SOIL SCIENCE ANNUAL
2023, 74(1), 163080

https://doi.org/10.37501/soilsa/163080

Metagenomics approaches to understanding soil health in environmental research – a review

Juan D. Duque-Zapata[1*], Jaime Eduardo Muńoz Florez[1], Diana López-Alvarez[1,2*]

[1] Universidad Nacional de Colombia, Facultad de Ciencias Agropecuarias, Departamento de Ciencias Biológicas, 763533, Palmira, Colombia
[2] Johns Hopkins University, School of Medicine. Department of Neurology, Pathology, 21287, Baltimore, USA

* PhD Candidate. Juan Diego Duque, email: jduquez@unal.edu.co, PhD. Diana López-Alvarez, email: dilopezal@unal.edu.co, ORCID iD: Duque-Zapata, J.D: https://orcid.org/0000-0001-5496-2502; Muńoz, J.E.: https://orcid.org/0000-0002-8237-0499; López-Alvarez, D: https://orcid.org/0000-0002-7734-8481

Received: 2023-01-10
Accepted: 2023-04-05
Published online: 2023-04-05
Associated editor: J. Wyszkowska

Keywords:

Bioindicator
Soil metagenomics
Metataxonomy
Whole genome sequencing

Abstract

Given the importance of soil as a supplier of nutrients and water for different ecosystems, understanding soil health and quality is necessary for its preservation. Microorganisms, due to their high abundance and their relationship with the degradation of organic matter and biogeochemical cycles, have a rapid response to environmental changes and thus are a discriminating factor that can be used as bioindicators of soil health. However, 97% of microorganisms are unculturable, leaving a gap in their taxonomic and functional knowledge. The development of metagenomics has reduced this problem through the direct extraction of DNA from soil, allowing the characterization of such non-culturable microorganisms, this technique can be considered one of the most impactful in soil health, given that it allows for an exploration of the biodiversity, the community structure, and the potential functions of the microbial communities from distinct environments. In addition to this, metagenomics have had an impact in different areas such as "OneHealth" or EcoGenomics allowing the formation of international projects. The aim of this paper is to show how metagenomics can be used as a technique to assess soil quality and health through the taxonomic and functional identification of the microorganisms present in the soil.

1. Introduction

Soil is responsible for providing Earth's distinct ecosystems with vital services, including being one of the largest deposits of nutrients and water for plants, regulating gas emissions, and cycling and recycling elements and molecules that are essential to life (Haygarth and Ritz, 2009). However, with the effects of climate change (e.g., long periods of drought, intense flooding) and anthropogenic activities (e.g., livestock grazing, mining, agriculture), soil fragmentation and multifunctionality have accelerated (Schloter et al., 2018), making it necessary to create different strategies that minimize these impacts and protect the soils. One of these strategies that have proven impactful in recent decades is the employment of bioindicators to characterize variations in soil health, which provides additional information to the physicochemical indicators that often are not able to fully reflect how soil health is affected, for example, exhibit the indirect biotic effects of pollutants (Alarcón Gutiérrez et al., 2021; Zaghloul et al., 2020).

A diversity of bioindicators have been utilized in environmental studies. Earthworms are good proxies for the extent of soil degradation due to their sensibility to anthropogenic alterations (Moreiera et al., 2012). Beetle diversity has been shown to be affected by environmental variations and thus reflects the degree of ecosystem deterioration (Menta and Remelli, 2020). Nematode communities, given their capacity to respond to changes in the soil, can provide information on the internal functioning of soils (such as the food chain) and how variations in this functioning can affect soil health (Martin et al., 2022). Finally, microorganisms are useful bioindicators, despite their relative omittance from most discussions of soil bioindicators (Schloter et al., 2018). This last point is evidenced by the proposed methods from the International Standardization Organization (ISO) for analyzing soil quality, of which only 7 of the more than 50 evaluated methods are related to microorganisms and their function in soils (ISO 14240-1:1997, ISO 16072:2002, ISO/TS 10832:2009, ISO/TS 29843-1:2010, ISO 17601:2016, ISO 18400-206:2018, ISO 11063:2020, https://www.iso.org/committee/54366/x/catalogue/), with only one (ISO 11063:2020) involving the use of genomic techniques to extract DNA directly from the soil. The present article provides a revision of metagenomic techniques and how they can be used to assess soil quality and health by identifying which microorganisms are present in soil samples.

2. Microorganisms: indicators of soil health

The biological transformations offered by soils, such as the degradation and contribution of organic material and biochemical processes, are carried out by the biological diversity present in the soil, including plants, invertebrates, arthropods, and microorganisms (i.e., bacteria, fungi, and algae); however, between 80–90% of these processes depend on activities generated by microorganisms, leading to this group's high diversity in soils (Nannipieri et al., 2003; Nesme et al., 2016).

Their high abundance and diversity present in soils permit soil microbiome which can be defined as the communities of microorganisms and their genetic material that reside in the soil (Fierer, 2017; Wang et al., 2021) to have a rapid response to environmental changes, leading those individuals that possess better adaptations to persist in the soil (Ezeokoli et al., 2020). A variety of studies have demonstrated this: In 2006, soil bacterial diversity was found to depend less on temperature or latitude, but more on variables closely related to soil, such as pH, with greater diversity in neutral soils and less diversity in acidic soils (Fierer and Jackson, 2006). Feng et al. (2018) showed how microbial communities in cadmium (Cd) contaminated soils are altered with respect to non-contaminated soils. Specifically, Cd-contaminated soils showed reduced microbial diversity with distinct species makeups and community structures (Feng et al., 2018). On the other hand, Bhowmik et al. (2019) found that between 62–90% of the variability in soil microbial populations is the product of distinct land use management (Bhowmik et al., 2019). These characteristics allow for populational changes in soil microorganisms to be employed as a discriminant for soil health, according to Hatten and Liles and the US Natural Resources Conservation Service, soil health can be defined as the continued capacity of soil to function as a vital living ecosystem to sustain plants, animals, primary productivity and ecological biodiversity (Hatten and Liles, 2019; USDA, 2022) making microorganisms an excellent bioindicator.

3. Discovering the microbial diversity in soils

The study of soil bacteria throughout the years has witnessed two "golden ages" that greatly advanced the knowledge of these organisms, as well as their importance and functions in soils. The first was marked by the discovery and recognition of soil microorganisms' importance in the vital nutrient cycling (Nannipieri et al., 2014). One of the characteristics of this first period was the need to cultivate and isolate samples for identification purposes; however, an estimated ~97% of bacterial and archaebacterial species cannot be cultivated using these methods, leading to ~70% of known bacterial phyla that do not possess any representatives capable of being cultivated (Jiao et al., 2021; Solden et al., 2016). Some of the causes of these complications include a lack of knowledge about the growing requirements of different microorganisms, such as their nutritional necessities, the physiochemical conditions of their natural environments, and the symbiotic or parasitic relationships that are maintained in a microbial community (Tang, 2019). Despite the advances gained from this first "golden age," the microbial world of soils remained a great enigma to researchers.

The present-day ability to learn and make discoveries of this relatively unexplored world was made possible due to new techniques that arose from the molecular revolution, bringing us into the second "golden age" of soil microbiology. Among these new techniques, metagenomics has allowed the extraction of nucleic acids directly from the soil, making it possible to characterize the "non-culturable" microorganisms (Nannipieri et al., 2014). In short, the methodology of this technique consists of extracting DNA or RNA directly from an environmental sample (Handelsman, 2005), creating a library that contains the genomes of every microbe that is found in a study site that can then be sequenced for bioinformatic analyses, such as taxonomic assignments, analyses of abundance, and the identification of potential functioning for select genes (S. Liu et al., 2022). This wealth of information afforded by metagenomics has led to it being proposed as a bioindicator tool for soil health (Torres et al., 2019).

4. Metagenomics and its relationship with soil health

Soil health can be defined as the continual capacity of soil to function as a living system central to sustaining the productivity of plants and animals, which ultimately promotes the well-being of every living being (Natural Resources Conservation Services, 2012). As such, microbial diversity is essential for the adequate functioning of soils; however, little information exists on how microorganisms can help to determine soil health. Since the origin of metagenomics, many advances have been made in the study and knowledge of structural and functional microorganisms diversity, species identification, the characterization of new genes (Nacke et al., 2011), and discovering enzymatic activities and active compounds (Craig et al., 2010). However, metagenomic studies have not been used to study soil health until recently (Kaushik et al., 2021). The discovery through metagenomics of the enormous diversity of non-culturable microorganisms, linked to the present-day knowledge of new genomes, has allowed the association of specific members to these microbial communities with transformations that soils may be experiencing (Long et al., 2016). For example, differences in taxonomic groups of microorganisms are a reliable variable for studying the impact of anthropogenic activities in soils. Conversion of natural ecosystems to agricultural land can lead to alterations of soil microbiome, as in the case of forest soils converted to grassland which showed higher diversity in bacteria and fungi compared to undisturbed forest soils as well as the introduction of bacterial taxa associated with agricultural activities such as ammonia-oxidizing bacteria (Navarrete et al., 2023). Similarly, the use of fertilizers, pesticides, and other chemicals for example long-term use of nitrogen fertilizers, which reduce the abundance of the phyla Proteobacteria, Actinobacteria, Acidobacteria, and Chloroflexi being higher was higher in short-term N application (Xu et al., 2022).

Soil metagenomics can provide crucial information on the adaptations and interactions between microorganism com-

munities and how these can be influenced by changes in the environment (Bonomo et al., 2022). To do this, metagenomics can be classified into two types: guided and shotgun. Guided metagenomics, also known as "metabarcoding," searches for taxonomic identification at a broad scale (hence the prefix "meta") through the analysis of DNA sequences of one or several genes (Deiner et al., 2017). This technique tends to be used for studying the phylogenetic diversity and relative abundance of a concrete gene (molecular marker) in an environmental sample (Techtmann and Hazen, 2016); these specified regions of a gene allow the identification of distinct taxonomic groups. The markers used vary depending on the type of organism that is being analyzed (Table 1). In the animal kingdom, the most commonly used marker is Cytochrome Oxidase I, or COI (Hebert et al., 2003), which is present in the mitochondrial DNA. In plants, given that the mitochondrial DNA has a low nucleotide substitution rate (Fazekas et al., 2008), markers in the chloroplast DNA have been proposed, specifically rbcL and matK being the most used (Hollingsworth et al., 2011). In the specific case of microorganisms, the most commonly used markers are the rRNA genes 16S for bacteria (Clarridge, 2004) and ITS ("Internal Transcribed Spacer") for fungi, which have become the favorite genes due to their high success rate in amplification (Dentinger et al., 2011).

Even though metabarcoding presents several difficulties, such as being limited to the use of PCR or the bias that can occur in the bioinformatics analyses leading to diversity estimations (Techtmann and Hazen, 2016), it has a great advantage in its capacity to provide a complete catalogue of the possible microbial taxa present in a given sample, allowing a more complete understanding of the changes in its diversity, for example, before and after an observed perturbation. Furthermore, this technique minimizes the number of reads associated with a possible host from a collected sample (Pearman et al., 2020).

The second technique, shotgun metagenomics, is implemented to determine the total genomic content from a sample through the preparation of sequencing libraries (Techtmann and Hazen, 2016). Furthermore, this technique can be utilized for the identification of the functional potential of microbial communities. Similar to metabarcoding, shotgun metagenomics also possesses its limitations, including sequencing depth, which should be high enough to include coverage of the entire genomic contents of each microorganism from the analyzed sample and achieve an integral analysis of the functional potential (Delmont et al., 2012). Despite these limitations, this technique's non-discriminant methodology (i.e., sequencing everything present in the sample) allows for the assignment of taxonomic identities, offering a much more robust result, and achieving the reconstruction of complete genomes and genes and the inference of distinct metabolic routes.

5. Bioinformatic pipelines used for metagenomic analyses in soil microorganism communities

Shotgun metagenomics and metabarcoding have distinct advantages and disadvantages. These differences are primarily the product of the techniques and bioinformatic methodologies that each employs.

5.1. Metabarcoding

The first step is to perform quality control, a set of procedures to ensure the accuracy and reliability of data generated from the sequencing instrument (e.g Illumina), that may include the removal of primers involved in the amplification and/or discarding short reads and those with low quality (e.g., Phred < 20). Several programs exist to carry out this step, with Trimmomatic (Bolger et al., 2014) being one of the most used (Fig. 1). The second step is the processing of reads, which are either single-end or paired-end. In the case of paired-end reads, forward and reverse reads are overlapped, and several parameters can be defined, such as the minimum and maximum lengths of the combined reads. Similarly, the quality values from the previous step can be used to perform cuts in the reads during the assembly process, which can be carried out in PEAR (Zhang et al., 2014).

Table 1
Primary molecular markers utilized in metabarcoding studies

Taxonomic Group	Marker	Most used primers	Genomic Source	Approximate Number of GenBank Accessions
Animals	COI	LCO1490/HCO2198	Mitochondrial	3,017,967
Plants	matK	390F and 1326R	Chloroplast	250,985
	rbcL	rbcLa-F/ rbcLa-R	Chloroplast	367,787
	psbA-trnH	psbA3-f/ trnHf_05	Chloroplast	119,808
Bacteria	16S	V3-V4 region	Ribosomal	44,377,078
	rpoB	rpoB-f1/rpoB-r1	Ribosomal	1,412,251
	cpn60	H729/ H730	Ribosomal	25,739
Fungi	ITS	ITS1F/ITS2	Ribosomal	42,245,580
	18S	EukA/EukB	Ribosomal	1,319,704

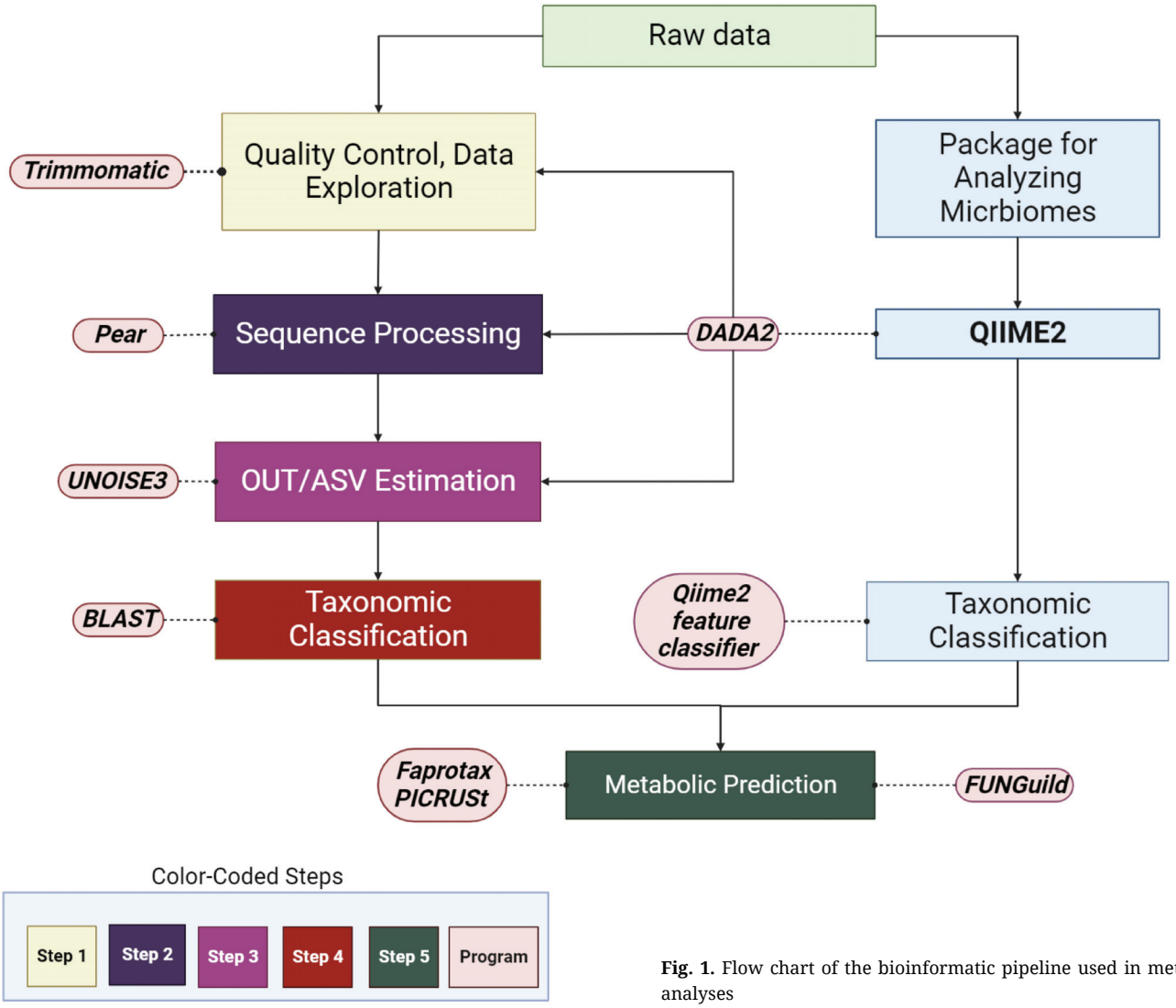

Fig. 1. Flow chart of the bioinformatic pipeline used in metabarcoding analyses

The third step is the estimation of OTUs ("Operational Taxonomic Units"), which consists of assigning a numeric code to each stack of identified sequences that may correspond to distinct species, genera, families, or the identified taxonomic rank. This is often done using UNOISE3 (Nearing et al., 2018) or Deblur (Amir et al., 2017). OTUs have been most utilized in microbial ecology due to their high percentage of identifications. Despite this, they are prone to showing incorrect information, such as overestimating the number of trueassignments or not having sufficient power to detect small variations in reads, making them relatively ineffective for discriminating between closely related, but distinct, taxa (Pérez-Cobas et al., 2020). Due to this, another type of assembly called ASVs ("Amplicon Sequence Variants") has been employed. To avoid the use of different programs and minimize the risks of incompatibility between the results of distinct steps, the program QIIME2 (Boylen et al., 2019) was created, which is an analysis package that performs all of the previously described steps. Within QIIME2, the algorithm DADA2 determines the quality of sequences, assembles forward and reverse sequences, and filters out low-quality data, providing clean data (Callahan et al., 2016) in an archive with the ASVs. Using this, the function *feature classifier* can be used to perform taxonomic classification of the obtained ASVs. This function can use different reference bases, such as Silva or GreenGenes for bacteria, or Unite for fungi (Bolyen et al., 2019).

Additionally, the obtained results can be utilized to perform a prediction of the functional potential of the identified microbiome through programs such as PICRUSt or Faprotax for Bacteria (Louca et al., 2016; Langille et al., 2013) these programs use a database of known functional gene profiles from reference genomes to estimate the functional composition of the soil samples. For fungal community, FUNguild is the preferred program (Nguyen et al., 2016), this program employs an algorithm to identify the ecological functions of fungi based on their taxonomic affiliations, using a reference database of functional guilds to assign functions to fungal sequences.

5.2. Shotgun Metagenomics

Using shotgun metagenomics that analyzes the complete genome, both the taxonomic composition and the functional potential of soil microorganism communities can be under-

stood. Before the sequencing step in shotgun metagenomics, the complete genome must first be digested and then sheared into small fragments of equal size (Fig. 2). The first step in the bioinformatics pipeline is the quality control of the raw data, which may be done Trimmomatic (Bolger et al., 2014), as in the metabarcoding pipeline. After this follows the assembly, the process of reconstructing complete DNA sequences from the fragment generated by sequencing technology, for which several assembly programs exist, including MegaHit (Li et al., 2015), MetaVelvet (Namiki et al., 2012), and metaSPAdes (Nurk

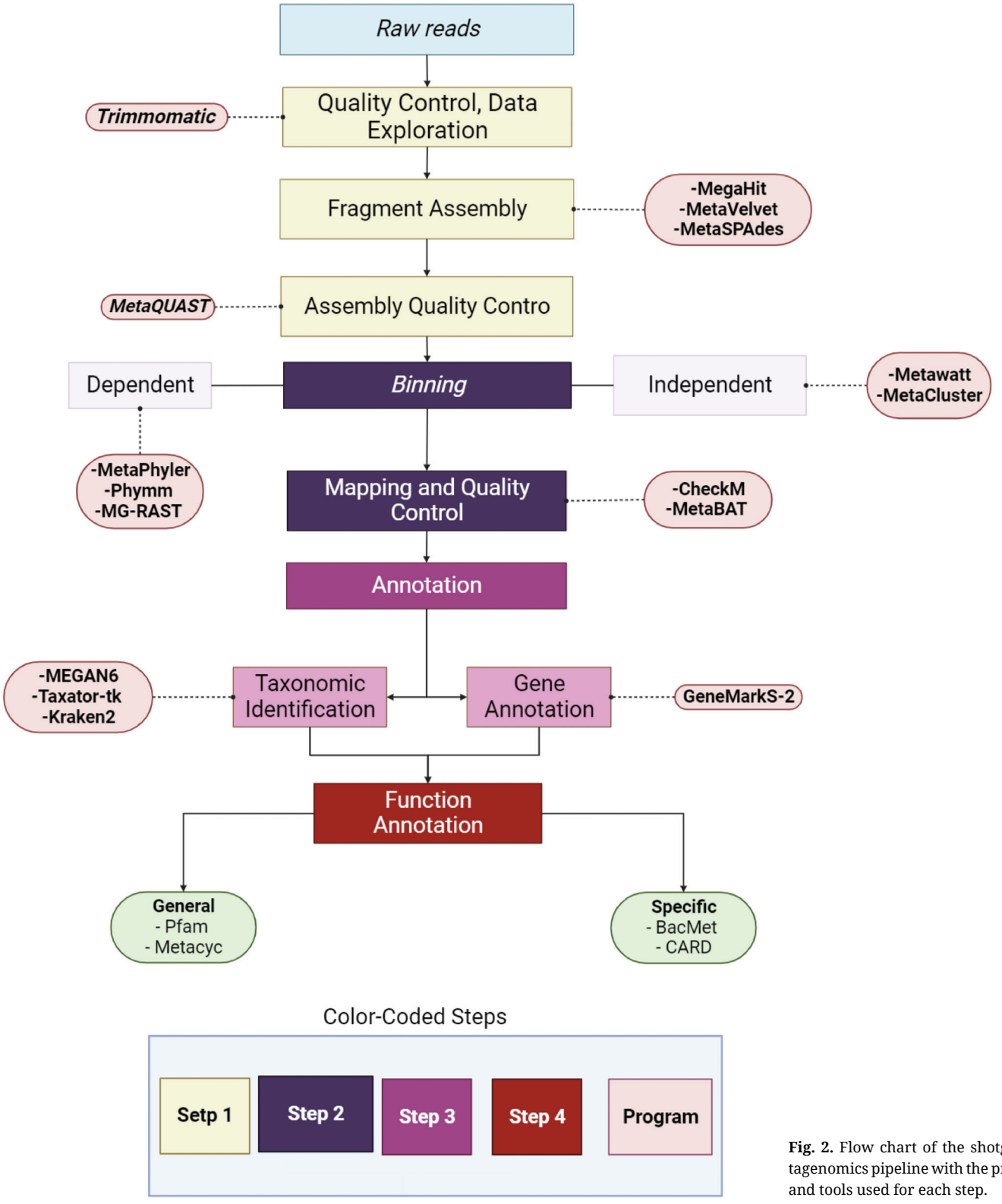

Fig. 2. Flow chart of the shotgun metagenomics pipeline with the programs and tools used for each step.

et al., 2017). Choosing an assembly program depends on two factors: 1) the format in which the input sequences are found (e.g., .fastq, .fastq.gz, .fasta, .fasta.gz), and 2) the algorithm that will be used to perform the genome assembly: Greedy, Overlap-Layout Consensus (OLC Graphics), or Bruijn graphics (Nagarajan and Pop, 2013). Next, quality control of the resulting assembly can be performed using MetaQUAST (Mikheenko et al., 2016).

Before the annotation step, an optional step called "binning" consists of grouping all the reads (i.e., the "contigs" assembled in the previous step) that pertain to the same taxonomic assignments. These groupings can then be classified taxonomically and characterized based on their functions. The program used for this process depends on the desired "binning" method, which may be dependent on the taxonomy (supervised) or be independent (unsupervised) (Pérez-Cobas et al., 2020). Those dependent on taxonomy use reference genomes to map the contigs, employing the alignment of the metagenomic sequences with a reference. This may be done in MetaPhyler (Liu et al., 2010). This can be done by taking into account nucleotide composition (e.g., GC content) and comparing it to the references, for which TACOA or Phymm (Brady and Salzberg, 2009; Diaz et al., 2009) may be utilized. Additionally, a hybrid of these two methods is offered in PhymmBL or MG-RAS (Brady and Salzberg, 2009; Keegan et al., 2016).

Alternatively, methods independent of taxonomy do not use reference genomes, but rather use other innate characteristics of the reads such as nucleotidic composition or abundance, and may be performed in the programs Metawatt (Strous et al., 2012) or AbundanceBin (Wu and Ye, 2011). However, this method is not very reliable for determining identifications when the contig or DNA fragment lengths are too short, or for determining a taxonomic identification at a high level, given that it may have problems differentiating genomes that are highly similar in composition. To diminish these errors, several hybrid methods have been designed that combine nucleotide composition with abundance, such as those employed in MetaCluster or MetaBAT (Kang et al., 2019; Y. Wang et al., 2012). After this, the results of the binning step are mapped to obtain longer reads, which requires a quality control step to guarantee the integrity of the mapped genome. This can be done using CheckM (Parks et al., 2015) or within MetaBAT.

The next step is the annotation of the genome, which involves identifying and characterizing the functional and taxonomic features of the obtained sequences, the annotation process can be divided into two parts: taxonomic identification and gene identification and annotation. For taxonomic identification, different programs including MEGAN6 (Huson et al., 2016) or Taxator-tk (Dröge et al., 2015) may be used to detect organisms at different taxonomic levels. However, these programs base identification on direct alignment, which can be a slow process. Other methods have arisen to reduce this processing time, such as the algorithm used in Kraken (Wood and Salzberg, 2014) that associates short genomic sub-chains, or k-mers (short sequences of length k nucleotides), with taxa. Kraken2 improves one of its predecessor's largest problems, memory usage, and can easily exceed 100 GB leading to faster taxonomic identification (Wood et al., 2019). It still may present some false positives, for which a parallel version called KrakenUniq (Breitwieser et al., 2018) was created specifically for the diagnosis of health disorders. The second part, comprising the identification and annotation of genes, can be done using geneMarkS-2 (Lomsadze et al., 2018), which is used to predict both typical and atypical (e.g., from horizontal transfer) prokaryotic genes. Additionally, this program can identify the characteristic mechanisms that are based on the comparative evaluation of universal orthologs from a single copy called BUSCO ("Benchmarking Universal Single-Copy Orthologs") (Simão et al., 2015). Orthologous genes generally perform the same function, so BUSCO employs databases that contain these types of genes, such as OrthoDB (www.orthodb.org) to identify and annotate possible genes with a high confidence value.

The final step involves functional annotation. This process collects information on the identified genes to understand their molecular function, their biological role, and their possible subcellular level. This can be done using Blast (Madden, 2002) or databases such as Pfam (Mistry et al., 2021) and Metacyc (Caspi et al., 2020). If the study is focused on previously determined genes or functions, some specifically curated databases exist, including CARD (Alcock et al., 2019), specific to genes involved in antibiotic resistance, or BacMet (Pal et al., 2014), specific to genes resistant to metals and antibacterial biocides.

Executing each one of the shotgun metagenomics steps separately requires a demanding use of time. As such, integrative environments that combine the assembly, quality control, contaminant identification, and functional annotation steps may be useful, such as MOCAT2 (Kultima et al., 2016), MetAMOS (Treangen et al., 2013), or MAGNETO (Churcheward et al., 2022).

6. The use of metagenomics in approximations of "OneHealth" and EcoGenomics

Metagenomics has had an impact on the discovery and analysis of microorganisms in different areas such as "OneHealth" (https://www.cdc.gov/onehealth/) and EcoGenomics allowing the formation of international groups/projects such as TerraGenome (Vogel et al., 2009) and the Earth Microbiome Project (EMP) (http://www.earthmicrobiome.org).

6.1. Implications in an approximation of One Health

From the end of the last century and the beginning of the XXI century, diverse studies have been performed under a unique focus deemed One Health. Initially, this term referenced the necessity to unify human medicine with veterinary medicine to combat zoonotic diseases (Evans and Leighton, 2014). Starting in 2008, this term was modified and now covers all interactions between animals, human beings, plants, and microorganisms, and how these have repercussions on the equilibrium of distinct ecosystems and their effects on health (Gerner-Smidt et al., 2019). One of the central pillars of One Health is to detect,

prevent, and control risks that affect the health and well-being of ecological, urban, and rural populations (Li, 2017). Microorganisms play a vital role in the health of the different actors within ecosystems. They can be harmful to hosts, be virulence factors, and/or contain genes resistant to illnesses and medications. As previously mentioned, the majority of microorganisms are non-culturable, and thus metagenomics plays an important role in the advances and knowledge of "One Health" by permitting the exploration of how microorganisms interact with their surroundings. As such, metagenomic techniques have been used in epidemiological surveillance, as demonstrated in the Human Microbiome Project which has been studying the interactions between human beings and their associated microbiomes for over 10 years (Human Microbiome Project, 2019). Using metagenomics, this project has characterized the intestinal microbiome of healthy individuals to compare with others suffering from distinct types of medical problems. At the level of specific illnesses, metagenomics has been employed to identify divergent regions of non-coding RNA in *Listeria monocytogenes*, the bacteria responsible for Listeria infection. In 2012, this project found more than 70 RNAs that could be involved in inhibiting the expression of different operons that could be used as regulators in this bacterium (Wurtzel et al., 2012). Another study analyzed and detected variations in the H1N1 flu, obtaining a genomic coverage of 97%, and found that 90% of the H1N1 genome could be assembled *de novo* (i.e., without using reference sequences/genome), leading to this technique as a preventative strategy and diagnosis of new outbreaks of the sickness (Greninger et al., 2010).

In the same way, metagenomics has allowed researchers to analyze and counteract one of the largest problems in public health: antimicrobial resistance, caused in large part by ARGs or resistomes are groups of genes that confer resistance to antibiotics (Wang et al., 2020). With full genome sequencing, the diverse structures and functions of pathogenetic microorganisms that present antimicrobial resistance have been studied, leading to the identification of new ARGs (Wang et al., 2020).

6.2. Implications in ecogenomics

Ecological genomics, or ecogenomics, aims to understand how the functioning of genes and/or the genome affects the interactions between organisms and their natural environments (Ungerer et al., 2008). Recently, Baksay et al. (2022) demonstrated how metabarcoding can be used as a tool for evaluating and studying the interactions between plants and pollinators. For this, they performed metabarcoding on grains of pollen from different insects, which they found to be more efficient than microscopy for identifying species of plants, and recovered a positive relationship between the number of sequences, pollen quantity, and pollinator visitation frequency (Baksay et al., 2022).

Metagenomics as an integrative technique has also allowed researchers to study the importance of the relationship between microorganisms and other living beings such as plants and animals. Shumo et al. (2021) analyzed the bacterial species present in the intestines of black fly larvae (*Hermetia illucens*), one of the main alternative foods for animals including chickens and fish. The researchers compared black flies fed with chicken excrement and kitchen scraps. They found bacteria of the genera *Providencia* and *Bordetella* to be the most dominant, which are considered causal pathogens for illnesses such as whooping cough. The study concluded that it is necessary to employ biosecurity strategies during the harvest of black flies to avoid the propagation of zoonotic diseases (Shumo et al. 2021). Similarly, metagenomic techniques have been used to analyze the feces of wild animals, such as the Canadian reindeer (*Rangifer tarandus*), to characterize their diets and formulate conservation strategies that include the proliferation of consumed food items. The results found the presence of a diversity of fungi in the genus *Lichenoconium*, which are characteristically associated with lichens known to be consumed by these animals. On the other hand, a primary consumption of coniferous trees such as yews (*Taxus* spp.), cherry wood (*Cronus* spp.), and maple (*Acer* spp.) was observed, suggesting that assuring the presence of these other food sources, in addition to lichen, is necessary for the conservation of the reindeer (Mitchell et al., 2022).

In plants, Su et al. (2022) performed the first metagenomic study on the phyllosphere (surface part of leaves) in rice. This study produced 1.34 terabases of sequenced data and 569 assembled genomes with over 50% completeness. The majority of these metagenomes pertain to bacteria in the classes *Alphaproteobacteria*, *Gammaproteobacteria*, and *Bacteroidia*. The authors concluded that the obtained information provides a start for phytopathology studies and the recognition of microorganisms that colonize these exterior parts of plants (Su et al., 2022). Senn et al. (2022) analyzed the composition and function of microorganisms associated with the radicular zone of the medicinal plant *Datura inoxia*, and found that the principal genera associated with the plant's rhizosphere were *Flavobacterium*, *Pedobacter*, and *Paenibacilus*, among others. Bacteria of the genus *Flavobacterium* are known to possess a tyrosine ammonia-lyase gene, which has been employed to optimize the production of aromatic compounds of pharmaceutical interest. They also found precursors to vegetal growth with antioxidant and phytoprotective compounds (Senn et al., 2022). On the other hand, a study in 2021 performed a metagenomic analysis of promotor genes involved with vegetal growth and the carbon cycle in the rhizosphere of corn fields with different land use histories. They found differences in component families and genera of the two types of soils analyzed and, of the 14 most abundant families, eight were present in soils that were previously pastureland, including Micromonosporaceae, *Nocardioidaceae*, and *Microbacteriaceae*. Six bacterial families, including *Geodermatophilaceae*, *Pseudonocardiaceae*, were present only in intensively farmed soils. As for the genes involved in vegetal growth, the authors found genes associated with nitrogen fixation, nitrification, and denitrification (nirK, nirS, and norB), as well as the potassium cycle. Thirty-four additional genes were recognized for their importance in the carbon cycle, including those related to carbohydrate metabolism, carbon fixation, and the breakdown of starch and methane (Chukwuneme et al., 2021).

6.3. The international projects TerraGenome and Earth Microbiome

Studying soil microorganisms through metagenomics is important for creating strategies aimed at the conservation, production, and well-being of animals, plants, and ecosystems. Given this importance, dedicated projects have been created to document the diversity and metabolic functions using metagenomics, such as the "Earth Microbiome Project" (EMP) and "TerraGenome." Both projects have the same goal – to increase our understanding of soil microorganisms using genomic techniques – albeit with different approaches. TerraGenome (Vogel et al., 2009) is an international consortium between 23 counties that was proposed under the same scheme of projects as the Human Genome Project (HGP), to construct reference genomes that can be used in future genomic projects focused on soils. Researchers involved in this project performed metagenomic analyses for ten years on soil samples from the same locality in Rothamstead, UK, where environmental conditions and land use is well-documented for over 150 years.

On the other hand, EMP originated in 2010 in the United States (Gilber et al., 2010) as an open collaborative project with the objective of studying the microbial composition in distinct ecosystems and surroundings across the globe, with the key aspect of using a set of standardized protocols to assure that no bias would exist that might compromise comparisons across multiple microbial communities (Gilbert et al., 2014). In 2017, the first results of this project were published, including 27,752 samples analyzed from 7 continents and 43 countries, with over two billion sequences, 308 identified species, over 60 published articles, and nearly 2000 publications that have used the project's proposed protocols (Thompson et al., 2017). Given the favorable reception that this project received and the high demand of the obtained results, a second phase of the project deemed "Earth Microbiome Project 500" (EMP500) was established that aims to perform the metagenomic sequencing and metabolic profiling of 500 microbial communities from around the world (https://earthmicrobiome.org/emp500/).

7. Limitations of soil metagenomics

While soil metagenomics has the potential to revolutionize our understanding of soil microbiomes, there are significant challenges that need to be addressed.

One of these challenges is the physicochemical properties of the soil. These properties influence the composition and function of soil microbiomes, however many studies on soil metagenomics only measure a limited set of parameters such as pH, soil organic carbon, moisture, and temperature, leaving out elements of relevance such as potassium, phosphorus, or iron. Additionally, these properties may vary with soil depth (e.g organic carbon, total nitrogen) and change with seasonal fluctuations or environments (Leite et al., 2022; Shi et al., 2020) so it is uncertain whether the findings of studies carried out on a specific soil or ecosystem can be extrapolated to identify the microorganisms at a global level (Frąc et al., 2018).

Even though soil metagenomics is generating a vast amount of sequencing data and despite the recommendation to use curated databases such as UNITE or SILVA for analysis, the sheer volume of data generated has made it increasingly challenging to carry out research that can be replicated (Katz et al., 2022). In addition, there remains a significant number of sequences or taxonomic units (OTUs/ASVs) for which no information is currently available (Frąc et al., 2022).

8. Conclusions

Through a variety of techniques and projects outlined above, utilizing the abundance and diversity of microorganisms has been shown to be useful for evaluating the state of soil health. Metagenomics can be considered one of the most impactful techniques in this field, given that it allows for an exploration of the biodiversity, the community structure, and the potential functions of the microbial communities from distinct environments.

Enriching our understanding of specific groups of microorganisms is useful not only for evaluating soil health, but also for distinct fields such as bioremediation, agriculture, and human health, given that it may help identify and broaden our understanding of the different mechanisms of resistance seen in some microorganisms.

As shown above, a large number of steps in a variety of programs exist to carry out analyses of metagenomic data, whether from the guided or shotgun approaches. Nonetheless, in order to understand biodiversity and mitigate possible errors during the assembly and annotation of genomes, we suggest using programs and tools integrated into bioinformatic workflows such as QIIME2, MOCAT2, and MetAMOS, among others.

It is crucial to incorporate soil physicochemical properties in all soil metagenomic studies to facilitate comparison between studies and improve our understanding of soil microbiomes.

Acknowledgments

This research was funded by the Patrimonio Autónomo Fondo Nacional de Financiamiento para la ciencia, la tecnología, y la innovación Francisco José de Caldas of Colombia, program title: Relaciones multiescalares de la biodiversidad en gradientes altitudinales del Bosque Tropical, code: 1106- 852-70306, contract number: 491-2020, and National Institutes of Health of the United States (NIH, No. R01 NS110122).

References

Alarcón Gutiérrez, E., Hernández, C., Gardner, T., García Pérez, J.A., Caballero, M., Perroni, Y., Farnet da Silva, A.M.A., Gaime Perraud, I., Barois, I., 2021. Soil bioindicators associated to different management regimes of *Cedrela odorata* plantations. Madera y Bosques 27(1), e2711912. https://doi.org/10.21829/myb.2021.2711912

Alcock, B.P., et al., 2019. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research 48(D1), D517-D525. https://doi.org/10.1093/nar/gkz935

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z., Kightley, E.P., Thompson, L.R., Hyde, E.R., Gonzalez, A., Knight, R., 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. Msystems *2*(2), e00191-16. https://doi.org/10.1128/mSystems.00191-16

Baksay, S., Andalo, C., Galop, D., Burrus, M., Escaravage, N., Pornon, A., 2022. Using Metabarcoding to Investigate the Strength of Plant-Pollinator Interactions From Surveys of Visits to DNA Sequences. Frontiers in Ecology and Evolution 10, 735588. https://doi.org/10.3389/fevo.2022.735588

Bhowmik, A., Kukal, S.S., Saha, D., Sharma, H., Kalia, A., Sharma, S., 2019. Potential indicators of soil health degradation in different land use-based ecosystems in the shiwaliks of northwestern India. Sustainability 11(14), 3908. https://doi.org/10.3390/su11143908

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolyen, E., et al., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bonomo, M.G., Calabrone, L., Scrano, L., Bufo, S.A., di Tomaso, K., Buongarzone, E., Salzano, G., 2022. Metagenomic monitoring of soil bacterial community after the construction of a crude oil flowline. Environmental Monitoring and Assessment 194(2), 48. https://doi.org/10.1007/s10661-021-09637-3

Brady, A., Salzberg, S.L., 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nature Methods 6(9), 673–676. https://doi.org/10.1038/nmeth.1358

Breitwieser, F.P., Baker, D.N., Salzberg, S.L., 2018. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biology 19(1), 198. https://doi.org/10.1186/s13059-018-1568-0

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong, W.K., Paley, S., Subhraveti, P., Karp, P.D., 2020. The MetaCyc database of metabolic pathways and enzymes – a 2019 update. Nucleic Acids Research 48(D1), D445–D453. https://doi.org/10.1093/nar/gkz862

Chukwuneme, C.F., Ayangbenro, A.S., Babalola, O.O., 2021. Metagenomic analyses of plant growth-promoting and carbon-cycling genes in maize rhizosphere soils with distinct land-use and management histories. Genes 12(9), 1431. https://doi.org/10.3390/genes12091431

Churcheward, B., Millet, M., Bihouée, A., Fertin, G., Chaffron, S., 2022. MAGNETO: An Automated Workflow for Genome-Resolved Metagenomics. Msystems 7(4), e0043222. https://doi.org/10.1128/msystems.00432-22

Clarridge, J.E., 2004. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical* Microbiology Reviews 17(4), 840–862. https://doi.org/10.1128/CMR.17.4.840-862.2004

Craig, J.W., Chang, F.Y., Kim, J.H., Obiajulu, S.C., Brady, S.F., 2010. Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria. Applied and Environmental Microbiology 76(5), 1633–1641. https://doi.org/10.1128/AEM.02169-09

Deiner, K., et al., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26(21), 5872–5895. https://doi.org/10.1111/mec.14350

Delmont, T.O., Simonet, P., Vogel, T.M., 2012. Describing microbial communities and performing global comparisons in the 'omic era. The ISME Journal 6(9), 1625–1628. https://doi.org/10.1038/ismej.2012.55

Dentinger, B.T.M., Didukh, M.Y., Moncalvo, J.M., 2011. Comparing COI and ITS as DNA Barcode Markers for Mushrooms and Allies (Agaricomycotina). PloS ONE 6(9), e25081. https://doi.org/10.1371/journal.pone.0025081

Diaz, N.N., Krause, L., Goesmann, A., Niehaus, K., Nattkemper, T.W., 2009. TACOA – Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. BMC Bioinformatics 10(1), 56. https://doi.org/10.1186/1471-2105-10-56

Dröge, J., Gregor, I., McHardy, A.C., 2015. Taxator-tk: precise taxonomic assignment of metagenomes by fast approximation of evolutionary neighborhoods. Bioinformatics 31(6), 817–824. https://doi.org/10.1093/bioinformatics/btu745

Evans, B.R., Leighton, F.A., 2014. A history of One Health. Revue Scientifique et Technique de l'OIE 33(2), 413–420. https://doi.org/10.20506/rst.33.2.2298

Ezeokoli, O.T., Bezuidenhout, C.C., Maboeta, M.S., Khasa, D.P., Adeleke, R.A., 2020. Structural and functional differentiation of bacterial communities in post-coal mining reclamation soils of South Africa: bioindicators of soil ecosystem restoration. Scientific Reports 10(1), 1759. https://doi.org/10.1038/s41598-020-58576-5

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., Percy, D.M., Hajibabaei, M., Barrett, S.C.H., 2008. Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PloS ONE 3(7), e2802. https://doi.org/10.1371/journal.pone.0002802

Feng, G., Xie, T., Wang, X., Bai, J., Tang, L., Zhao, H., Wei, W., Wang, M., Zhao, Y., 2018. Metagenomic analysis of microbial community and function involved in cd-contaminated soil. BMC Microbiology 18(1), 1–13. https://doi.org/10.1186/s12866-018-1152-5

Fierer, N., 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology 15(10), 579–590. https://doi.org/10.1038/nrmicro.2017.87

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America 103(3), 626–631. https://doi.org/10.1073/pnas.0507535103

Frąc, M., Hannula, E.S., Bełka, M., Salles, J.F., Jedryczka, M., 2022. Soil mycobiome in sustainable agriculture. Frontiers in Microbiology 13. 1033824. https://doi.org/10.3389/fmicb.2022.1033824

Frąc, M., Hannula, S.E., Belka, M., Jędryczka, M., 2018. Fungal biodiversity and their role in soil health. Frontiers in Microbiology 9, 707. https://doi.org/10.3389/fmicb.2018.00707

Gerner-Smidt, P., Besser, J., Concepción-Acevedo, J., Folster, J.P., Huffman, J., Joseph, L.A., Kucerova, Z., Nichols, M.C., Schwensohn, C.A., Tolar, B., 2019. Whole Genome Sequencing: Bridging One-Health Surveillance of Foodborne Diseases. Frontiers in Public Health 7, 172. https://doi.org/10.3389/fpubh.2019.00172

Gilbert, J.A., Jansson, J.K., Knight, R., 2014. The Earth Microbiome project: successes and aspirations. BMC Biology 12(1), 69. https://doi.org/10.1186/s12915-014-0069-1

Gilbert, J.A. et al., 2010. Meeting Report: The Terabase Metagenomics Workshop and the Vision of an Earth Microbiome Project. Standards in Genomic Sciences 3(3), 243–248. https://doi.org/10.4056/sigs.1433550

Greninger, A.L. et al.., 2010. A Metagenomic Analysis of Pandemic Influenza A (2009 H1N1) Infection in Patients from North America. PloS ONE, 5(10), e13381. https://doi.org/10.1371/journal.pone.0013381

Hatten, J., Liles, G., 2019. A 'healthy' balance – The role of physical and chemical properties in maintaining forest soil function in a changing world. Developments in Soil Science 36, 373–396. https://doi.org/10.1016/B978-0-444-63998-1.00015-X

Haygarth, P.M., Ritz, K., 2009. The future of soils and land use in the UK: Soil systems for the provision of land-based ecosystem services. Land Use Policy 26, 187–197. https://doi.org/10.1016/j.landusepol.2009.09.016

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and Using a Plant DNA Barcode. PloS ONE 6(5), e19254. https://doi.org/10.1371/journal.pone.0019254

Human Microbiome Project., 2019. The Integrative Human Microbiome Project. Nature 569(7758), 641–648. https://doi.org/10.1038/s41586-019-1238-8

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., 2016. MEGAN Community Edition – Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLOS Computational Biology 12(6), e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Jiao, J.Y., Liu, L., Hua, Z.S., Fang, B.Z., Zhou, E.M., Salam, N., Hedlund, B.P., Li, W.J., 2021. Microbial dark matter coming to light: Challenges and opportunities. National Science Review 8(3), nwaa280. https://doi.org/10.1093/nsr/nwaa280

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359. https://doi.org/10.7717/peerj.7359

Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J.R., O'Sullivan, C., 2022. The Sequence Read Archive: a decade more of explosive growth. Nucleic Acids Research 50(D1), D387–D390. https://doi.org/10.1093/nar/gkab1053

Kaushik, P., Singh Sandhu, O., Singh Brar, N., Kumar, V., Singh Malhi, G., Kesh, H., Saini, I., 2021. Soil Metagenomics: Prospects and Challenges. In Mycorrhizal Fungi – Utilization in Agriculture and Industry. IntechOpen. https://doi.org/10.5772/intechopen.93306

Keegan, K.P., Glass, E.M., Meyer, F., 2016. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. Microbial environmental genomics (MEG) 207–233. https://doi.org/10.1007/978-1-4939-3369-3_13

Kultima, J.R., Coelho, L.P., Forslund, K., Huerta-Cepas, J., Li, S.S., Driessen, M., Voigt, A.Y., Zeller, G., Sunagawa, S., Bork, P., 2016. MOCAT2: a metagenomic assembly, annotation and profiling framework. Bioinformatics 32(16), 2520–2523. https://doi.org/10.1093/bioinformatics/btw183

Langille, M.G. et al., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31(9), 814–821. https://doi.org/10.1038/nbt.2676

Leite, M.F.A., van den Broek, S.W.E.B., Kuramae, E.E., 2022. Current Challenges and Pitfalls in Soil Metagenomics. Microorganisms 10(10), 1900. https://doi.org/10.3390/microorganisms10101900

Li, A.M., 2017. Ecological determinants of health: food and environment on human health. Environmental Science and Pollution Research 24(10), 9002–9015. https://doi.org/10.1007/s11356-015-5707-9

Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Liu, B., Gibbons, T., Ghodsi, M., Pop, M., 2010. MetaPhyler: Taxonomic profiling for metagenomic sequences. 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 95–100. https://doi.org/10.1109/BIBM.2010.5706544

Liu, S., Moon, C.D., Zheng, N., Huws, S., Zhao, S., Wang, J., 2022. Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. Microbiome 10(1), 76. https://doi.org/10.1186/s40168-022-01272-5

Lomsadze, A., Gemayel, K., Tang, S., Borodovsky, M., 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. Genome Research 28(7), 1079–1089. https://doi.org/10.1101/gr.230615.117

Long, P.E., Williams, K.H., Hubbard, S.S., Banfield, J.F., 2016. Microbial Metagenomics Reveals Climate-Relevant Subsurface Biogeochemical Processes. Trends in Microbiology 24(8), 600–610. https://doi.org/10.1016/j.tim.2016.04.006

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

Madden, T., 2002. The BLAST Sequence Analysis Tool. In The NCBI Handbook.

Martin, T., Wade, J., Singh, P., Sprunger, C.D., 2022. The integration of nematode communities into the soil biological health framework by factor analysis. Ecological Indicators 136, 108676. https://doi.org/10.1016/j.ecolind.2022.108676

Menta, C., Remelli, S., 2020. Soil Health and Arthropods: From Complex System to Worthwhile Investigation. Insects 11(1), 54. https://doi.org/10.3390/insects11010054

Mikheenko, A., Saveliev, V., Gurevich, A., 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32(7), 1088–1090. https://doi.org/10.1093/bioinformatics/btv697

Mistry, Jet al., 2021. Pfam: The protein families database in 2021. Nucleic Acids Research 49(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mitchell, G., Wilson, P.J., Manseau, M., Redquest, B., Patterson, B.R., Rutledge, L.Y., 2022. DNA metabarcoding of faecal pellets reveals high consumption of yew (*Taxus* spp.) by caribou (*Rangifer tarandus*) in a lichen-poor environment. FACETS 7, 701–717. https://doi.org/10.1139/facets-2021-0071

Moreira, F.M.S., Huising, J.E., Bignell, D.E., 2012. Manual de biología de suelos tropicales. Muestreo y caracterización de la biodiversidad bajo suelo. Instituto Nacional de Ecologia INE.

Nacke, H., Will, C., Herzog, S., Nowka, B., Engelhaupt, M., Daniel, R., 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. FEMS Microbiology Ecology 78(1), 188–201. https://doi.org/10.1111/j.1574-6941.2011.01088.x

Nagarajan, N., Pop, M., 2013. Sequence assembly demystified. Nature Reviews Genetics 14(3), 157–167. https://doi.org/10.1038/nrg3367

Namiki, T., Hachiya, T., Tanaka, H., Sakakibara, Y., 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Research 40(20), e155–e155. https://doi.org/10.1093/nar/gks678

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G., 2003. Microbial diversity and soil functions. European Journal of Soil Science 54(4), 655–670. https://doi.org/10.1046/j.1351-0754.2003.0556.x

Nannipieri, P., Pietramellara, G., Renella, G., 2014. Omics in Soil Science.

Natural Resources Conservation Services., 2012. Soil Health.

Navarrete, A.A., Aburto, F., González-Rocha, G., Guzmán, C.M., Schmidt, R., Scow, K., 2023. Anthropogenic degradation alter surface soil biogeochemical pools and microbial communities in an Andean temperate forest. Science of the Total Environment 854(1), 158508. https://doi.org/10.1016/j.scitotenv.2022.158508

Nearing, J.T., Douglas, G.M., Comeau, A.M., Langille, M.G.I., 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 6, e5364. https://doi.org/10.7717/peerj.5364

Nesme, J. et al., 2016. Back to the future of soil metagenomics. Frontiers in Microbiology 7, 73. https://doi.org/10.3389/fmicb.2016.00073

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., Schilling, J.S., Kennedy, P.G., 2016. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology 20, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Research 27(5), 824–834. https://doi.org/10.1101/gr.213959.116

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., Larsson, D.G.J., 2014. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Research 42(D1), D737–D743. https://doi.org/10.1093/nar/gkt1252

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 25(7) 1043–1055. https://doi.org/10.1101/gr.186072.114

Pearman, W.S., Freed, N.E., Silander, O.K., 2020. Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads. BMC Bioinformatics 21(1), 220. https://doi.org/10.1186/s12859-020-3528-4

Pérez-Cobas, A.E., Gomez-Valero, L., Buchrieser, C., 2020. Metagenomic approaches in microbial ecology: An update on whole-genome and marker gene sequencing analyses. Microbial Genomics 6(8), 1–22. https://doi.org/10.1099/mgen.0.000409

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

Senn, S., Pangell, K., Bowerman, A.L., 2022. Metagenomic Insights into the Composition and Function of Microbes Associated with the Root-zone of Datura inoxia. BioTech 11(1). https://doi.org/10.3390/BIOTECH11010001

Shi, Y., Su, C., Wang, M., Liu, X., Liang, C., Zhao, L., Zhang, X., Minggagud, H., Feng, G., Ma, W., 2020. Modern Climate and Soil Properties Explain Functional Structure Better Than Phylogenetic Structure of Plant Communities in Northern China. Frontiers in Ecology and Evolution 8, 531947. https://doi.org/10.3389/fevo.2020.531947

Shumo, M., Khamis, F.M., Ombura, F.L., Tanga, C.M., Fiaboe, K.K.M., Subramanian, S., Ekesi, S., Schlüter, O.K., van Huis, A., Borgemeister, C., 2021. A Molecular Survey of Bacterial Species in the Guts of Black Soldier Fly Larvae (*Hermetia illucens*) Reared on Two Urban Organic Waste Streams in Kenya. Frontiers in Microbiology 12, 687103. https://doi.org/10.3389/fmicb.2021.687103

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Solden, L., Lloyd, K., Wrighton, K., 2016. The bright side of microbial dark matter: Lessons learned from the uncultivated majority. Current Opinion in Microbiology 31, 217–226. https://doi.org/10.1016/j.mib.2016.04.020

Strous, M., Kraft, B., Bisdorf, R., Tegetmeyer, H.E., 2012. The Binning of Metagenomic Contigs for Microbial Physiology of Mixed Cultures. Frontiers in Microbiology 3, 410. https://doi.org/10.3389/fmicb.2012.00410

Su, P. et al, 2022. Recovery of metagenome-assembled genomes from the phyllosphere of 110 rice genotypes. Scientific Data, 9(1), 254. https://doi.org/10.1038/s41597-022-01320-7

Tang, L., 2019. Culturing uncultivated bacteria. Nature Methods 16(11), 1078–1078. https://doi.org/10.1038/s41592-019-0634-1

Techtmann, S.M., Hazen, T.C., 2016. Metagenomic applications in environmental monitoring and bioremediation. Journal of Industrial Microbiology and Biotechnology 43(10), 1345–1354. https://doi.org/10.1007/s10295-016-1809-8

Thompson, L.R. et al., 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551(7681), 457–463. https://doi.org/10.1038/nature24621

Torres, G.G., Figueroa-Galvis, I., Muńoz-García, A., Polanía, J., Vanegas, J., 2019. Potential bacterial bioindicators of urban pollution in mangroves. Environmental Pollution 255, 113293. https://doi.org/10.1016/j.envpol.2019.113293

Treangen, T.J., Koren, S., Sommer, D.D., Liu, B., Astrovskaya, I., Ondov, B., Darling, A.E., Phillippy, A.M., Pop, M., 2013. MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. Genome Biology 14(1), R2. https://doi.org/10.1186/gb-2013-14-1-r2

Ungerer, M.C., Johnson, L.C., Herman, M.A., 2008. Ecological genomics: understanding gene and genome function in the natural environment. Heredity 100(2), 178–183. https://doi.org/10.1038/sj.hdy.6800992

USDA, Natural Resources Conservation Service, 2022. What is Soil Health? Https://Www.Nrcs.Usda.Gov/Wps/Portal/Nrcs/Main/Soils/Health/

Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M., van Elsas, J.D., Bailey, M.J., Nalin, R., Philippot, L., 2009. TerraGenome: a consortium for the sequencing of a soil metagenome. Nature Reviews Microbiology 7(4), 252–252. https://doi.org/10.1038/nrmicro2119

Wang, M. et al., 2021. Soil Microbiome Structure and Function in Ecopiles Used to Remediate Petroleum-Contaminated Soil. Frontiers in Environmental Science 9, 624070. https://doi.org/10.3389/fenvs.2021.624070

Wang, S., Yan, Z., Wang, P., Zheng, X., Fan, J., 2020. Comparative metagenomics reveals the microbial diversity and metabolic potentials in the sediments and surrounding seawaters of Qinhuangdao mariculture area. PLOS ONE 15(6), e0234128. https://doi.org/10.1371/journal.pone.0234128

Wang, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample. Bioinformatics 28(18), i356–i362. https://doi.org/10.1093/bioinformatics/bts397

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biology 20(1), 257. https://doi.org/10.1186/s13059-019-1891-0

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

Wu, Y.W., Ye, Y., 2011. A Novel Abundance-Based Algorithm for Binning Metagenomic Sequences Using l -tuples. Journal of Computational Biology 18(3), 523–534. https://doi.org/10.1089/cmb.2010.0245

Wurtzel, O., Sesto, N., Mellin, J.R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., Sorek, R., 2012. Comparative transcriptomics of pathogenic and non-pathogenic Listeria species. Molecular Systems Biology 8(1), 583. https://doi.org/10.1038/msb.2012.11

Xu, A., Li, L., Xie, J., Zhang, R., Luo, Z., Cai, L., Liu, C., Wang, L., Anwar, S., Jiang, Y., 2022. Bacterial Diversity and Potential Functions in Response to Long-Term Nitrogen Fertilizer on the Semiarid Loess Plateau. Microorganisms 10(8), 1579. https://doi.org/10.3390/microorganisms10081579

Zaghloul, A., Saber, M., Gadow, S., Awad, F., 2020. Biological indicators for pollution detection in terrestrial and aquatic ecosystems. Bulletin of the National Research Centre 44(1), 127. https://doi.org/10.1186/s42269-020-00385-x

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30(5), 614–620. https://doi.org/10.1093/bioinformatics/btt59