia paludis</i>	Free-living nitrogen-fixing bacteria for enhancing the fertility of soil	BLASTn and Mega-BLAST	[39]

Moreover, advanced high-throughput techniques including big data analysis along with artificial intelligence (AI) can help in better understanding of intricacies of microbial patterns. Rapid growth of high-throughput sequencing techniques has generated a huge amount of microbial data, necessitating advanced and continuously evolving bioinformatics tools to process and interpret these datasets. Parallel growth of artificial intelligence has led to development of powerful tools based on machine learning (ML) and neural network models. These ML/AI-driven models are accelerating the study of complex, unexplored metagenomic data, finding hidden patterns, predicting microbial functions, and improving soil and agricultural sustainability. One of the most applicable strategies of ML in microbiome research is metagenomic binning that leads to accurate recovery of metagenome-assembled genomes. In addition, ML methods are applied for taxonomic identification and functional annotation, improving our understanding of the potential of complex microbial communities [45–47].

It is quite evident that advanced sequencing technologies have played a pivotal role in advancing soil microbiome research and deepening our understanding of complex soil

ecosystems and interactions. Critical evaluation of widely used strategies such as specific ribosomal RNA gene sequencing using modern platforms including PacBio and Oxford Nanopore demonstrated their effectiveness and applications of diverse microorganisms in agriculture. However, the integration of bioinformatics and artificial-intelligence-based analytical tools further enhances data interpretation and analysis, accelerating the comparative assessment of all methods depending on the objective and availability of resources. Thus, the ability of all available techniques to accurately characterize microbial diversity and function provides innovative microbiome-based solutions for sustainable agriculture.

The major challenge in rapidly changing agriculture with advances in artificial intelligence is to provide safe, nutritious and robust food supply with a total increase of 35% to 56% in food demand [48]. Digital and precision agriculture was the topic of discussion at the 30th annual conference of the International Society for Computational Biology (ISCB) [49]. Digital agriculture (DA) is a novel technology utilizing AI/ML approaches for large-scale predictive modeling [50]. DA along with computational biology has been used for monitoring soil characteristics, enabling microbiome-based approaches for sustainable agriculture [51]. The scalability of the agriculture sector in terms of profitability and productivity integration of biological datasets relevant to microbiomes and host genetics along with epidemiological and environmental data poses a significant challenge in DA. However, enhancement of scaling of genomics data, depending on data storage, better annotation and curated sustainable databases, requires more rapid and optimized algorithms. It has also necessitated imparting of training of skills for computational biologists and scientists. Moreover, cost-effectiveness of using computational tools is another major obstacle. The National Artificial Intelligence Research Resource (NAIRR) and artificial intelligence on demand (AI4Europe) by the European Union are steps in the right direction. Lastly, managing finances is another bottleneck and national research workshops, USDA, ARS and NIFA grants, and Horizon Europe 2028–2034 can be used for funding [49].

Multimics technology has advanced the research but major challenges are analytical and ecological limitations. Currently available omics tools fail to detect low-abundance yet functionally significant microorganisms. Future predictions like CRISPR advancements and Cas-13-mediated RNA targeting will expand microbial enhancement. Concurrently, artificial intelligence (AI) models optimize microbe–crop–environment matching by using microbiome data. Inoculant optimization, synthetic microbial consortia (SynCom) and CRISPR-based genome editing can be used for precise agriculture. However, regulatory gaps, ethical concerns and data transparency are some of the major concerns [41]. The application of omics-guided microbial consortia shows enhanced effects as compared to single-strain inoculants, promoting greater biomass yield and stress resilience. Integration of the sequencing data with other omics and clinical data complicates the process, necessitating advanced statistical methods. Advanced sequencing technology has resulted in reduced errors and high accuracy. Development of robust algorithms for error correction and data integration is crucial. Machine learning and AI can be employed for enhancing data analysis capabilities. Using cloud computing, high-performance computing clusters and parallel processing can bridge the gap between the experimental and computational aspects of genomic research [17]. Advancements in technologies like single-cell genomics and transcriptomics have improved our understanding of microbial communities in agriculture and environmental science. Single-cell genomic studies involve individual microbial genome studies without the need for culturing them, helping in discovering unexplored microbial diversity and ecological functions. However, metatranscriptomics analyze active gene expression, revealing real-time metabolic activity. Integration of machine learning (ML) and artificial intelligence (AI) helps to analyze complex datasets and thus transform agriculture.

4. Bioinformatics and Growth in Microbial Databases

It has been observed that there is exceptional growth of massive biological data due to advanced next-generation sequencing technologies. Dealing with biological data, particularly microbial omics data, namely genomics, metagenomics, transcriptomics, and proteomics, is a pressing need requiring proper management, utilization, and updating of the data for future predictions. According to Grand View Research, the global bioinformatics market is expanding and expected to reach USD 28.5 billion by 2030, with a compound annual growth rate of 13.7%.

The research work of comparative genomics, structural genomics, and functional genomics relies on databases. Analyzing similarities and differences between two or more genomes requires database traversal. They are required to compare the characteristics between species, to identify and analyze gene families in crop genomes, to discover stress-inducible promoter regions, etc. Structural genomics involves revealing the three-dimensional structure of genes. DNA sequencing provides data on genome features, coding and non-coding genes, and repetitive elements for which the databases are required. Functional genomics includes the study of gene functions and interactions within regulatory networks, analysis of gene expression, and identification of overexpressed genes which also depends on databases [52]. Phytozome, NCBI, Ensembl Plants, PlantGDB, and UCSC are major web-based genome browsers that have huge databases of genomes. Phytozome is the JGI's Plant Comparative Genomics portal, which lists 439 phytozome genomes (accessed on 26th December 2025).

Here, the role of computer science and engineering in building a specific database is evident with the growing microbial databases.

4.1. Trends in Databases

In computer science and engineering, four database models have been employed for constructing databases, namely the relational model, entity-relationship model, semi-structured model and object-based data model [53]. The evolution of computer science has directly dictated the trajectory of database development, moving from the rigid, tree-like hierarchical and graph-based network models of the 1960s to more flexible systems. A major turning point occurred in the 1970s when E.F. Codd proposed the relational model, providing a mathematical foundation that remains industry-standard today. As the Internet evolved toward the late 20th century, the surge in unstructured data led to the birth of NoSQL and big data systems, allowing for massive horizontal scaling [54]. This evolution has enabled specialized tools like KRAKEN [55] and CLARK [56] which are used to classify microbial genomes. Computer-science-driven algorithms such as Needleman–Wunsch [57], Smith–Waterman [58], as well as BLAST [59] leverage these database foundations to perform global and local sequence alignments for identifying conserved regions in genes and proteins for expression analysis.

The architectural overview of microbial databases for agricultural microbes including bacteria, fungi and archaea can be illustrated as depicted in Figure 2 [60]. In the flowchart, the terminal nodes represent the relational database model, the middle layer highlights the entity-relationship aspect, and the topmost layer identifies the object-oriented functional layer of database utility.

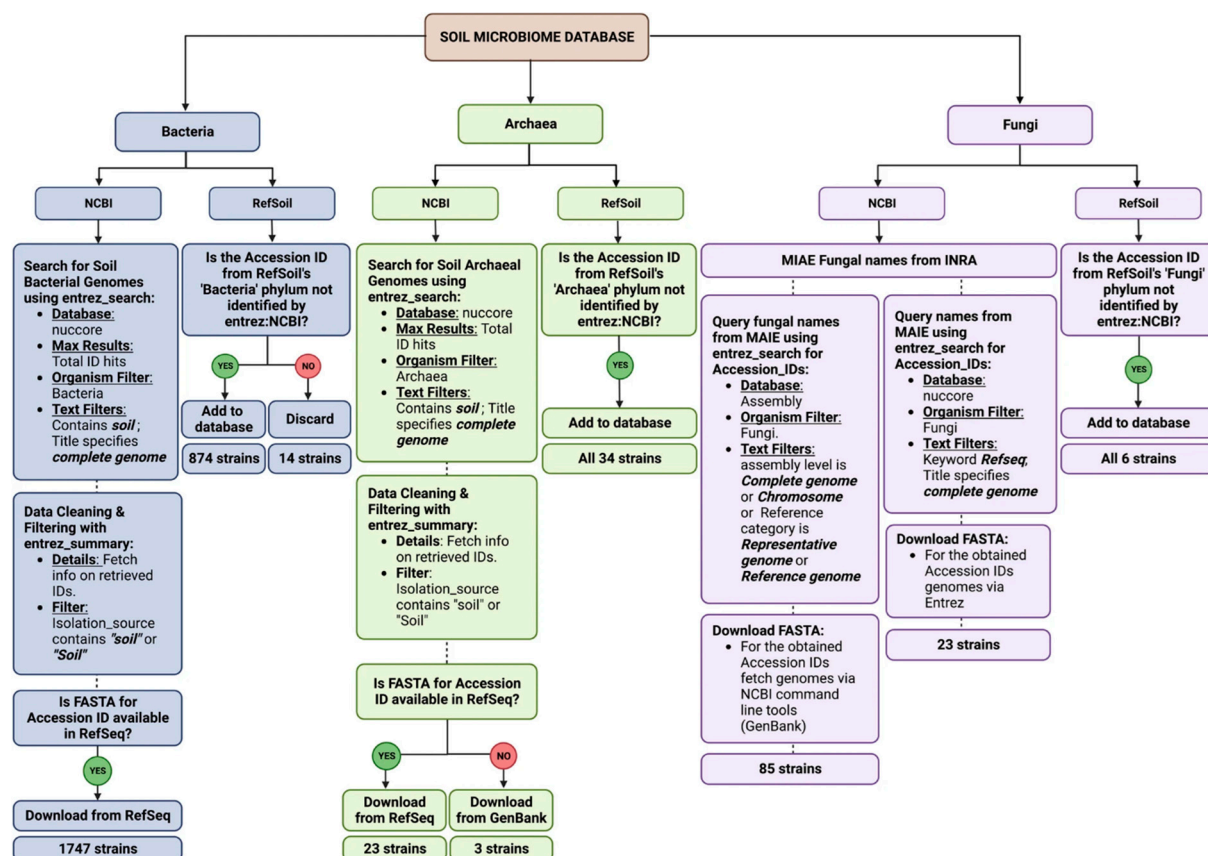


Figure 2. Using microbial databases for exploring the genome selection process (adapted from [60]). This schematic illustrates the steps and criteria employed to curate genomes of bacteria, archaea, and fungi from various databases with potential application in sustainable agriculture.

4.2. Features of Robust Databases

The creation of any database design is rooted in a structured computerized record-keeping system that allows users to update and retrieve information through four levels of abstraction: addressing the why, what, how, and physical existence of the data. This process begins with requirement analysis to define stakeholder constraints, followed by a conceptual stage that uses entity-relationship modeling to distinguish attributes from keys independently of specific software. Once the conceptual design is mapped to a relational model, the focus shifts to normalization, which organizes data to minimize redundancy and prevent anomalies. Finally, the physical level involves data definition languages, performance tuning, and security structures. Similarly, traditional relational databases offer the strength of ACID compliance: atomicity, consistency, isolation, durability, but they historically lacked the horizontal scalability of NoSQL systems [61]. Advanced security features like AES-256 encryption, role-based access control, and SSL/TLS standards now bolster these strengths by protecting sensitive medical and genomic datasets from theft or malicious usage [62].

4.3. Current Status of Microbial Databases

Current trends in database evolution are increasingly focused on environmental sustainability, particularly in mapping soil microbial diversity to improve sustainable agriculture [63].

Along a similar line, targeting bioinformatics analysis, the databases need to be constructed owing to usage by scientific communities, for example, for feasible soil enrichment by microbial actions. Their aim should not be limited to the perspectives of mere data collection. Instead, the focus should be oriented towards their significance in resilient

agriculture and global food security [63]. Certain research highlights the omics-driven potential in growth of plants for sustainable agriculture [1,2].

Researchers proposed a consortium named **AgBioData** for agriculture, supported by sustainable genomics and genetic databases [64]. It lists forty-five genetic, genomics, and breeding databases, and the work is supported to date by the National Science Foundation. There exists a Microbial Genome Database for Comparative Analysis, popularly known as **MBGD**, supported by the National Bioscience Database Center (NBDC), Japan Science Technology Agency (JST), which contains nearly 34k genomes [52]. To support the same cause of soil microbial diversity to improve sustainable agriculture, organizations such as the Indian Council of Agricultural Research (ICAR) manage a sophisticated network of databases, including the **KrishiKosh** open-access repository and the PGR Portal, which focus on germplasm, soil health, and pest management [65]. Current research infrastructure is heavily supported by specialized repositories such as **GenBank**, the RCSB **Protein Data Bank** (PDB), and **Phytozome**, which currently hosts 439 plant genomes [66]. In India, the Advanced Supercomputing Hub for OMICS Knowledge in Agriculture (ASHOKA) provides a high-performance environment with 448 cores and 150 TB of storage to support genomics and proteomics research [67].

4.4. Challenges and Future Perspectives for Current Databases

The present status of database development is defined by a hybrid approach that combines the reliability of relational models with the elastic scalability of the cloud, seen in modern systems like Snowflake and CockroachDB. For example, the integration of machine learning with in-database analytics allows for parallel processing and graph analytics, enabling researchers to perform complex calculations and fraud detection directly where the data resides, significantly accelerating the training of models for scientific discovery.

A significant trend involves using these datasets to examine the functional genomics of soil microbes. By coupling machine learning with these agricultural databases, scientists can now better understand the links between soil microbial ecosystems and environmental stress, creating a “molecular blueprint” for high-yielding crops that can survive harsh weather conditions in different regions.

The strengths and weaknesses of modern databases are often defined by the specific biological or computational problems they aim to solve. For instance, metataxonomics is highly effective for studying a wide range of organisms via ribosomal RNA conservation but fails to capture gene content outside of targeted markers. Conversely, metagenomics data offers the strength of de novo genome assembly and the ability to identify viruses, though it suffers from the weakness of requiring extremely high read counts and extensive reference genomes for classification, which is again dependent on database size.

5. Big Data Analytics (BDA) in Microbial Diversity and Sustainable Agriculture

5.1. Big Data in Microbial Diversity

Big data (BD) in microbial diversity research represents the paradigm shift in agricultural science. The integration of BD will facilitate the handling, management, and utilization of the large volume of data gathered from diverse sources related to microorganisms in soil and plants. Microbial diversity plays a crucial and complex role in regulating plant resistance, ecosystem function, soil health, nutrient regulation, and more. All these factors are essential for the sustainability of agriculture [68,69]. Earlier, researchers used the culture method (inoculation), which involves collecting a small portion of the microorganism and conducting further studies. But with the advent and advancement of

technological developments in sequencing and computational analytics, there is a shift from descriptive to data-driven ecological modeling. It is observed that researchers are now generating substantial amounts of raw data across fields of microbial diversity, such as plant genomics, taxonomy, metagenomics, and metabolomics, to characterize soil properties in terms of microbial structure and function. These large datasets will provide valuable information on microbial contributions and will further build the scientific basis for sustainable agricultural systems [70,71].

Big data refers not only to the amount of data but also the five Vs: volume, velocity, variety, veracity, and value (Figure 3). With advances in technology, it is possible to minimize the time and cost required for sequencing (whole-genome and metagenomic). Before implementing BDA in the agricultural domain, data analysis must be performed on soil samples obtained from different agricultural lands, ranging from traditional monoculture to organic polyculture [72–74]. Certain environmental conditions and different types of datasets provide varying levels of detail, such as metatranscriptomics and metaproteomics, which provide information on microbial activities. Environmental conditions include many parameters, such as soil moisture content, humidity, pH, crop type, season, crop yield, climatic conditions, fertilizers utilized, etc. All these factors will lead to the creation of multiomics datasets. Now, to analyze this complex multiomics dataset, computational tools and techniques such as ML, DL, bioinformatics, modeling, and statistical analysis are needed [74,75].

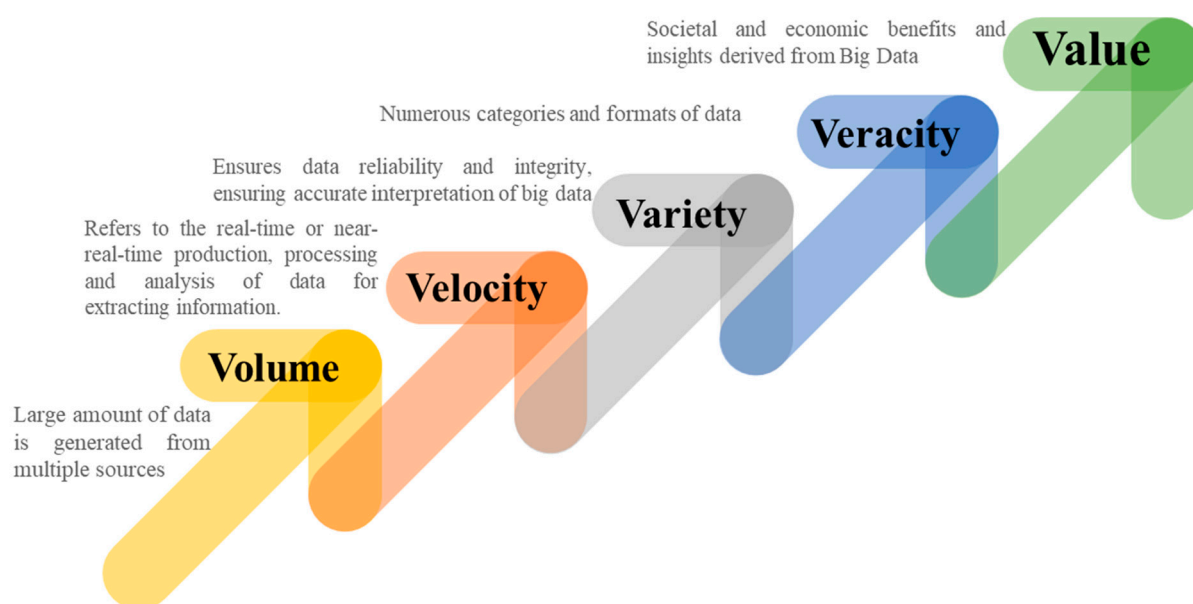


Figure 3. Key Components, i.e., 5V representation in Big Data Analytics.

5.2. Sources of Big Data and Techniques Required to Handle Big Data from Microbiomes

In the context of microbial diversity, the main contribution of BD is to examine the structure of the soil microbiome and its association with markers of soil health. Some key components that impact soil health include water retention, crop yield, fertilizer use, and carbon sequestration. By examining extensive metagenomic databases, researchers and scientists can identify microbial species and their functions (nitrogen fixation, phosphorus fixation, and decomposition of organic matter). Research uses BDA to evaluate and identify the changes observed in microbial communities in terms of their structure and their function by comparing soil samples that have been treated with different crops, organic amendments and reduced tillage. Apart from this, researchers also implement different models to increase agricultural productivity and better understand how microbially

mediated nutrient cycling can be improved by applying various techniques to achieve sustainable practices. Furthermore, by using computational frameworks and techniques such as image processing, GIS, big data, ML, DL, NN, and AI, ecological network analysis provides insights into the associations among protists, fungi, bacteria, and archaea [76,77]. Different big data sources are needed for each agricultural application to address the issue it seeks to solve (Table 3). While GIS and satellite-based remote sensing data are widely employed, information from static databases and datasets is used in almost every agricultural area. Research pertaining to crops, soil, and animals uses field-based ground sensors, while weather stations are used for land mapping, weather and climate change applications, farmer decision making, insurance, and finance [74].

Table 3. Big Data sources and analytical methods in the sector of sustainable agriculture.

Sr. No.	Areas	Sources	Techniques
1.	Microbial biodiversity	GIS geospatial data, historical information and databases like the SER database.	Statistical methods (Bayesian belief networks).
2.	Soil	Ground-based sensors (measuring moisture, salinity, and electrical conductivity), optical cameras, AGRIC soils like historical databases.	ML algorithms such as K-means and farthest first clustering.
3.	Climate variability and weather patterns	Weather stations, satellite remote sensing, geospatial data, historical records, and surveys.	ML algorithms, statistical analysis, modeling, cloud platforms, MapReduce analytics, and GIS geospatial analysis.
4.	Crops	Ground-based metabolites sensors, satellite remote sensing, and historical datasets (including land use and yield statistics).	ML (scalable vector machines, K-means clustering), wavelet filtering, NDVI, Fourier transform.
5.	Remote sensing	Remote sensing (aircraft, drones, satellite) and historical datasets (WMO weather, MODIS surface reflectance, Earth land surface images, radar-altimetry reservoir heights, web-based and geospatial data).	Cloud platforms, GIS, image processing, statistical analysis, NDVI, decision support, big data storage, web/community portals, MapReduce, mobile apps, computer vision, and AI.

5.3. Applications of BDA in Microbial Diversity and Sustainable Agriculture

Another application where BD is essential is rhizosphere-level interaction between microbes and plants. Microbial populations influenced by root exudates and environmental conditions are found in the rhizosphere, a dynamic interface between plant roots and soil. By integrating genetic data from microorganisms and plants, scientists can identify genes that contribute to advantageous interactions, such as tolerance to abiotic stresses (e.g., drought and salinity) or resistance to disease. These findings help build microbiome-informed breeding techniques that select crop types for their capacity to form advantageous microbial associations, in addition to yield traits [52,78].

Big data also helps manage biological pests and diseases, thereby enhancing nutrient efficiency. The yield of the crop is significantly reduced by soil-borne diseases caused by chemical pesticides, leading to ecosystem contamination and harming living creatures. Early identification of harmful microorganisms and their evaluation under different environmental conditions are managed using various predictive modeling techniques that investigate disease development. Furthermore, BDA facilitates the easy identification of antagonistic pathogens that can combat diseases by synthesizing antibiotics. Using pathogens as bioinoculants will benefit sustainable agriculture by lowering the use of chemical inputs without degrading crop health [63,76,78].

One of the great challenges that agriculturists will face is weather conditions, which lead to crop loss and increased pest and disease pressure. In this regard, BD techniques are used to track changes in the composition and function of microorganisms over time and in response to climate change. Finding microbial consortia that are climate-resilient is only possible by collecting datasets across both longitudinal and geographical locations. Integration of pathogenic data with climatic and remote sensing data will facilitate the development of adaptive methods suitable for particular agroecological zones.

Precision agriculture also uses BDA to maximize the efficient use of resources by using the adaptive management method at specific sites. Large datasets of microbial and agronomic sectors, when collected using sensors or GIS mapping and trained with ML models, yield more accurate results. These outcomes will help farmers make well-informed decisions about the optimal and accurate use of fertilizers, forecast crop performance, and assess pesticide and water retention, among other measures [74]. This synergy between microbial diversity and digital technology will help maximize outcomes while reducing environmental impacts (Figure 4).

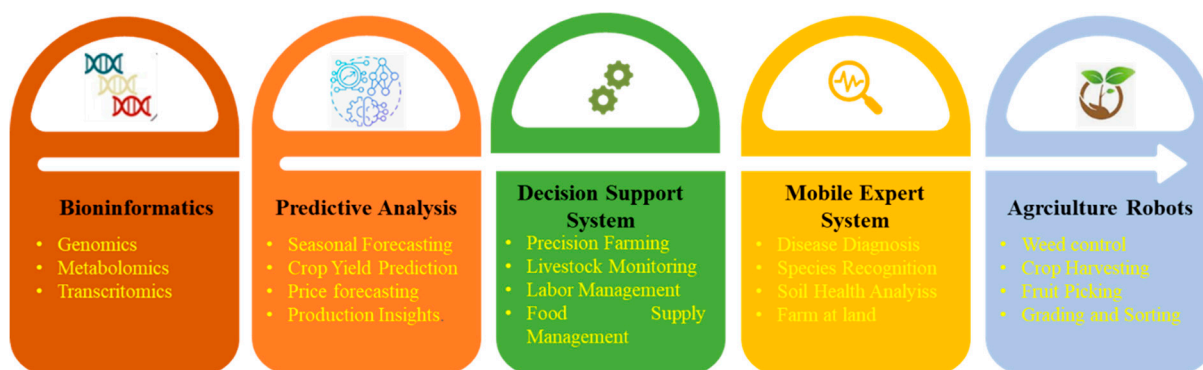


Figure 4. Applications of Big Data Across Different Sectors.

5.4. Case Studies Employing BDA in Sustainable Agriculture

The application of metagenomic network analysis to attract crop-beneficial microbial species in agricultural soils was shown in one of the most significant research projects. Researchers used microbial interaction networks, metabolic pathway reconstruction, and high-throughput sequencing data to find chemicals that can specifically enhance beneficial rhizosphere microorganisms [79]. In order to forecast plant–microbe interactions and enhance soil sustainability, this study emphasized the significance of computational frameworks such as metagenome assembly, genome annotation pipelines, and ecological network modeling. The researchers also demonstrated the use of bioinformatics in precision microbiome engineering by introducing algorithmic pipelines that convert large-scale metagenomic data into practical agricultural solutions [79].

The extensive meta-analysis of European agricultural soil metagenomes, which examined thousands of microbial taxa from various soil habitats, was another significant advancement [80]. The identification of core microbiomes and archaeal groups like Thaumarchaeota that are crucial for nitrogen transformation and nutrient cycling in agricultural fields was made possible by sophisticated bioinformatics tools like metagenome-assembled genomes, taxonomic classifiers, and machine-learning-assisted annotation systems. This study explored how big data frameworks may be used to combine heterogeneous environmental variables in order to examine microbial diversity trends across geographically different farming systems. The work highlighted how contemporary computational metagenomics might reveal functional potential and hidden microbial diversity in agricultural soils that are undetectable by conventional culturing techniques [80].

Similar to this, another case study used shotgun metagenomics and culturomics to examine arid farming systems in the Sinai Peninsula [81]. To describe rhizosphere microbial populations under severe environmental stress, researchers used comparative microbial profiling techniques, functional gene annotation databases, and next-generation sequencing platforms. The results revealed that genes linked to stress tolerance, nutrient uptake, and plant growth promotion are present in microbial taxa that have adapted to arid environments, allowing sustainable agriculture in marginal areas. A significant development in agricultural microbial ecology was made possible by the combination of culturable microbial databases with metagenomic data, which allowed for the functional and taxonomic validation of microbial communities [81].

Another significant addition was termite mound soils, which are recognized as repositories of agriculturally beneficial bacteria [82]. This case study also demonstrated how environmental metagenomics and functional annotation might aid in the identification of naturally existing biofertilizers and microbial inoculants for sustainable farming systems. During the same time, breakthroughs in genomic cataloguing changed soil microbiome research by enabling the creation of large-scale microbial genome archives [82].

A major study reassembled over 40,000 metagenome-assembled genomes from over 3000 soil metagenomes, resulting in one of the most comprehensive genomic inventories of soil microbial diversity [83]. Researchers used cloud computing, large-scale sequence assembly, and high-throughput bioinformatics processes to find previously unknown microbial taxa and functional genes involved in nutrient turnover, carbon sequestration, and soil resilience. These genomic catalogues are crucial resources for agricultural biotechnology because they make it easier to identify beneficial microbial genes that may be used to improve crops and protect the environment [83].

Another pertinent case study investigated the effects of sea sediment treatment on agricultural soils using amplicon sequencing and microbial diversity analysis [84]. Bioinformatics techniques such as QIIME-based taxonomic profiling, diversity measures, and microbial community comparison algorithms helped researchers assess how different soil amendments affect microbial ecosystem composition. The study found that marine sediments altered bacterial diversity and increased nutrient-associated microbial communities, providing insights into sustainable soil management approaches [84]. This study demonstrated the benefits of combining sequencing technologies with statistical and ecological modeling methodologies for evaluating agricultural solutions.

Recently, researchers used machine-learning-based bioinformatics tools to evaluate microbial diversity in important river ecosystems linked to agricultural landscapes [85]. To improve microbial identification and ecological prediction, high-throughput sequencing datasets were evaluated with artificial intelligence algorithms, supervised learning models, and microbial classification systems. The study determined that machine learning methods outperform traditional analytical approaches for big and complicated environmental datasets, notably for finding microbial biomarkers and ecosystem health indicators. Riverine microbial communities have a significant impact on irrigation quality, nutrient flow, and agroecosystem sustainability, making such insights extremely important for agricultural microbial ecosystem study [85].

Another study examined soybean endosphere microbiomes to identify microbial signatures connected with plant health and disease [86]. The researchers used metagenomic sequencing, microbial taxonomic profiling, and functional annotation processes to compare healthy and sick soybean tissues. Bioinformatics analysis revealed changes in microbial community composition and identified beneficial microorganisms associated with disease suppression and increased crop resilience. This case study demonstrated the expanding importance of plant microbiome analytics in precision agriculture, where microbial indicators can predict crop health and optimize disease management tactics [86].

Between 2024 and 2025, the intersection of artificial intelligence, deep learning, and systems biology radically altered agricultural microbial ecosystem research [87]. Literature surveys highlighted the rise of AI-powered microbial analytics capable of analyzing massive multiomics information derived from agricultural soils and plant microbiomes. Neural networks, deep learning algorithms, and computational frameworks inspired by massive language models are increasingly being used by researchers to predict microbial interactions, functional pathways, and ecosystem reactions in the face of environmental change. These AI-based systems tackled significant issues related to microbiological data complexity, sparsity, and non-linearity, allowing for more accurate predictions of soil health and agricultural output [88]. Furthermore, recent research has highlighted the usefulness of computer platforms in modeling microbial ecosystems and forecasting ecological interactions. Researchers have used tools such as COMETS and MetConSIN to simulate microbial metabolic exchanges, nutrient diffusion, and species interactions in agricultural settings. These systems biology frameworks combine genome-scale metabolic models with environmental variables to forecast how microbial communities will respond to changes in soil nutrients, climate, and agricultural practices. These predictive modeling tools are becoming increasingly significant in the design of sustainable farming systems and microbiome-based agricultural interventions.

Concurrently, developments in metagenome-assembled genome analysis and computational metagenomics have improved the ability to reconstruct uncultured microbial genomes from agricultural soils. Reviews of computational tools published during this time period described the rapid progress of assembly algorithms, genome binning software, and annotation platforms capable of handling terabytes of sequencing data [88]. These methods have helped researchers find new microbial species, metabolic capabilities, and ecological activities that are important for nutrient cycling, disease suppression, and environmental resilience in agricultural environments.

Collectively, the case studies discussed above show that bioinformatics techniques and big data analytics are changing agricultural microbial ecosystem research [89]. The combination of metagenomics, machine learning, artificial intelligence, ecological network analysis, and systems biology has provided hitherto unknown insights into microbial diversity, functional connections, and ecosystem dynamics. Through data-driven crop management techniques, predictive soil health analytics, and precision microbiome engineering, these technology advancements are assisting the development of sustainable agriculture practices [89]. To maximize agricultural productivity while maintaining environmental sustainability and microbial biodiversity, future research is anticipated to further combine multiomics information, real-time environmental monitoring, and AI-powered predictive models.

5.5. Challenges of BDA in Microbial Diversity and Sustainable Agriculture

There are some challenges, too, when implementing BDA in microbial diversity (Figure 5). To handle large, diverse sequencing data, a robust data management infrastructure and well-standardized, established procedures and protocols are required. Cross-study comparisons done by various researchers may lead to biased outcomes. This biased result may be due to differences in soil samples, sampling techniques, sequencing platforms, DNA extraction procedures, or the bioinformatics pipelines used. Initial efforts focus on creating open-access repositories and standard procedures and methodologies to maximize the use of microbial data and support further research. Interdisciplinary research in ecology, computational science, microbiology, plant genomics, etc. is encouraged through collaborative work, which provides a useful framework for data exchange and reproducibility [63,90].

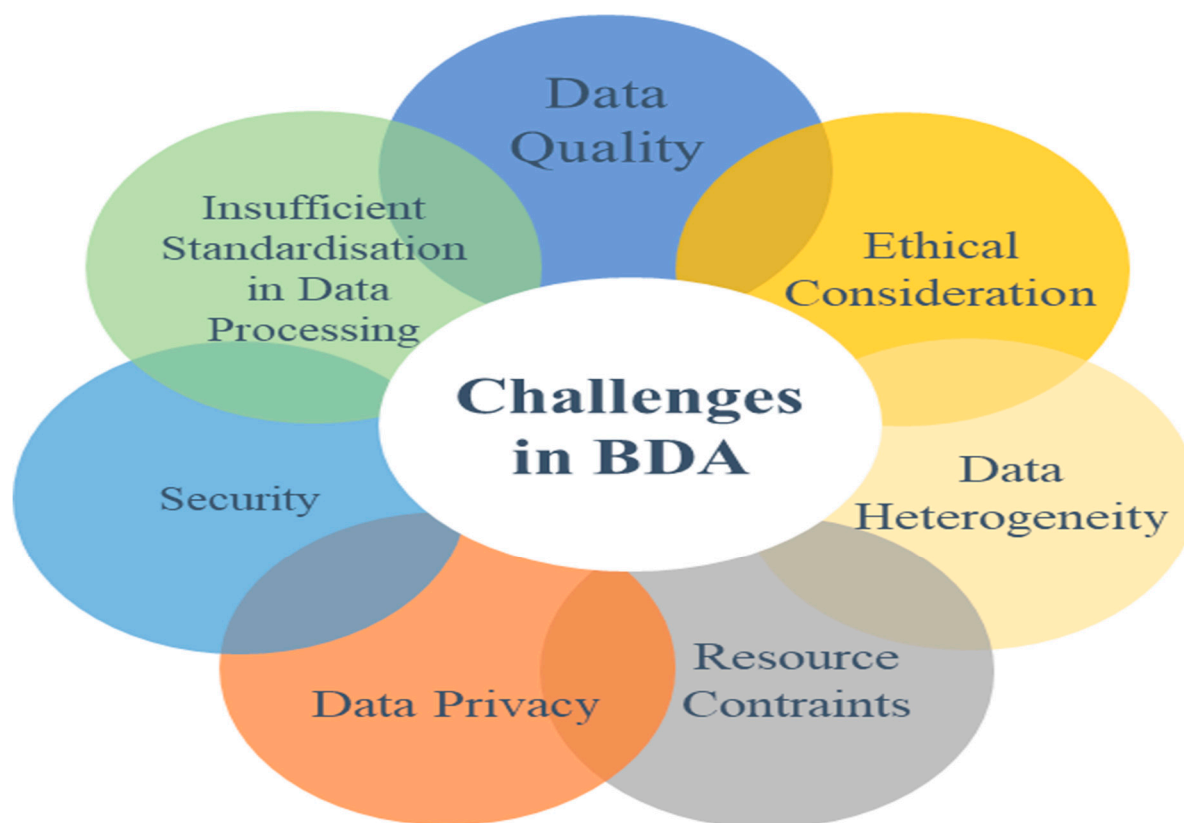


Figure 5. Challenges in implementing BDA in agricultural sector.

Advancements in the area of microbiome-based agriculture also raise ethical and socioeconomic issues. The adoption of new technology in developing nations may hinder small farmers, whereas large agricultural industries and research laboratories use advanced technologies for sequencing and computer resources. Ensuring access to microbial technology aligns with broader sustainability goals, such as rural development and food security [63,75].

5.6. Future Perspective and Conclusion

The key focus of future study in microbiome-based agriculture research is to shift from descriptive to predictive, data-driven modeling. Advances in AI, ML, DL, and bioinformatics lead to the development of a mechanistic model that relates to the dynamics of microbial communities. Research can use biofertilizers and biopesticides to make a synthetic microbial consortium using data-driven methods. In addition, incorporating multi-omics data with environmental factors, along with their phenotypic datasets, will improve understanding of resilience and redundancy for researchers in the microbial sector. Further increasing computational power may also improve the model's ability to capture microbial interactions under various conditions.

In this context, we can conclude that advances in big data are crucial for microbiome-based sustainable agriculture research. By using advanced sequencing methods, tools, and techniques, computational analysis, and integrating multiomics data, researchers can decipher and understand the complex associations that exist between pathogens, plants, soil, and environmental conditions for crop production. Understanding these factors and the strategies implemented to achieve the goal will improve soil health, meet nutrient requirements, inactivate pathogens, and strengthen climatic resilience. Still, there are issues in implementing BDA in microbial datasets, such as heterogeneous data, varied data formats, data standardization, infrastructure, and resource utilization. But this can be one of the directions in microbiome science in which researchers can proceed to develop accurate

and reliable sustainable agricultural systems. Using microbial big data is not just a scientific breakthrough but also a strategic necessity for achieving long-term agricultural sustainability, as the world's food supply demands rise amid environmental limits.

6. Conclusion and Future Direction

Bioinformatics and big data analytics are the frontier domains in the agriculture sector. These tools are the backbone for processing massive biological data that is indispensable for sustainable agriculture. Application of bioinformatics and big data analytics in the era of NGS is a key driving force to understand and predict the biological phenomena at the plant–microbe interface in the agriculture ecosystem setting.

The first key challenge associated with bioinformatics and big data analytics is the development of analytical techniques to process the massive diverse datasets obtained from agricultural ecosystems. Thus, robust algorithms and methods are needed to understand and process the complex datasets emanating from genomics patterns, nutrient recyclability, climatic patterns, and resource availability in agricultural ecosystems. An interdisciplinary approach based on teams of computer science engineers, biologists, and statisticians is needed to translate the data into practical solutions.

The other challenge is the restriction of data sharing due to competition in the agriculture sector. Proprietary datasets have been utilized by specific industries and this will hinder big data analytics in the agriculture sector. This will encourage public–private partnerships to share the data emanating from the agriculture sector and apply big data analytics.

Undoubtedly, bioinformatics and big data on microbial biodiversity will improve the understanding of complex systems for the betterment of agricultural ecosystems. The integration of emerging technologies with traditional agricultural practices will face synchronization issues at the ground level. Thus, it may be concluded that bioinformatics and big data analytics in integrated microbe–agriculture ecosystems opens up a plethora of opportunities for better crop yield and sustainable ecosystems.

Author Contributions: M.L.V.: conceptualization, methodology, software, data curation, formal analysis, validation, writing—original draft. M.T.: software, data curation, writing—original draft, writing—review and editing. P.A.: visualization, writing—review and editing, formal analysis, writing—original draft. P.S.: formal analysis, writing—original draft. S.D.: validation, formal analysis, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The authors have nothing to report.

Acknowledgments: Authors are grateful to the Director of the Indian Institute of Information Technology Una, Himachal Pradesh, India, for providing the necessary facilities to pursue the present work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sahoo, A.; Yadav, G.; Mehta, T.; Meena, M.; Swapnil, P. Omics-Driven Insights into Plant Growth-Promoting Microorganisms for Sustainable Agriculture. *Discov. Sustain.* **2025**, *6*, 659. <https://doi.org/10.1007/s43621-025-01582-2>.
2. Renu; Gupta, S.; Sharma, T.R. (Eds.) Recent Advances in Methods and Techniques of Meta-omics for Crop Improvement. In *Meta-Omics in Crop Improvement*; Springer: Singapore, 2025; pp. 1–12. https://doi.org/10.1007/978-981-95-2294-1_1.

3. Reverchon, F.; Rosaura, G.; Alfaro-García, J.; Méndez-Bravo, G.A.A.; Monribot-Villanueva, J.L. Unraveling Plant–Microbe Interactions with Multiomics: State of the Art and Opportunities for Sustainable Agriculture. In *Soil Microorganisms for Plant Growth Promotion and Soil Health*; Santoyo, G., Ed.; Elsevier: Amsterdam, The Netherlands, 2026; pp. 529–555. <https://doi.org/10.1016/B978-0-443-34055-0.00025-4>.
4. Saggu, S.K.; Kumar, M.; Kumar, S. Metagenomics and Its Impact on Environmental and Therapeutic Microbiology. *Arch. Microbiol.* **2026**, *208*, 24. <https://doi.org/10.1007/s00203-025-04588-6>.
5. Chauhan, S.; Chen, C.W.; Dong, C.D. Synergistic Bioremediation of Polycyclic Aromatic Hydrocarbons and Heavy Metals: A Comprehensive Review. *Rev. Environ. Sci. Biotechnol.* **2026**, *25*, 9. <https://doi.org/10.1007/s11157-025-09759-x>.
6. Miglani, G.; Kaur, M. Multi-Omics and Crop Improvement for Sustainable Agriculture. In *Sustainable Landscape Planning and Natural Resources Management: Omics and Genome Editing*; Springer: Cham, Switzerland, 2025; pp. 15–35. https://doi.org/10.1007/978-3-031-81639-0_2.
7. Reznikova, D.A.; Barannikova, M.V.; Shnakhova, L.M.; Mitkin, N.A.; Vatlin, A.A. Next-Generation Sequencing Approaches for Soil Microbiome Research. *Front. Soil Sci.* **2026**, *5*, 1706999. <https://doi.org/10.3389/fsoil.2025.1706999>.
8. Nadarajah, K.; Abdul Rahma, N.S.N. The Microbial Connection to Sustainable Agriculture. *Plants* **2023**, *12*, 2307. <https://doi.org/10.3390/plants12122307>.
9. Aghahari, V.; Aghahari, V.; Chou, M.L.; Chew, C.H.; Noll, J.; Burnouf, T. Intelligent Micro-/Nanorobots as Drug and Cell Carrier Devices for Biomedical Therapeutic Advancement: Promising Development Opportunities and Translational Challenges. *Biomaterials* **2020**, *260*, 120163. <https://doi.org/10.1016/j.biomaterials.2020.120163>.
10. Zhong, C.; Chen, C.; Wang, L.; Ning, K. Integrating Pan-Genome with Metagenome for Microbial Community Profiling. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1458–1466. <https://doi.org/10.1016/j.csbj.2021.02.021>.
11. Yusuf, A.; Li, M.; Zhang, S.-Y.; Odedishemi-Ajibade, F.; Luo, R.-F.; Wu, Y.-X.; Zhang, T.-T.; Ugya, Y.A.; Zhang, Y.; Duan, S. Harnessing Plant–Microbe Interactions: Strategies for Enhancing Resilience and Nutrient Acquisition for Sustainable Agriculture. *Front. Plant Sci.* **2025**, *16*, 1503730. <https://doi.org/10.3389/fpls.2025.1503730>.
12. Kamboj, S.S.; Dinkar, S.; Kumar, S.; Girish, M.; Deepak, K.; Singh, S. Soil Microbiome and Nutrient Cycling: Implications for Sustainable Agriculture. *J. Sci. Res. Rep.* **2025**, *31*, 241–251. <https://doi.org/10.9734/jsrr/2025/v31i93487>.
13. Purohit, H.J.; Pandit, P.; Pal, R.; Warke, R.; Warke, G.M. Soil Microbiome: An Intrinsic Driver for Climate Smart Agriculture. *J. Agric. Food Res.* **2024**, *18*, 101433. <https://doi.org/10.1016/j.jafr.2024.101433>.
14. Ebeed, H.T.; Ceasar, S.A. Genomics and Bioinformatics Applications in Crop Stress: Unraveling Molecular Mechanisms for Sustainable Agriculture. *S. Afr. J. Bot.* **2024**, *170*, 38–47. <https://doi.org/10.1016/j.sajb.2024.05.015>.
15. Mishra, D.P.; Badajena, J.C.; Nayak, S.K.; Baliyarsingh, B. Bioinformatics: A Tool for Sustainable Agriculture. In *Advances in Agricultural and Industrial Microbiology*; Nayak, S.K.; Baliyarsingh, B.; Singh, A.; Mannazzu, I.; Mishra, B.B., Eds.; Springer: Singapore, 2022. https://doi.org/10.1007/978-981-16-9682-4_13.
16. Chetal, G.; Verma, M.L. In Silico Approaches for the Quest of Novel Enzymes. In *Value-Addition in Agri-Food Industry Waste Through Enzyme Technology*; Academic Press: Cambridge, MA, USA, 2023; pp. 65–78. <https://doi.org/10.1016/B978-0-323-89928-4.00017-1>.
17. Dweh, T.J.; Pattnaik, S.; Sahoo, J.P. Assessing the Impact of Metagenomic Tools on Current Cutting-Edge Genome Engineering and Technology. *Int. J. Biochem. Mol. Biol.* **2023**, *14*, 62–75. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10509535/> (accessed on).
18. Tao, Y.; Xun, F.; Zhao, C.; Mao, Z.; Li, B.; Xing, P.; Wu, Q.L. Improved Assembly of Metagenome-Assembled Genomes and Viruses in Tibetan Saline Lake Sediment by HiFi Metagenomic Sequencing. *Microbiol. Spectr.* **2023**, *11*, e0332822. <https://doi.org/10.1128/spectrum.03328-22>.
19. Nafea, A.M.; Wang, Y.; Wang, D.; Salama, A.M.; Aziz, M.A.; Xu, S.; Tong, Y. Application of Next-Generation Sequencing to Identify Different Pathogens. *Front. Microbiol.* **2024**, *14*, 1329330. <https://doi.org/10.3389/fmicb.2023.1329330>.
20. Larson, N.B.; Oberg, A.L.; Adjei, A.A.; Wang, L. A Clinician’s Guide to Bioinformatics for Next-Generation Sequencing. *J. Thorac. Oncol.* **2023**, *18*, 143–157. <https://doi.org/10.1016/j.jtho.2022.11.006>.
21. Notario, E.; Visci, G.; Fosso, B.; Gissi, C.; Tanaskovic, N.; Rescigno, M.; Marzano, M.; Pesole, G. Amplicon-Based Microbiome Profiling: From Second- to Third-Generation Sequencing for Higher Taxonomic Resolution. *Genes* **2023**, *14*, 1567. <https://doi.org/10.3390/genes14081567>.
22. Nwachukwu, B.C.; Babalola, O.O. Metagenomics: A Tool for Exploring Key Microbiomes with the Potential for Improving Sustainable Agriculture. *Front. Sustain. Food Syst.* **2022**, *6*, 886987. <https://doi.org/10.3389/fsufs.2022.886987>.

23. Ermini, L.; Driguez, P. The Application of Long-Read Sequencing to Cancer. *Cancers* **2024**, *16*, 1275. <https://doi.org/10.3390/cancers16071275>. PMID: 38610953; PMCID: PMC11011098.
24. De Coster, W.; D'Hert, S.; Schultz, D.T.; Cruts, D.T.; Van Broeckhoven, C. NanoPack: Visualizing and Processing Long-Read Sequencing Data. *Bioinformatics* **2018**, *34*, 2666–2676. <https://doi.org/10.1093/bioinformatics/bty149>.
25. Su, Y.; Yu, Z.; Jin, S.; Ai, Z.; Yuan, R.; Chen, X.; Xue, Z.; Guo, Y.; Chen, D.; Liang, H.; et al. Comprehensive Assessment of mRNA Isoform Detection Methods for Long-Read Sequencing Data. *Nat. Commun.* **2024**, *15*, 3972. <https://doi.org/10.1038/s41467-024-48117-3>.
26. Perdomo, J.E.; Ahsan, M.U.; Liu, Q.; Fang, L.; Wang, K. LongReadSum: A Fast and Flexible Quality Control and Signal Summarization Tool for Long-Read Sequencing Data. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 556–563. <https://doi.org/10.1016/j.csbj.2025.01.019>.
27. Lo, L.S.H.; Xu, Z.; Lee, S.S.; Lau, W.K.; Qiu, J.-W.; Liu, H.; Qian, P.-Y.; Cheng, J. How Elevated Nitrogen Load Affects Bacterial Community Structure and Nitrogen Cycling Services in Coastal Water. *Front. Microbiol.* **2022**, *13*, 1062029. <https://doi.org/10.3389/fmicb.2022.1062029>.
28. Aplakidou, E.; Vergoulidis, N.; Chasapi, M.; Venetsianou, N.K.; Kokoli, M.; Panagiotopoulou, E.; Iliopoulos, I.; Karatzas, E.; Pafilis, E.; Georgakopoulos-Soares, I.; et al. Visualizing Metagenomic and Metatranscriptomic Data: A Comprehensive Review. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 2011–2033. <https://doi.org/10.1016/j.csbj.2024.04.060>.
29. Tong, Y.; Zheng, X.; Hu, Y.; Wu, J.; Liu, H.; Deng, Y.; Lv, W.; Yao, H.; Chen, J.; Ge, T. Root Exudate Mediated Plant–Microbiome Interactions Determine Plant Health during Disease Infection. *Agric. Ecosyst. Environ.* **2024**, *370*, 109056. <https://doi.org/10.1016/j.agee.2024.109056>.
30. Jain, S.; Sarsaiya, S.; Singh, R.; Gong, Q.; Wu, Q.; Shi, J. Omics Approaches in Understanding the Benefits of Plant–Microbe Interactions. *Front. Microbiol.* **2024**, *15*, 1391059. <https://doi.org/10.3389/fmicb.2024.1391059>.
31. Crandall, S.G.; Gold, K.M.; Jiménez-Gasco, M.M.; Filgueiras, C.C.; Willett, D.S. A Multi-Omics Approach to Solving Problems in Plant Disease Ecology. *PLoS ONE* **2020**, *15*, e0237975. <https://doi.org/10.1371/journal.pone.0237975>.
32. Dubey, R.K.; Tripathi, V.; Prabha, R.; Chaurasia, R.; Singh, D.P.; Rao, C.S.; El-Keblawy, A.; Abhilash, P.C. Metatranscriptomics and Metaproteomics for Microbial Communities Profiling. In *Unravelling the Soil Microbiome*; Springer Briefs in Environmental Science; Springer: Cham, Switzerland, 2020; pp. 51–60. https://doi.org/10.1007/978-3-030-15516-2_5.
33. Samantara, K.; Shiv, A.; de Sousa, L.L.; Sandhu, K.S.; Priyadarshini, P.; Mohapatra, S.R. A Comprehensive Review on Epigenetic Mechanisms and Application of Epigenetic Modifications for Crop Improvement. *Environ. Exp. Bot.* **2021**, *188*, 104479. <https://doi.org/10.1016/j.envexpbot.2021.104479>.
34. Calvo, F.P.; Nelson, L.; Kloepper, J.W. Agricultural Uses of Plant Biostimulants. *Plant Soil* **2014**, *383*, 3–41. <https://doi.org/10.1007/s11104-014-2131-8>.
35. Roque, J.; Brito, Â.; Rocha, M.; Pissarra, J.; Nunes, T.; Bessa, M.; Vieira, J.; Vieira, C.P.; Melo, P.; Tamagnini, P. Isolation and Characterization of Soil Cyanobacteria and Microalgae and Evaluation of Their Potential as Plant Biostimulants. *Plant Soil* **2023**, *493*, 115–136. <https://doi.org/10.1007/s11104-023-06217-x>.
36. Wadduwage, J.; Liu, H.; Egidi, E.; Singh, B.K.; Macdonald, C.A. Effects of Biostimulant Application on Soil Biological and Physicochemical Properties: A Field Study. *J. Sustain. Agric. Environ.* **2023**, *2*, 285–300. <https://doi.org/10.1002/sae2.12057>.
37. Macias-Benitez, S.; Garcia-Martinez, A.M.; Caballero Jimenez, P.; Gonzalez, J.M.; Tejada, M.T.; Parrado Rubio, M.T.J. Rhizospheric Organic Acids as Biostimulants: Monitoring Feedbacks on Soil Microorganisms and Biochemical Properties. *Front. Plant Sci.* **2020**, *11*, 633. <https://doi.org/10.3389/fpls.2020.00633>.
38. Selvaraj, V.K.; Patel, J.S.; Hui, J.P.M.; Zhang, J.; Berrué, F.; Prithiviraj, B. *Ascophyllum nodosum* Based Plant Biostimulant Shapes the Bacterial Community in the Rhizosphere of Corn. *BMC Plant Biol.* **2025**, *25*, 1263. <https://doi.org/10.1186/s12870-025-07270-7>.
39. Buisset, M.; Soust, M.; Scott, P.T. The Isolation of Free-Living Nitrogen-Fixing Bacteria and the Assessment of Their Potential to Enhance Plant Growth in Combination with a Commercial Biostimulant. *Microbiol. Res.* **2025**, *16*, 69. <https://doi.org/10.3390/microbiolres16030069>.
40. Regueira-Iglesias, A.; Balsa-Castro, C.; Blanco-Pintos, T.; Tomás, I. Critical Review of 16S rRNA Gene Sequencing Workflow in Microbiome Studies: From Primer Selection to Advanced Data Analysis. *Mol. Oral Microbiol.* **2023**, *38*, 347–399. <https://doi.org/10.1111/omi.12434>.
41. Clagnan, E.; Costanzo, M.; Visca, A.; Di Gregorio, L.; Tabacchioni, S.; Colantoni, E.; Sevi, F.; Sbarra, F.; Bindo, A.; Nolfi, L.; et al. Culturomics- and Metagenomics-Based Insights into the Soil Microbiome Preservation and Application for Sustainable Agriculture. *Front. Microbiol.* **2024**, *15*, 1473666. <https://doi.org/10.3389/fmicb.2024.1473666>.

42. Koner, S.; Chen, J.-S.; Hsu, B.-M.; Rathod, J.; Huang, S.-W.; Chien, H.-Y.; Hussain, B.; Chan, M.W. Depth-Resolved Microbial Diversity and Functional Profiles of Trichloroethylene-Contaminated Soils for Biolog EcoPlate-Based Biostimulation Strategy. *J. Hazard. Mater.* **2022**, *424*, 127266. <https://doi.org/10.1016/j.jhazmat.2021.127266>.
43. Huang, Y.; Sheth, R.U.; Zhao, S.; Cohen, L.A.; Dabaghi, K.; Moody, T.; Sun, Y.; Ricaurte, D.; Richardson, M.; Velez-Cortes, F.; et al. High-Throughput Microbial Culturomics Using Automation and Machine Learning. *Nat. Biotechnol.* **2023**, *41*, 1424–1433. <https://doi.org/10.1038/s41587-023-01674-2>.
44. Patra, D.M.; Mohanty, S.; Subhadarsani, S.; Sahoo, J. Harnessing the Plant Microbiome: A Genomic and Metagenomic Perspective for Sustainable Agriculture. *Acad. Mol. Biol. Genomics* **2025**, *2*, 1–11. <https://doi.org/10.20935/AcadMolBioGen7892>.
45. Liang, Q.; Bible, P.W.; Liu, Y.; Zou, B.; Wei, L. DeepMicrobes: Taxonomic Classification for Metagenomics with Deep Learning. *NAR Genom. Bioinform.* **2020**, *2*, lqaa009. <https://doi.org/10.1093/nargab/lqaa009>.
46. Liu, M.; Li, Y.; Li, H. Deep Learning to Predict the Biosynthetic Gene Clusters in Bacterial Genomes. *J. Mol. Biol.* **2022**, *434*, 167597. <https://doi.org/10.1016/j.jmb.2022.167597>.
47. Antonielli, L.; Pfeiffer, S.; Sessitsch, A. Microbial Community Prediction in Plant-Soil Systems Using Machine Learning. 2024. Available online: <https://ai4soilhealth.eu/event/international-conference-artificial-intelligence-for-soil-health/> (accessed on 9 May 2026).
48. van Dijk, M.; Morley, T.; Rau, M.L.; Saghai, Y. A Meta-Analysis of Projected Global Food Demand and Population at Risk of Hunger for the Period 2010–2050. *Nat. Food* **2021**, *2*, 494–501. <https://doi.org/10.1038/s43016-021-00322-9>.
49. Gomes-Neto, J.C.; Crook, A.; Hestrin, R.; Li, G.; Liew, C.S.; Rosa, G.; Singh, K.D.; Tuggle, C.K.; Summers, K.L.; Valdes, C.; et al. Challenges and Opportunities: Computational Biology and the Future of Agriculture. *Bioinform. Adv.* **2026**, *6*, vbag003. <https://doi.org/10.1093/bioadv/vbag003>. PMID: 41716220; PMCID: PMC12916170.
50. Hoyos-Villegas, V.; Jackson, M.; Vargas-Cedeño, M.; Farmer, E.E.; Hanneman, M.; Mazis, A.; Singh, K.D.; Sangjan, W.; McNair, M.; Sankaran, S.; et al. Affordable Phenomics: Expanding Access to Enhancing Genetic Gain in Plant Breeding. *Plant Phenom. J.* **2025**, *8*, e70034. <https://doi.org/10.1002/ppj2.70034>.
51. Rajak, P.; Ganguly, A.; Adhikary, S.; Bhattacharya, S. Internet of things and smart sensors in agriculture: Scopes and challenges. *J. Agric. Food Res.* **2023**, *14*, 100776. <https://doi.org/10.1016/j.jafr.2023.100776>.
52. Uchiyama, M.; Mihara, M.; Nishide, H.; Chiba, H.; Takayanagi, M.; Kawai, M.; Takami, H. MBGD: Microbial genome database for comparative analysis featuring enhanced functionality to characterize gene and genome functions through large-scale orthology analysis. *J. Mol. Biol.* **2025**, *437*, 168957. <https://doi.org/10.1016/j.jmb.2025.168957>.
53. Silberschatz, A.; Korth, H.F.; Sudarshan, S. *Database System Concepts*; McGraw-Hill: New York, 2002; Volume 5.
54. Elmasri, R.; Navathe, S.B. *Fundamentals of Database Systems*, 7th ed.; Pearson Education: Bengaluru, India, 2016. Available online: <https://surl.li/cvrafb> (accessed on 9 May 2026).
55. Lu, J.; Salzberg, S.L. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* **2020**, *8*, 124. <https://doi.org/10.1186/s40168-020-00900-2>. PMID: 32859275; PMCID: PMC7455996.
56. Ounit, R.; Wanamaker, S.; Close, T.J.; Lonardi, S. CLARK: Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genom.* **2015**, *16*, 236. <https://doi.org/10.1186/s12864-015-1419-2>.
57. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453. [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4).
58. Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.* **1981**, *147*, 195–197. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
59. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
60. Edwin, N.R.; Fitzpatrick, A.H.; Brennan, F.; Abram, F.; O’Sullivan, O. An in-depth evaluation of metagenomic classifiers for soil microbiomes. *Environ. Microbiome* **2024**, *19*, 19. <https://doi.org/10.1186/s40793-024-00561-w>.
61. Dritsas, E.; Trigka, M. A survey on database systems in the big data era: Architectures, performance, and open challenges. *IEEE Access* **2025**, *13*, 95068–95084. <https://doi.org/10.1109/ACCESS.2025.3572059>.
62. Ganesh, R.; Khan, B.U.I.; Khan, A.R.; Kamsin, A.B. A panoramic survey of the advanced encryption standard: From architecture to security analysis, key management, real-world applications, and post-quantum challenges. *Int. J. Inf. Secur.* **2025**, *24*, 216. <https://doi.org/10.1007/s10207-025-01116-x>.
63. Parveen, S.; Shafi, Z.; Shahid, M.; Iqbal, M.Z.; Naznine, F.; Ansari, M.I. Omics-driven insights into soil microbial diversity and phytopathogen interactions for sustainable agriculture and food security. *J. Basic Microbiol.* **2026**, *66*, e70155. <https://doi.org/10.1002/jobm.70155>.

64. Harper, L.; Campbell, J.; Cannon, E.K.S.; Jung, S.; Poelchau, M.; Walls, R.; Andorf, C.; Arnaud, E.; Berardini, T.Z.; Birkett, C.; et al. AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture. *Database* **2018**, 2018, bay088. <https://doi.org/10.1093/database/bay088>.
65. Veeranjanyulu, K. Krishikosh: An institutional repository of National Agricultural Research System in India. *Library Manag.* **2014**, 35, 345–354. <https://doi.org/10.1108/LM-08-2013-0072>.
66. Chavhan, R.L.; Hinge, V.R.; Wankhade, D.J.; Deshmukh, A.S.; Mahajan, N.; Kadam, U.S. Bioinformatics for molecular breeding and enhanced crop performance: Applications and perspectives. In *Bioinformatics for Plant Research and Crop Breeding*; Wiley: Hoboken, NJ, USA, 2024; pp. 21–74. <https://doi.org/10.1002/9781394209965.ch2>.
67. ASHOKA. 2024 Indian Council of Agricultural Research (ICAR). Ashoka Brochure. 2024. Available online: https://icar.org.in/sites/default/files/Circulars/Brochure_ASHOKA_A4.pdf (accessed on 9 May 2026).
68. Hussein, J.B.; Workneh, T.S.; Kassim, A.; Ntsowe, K.; Melesse, S.F.; El-Mesery, H.S. A review on the impact of big data analytics in transforming agricultural practices, food processing, and preservation strategies. *Appl. Food Res.* **2025**, 5, 101234. <https://doi.org/10.1016/j.afres.2025.101234>.
69. Osinga, S.A.; Paudel, D.; Mouzakitis, S.A.; Athanasiadis, I.N. Big data in agriculture: Between opportunity and solution. *Agric. Syst.* **2022**, 195, 103298. <https://doi.org/10.1016/j.agsy.2021.103298>.
70. Navas-Molina, J.A.; Hyde, E.R.; Sanders, J.G.; Knight, R. The microbiome and big data. *Curr. Opin. Syst. Biol.* **2017**, 4, 92–96. <https://doi.org/10.1016/j.coisb.2017.07.003>.
71. Pandey, N. Artificial Intelligence and Big Data Analysis. In *Microbes of Medical Importance*; IIP Series: Fresno, CA, USA, 2024; Volume 4, pp. 569–581. <https://www.doi.org/10.58532/nbennurmmch24>.
72. Lobanov, V.; Gobet, A.; Joyce, A. Ecosystem-specific microbiota and microbiome databases in the era of big data. *Environ. Microbiome* **2022**, 17, 37. <https://doi.org/10.1186/s40793-022-00433-1>.
73. Vohra, S. Big data analytics in bioinformatics: Transforming genomics and proteomics for advanced biological insights. In *2024 International Conference on Big Data Analytics in Bioinformatics (DABCon)*; IEEE: New York, NY, USA, 2024; pp. 1–5. <https://doi.org/10.1109/DABCon63472.2024.10919359>.
74. Delgado, A.; Short, N.M., Jr.; Roberts, D.P.; Vandenberg, B. Big data analysis for sustainable agriculture on a geospatial cloud framework. *Front. Sustain. Food Syst.* **2019**, 3, 54. <https://doi.org/10.3389/fsufs.2019.00054>.
75. Bhat, S.A.; Huang, N.F. Big data and AI revolution in precision agriculture: Survey and challenges. *IEEE Access* **2021**, 9, 110209–110222. <https://doi.org/10.1109/ACCESS.2021.3102227>.
76. Zaborowicz, M.; Frankowski, J. Big data analytics and machine learning for smart agriculture. *Agriculture* **2025**, 15, 757. <https://doi.org/10.3390/agriculture15070757>.
77. Ahmed, N.; Shakoor, N. Advancing agriculture through IoT, big data, and AI: A review of smart technologies enabling sustainability. *Smart Agric. Technol.* **2025**, 10, 100848. <https://doi.org/10.1016/j.atech.2025.100848>.
78. Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F.X. A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* **2017**, 143, 23–37. <https://doi.org/10.1016/j.compag.2017.09.037>.
79. Sasse, M.H.B.; de Oliveira, J.C.P.; Berihu, M.; Somera, T.S.; Malik, A.; Medina, S.; Piombo, E.; Tal, O.; Cohen, M.; Ginatt, A.; et al. A framework for the targeted recruitment of crop-beneficial soil taxa based on network analysis of metagenomics data. *Microbiome* **2023**, 11, 8. <https://doi.org/10.1186/s40168-022-01438-1>.
80. Nelkner, J.; Huang, L.; Lin, T.W.; Schulz, A.; Osterholz, B.; Henke, C.; Blom, J.; Pühler, A.; Sczyrba, A.; Schlüter, A. Abundance, classification and genetic potential of Thaumarchaeota in metagenomes of European agricultural soils: A meta-analysis. *Environ. Microbiome* **2023**, 18, 26. <https://doi.org/10.1186/s40793-023-00479-9>.
81. Lian, W.-H.; Mohamad, O.A.A.; Dong, L.; Zhang, L.-Y.; Wang, D.; Liu, L.; Han, M.-X.; Li, S.; Wang, S.; Antunes, A.; et al. Culturomics- and metagenomics-based insights into the microbial community and function of rhizosphere soils in Sinai desert farming systems. *Environ. Microbiome* **2023**, 18, 4. <https://doi.org/10.1186/s40793-023-00463-3>.
82. Akinola, O.A.; Babalola, O.O.; Adeleke, A.O. Metagenomic insights into termite mound-associated microbial communities and agricultural sustainability. *Microbiol. Insights* **2023**, 16, 11786361231171862. <https://doi.org/10.1177/11786361231171862>.
83. Geydirici, I.; Martínez-Cuesta, R.; Bibinger, S.; Espíndola-Hernández, P.; Gschwendtner, S.; Rodríguez-Berbel, N.; Schulz, S.; Wicht, A.-C.; Schloter, M. Digging deeper into soil metagenomics: Opportunities and limitations for studying the genomic potential of soil bacteria and archaea. *Biol. Fertil. Soils* **2026**, 62, 641–648. <https://doi.org/10.1007/s00374-026-02004-9>.
84. Núñez-Gómez, D.; Melgarejo, P.; Martínez-Nicolás, J.J. et al. Effects of marine sediment as agricultural substrate on soil microbial diversity: an amplicon sequencing study. *Environ. Microbiome* **2023**, 18, 69. <https://doi.org/10.1186/s40793-023-00519-4>.

85. Sharma, P.; Verma, R.; Kumar, A.; Sharma, S. Machine learning and bioinformatics approaches for microbial diversity analysis in Indian river ecosystems. *Genes* **2023**, *14*, 1082. <https://doi.org/10.3390/genes14051082>.
86. Popescu, S.C.; Tomaso-Peterson, M.; Wilkerson, T.; Bronzato-Badial, A.; Wesser, U.; Popescu, G.V. Metagenomic analyses of the soybean root mycobiome and microbiome reveal signatures of the healthy and diseased plants affected by taproot decline. *Microorganisms* **2022**, *10*, 856. <https://doi.org/10.3390/microorganisms10050856>
87. Rizzi, S.; Saroglia, G.; Kalemi, V.; Rimoldi, S.; Terova, G. Artificial intelligence in microbiome research and beyond: connecting human health, animal husbandry, and aquaculture. *Appl. Sci.* **2025**, *15*, 9781. <https://doi.org/10.3390/app15179781>
88. Dukovski, J.; Bajić, M.; Živković, L.; Spasić, I. Computational frameworks for modeling microbial interactions in agricultural ecosystems. *arXiv* **2023**, arXiv:2307.02624. <https://doi.org/10.48550/arXiv.2307.02624>.
89. Quince, C.; Hallam, S.J.; Gilbert, J.A.; van der Gast, A. Advances in computational metagenomics and metagenome-assembled genome analysis. *arXiv* **2021**, arXiv:2109.00809. <https://doi.org/10.48550/arXiv.2109.00809>.
90. Sengupta, P.; Sivabalan, S.K.M.; Mahesh, A.; Palanikumar, I.; Kuppa Baskaran, D.K.; Raman, K. Big data for a small world: A review on databases and resources for studying microbiomes. *J. Indian Inst. Sci.* **2023**, *103*, 891–907. <https://doi.org/10.1007/s41745-023-00370-z>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.