Waikato Regional Council Technical Report 24/21

Review of the potential use of metagenomics for monitoring soil health

ISSN 2230-4363 (Online)
waikatoregion.govt.nz

Prepared by	Eva Biggs, Ziva Louisson, Jo Cavanagh Manaaki Whenua – Landcare Research Gavin Lear - University of Auckland
For	Waikato Regional Council Private Bag 3038 Waikato Mail Centre HAMILTON 3240
Publication date	March 2026
Document ID	30190258

	Name	Date
Peer Reviewer	Matthew Taylor	June 2025
Approving Manager	Mike Scarsbrook	February 2026

Manaaki Whenua
Landcare Research

Review of the potential use of metagenomics for monitoring soil health

Prepared for: Waikato Regional Council

August 2024

Review of the potential use of metagenomics for monitoring soil health

Contract Report: LC4505

Eva Biggs, Ziva Louisson, Jo Cavanagh

Manaaki Whenua – Landcare Research

Gavin Lear

University of Auckland

Reviewed by:	*Approved for release by:*
Danielle Middleton Senior Researcher – Research Priority Area Leader Manaaki Whenua – Landcare Research	Graham Sevicke-Jones Science General Manager Manaaki Whenua – Landcare Research

---

Disclaimer

This report has been prepared by Landcare Research New Zealand Ltd for Waikato Regional Council. If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

Contents

Summary

Project and client

- Waikato Regional Council contracted Manaaki Whenua – Landcare Research and the University of Auckland to undertake a review of the recent use of metagenomics to inform soil functioning and its application in soil environmental monitoring.

Objectives

- To review the national and international literature on the use and evaluation of metagenomics in soil quality monitoring to provide an assessment of the use of eDNA (environmental DNA) approaches for monitoring soils in New Zealand, focusing on metagenomics.
- To provide recommendations on the next steps for further investigating the integration of metagenomics methods into soil monitoring programmes.

Results

- 'Omics' techniques have become increasingly popular for analysing soil communities, their functioning and metabolites, and eDNA metabarcoding approaches have been widely used since the mid-2000s.
- In recent years extensive research on using metabarcoding approaches to inform regional council state of the environment (SOE) soil quality monitoring has been undertaken. However, metabarcoding approaches have limitations for monitoring soil quality: they provide extensive information about what is in the soils, but not how well soils are functioning and what is enabling various soil processes.
- Untargeted sequencing of all genomes in a sample through 'shotgun' metagenomic sequencing (hereafter referred to as metagenomics) provides greater information on soil functions and offers far greater potential for use in environmental monitoring.
- In metagenomic analysis there are four main steps: sampling, DNA extraction, sequencing, and bioinformatic analysis. After DNA has been extracted a decision is needed on whether short-read or long-read sequencing is more appropriate. There are three technologies that can achieve this, each with advantages and disadvantages. The final choice depends on the research question, computational and financial resources, and the bioinformatic expertise available.
- Bioinformatic processes can also be done in different ways. Sequenced reads can either be analysed by assembling them into metagenome-assembled genomes, or by directly analysing sequence reads after quality control and inferring biological function on unassembled data sets.
- The cost of sequencing technologies is steadily decreasing, enabling increased sequencing depth (how many times a nucleotide is sequenced in a given sample) and more accurate long-read sequencing. Also, the availability of high-performance computing clusters is enhancing the capability to process and analyse large data sets efficiently.
- Metagenomic analysis generates gene abundance matrices. These matrices show the presence or abundance of specific genes (or taxa) in each sample, allowing

researchers to compare patterns of gene distribution across various samples. This helps to identify functional genes or pathways that are more or less abundant and potentially linked to specific environmental factors or treatments.

- However, comparing results across multiple studies is challenging due to the lack of standardisation in techniques. Various laboratory and bioinformatic techniques are available, and the chosen methods can vary greatly among research groups, influencing the results obtained.
- The metagenomics approach is an underexplored avenue to explore soil health through analysing phylogenetic or functional genes. Since metagenomics involves collecting data on all the genetic material present in a sample, it can be used to quantify the presence of DNA sequence barcodes (e.g. 16S rRNA genes, similar to metabarcoding), and therefore the taxonomic composition of a community. It is important to use the correct tool as some metagenomic profilers report relative sequence abundance while others report relative taxonomic abundance.
- There are several case study examples of the use of metagenomics, including for:
  - nutrient cycling (carbon, nitrogen, phosphorus)
  - contaminant responses (metals, organic contaminants)
  - antibiotic resistance and production
  - pathogens – providing an understanding of the conditions that support disease-suppressive capacities

Conclusions and next steps/recommendations

- For councils to effectively adopt metagenomics or other omics techniques for soil health monitoring, it is crucial to standardise protocols for sample collection, data analysis, and storage. This requires a coordinated research effort to determine the appropriate scale for using metagenomics in soil health monitoring. Also, the necessary bioinformatic expertise is essential for analysing data obtained from sequencing providers.
- Integrating artificial intelligence (AI) in metagenomic data analysis presents significant opportunities to enhance the processing and interpretation of complex data sets. AI developments can automate the processing of sequence data and the identification of biologically meaningful features, reducing the need for human intervention and making functional data more accessible to end-users, including councils.
- National investment and cohesive research programmes are needed to develop the required standardisation and to conduct case studies for testing and validating metagenomic soil health indicators on larger (including regional and national) scales. Although the interpretation of data from metagenomic and metabarcoding analyses is still in its early stages, investing in metagenomics offers substantial value by providing deeper insights into soil functioning. Progress towards standardising sample collection and sequencing approaches will help generate comparable data sets and facilitate wider testing and validation of metagenomic data for routine soil health monitoring.

1 Introduction

Regional council state of the environment (SOE) soil quality monitoring is the primary means by which councils monitor the health of soils in their region. Monitoring has been conducted since the early 2000s and currently comprises measuring seven primary soil quality indicators (carbon, nitrogen, pH, anaerobically mineralisable nitrogen, Olsen P, bulk density, and macroporosity). Some reviews and evaluations of the indicators and/or 'target values' being used have been undertaken (Mackay et al. 2013; Cavanagh et al. 2023).

Molecular analyses of soil offer the opportunity to gain a wealth of information from single samples, and eDNA metabarcoding approaches have been widely used since the mid-2000s. In recent years, extensive research on using metabarcoding approaches to inform regional council SOE soil quality monitoring has been undertaken, primarily in partnership with Waikato Regional Council (Hermans et al. 2016; Hermans et al. 2020 a, b, c; Mathieu et al. 2020).This research has shown that analysis of metabarcoding data, which typically provide information on organisms' taxonomic (e.g. species) identity, are to some extent consistent with soil health indicators used by regional councils (Hermans et al. 2016; Hermans et al. 2020a).

However, metabarcoding approaches have limitations for monitoring soil quality: they provide extensive information about what is in the soils, but not how well soils are functioning and what is enabling various soil processes. In contrast, metagenomics methods analyse the entire genetic material recovered from an environmental sample (e.g. genes related to metal resistance, carbon and nitrogen cycling, degradation of pesticide compounds, and pathogenesis). In this way metagenomics provides greater information on soil functions and offers far greater potential for use in environmental monitoring.

Waikato Regional Council contracted Manaaki Whenua – Landcare Research and the University of Auckland to undertake a review to ascertain the state of knowledge on the recent use of metagenomics to inform soil functioning and its use in environmental monitoring. The next steps to fill identified knowledge gaps in a New Zealand context are then outlined.

2 Background

2.1 What is soil health?

The concept of soil 'health' has gained popularity for describing the state and desired condition of soil. In the European Union, for example, the Soil Health and Food mission board has established a goal to ensure that 75% of European soils are healthy or significantly improved by 2030. Historically, soil 'quality' was the commonly used term, and there is considerable debate in the literature regarding the distinction between soil health and soil quality, with many experts considering these terms to be synonymous (Bünemann et al. 2018; Creamer et al. 2022).

The definition of soil health as 'the continued capacity of a soil to function as a vital living ecosystem that sustains plants, animals and humans', originating from the USDA Natural Resources Conservation Service[1], has been used in several recent papers (Creamer et al. 2022; Liptzin et al. 2023). Taylor (2024) on page 1 has proposed the following definitions for SOE soil quality monitoring:

> ***soil health***: *the condition of the soil and whether it has been degraded. This assesses the soil's ability to carry out functions and provide services, irrespective of its land use.*
>
> ***soil quality***: *How well the soil performs its intended functions, i.e. suitability for its specific land use.*

Internationally, an ecosystem services framework is increasingly used to assess soil quality and health. This framework groups individual soil functions to determine the extent to which soils contribute to ecosystem services (Bünemann et al. 2018). (Schulte et al. 2014) define the most important functions in agroecosystems as:

- primary productivity
- nutrient cycling
- water regulation and purification
- carbon sequestration
- habitat for above- and below-ground biodiversity.

In the context of agricultural systems, Creamer et al. (2022) use four functions; water regulation and purification, nutrient cycling, carbon and climate regulation and pest and disease regulation. These functions are further divided into sub-functions to link them to specific processes, noting that some processes are relevant for multiple sub-functions or functions. For example, they identified the sub-function of nutrient assimilation within the broader function of nutrient cycling. Nutrient cycling involves processes such as mineralisation, immobilisation, and uptake by plants. These processes are crucial for maintaining soil fertility, which in turn supports plant growth and ecosystem productivity.

An alternative approach adopted by the European Environmental Agency (Baritz et al. 2021) gives greater consideration to the context in which soil indicators are likely to be used: in the management of land or soil to reduce soil degradation arising from soil threats, including soil erosion, sealing, and contamination (including excess nutrients).

Regardless of whether 'soil health' or 'soil quality' is used, there is a need to measure certain properties (commonly called indicators) that provide insights into the state or condition of the soil.

---

[1] https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/national/soils/health/?cid=stelprdb1048783

2.2 Objectives for state of the environment soil quality monitoring

The statutory requirements under the Resource Management Act 1991 (RMA) led to the development of a national soil quality monitoring programme in New Zealand. This began with the Sustainable Management Fund project 'Implementing soil quality indicators for land,' which collected new soil quality data from approximately 500 sites across New Zealand between 1999 and 2001 (Sparling et al. 2000; Sparling et al. 2001b, a). This project, known as the 500 Soils project, built on earlier work from the Sustainable Management Fund Project 'Trialling soil quality indicators for land' (Sparling & Schipper 1997; Sparling et al. 1998). Seven key soil properties were analysed: soil pH, bulk density, microporosity, Olsen P, total carbon, total nitrogen, and mineralisable nitrogen.

A subsequent review by Hill et al. (2003) on page 3 helped further develop the programme and identified the following objectives for a national soil quality monitoring programme:

- *to provide an early-warning system for the effects of primary land uses on long-term soil quality (physical, chemical, biological)*
- *to track and identify issues related to the effects of land use on long-term soil quality (which may also be district/area-specific)*
- *to utilize the results for state of the environment reporting and policy development*
- *where possible, to integrate soil quality monitoring with other regional monitoring (e.g. water, especially groundwater).*

Hill et al. (2003) noted that the soil quality monitoring programme aimed to address the seven main soil management issues identified by the National Land Monitoring Forum: structural decline, nutrient depletion, carbon depletion, nutrient saturation, biological activity change, pH change, and contaminants. The Land Monitoring Forum developed a guide for soil quality monitoring (LMF 2009), adopting Hill et al.'s (2003) objectives.

While earlier studies focused on soil structure, pH, and concentrations of carbon, nitrogen, and phosphorus, more recently regional councils initiated the Monitoring of Soil Quality and Trace Element National Environmental Monitoring Standard (SQNEMS) to address the following objectives:

- *to provide a representative assessment of the quality of regional soil resource states and trends over time*
- *to assess soil quality across various land uses and soils representative of regional soil resources*
- *to provide an early warning system to identify the effects of primary land uses on long-term soil quality (physical, chemical, biological traits, including soil trace elements)*
- *to detect spatial and temporal changes in soil quality and soil trace elements*
- *to integrate with other regional monitoring (e.g. groundwater monitoring)*
- *to collect scientifically robust data*
- *to provide data that can be aggregated for national reporting.*

The inclusion of trace element monitoring identified the accumulation characteristics of diffuse soil contaminants under different land uses, such as elevated soil cadmium concentrations under dairying and pasture, and elevated zinc in soils under pasture. In the Waikato Region, concentrations of chromium, mercury, and nickel were found to be elevated and trending upwards under annual cropping, whereas soil concentrations of lead tended to be either stable or reducing. Ongoing monitoring is required to confirm future trends in the concentrations of elements linked to phosphate fertilisers and agricultural chemicals (Taylor et al. 2011).

2.3 Biological indicators in state of the environment soil quality monitoring

Understanding soil biology is key to understanding the true state of a soil’s health or quality, and to connecting physiochemical parameters to changes in soil function and potential vegetative performance (Bünemann et al. 2018). The biological component of soil encompasses populations of soil biota, including micro-organisms (bacteria, fungi, protozoa, archaea, algae) and soil fauna (which can be divided into macro-, meso-, and microfauna).

Soil biology drives many soil functions, including organic matter decomposition, nutrient cycling, and the maintenance of soil structure. Soil biological populations also form symbiotic relationships with plant roots, suppress pests and pathogens, and influence the weathering and solubilisation of minerals (Sparling 1997). Because soil biological populations are sensitive to changes in the soil environment that can arise through land management (e.g. nutrient loading, structural degradation, intensification, or contaminant concentrations), they have the potential to be used as soil quality indicators (Bhaduri et al. 2022).

Perhaps most significantly, it is only through direct assessment of soil biology that biodiversity in soils can be assessed: soils are estimated to hold between 25% and 59% of the world’s biodiversity (UNEP & FAO, accessed July 2024; Anthony et al. 2023). However, soil biodiversity has typically received little attention in soil quality monitoring programmes.

Using biological indicators can include direct assessments of soil biology. This may involve:

- counting the organisms present in soils (e.g. earthworm abundance; Schon et al. 2023)
- comparing the relative abundances or diversity of organisms present (e.g. a high ratio of bacterial- to fungal-feeding nematodes may indicate nutrient cycling is occurring rapidly through bacterial decomposition (Yeates 2003)
- biomass assessments
- assessment of biological processes (e.g. via assessment of soil enzyme activities; Sapsford & Dickie 2023).

Chemical measures of biological relevance may also be examined (e.g. plant-available nitrogen, or the evolution of carbon dioxide or nitrogen gas).

The idea of incorporating soil biological indicators has been considered from the inception of soil quality monitoring programmes in New Zealand (Sparling 1997, Table 1). Re-evaluations of potential biological indicators were undertaken in 2011 (Mackay et al 2013) as well as in more recent reports (Stevenson 2022; Cavanagh et al. 2023). Most recently, research has focused on the potential use of metabarcoding approaches for developing soil biological quality indicators (e.g. Hermans et al. 2016; Hermans et al. 2020; Holdaway et al. 2017; Louisson et al. 2023 (see further discussion on this work in section 4.1. a, c).

Internationally, Griffiths et al. (2018) and Zwetsloot et al. (2022) provide an evaluation of and approaches to using soil biological indicators. However, selecting appropriate indicators and interpreting the results remain the biggest barriers to utilising soil biological indicators in New Zealand soil quality monitoring programmes.

Table 1. Biological measures trialled as soil quality indicators in initial state of the environment monitoring trials

Indicator	Soil quality information	Method*
CO_2 efflux	Soil respiration and microbial activity	7-day incubation at 25°C
Microbial biomass	Active soil organic fraction	Fumigation-extraction
Potentially mineralisable nitrogen	Readily mineralised nitrogen reserves	Waterlogged incubation

Source: (Sparling et al. 1998)

* Details of all methods are given in Sparling (1997)

Notes: Potentially mineralisable nitrogen was carried forward as a key property to include in soil quality monitoring.

3 Objectives

Our objectives were to review the national and international literature on the use and evaluation of metagenomics in soil quality monitoring to provide:

- an assessment of the state of knowledge of the use of eDNA approaches for monitoring the state of soils in New Zealand (i.e. what we have learned from prior metabarcoding analyses).
- a summary of metagenomics methods, addressing what is metagenomics, how is metagenomics work completed, what are the realistic outputs of metagenomics analyses, and how do they differ / improve on the outputs of prior metabarcoding studies?
- a summary of the current use of metagenomics for analysing soil processes
- an overview of the relationship between soil metagenomics methods and soil functioning (e.g. relationships between soil nitrogen fluxes and the abundances of nitrogen-cycling genes) and perceived advantages over metabarcoding approaches for assessing the health/functioning of soils
- avenues for the application of metagenomics to environmental monitoring of soils: what, realistically, are the pathways for implementation and interpretation from various perspectives (i.e. landowners, councils, scientists)?

- identification of any key knowledge gaps and barriers to implementation, including, where possible, reference to study scale and costs
- recommendations for the next steps for further investigation and integration of metagenomics methods into soil monitoring programmes, based on the outcomes of our prior review of the literature.

4 Current use of omics techniques for analysing soil processes

In recent years 'omics' techniques have become increasingly popular to analyse soil communities, their functioning and metabolites (Figure 1). DNA-based approaches are an underexplored avenue that offer the opportunity to enhance our assessments of soil health through analysing phylogenetic or functional genes. This can be achieved by examining changes in the abundance of targeted marker genes (such as the 16S rRNA gene) to quantify the taxonomic composition of prokaryotic communities. Alternatively, untargeted sequencing of all genomes in a sample through shotgun metagenomic sequencing (hereafter referred to as metagenomics) offers an overview of all the taxa and functional genes identified in a sample.

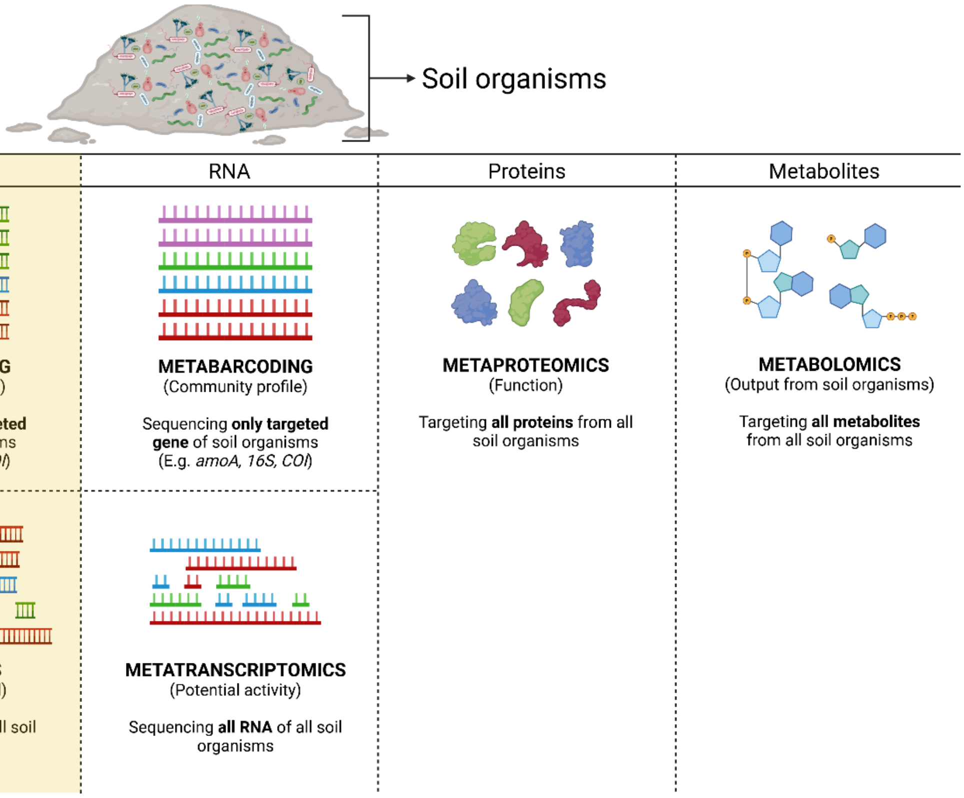

Figure 1. An overview of omics techniques used in analysing soil communities.
Notes: The figure showcases the target of each omics technique (DNA, RNA, proteins or metabolites) and what each technique measures (community profile, potential activity, function or output). This review focuses on DNA-based techniques, metabarcoding, and metagenomics.

4.1.1 Metabarcoding

Metabarcoding is a high-throughput DNA-sequencing approach to analyse the abundance and diversity of specific genetic markers, often in order to analyse taxonomic diversity (Lear et al. 2018). Metabarcoding usually involves amplifying and sequencing ribosomal RNA (rRNA) genes for taxonomic identification, including 16S genes for prokaryotes (Caporaso et al. 2012), cytochrome *c* oxidase subunit I (*COI*) genes in animals (Leray et al. 2013), or the nuclear internal transcribed spacer (ITS) region of fungi (Ihrmark et al. 2012). Various other genetic regions may be targeted to analyse the diversity of specific functional genes, such as *amoA* gene for ammonia oxidation (Herbold et al. 2015).

In metabarcoding studies, DNA is first extracted from the sample material of interest (in this case soil; Lear et al. 2018) before the intended barcoding region is amplified by polymerase chain reaction (PCR) amplification and sequenced by high-throughput DNA sequencing (e.g. Illumina). The targeted genetic regions, known as ‘barcode regions’ (Hebert et al. 2003), are chosen because they contain sequences that are variable, allowing them to be used to assess the diversity of the target region sequences.

Metabarcoding has become the primary technique for microbial biodiversity assessments, although it is less common for larger organisms. The Earth Microbiome Project (https://earthmicrobiome.org/) is a prominent example, cataloguing microbial diversity globally (Thompson et al. 2017). National-scale inventories of soil microbial diversity have since been conducted across Britain (Griffiths et al. 2011), France (Karimi et al. 2018), and Australia (Bissett et al. 2016), as well as across various regions of New Zealand (Hermans et al. 2020a) (Figure 2).

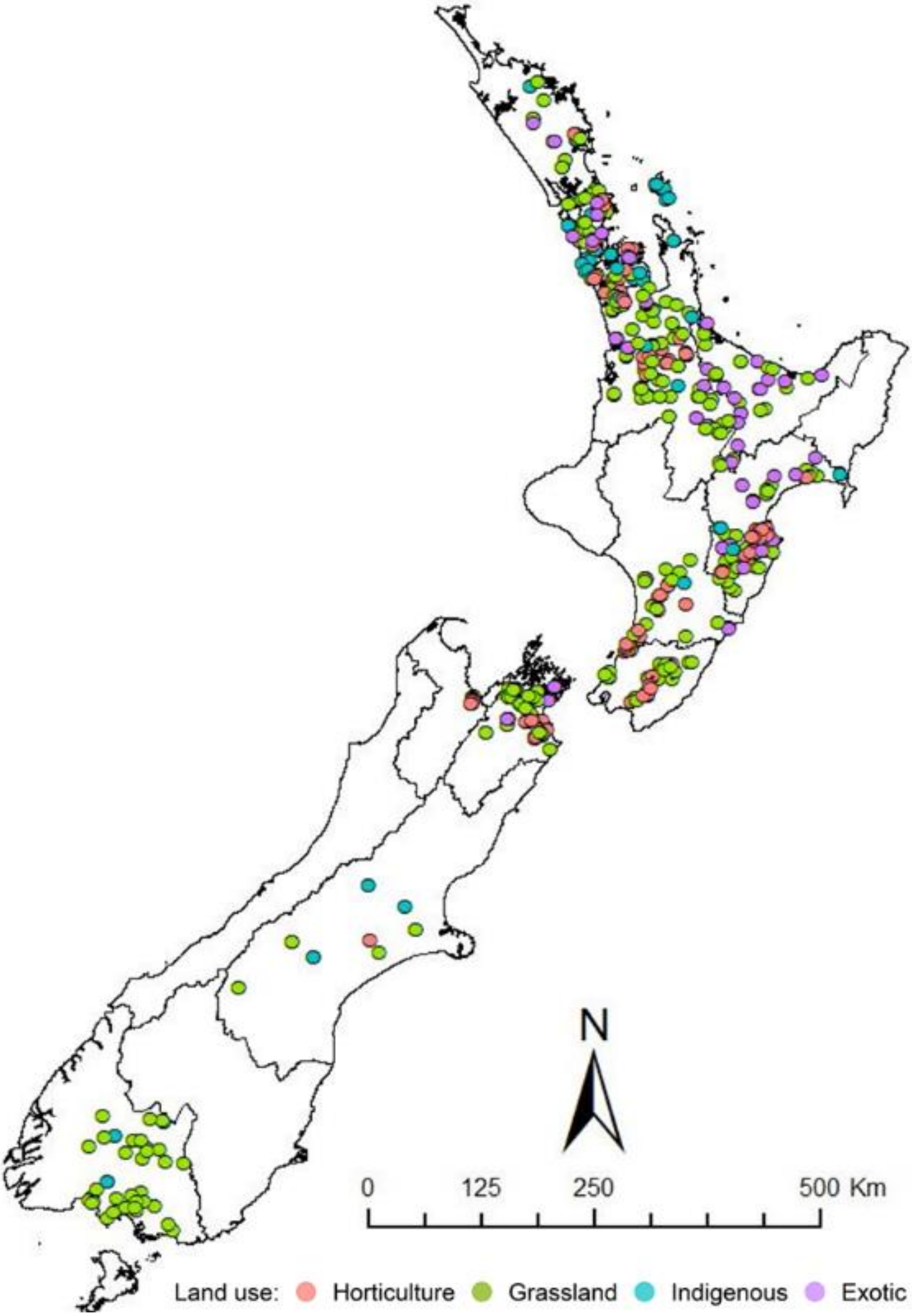

Figure 2. Map of national-scale sampling from Hermans et al. 2020a showing the range of land uses examined in their study.

In addition to mapping the distributions of microbial taxa, many studies have used metabarcoding to assess the impacts of agricultural practices on soil microbial communities (Hartmann et al. 2015; Wu et al. 2021; Romdhane et al. 2022). However, approaches to predict soil attributes from the analysis of metabarcoding data remain in their infancy. Hermans et al. (2016) demonstrated, from the analysis of 110 sites, a correlation between microbial community composition and soil physiochemical properties. In particular, the authors noted strong correlations between the relative abundances of specific taxa and both soil pH and concentrations of Olsen P in the soil (often used as a measure of plant-available phosphorus).

Subsequent studies have reinforced the potential of metabarcoding for developing novel soil biological quality indicators (Holdaway et al. 2017; Delgado-Baquerizo et al. 2020; Hermans et al. 2020a; Louisson et al. 2023). As a proof of concept, Hermans et al. (2020a) analysed over 3,000 soil samples from 600 sites across New Zealand using random forest models. After stratified random sampling, bacterial community composition data from 20% of the sites predicted land use and soil physiochemical variables with over 80% accuracy. Fernandez et al. (2016) used a similar approach to confirm that soil bacterial community composition can even be used to predict farming systems (i.e. conventional versus organic), and soil chemical and biological attributes to a different extent.

4.2 Limitations of metabarcoding: why metagenomics?

Although metabarcoding has proven the importance of microbial data, it has its limitations. The accessibility of amplicon sequencing has led to a more taxon-centric approach to defining microbial bioindicators of soil health, as outlined in the previous section. Understanding how the abundance of certain taxa or the composition of soil microbial communities responds to environmental stress is valuable, but it does not directly indicate how ecosystem processes regulated by microbes are affected. Since many soil microorganisms remain uncultured, the functional potential inferred from amplicon sequencing is largely putative, via predictive tools such as PICRUSt2 (Langille et al. 2013).

Metagenomics offers significant advantages because it captures the entire genetic content of a sample, allowing for direct analysis of changes in the abundance of functional genes and metabolic pathways. Currently, the incorporation of microbial functional data into soil ecosystem monitoring remains largely unexplored, and the use of metabarcoding to identify bioindicators in small-scale studies may not be consistent across different times and environments. Furthermore, environments monitored by councils may not represent extreme conditions, making the detection of significant differences even more challenging using metabarcoding.

5 Metagenomics

5.1.1 Overview of the metagenomics workflow

Metabarcoding is primarily used to characterise taxonomic diversity in soil, since it is based on the amplification of a targeted genetic region. Metagenomics, on the other hand, involves the untargeted sequencing of all genetic material in a sample, allowing us to simultaneously characterise both taxonomic diversity and the functional potential of a microbial community in soil (Figure 1) (Quince et al. 2017).

There have been several reviews describing the metagenomics workflows (Quince et al. 2017; Zhou et al. 2022; Aizpurua et al. 2023). Briefly, DNA gets extracted, fragmented, and directly sequenced without the need for PCR amplification, which avoids polymerase-induced errors and primer biases. One of the main metrics in metagenomic sequencing is sequencing depth. This is defined by the number of times each nucleotide is sequenced in each sample. If we have high sequencing depth, we can even detect organisms present at low abundances.

Before sequencing it is important to decide whether short-read or long-read sequencing is more appropriate. Researchers can choose between several different sequencing strategies, facilitated by Illumina for short-read sequencing, where individual pieces of sequences are 150–300 base pairs long (Segerman 2020), and by Oxford Nanopore Technologies (Nanopore) and Pacific Biosciences Sequel (PacBio) for long-read sequencing, where we can consistently achieve a continuous sequence length of over 10,000 base pairs (Agustinho et al. 2024). There are advantages and disadvantages in all three technologies, and which one to choose will depend on the research question, computational resource availability, and financial constraints (Tao et al. 2023; Eisenhofer et al. 2024) (Figure 3, top).

Bioinformatic processes can also follow different workflows, depending on the research question and the available computational resources and bioinformatic expertise. Sequenced reads can either be analysed by assembling them into metagenome-assembled genomes, or by directly analysing sequence reads after quality control and inferring biological function based on unassembled data sets (Van der Jeugt et al. 2022) (Figure 3, bottom).

It is important to note that ongoing advances in sequencing technologies are making these tools more accessible. The cost of sequencing technologies is steadily decreasing, enabling increased sequencing depth and more accurate long-read sequencing. Also, the availability of high-performance computing (HPC) clusters is enhancing the capability to process and analyse large data sets efficiently. For example, New Zealand eScience Infrastructure (NeSi) is a provider of HPC resources that allows researchers to perform computationally intensive activities. These improvements are lowering the financial and computational barriers, allowing more comprehensive and detailed metagenomic studies. As a result, researchers can now undertake larger-scale projects, leading to more in-depth insights into microbial communities and their functional potentials.

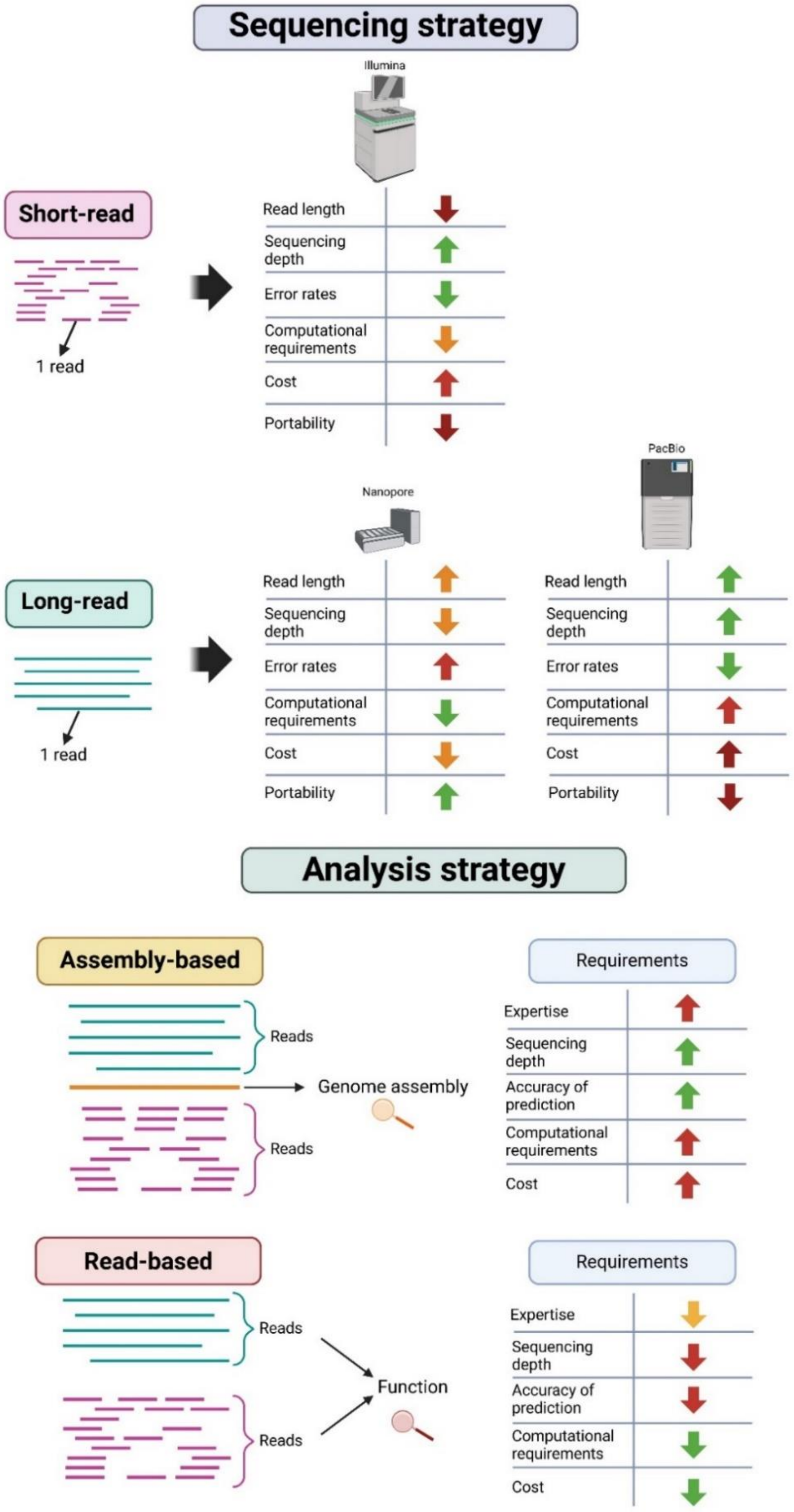

Figure 3. Visual guide representing differences among sequencing and bioinformatic analysis strategies.

Notes: Top: sequencing strategies and a comparison of the properties of each sequencing technology. Arrows indicate whether the attribute performance is lower or higher relative to other sequencing technologies in this figure, and the colours represent the relative performance among the strategies: green = high performance, orange = moderate performance, red = lower performance. Bottom: Bioinformatic analysis strategies and the requirements for each. Arrows and colours compare the two strategies, as above.

5.1.2 Interpreting metagenomic data outputs

Irrespective of which sequencing platform or processing method is used, the result is the generation of gene abundance matrices. These matrices show the presence or abundance of specific genes (or taxa) in each sample, allowing researchers to compare patterns of gene distribution across various samples. This helps identify functional genes or pathways that are more or less abundant and potentially linked to specific environmental factors or treatments.

Despite the vast amount of data generated by metagenomics, which can initially be challenging to navigate, it offers a comprehensive overview of microbial communities. A significant portion of these data include highly conserved housekeeping genes, which, although essential for basic cellular processes, may overshadow ecologically relevant genes (Quince et al. 2017). By focusing on ecologically significant genes, researchers can uncover valuable insights into microbial functions and their roles in different environments. Databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto 2000) and SEED (Overbeek et al. 2014) facilitate this process by allowing targeted analysis of specific genes or pathways. For example, examining nitrogen-cycling genes can reveal how land use and vegetation types influence key biogeochemical processes, providing crucial information for land management and ecosystem forecasting (Figure 4) (Louisson et al. 2023).

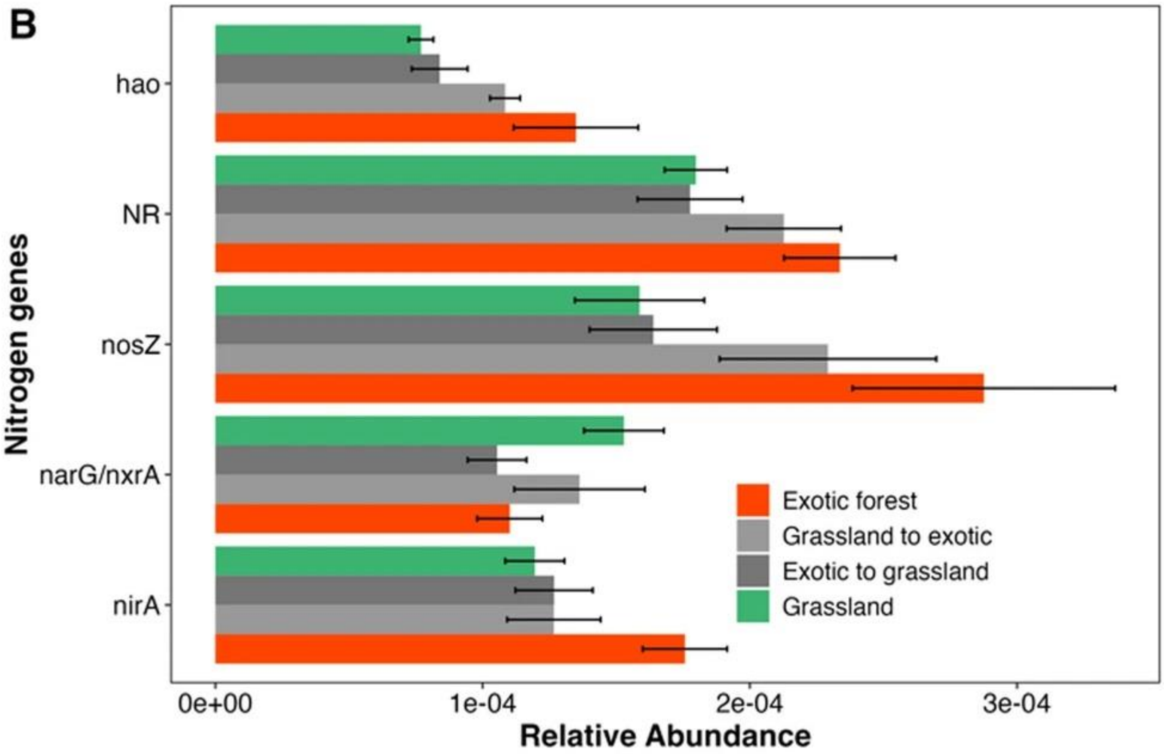

Figure 4. Figure from Louisson et al. 2023 displaying the mean relative abundance of functional genes involved in nitrogen cycling that significantly differ in relative abundance between grassland and exotic forest soils.

As more studies generate metagenomic sequence data, opportunities for comparative analyses increase. Combining data from multiple studies offers advantages such as incorporating a range of environments and conditions while optimising resources. Meta-analyses can provide a broader ecological perspective, helping to identify consistent ecological patterns across different scales (e.g. country or global) and land uses.

However, comparing results across multiple studies is challenging due to the lack of standardisation in techniques. Various laboratory and bioinformatic techniques are available, and the chosen methods can vary greatly among research groups, influencing the results obtained (Szóstak et al. 2022). Therefore, as more metagenomic data are generated, there is a need for a standardised approach to analysing samples to maximise the benefits of cross-study analyses. Previous research has highlighted the need to standardise the analysis of sequence data from environmental samples, recommending robust methods for a standardised approach (Lear et al. 2018). Implementing such guidelines will help produce more comparable and reproducible results across studies, ultimately advancing our understanding of microbial diversity and functioning.

Further interpretation of gene abundance matrices is required to enable their use in the context of soil health monitoring, and artificial intelligence (AI) offers opportunities to enhance the processing and interpretation of these complex data sets (He et al. 2022). AI is already applied in tasks such as *de novo* genome assembly, but some manual curation is still involved (Hillmann et al. 2020; Adams et al. 2023). As AI technologies advance, fully automated pipelines for processing metagenomic data, from raw reads to functional annotation and ecological interpretation could be developed. This will allow end-users and councils to utilise functional data without needing bioinformatic expertise, making the processing of large amounts of data more accessible and leaving capacity for a more hands-on approach.

6 Integrating metagenomics and soil functioning

6.1 Relationship between metagenomics and soil functioning

The metagenomics approach is an underexplored way to explore soil health through analysing phylogenetic or functional genes. Since metagenomics involves collecting data on all the genetic material present in a sample, it can be used to quantify the presence of DNA sequence barcodes (e.g. 16S rRNA genes, similar to metabarcoding), and therefore the taxonomic composition of a community.

While understanding how the abundance of certain taxa or the composition of soil microbial communities responds to environmental stress does offer important information, it does not directly indicate how important ecosystem processes regulated by microbes may be affected. The untargeted sequencing of all genomes in a sample through metagenomic sequencing is primarily used to provide an overview of the functional genes present in a sample (i.e. its functional potential more than its taxonomic diversity).

The activity of microorganisms is important for maintaining the functioning of key ecosystem services. However, the few methods that have been successfully applied to assess soil functioning tend to include measures of enzyme activities, rates and extents of organic matter degradation, and measures of basal respiration. An important consideration is that DNA-based sequencing methods such as metagenomics do not distinguish between active and inactive genes. Therefore, metagenomic data are not a direct measure of the metabolic activity of a community, but rather its metabolic potential (Moran 2009). However, as highlighted by Thiele-Bruhn et al. (2020), it is the potential functioning of microbial communities, rather than actual activity levels, that probably provides a more useful, cumulative measure of soil attributes compared to the analysis of activity levels. The latter may be more affected by short-term properties, including nutrient availability, current pedoclimatic conditions, and soil cultivation activities. Thus, analysis of soil metagenomics data, which provides information about microbial functional potential rather than actual activity, appears well placed to provide information on soil quality and health.

6.2 Case subjects (studies) for use of metagenomics

6.2.1 Carbon cycling (including fluxes of carbon dioxide and methane)

Metagenomics can provide useful insights into the potential degradation of soil carbon (C) stocks, and the evolution of gases, including carbon dioxide (CO_2) and methane (CH_4). For example, metagenomics methods can be used to assess the potential activity of cellulases (e.g. *celA*, *celB*), ligninases (e.g. the lignin peroxidase *lignP* and the manganese peroxidase *mnp*) and hemicellulases (e.g. the xylanase *xylA* and the mannanase *manA*) to degrade cellulose, lignin, and hemicellulose, respectively. By analysing the abundances of these genes, researchers can infer the microbial decomposition of organic C, and microbial contributions to CO_2 production.

For example, Liu et al. (2020) found that respiration rates were positively correlated with the elevated relative abundances of genes encoding indole pyruvate ferredoxin oxidoreductase, an essential enzyme in C metabolism that produces CO_2. Further, respiration rates were negatively correlated with the elevated relative abundance of gene predictors relating to UDP-N-acetylmuramate dehydrogenase, which is an essential enzyme in sugar metabolism, and assimilates CO_2 to form carbohydrate. Studies such as these provide the initial steps in linking the presence and abundance of microbial functional genes to soil microbial processes, and thus to soil quality and health attributes.

Metagenomics has also been used to assess soil CH_4 production in various ecosystems (Abraham et al. 2020; Zhou et al. 2021; Frey et al. 2022). However, links between abundances of methanogens and methanogenesis genes appear only weakly correlated with CH_4 fluxes (Zhou et al. 2021). Such findings suggest that other microbial guilds (or their related genes) need to be better accounted for to predict CH_4 fluxes at landscape scales more accurately (Hartman et al. 2024).

6.2.2 Nitrogen cycling

Metagenomics methods have proven particularly useful in assessing the impact of land management practices on soil nitrogen cycling. Compared to the carbon cycle, the nitrogen cycle is less complex. Nitrogen (N) exists in fewer forms (e.g. NH_3, NO_2^-, NO_3^-, N_2) and its transformation thus involves a smaller number of specific processes (e.g. N fixation, ammonia oxidation, denitrification) that are mediated by distinct functional genes, such as *nifH*, *amoA*, and *nosZ*. In certain cases the presence of specific taxa, such as *Nitrosomonas* (nitrification) or *Rhizobium* (nitrogen fixation), can also be directly linked to functional roles in N cycling.

Metagenomics has been used to assess not only the impacts of land management on the potential of soils to store and cycle N, but also to ascertain the ecological changes occurring in the soil. For example, Zhao et al. (2023) questioned how the biomass harvest of *Miscanthus* (a C_4 perennial grass) did not contribute to soil N losses, even over prolonged (15-year) periods. Analysis of the metagenomic data revealed changes in soil N-cycling in the soil under *Miscanthus* cropping and alterations in dominant microbial contributors to the soil N cycle. In particular, the potential for soil N fixation was dramatically enhanced; the L-glutamine to L-glutamate step was significantly promoted, whereas denitrification processes were diminished. Further analysis of the data revealed the genus *Anaeromyxobacter* to be the main contributor to soil N fixation potential.

The authors determined that increased N fixation, principally by *Anaeromyxobacter*, along with decreased soil nitrification and denitrification, helped to sustain soil N balances over a period of 15 years. The results of studies such as these help to inform the sustainability of agricultural practices and the best approaches for realising soil N and biomass production.

6.2.3 Phosphorous cycling

Unlike N, soil phosphorus (P) exists in many forms, complicating attempts to assess soil P cycling potential through the analysis of soil DNA. Previous studies have revealed a strong relationship between the abundances of certain soil-dwelling bacteria (i.e. members of the Gram-negative genus *Bradyrhizobium*) and concentrations of available soil P (Hermans et al. 2016). However, it is unclear if such trends are truly related to soil P, or soil N concentration, since the concentrations of both nutrients tend to be correlated in agricultural soils.

Microbes in soil have a great potential to increase soil P availability, and a positive correlation between the abundance of microbial genes encoding high-affinity P uptake under conditions of low P bioavailability has been reported (Bergkemper et al. 2016). Microbial genes involved in the soil P cycle can be broadly determined to encode four distinct traits: inorganic phosphate solubilisation (e.g. *gcd*, *ppa*, and *ppx*), organic P mineralisation (e.g. *phoA* and *phoD*), transporters (e.g. *pit*, *pstA*, *pstB,* and *ugpQ*), and regulatory genes (e.g. *phoB* and *phoR*) (Wu et al. 2022).

From the analysis of nearly 600 samples, Oliverio et al. (2020) confirmed that the relative abundances of approximately 80 phosphate-related functional genes could be correlated with P availability. Specifically, their results suggest that when phosphate is limited in soil, a primary adaptive strategy employed by bacteria is to catabolise organophosphates to acquire P by CP-lyase or phosphonate pathways. Identification of P-responsive taxa and pathways indicates their potential to serve as useful indicators of P availability, either to predict differences in soil P availability from the analysis of DNA data or to complement more traditional analyses.

6.2.4 Heavy metals

Metagenomics has been shown to be a powerful tool for understanding the prevalence and biological availability of heavy metals in soils and for assessing their likely impact on soil microbial communities and processes. Of particular relevance is the analysis of a collection of genes and genetic elements within microbial communities that confer resistance to heavy metals, often referred to as the *'heavy metal resistome'.* Within the metagenome, this includes genes encoding for proteins involved in processes such as metal sequestration, efflux, transformation, and detoxification, allowing microbes to survive and adapt to environments with varying heavy metal concentrations.

Components of the heavy metal resistome include the following.

- *Efflux pumps:* these function to transport heavy metals out of cells to reduce intracellular toxicity. Examples of target genes of interest are those encoding the CzcCBA cation efflux involved in resistance to zinc, cadmium, and cobalt (Salam et al. 2020), and the *pbrA* gene encoding a Zn(II)/Cd(II)/Pb(II)-translocating ATPase (Maunsell et al. 2006).
- *Metal-sequestration proteins:* these bind and sequester heavy metals, reducing their availability and toxicity. Examples include metallothionein, which bind metals such as cadmium and copper (Li et al. 2020), and microbial chelating agents, or siderophores (Li et al. 2020).
- *Metal transforming enzymes:* these transform metals ions into less toxic forms, often through reduction or oxidation processes. Examples include *merA*, which encodes a mercury reductase that reduces Hg(II) to less toxic Hg(0), and *arsC* (Oregaard & Sørensen 2007), which encodes an arsenate reductase that converts arsenate (As(V)) to arsenite (As(III)) (Villegas-Torres et al. 2011).
- *Metal-specific sensors and regulators:* these mediate the expression of resistance genes in response to the presence of metals. Examples include the ArsR repressor protein that controls the *ars* operon involved in arsenic resistance (Gao et al. 2023), and CadC, which regulates cadmium resistance genes (Kim et al. 2016).

The diversity and abundance of resistome genes may serve as biomarkers for the extent of heavy metal pollution and its impacts on microbial communities.

6.2.5 Organic pollutants

Metagenomics has allowed researchers to predict the prevalence of bioavailable contaminants in soil, and to identify metabolic pathways and potential bioremediation strategies for various soil pollutants. For example, Jeffries et al. (2018) performed metagenomics sequencing of community DNA on soils with a history of pesticide usage and recorded shifts in the abundances of taxa and functional genes present. They found that differences in pesticide degradation potential largely reflected the legacy impact of pesticide usage, again marking a direct link between land management practices and soil microbial functional potentials. Remarkably the authors assert that elevated pesticide degradation potentials could remain even after pesticide use was discontinued 13 years prior.

Metagenomics data can therefore be used as an indicator of degradation potential to support decision-making, including for determining optimal remediation strategies. Metagenomics has also been used to assess the presence and/or degradation potential of multiple organic pollutants, including hydrocarbons (Jurelevicius et al. 2022; Goma-Tchimbakala et al. 2023), pesticides (Sim et al. 2022), and plastic pollutants (Saleem et al. 2023).

6.2.6 Antibiotic resistance and production

In addition to providing a means to assess the potential of soils to store and cycle nutrients and contaminating substances, metagenomics offers the ability to quantify additional community traits, including the prevalence of antibiotic resistance, a pressing global threat to human and animal health. Increased antibiotic resistance in soils may be related to the disposal of both raw and treated sewage waste to land, and the growing reliance on antibiotics such as the ionophore monensin to treat stock animals (Carresi et al. 2024). Fertiliser use has also been reported to increase the abundance of genes encoding for resistance to tetracycline (e.g. *tetW*) and sulfonamide (e.g. *sul2*) (Xie et al. 2018). The concentrations of some heavy metals are also proposed to drive the distribution and spread of antibiotic-resistance genes in the environment (Zhao et al. 2019).

Wang et al. (2023) used metagenomics methods to confirm that zero tillage methods substantially reduce the abundance of resistance genes, including towards β-lactam, sulphonamide, and chloramphenicol antibiotics. Because microbial metabolic activity was significantly higher in the zero tillage soil, researchers postulate that the greater versatility and health of organisms in this treatment, as compared to under conventional tillage, could be due to antibiotic-carrying organisms being outcompeted. Su et al. (2015) reporting similar findings, highlighting the potential for metagenomics methods to clarify the impact of land management practices for levels of antibiotic resistance in soil.

Microbial enzymes are responsible for the production of a plethora of antibiotics within soil, so metagenomics can be used to quantify and characterise the diversity of biosynthetic gene clusters in soil. These gene clusters, encoding for the production of bacterial secondary metabolites, play critical roles in mediating biological interactions and communication, nutrient scavenging, and antibiotic production (Van Goethem et al. 2021). Indeed, natural bacterial secondary metabolites are the basis of more than half of all registered drugs (Cragg & Newman 2013).

Soil pH, already known to have a strong impact on soil microbial community composition, has been shown to have a strong influence on the abundance of biosynthetic gene clusters in agricultural soils, with soil pH management and climate factors shown to disrupt the biosynthetic potential of the soil microbiome (Zhang et al. 2024). While our understanding of the impact of land management on the production of, and resistance towards, antibiotics in soil remains in its infancy, metagenomics methods offer great potential to expand our understanding of the biosynthetic potential of the soil microbiome.

6.2.7 Pathogens

Metagenomics methods can provide insights into the presence of potential pathogens and their associated genes, and can be used to assess the soil types and environmental conditions that support disease-suppressive capacities. For example, Li et al. (2023) performed metagenomic sequencing of soil communities across a pH gradient ranging from 4.0 to 6.0. This revealed that reductions in soil pH affect bacterial community composition and the capacity of soils to combat fungal pathogens, including *Fusarium*, a globally important pathogen. They related this to a concomitant reduction in the synthesis of sulphur compounds, and traits related to soil sulphur metabolism also observed in the metagenomics data.

However, while multiple researchers have shown the huge potential of metagenomics for pathogen surveillance, the sequencing depth can be too low to confirm the presence of pathogens present in low abundance. Also, pathogen taxonomic assignments are biased towards species with their full genome data deposited in DNA sequence databases (Piombo et al. 2021). Thus, the reliable detection of pathogens from metagenomics data remains an aspirational aim.

7 There is more than one approach: multi-omics

Although we focus on the use of metagenomics in this review, it is important to note that there are numerous omics-based technologies that could contribute to improved biological measures of soil health and quality. The following are of most relevance for soil quality monitoring.

- *Metatranscriptomics:* instead of analysing DNA, this assesses gene expression profiles via the analysis of RNA transcripts. The advantage of metatranscriptomics is that it specifically targets the most active community members. By analysing RNA, it avoids the biases associated with relic DNA that persists in soil, with delayed functional and community responses (Bang-Andreasen et al. 2020). Examples of metatranscriptomics studies include assessments of the impact of nutrient amendments, shown to increase the abundance of copiotrophic bacteria (Bang-Andreasen et al. 2020), and observations of higher carbohydrate degradation potential among soil bacteria exposed to long-term (c. 30 years) warming (Roy Chowdhury et al. 2021).
- *Metaproteomics* entails the study of proteins, their structures, functions, and interactions. Metaproteomics and metabolomics (see below) are the most direct methods for identifying the functioning of environmental microbial communities, and have been used to demonstrate the impacts of biostimulants on soil rhizosphere communities (Mattarozzi et al. 2020).
- *Metabolomics* is the comprehensive analysis of metabolites within a biological sample. Although untargeted metagenomics offers rich and functional biochemical

data, it is a nascent field (Brown et al. 2024), and few studies have attempted to link soil metabolites with soil health status.

Although DNA-focused metagenomic studies are currently the most promising for developing novel soil biological indicators due to their stability and reliability, the integration of multiple omics methods, including metatranscriptomics and metabolomics, can provide a more comprehensive understanding of soil health. This integrated strategy can generate robust indicators that effectively inform the most critical outcomes in existing soil health frameworks (Sprunger & Martin 2023).

8 Conclusion

For councils to adopt metagenomics (or other omics techniques) for soil monitoring, a coordinated research effort is required. It is crucial to determine the relevant scale for the use of metagenomics in soil health monitoring, and to standardise sample collection, data analysis and storage protocols. Infrastructure and bioinformatic expertise are essential to undertake metagenomic analyses, and sequencing providers currently provide this to a certain level, notably the production of gene abundance matrices. However, further interpretation of these data is required to enable their use in the context of soil health monitoring, and this is where the use of artificial intelligence (AI) offers opportunities to enhance the processing and interpretation of these complex data sets.

National investment and cohesive research programmes are needed to develop the standardisation required and to conduct case studies to test and validate metagenomic soil health indicators for routine use by stakeholders on larger (including regional and national) scales. Although the interpretation of data generated through metagenomic and metabarcoding analyses is still largely in its infancy in relation to utilisation, investing in metagenomics offers enormous value because of the greater insight into the functioning of soil it provides. AI technologies are integral to enabling the use of these data and offers enormous value because of the greater insight into the functioning of soil it provides. In the interim, progress towards the standardisation of sample collection and sequencing approaches will help provide comparable data sets to enable the wider testing and validation of metagenomic data for soil health monitoring.

9 References

Abraham BS, Caglayan D, Carrillo NV, Chapman MC, Hagan CT, Hansen ST, et al. 2020. Shotgun metagenomic analysis of microbial communities from the Loxahatchee nature preserve in the Florida Everglades. Environmental Microbiome 15(1): 2.

Adams AK, Kristy BD, Gorman M, Balint-Kurti P, Yencho GC, Olukolu BA 2023. Qmatey: an automated pipeline for fast exact matching-based alignment and strain-level taxonomic binning and profiling of metagenomes. Briefings in Bioinformatics 24(6): bbad351.

Agustinho DP, Fu Y, Menon VK, Metcalf GA, Treangen TJ, Sedlazeck FJ 2024. Unveiling microbial diversity: harnessing long-read sequencing technology. Nature Methods 21: 1-13.

Aizpurua O, Dunn RR, Hansen LH, Gilbert MTP, Alberdi A 2023. Field and laboratory guidelines for reliable bioinformatic and statistical analysis of bacterial shotgun metagenomic data. Critical Reviews in Biotechnology: 1-19.

Anthony MA, Bender SF, van der Heijden MGA 2023. Enumerating soil biodiversity. Proceedings of the National Academy of Sciences 120(33): e2304663120.

Bang-Andreasen T, Anwar MZ, Lanzén A, Kjøller R, Rønn R, Ekelund F, et al. 2020. Total RNA sequencing reveals multilevel microbial community changes and functional responses to wood ash application in agricultural and forest soil. FEMS Microbiology Ecology 96(3): fiaa016.

Baritz R, Amelung W, Antoni V, Boardman J, Horn R, Prokop G, et al. 2021. Soil monitoring in Europe. Indicators and Thresholds for Soil Quality Assessments. https://doi.org/10.1109/ehb47216.2019.8969931.

Bergkemper F, Schöler A, Engel M, Lang F, Krüger J, Schloter M, et al. 2016. Phosphorus depletion in forest soils shapes bacterial communities towards phosphorus recycling systems. Environmental Microbiology 18(6): 1988–2000.

Bhaduri D, Sihi D, Bhowmik A, Verma BC, Munda S, Dari B 2022. A review on effective soil health bio-indicators for ecosystem restoration and sustainability. Frontiers in Microbiology 13.

Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, et al. 2016. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. GigaScience 5(1): 21.

Brown RW, Reay MK, Centler F, Chadwick DR, Bull ID, McDonald JE, et al. 2024. Soil metabolomics: current challenges and future perspectives. Soil Biology and Biochemistry 193: 109382.

Bünemann EK, Bongiorno G, Bai Z, Creamer RE, De Deyn G, de Goede R, et al. 2018. Soil quality – a critical review. Soil Biology and Biochemistry 120: 105–125.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. 2012. Ultra- high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6(8): 1621–1624.

Carresi C, Marabelli R, Roncada P, Britti D 2024. Is the use of Monensin another Trojan horse for the spread of antimicrobial resistance? Antibiotics 13(2): 129.

Cavanagh J, Drewry J, Thompson-Morrison H 2023. Review of methods and data used to develop target values for soil quality indicators. Contract Report: LC4305.

Cragg GM, Newman DJ 2013. Natural products: a continuing source of novel drug leads. Biochimica et Biophysica Acta (BBA) – General Subjects 1830(6): 3670–3695.

Creamer RE, Barel JM, Bongiorno G, Zwetsloot MJ 2022. The life of soils: integrating the who and how of multifunctionality. Soil Biology and Biochemistry 166: 108561.

Delgado-Baquerizo M, Reich PB, Trivedi C, Eldridge DJ, Abades S, Alfaro FD, et al. 2020. Multiple elements of soil biodiversity drive ecosystem functions across biomes. Nature Ecology & Evolution 4(2): 210–220.

Eisenhofer R, Nesme J, Santos-Bay L, Koziol A, Sørensen SJ, Alberdi A, et al. 2024. A comparison of short-read, HiFi long-read, and hybrid strategies for genome- resolved metagenomics. Microbiology Spectrum 12(4): e03590-23.

Fernandez AL, Sheaffer CC, Wyse DL, Staley C, Gould TJ, Sadowsky MJ 2016. Associations between soil bacterial community structure and nutrient cycling functions in long- term organic farm soils following cover crop and organic fertilizer amendment. Science of the Total Environment 566–567: 949–959.

Frey B, Varliero G, Qi W, Stierli B, Walthert L, Brunner I 2022. Shotgun Metagenomics of deep forest soil layers show evidence of altered microbial genetic potential for biogeochemical cycling. Frontiers in Microbiology 13.

Gao R, Ma B, Hu M, Fang L, Chen G, Zhang W, et al. 2023. Ecological drivers and potential functions of viral communities in flooded arsenic-contaminated paddy soils. Science of the Total Environment 872: 162289.

Goma-Tchimbakala EJCDA, Pietrini I, Goma-Tchimbakala J, Corgnati SP 2023. Use of shotgun metagenomics to assess the microbial diversity and hydrocarbons degrading functions of auto-mechanic workshops soils polluted with gasoline and diesel fuel. Microorganisms 11(3): 722.

Griffiths BS, Faber J, Bloem J 2018. Applying soil health indicators to encourage sustainable soil use: the transition from scientific study to practical application. Sustainability 10(9): 3021.

Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS 2011. The bacterial biogeography of British soils. Environmental Microbiology 13(6): 1642–1654

Hartman WH, Mesquita CPBd, Theroux SM, Morgan-Lang C, Baldocchi DD, Tringe SG 2024. Multiple microbial guilds mediate soil methane cycling along a wetland salinity gradient. mSystems 9(1): e00936-23.

Hartmann M, Frey B, Mayer J, Mäder P, Widmer F 2015. Distinct soil microbial diversity under long-term organic and conventional farming. The ISME Journal 9(5): 1177– 1194.

Hebert PD, Cywinska A, Ball SL, deWaard JR 2003. Biological identifications through DNA barcodes. Proceedings: Biological Sciences 270(1512): 313–321.

Herbold CW, Pelikan C, Kuzyk O, Hausmann B, Angel R, Berry D, et al. 2015. A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. Frontiers in Microbiology 6.

Hermans SM, Buckley Hannah L, Case Bradley S, Curran-Cournane F, Taylor M, Lear G 2016. Bacteria as emerging indicators of soil condition. Applied and Environmental Microbiology 83(1): e02826-16.

Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M, Lear G 2020a. Using soil bacterial communities to predict physico-chemical variables and soil quality. Microbiome 8(1): 79.

Hermans SM, Buckley HL, Curran-Cournane F, Taylor M, Lear G. 2020b. Temporal variation in soil bacterial communities can be confounded with spatial variation. FEMS Microbiolog Ecology. 96(12): fiaa192.

Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M, Lear G 2020b. Using soil bacterial communities to predict physico-chemical variables and soil quality. Microbiome 8(1): 79.

Hermans SM, Taylor M, Grelet G, Curran-Cournane F, Buckley HL, Handley KM, Lear G. 2020c. From pine to pasture: land use history has long-term impacts on soil bacterial community composition and functional potential, FEMS Microbiology Ecology, 94(4). fiaa041

He Q, Niu X, Qi R-Q, Liu M 2022. Advances in microbial metagenomics and artificial intelligence analysis in forensic identification. Frontiers in Microbiology 13: 1046733.

Hill RB, Sparling G, Frampton C, Cuff J 2003. National soil quality review and programme design. Technical Paper No. 75, Land. Wellington, Ministry for the Environment. Wellington. 34 p.

Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Knight R, Knights D 2020. SHOGUN: a modular, accurate and scalable framework for microbiome quantification. Bioinformatics 36(13): 4088-4090.

Holdaway RJ, Wood JR, Dickie IA, Orwin KH, Bellingham PJ, Richardson SJ, et al. 2017. Using DNA metabarcoding to assess New Zealand's terrestrial biodiversity. New Zealand Journal of Ecology 41(2): 251–262.

Ihrmark K, Bödeker IT, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, et al. 2012. New primers to amplify the fungal ITS2 region: evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiology Ecology 82(3): 666–677.

Jeffries TC, Rayu S, Nielsen UN, Lai K, Ijaz A, Nazaries L, et al. 2018. Metagenomic functional potential predicts degradation rates of a model organophosphorus xenobiotic in pesticide contaminated soils. Frontiers in Microbiology 9: 147.

Jurelevicius D, Pereira RDS, da Mota FF, Cury JC, de Oliveira IC, Rosado AS, et al. 2022. Metagenomic analysis of microbial communities across a transect from low to highly hydrocarbon-contaminated soils in King George Island, Maritime Antarctica. Geobiology 20(1): 98–111.

Kanehisa M, Goto S 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28(1): 27–30.

Karimi B, Terrat S, Dequiedt S, Saby NPA, Horrigue W, Lelièvre M, et al. 2018. Biogeography of soil bacteria and archaea across France. Science Advances 4(7): eaat1808.

Kim HJ, Lim JW, Jeong H, Lee S-J, Lee D-W, Kim T, et al. 2016. Development of a highly specific and sensitive cadmium and lead microbial biosensor using synthetic CadC- T7 genetic circuitry. Biosensors and Bioelectronics 79: 701–708.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31(9): 814–821.

Lear G, Dickie I, Banks J, Boyer S, Buckley HL, Buckley TR, et al. 2018. Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. New Zealand Journal of Ecology 42(1): 10–50A.

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology 10: 1–14.

Li X, Islam MM, Chen L, Wang L, Zheng X 2020. Metagenomics-guided discovery of potential bacterial metallothionein genes from the soil microbiome that confer Cu and/or Cd resistance. Applied and Environmental Microbiology 86(9): e02907-19.

Li X, Chen D, Carrión VJ, Revillini D, Yin S, Dong Y, et al. 2023. Acidification suppresses the natural capacity of soil microbiome to fight pathogenic Fusarium infections. Nature Communications 14(1): 5090.

Liptzin D, Rieke EL, Cappellazzi SB, Bean GM, Cope M, Greub KL, et al. 2023. An evaluation of nitrogen indicators for soil health in long-term agricultural experiments. Soil Science Society of America Journal 87(4): 868–884.

Liu Y-R, Delgado-Baquerizo M, Yang Z, Feng J, Zhu J, Huang Q 2020. Microbial taxonomic and functional attributes consistently predict soil CO_2 emissions across contrasting croplands. Science of the Total Environment 702: 134885.

LMF (Land Monitoring Forum) 2009. Land and soil monitoring: a guide for SoE and regional council reporting. https://envirolink.govt.nz/assets/Envirolink/Land20and20soil20monitoring-A-guide- for-SoE20and20regional20council20reporting.PDF.

Louisson Z, Hermans S, Buckley H, Case B, Taylor M, Curran-Cournane F, et al. 2023. Land use modification causes slow, but predictable, change in soil microbial community composition and functional potential. Environmental Microbiome 18(1): 30.

Mackay ADE, Dominati E,, Taylor MD 2013. Soil quality indicators: the next generation. Report prepared for Land Monitoring Forum of Regional Councils. AgResearch client report number: RE500/2012/025. AgResearch Ltd, Lincoln

Mathieu C, Hermans SM, Lear G, Buckley TR, Lee KC, Buckley HL 2020. A systematic review of sources of variability and uncertainty in eDNA data for environmental monitoring. Frontiers in Ecology and Evolution 8: 135.

Mattarozzi M, Di Zinno J, Montanini B, Manfredi M, Marengo E, Fornasier F, et al. 2020. Biostimulants applied to maize seeds modulate the enzymatic activity and metaproteome of the rhizosphere. Applied Soil Ecology 148: 103480.

Maunsell B, Adams C, O'Gara F 2006. Complex regulation of AprA metalloprotease in Pseudomonas fluorescens M114: evidence for the involvement of iron, the ECF sigma factor, PbrA and pseudobactin M114 siderophore. Microbiology 152(1): 29– 42.

Moran MA 2009. Metatranscriptomics: eavesdropping on complex microbial communities. Microbe 4(7): 7.

Oliverio AM, Bissett A, McGuire K, Saltonstall K, Turner BL, Fierer N 2020. The role of phosphorus limitation in shaping soil bacterial communities and their metabolic capabilities. mBio 11(5): 10.1128/mbio.01718-20.

Oregaard G, Sørensen SJ 2007. High diversity of bacterial mercuric reductase genes from surface and sub-surface floodplain soil (Oak Ridge, USA). The ISME Journal 1(5): 453–467.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Research 42(D1): D206–D214.

Piombo E, Abdelfattah A, Droby S, Wisniewski M, Spadaro D, Schena L 2021. Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. Microorganisms. 9(1):188.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology 35(9): 833–844.

Romdhane S, Spor A, Banerjee S, Breuil M-C, Bru D, Chabbi A, et al. 2022. Land-use intensification differentially affects bacterial, fungal and protist communities and decreases microbiome network complexity. Environmental Microbiome 17(1): 1.

Roy Chowdhury P, Golas SM, Alteio LV, Stevens JTE, Billings AF, Blanchard JL, et al. 2021. The transcriptional response of soil bacteria to long-term warming and short-term seasonal fluctuations in a terrestrial forest. Frontiers in Microbiology 12.

Salam LB, Obayori OS, Ilori MO, Amund OO 2020. Effects of cadmium perturbation on the microbial community structure and heavy metal resistome of a tropical agricultural soil. Bioresources and Bioprocessing 7: 1–19.

Saleem M, Yahya S, Razzak SA, Khawaja S, Ali A 2023. Shotgun metagenomics and computational profiling of the plastisphere microbiome: unveiling the potential of enzymatic production and plastic degradation. Archives of Microbiology 205(11): 359.

Sapsford SJ, Dickie IA 2023. Slow soil enzyme recovery following invasive tree removal through gradual changes in bacterial and fungal communities. Journal of Ecology 111(12): 2614–2626.

Schon NL, Fraser PM, Mackay AD 2023. Earthworms for inclusion as an indicator of soil biological health in New Zealand pastures. New Zealand Journal of Agricultural Research 66(3): 208–223.

Schulte RPO, Creamer RE, Donnellan T, Farrelly N, Fealy R, O'Donoghue C, et al. 2014. Functional land management: a framework for managing soil-based ecosystem services for the sustainable intensification of agriculture. Environmental Science & Policy 38: 45–58.

Segerman B 2020. The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. Frontiers in Cellular and Infection Microbiology 10: 527102.

Sim JXF, Drigo B, Doolette CL, Vasileiadis S, Karpouzas DG, Lombi E 2022. Impact of twenty pesticides on soil carbon microbial functions and community composition. Chemosphere 307: 135820.

Sparling GP 1997. Soil microbial biomass, activity and nutrient cycling as indicators of soil health. In: Pankhurst CE, Doube BM, Gupta V eds. Biological indicators of soil health, CAB International. Pp. 97–119.

Sparling GP, Schipper LA 1997. Trialling soil quality indicators for the State of the Environment Monitoring Annual Research Report for 1996/1997.

Sparling GP, Schipper LA, Macleod M 1998. Trialling soil quality indicators for the State of the Environment Monitoring Annual Research Report for 1997/1998.

Sparling GP, Rijkse W, Wilde H 2000. Implementing soil quality indicators for land: research report for 1998/1999. Contract Report 9900/108.

Sparling GP, Rijkse W, Wilde RH 2001a. Implementing soil quality indicators for land: research report for 2000/2001 and final report for MfE Project Number 5089. Contract Report LC0102/015.

Sparling GP, Rijkse W, Wilde RH 2001b. Implementing soil quality indicators for land: research report for 1999/2000. Contract Report 0001/059.

Sprunger CD, Martin TK 2023. Chapter Three: an integrated approach to assessing soil biological health. In: Sparks DL ed. Advances in agronomy. Academic Press. Pp. 131– 168.

Stevenson B 2022. Soil health indicators. Manaaki Whenua – Landcare Research Contract Report LC4166. Envirolink Report.

Su J-Q, Wei B, Ou-Yang W-Y, Huang F-Y, Zhao Y, Xu H-J, et al. 2015. Antibiotic resistome and its association with bacterial communities during sewage sludge composting. Environmental Science & Technology 49(12): 7356–7363.

Szóstak N, Szymanek A, Havránek J, Tomela K, Rakoczy M, Samelak-Czajka A, et al. 2022. The standardisation of the approach to metagenomic human gut analysis: from sample collection to microbiome profiling. Scientific Reports 12(1): 8470.

Tao Y, Xun F, Zhao C, Mao Z, Li B, Xing P, et al. 2023. Improved assembly of metagenome-assembled genomes and viruses in Tibetan saline lake sediment by HiFi metagenomic sequencing. Microbiology Spectrum 11(1): e03328-22.

Taylor 2024. Soil health and soil quality monitoring in NZ: a regional council perspective. Unpublished

Taylor M, Kim N, Hill R 2011. A trace element analysis of soil quality samples from the Waikato region. 24th Annual FLRC Workshop: Adding to the knowledge base for the nutrient manager. Pp. 8–10.

Thiele-Bruhn S, Schloter M, Wilke B-M, Beaudette LA, Martin-Laurent F, Cheviron N, et al. 2020. Identification of new microbial functional standards for soil quality assessment. Soil 6(1): 17–34.

Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551(7681): 457–463.

UNEP, FAO Soil biodiversity: contributions and threats. https://www.decadeonrestoration.org/stories/soil-biodiversity-contributions-and- threats, accessed on 29.7.2024

Van der Jeugt F, Dawyndt P, Mesuere B 2022. FragGeneScanRs: faster gene prediction for short reads. BMC Bioinformatics 23(1): 198.

Van Goethem MW, Osborn AR, Bowen BP, Andeer PF, Swenson TL, Clum A, et al. 2021. Long-read metagenomics of soil communities reveals phylum-specific secondary metabolite dynamics. Communications Biology 4(1): 1302.

Villegas-Torres MF, Bedoya-Reina OC, Salazar C, Vives-Florez MJ, Dussan J 2011. Horizontal arsC gene transfer among microorganisms isolated from arsenic polluted soil. International Biodeterioration & Biodegradation 65(1): 147–152.

Wang W, Shen P, Lu Z, Mo F, Liao Y, Wen X 2023. Metagenomics reveals the abundance and accumulation trend of antibiotic resistance gene profile under long-term no tillage in a rainfed agroecosystem. Frontiers in Microbiology 14: 1238708.

Wu J, Buckley H, Curry L, Stevenson B, Schipper LA, Lear G 2021. Livestock exclusion reduces the spillover effects of pastoral agriculture on soil bacterial communities in adjacent forest fragments. Environmental Microbiology 23(6): 2919–2936.

Wu X, Rensing C, Han D, Xiao K-Q, Dai Y, Tang Z, et al. F 2022. Genome-resolved metagenomics reveals distinct phosphorus acquisition strategies between soil microbiomes. mSystems 7(1): e01107-21.

Xie W-Y, Yuan S-T, Xu M-G, Yang X-P, Shen Q-R, Zhang W-W, et al. 2018. Long-term effects of manure and chemical fertilizers on soil antibiotic resistome. Soil Biology and Biochemistry 122: 111–119.

Yeates GW 2003. Nematodes as soil indicators: functional and biodiversity aspects. Biology and Fertility of Soils 37: 199–210.

Zhang Z, Zhang L, Zhang L, Chu H, Zhou J, Ju F 2024. Diversity and distribution of biosynthetic gene clusters in agricultural soil microbiomes. mSystems 9(4): e01263- 23.

Zhao C, Yue Y, Guo Q, Wu J, Song J, Wang Q, et al. 2023. Metagenomic insights into the alteration of soil N-cycling-related microbiome and functions under long-term conversion of cropland to Miscanthus. GCB Bioenergy 15(9): 1105–1118.

Zhao X, Wang J, Zhu L, Wang J 2019. Field-based evidence for enrichment of antibiotic resistance genes and mobile genetic elements in manure-amended vegetable soils. Science of the Total Environment 654: 906–913.

Zhou J, Theroux SM, Bueno de Mesquita CP, Hartman WH, Tian Y, Tringe SG 2021. Microbial drivers of methane emissions from unrestored industrial salt ponds. The ISME Journal 16(1): 284–295.

Zhou Y, Liu M, Yang J 2022. Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: methods, applications, challenges, and opportunities. Microbiological Research 260: 127023.

Zwetsloot MJ, Bongiorno G, Barel JM, di Lonardo DP, Creamer RE 2022. A flexible selection tool for the inclusion of soil biology methods in the assessment of soil multifunctionality. Soil Biology and Biochemistry 166: 108514.