SOIL ECOLOGY LETTERS

SEL

REVIEW
Volume 6 / Issue 3 / 230217 / 2024
https://doi.org/10.1007/s42832-023-0217-z

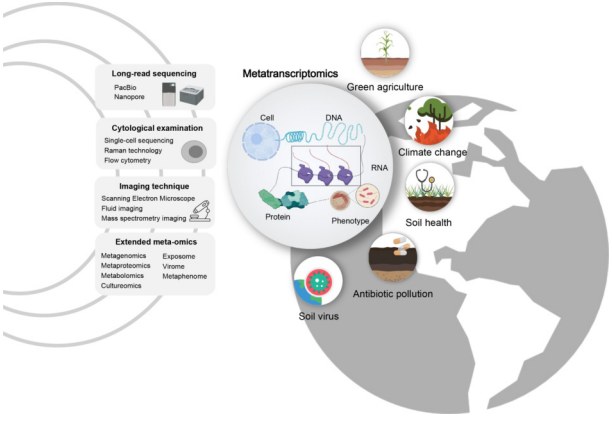
Soil microbial ecology through the lens of metatranscriptomics

Jingjing Peng[1,*,†], Xi Zhou[1,†], Christopher Rensing[2], Werner Liesack[3], Yong-Guan Zhu[4,5,*]

1 State Key Laboratory of Nutrient Use and Management, College of Resources and Environmental Sciences, National Academy of Agriculture Green Development, Key Laboratory of Plant-Soil Interactions, Ministry of Education, China Agricultural University, Beijing 100193, China
2 Institute of Environmental Microbiology, College of Resource and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, China
3 Max Planck Institute for Terrestrial Microbiology, Marburg 35043, Germany
4 State Key Laboratory of Urban and Regional Ecology, Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
5 Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
* Corresponding authors. E-mail: jingjing.peng@cau.edu.cn (J.J. Peng); ygzhu@rcees.ac.cn (Y.-G. Zhu)
† These authors contributed equally.
Received November 7, 2023; Accepted November 9, 2023

ABSTRACT

• Metatranscriptomics uncovers the dynamic expression of functional genes in soil environments, providing insights into the intricate metabolic activities within microbial communities.
• mRNA enrichment from soil samples remains a formidable challenge due to the presence of inhibitory compounds, low RNA yields, and sample heterogeneity.
• Soil metatranscriptomics unravels the expression levels of genes involved in the real-time molecular dialogues between plants and rhizobionts, uncovering the dynamics of nutrient exchange, symbiotic interactions, and plant-microbe communication.
• Metatranscriptomics unlocks the active expression of the soil resistome, elucidating the mechanisms of resistance dissemination under anthropogenic activities.
• Metatranscriptomics provides comprehensive data regarding the identification, quantification, and evolutionary history of RNA viruses.



Metatranscriptomics is a cutting-edge technology for exploring the gene expression by, and functional activities of, the microbial community across diverse ecosystems at a given time, thereby shedding light on their metabolic responses to the prevailing environmental conditions. The double-RNA approach involves the simultaneous analysis of rRNA and mRNA, also termed structural and functional metatranscriptomics. By contrast, mRNA-centered metatranscriptomics is fully focused on elucidating community-wide gene expression profiles, but requires either deep sequencing or effective rRNA depletion. In this review, we critically assess the challenges associated with various experimental and bioinformatic strategies that can be applied in soil microbial ecology through the lens of functional metatranscriptomics. In particular, we demonstrate how recent methodological advancements in soil metatranscriptomics catalyze the development and expansion of emerging research fields, such as rhizobiomes, antibiotic resistomes, methanomes, and viromes. Our review provides a framework that will help to design advanced metatranscriptomic research in elucidating the functional roles and activities of microbiomes in soil ecosystems.

Keywords metatranscriptomics, mRNA, MAGs, rhizobiont, resistome, virome

1 Metatranscriptomics in soil microbial ecology

Soil microbial communities are incredibly diverse and constantly evolving, which is crucial in regulating biogeochemical cycling in terrestrial ecosystems (Prosser, 2015). The recent advancements in molecular ecology, including **metagenomics** and **metatranscriptomics** (see Glossary), have substantially improved our understanding of soil microbiomes (Neri et al., 2022; Zhou et al., 2022). Metagenomics assesses the genetic potential of a microbial community. Still, it may be biased by necromass, while metatranscriptomics identifies the viable microbes and the functional genes expressed during sampling. A growing trend is to apply these meta-omics approaches simultaneously, linking genetic potential with the actual gene expression of a microbial community. Metatranscriptomics, in particular, has

proven to be a valuable tool in revealing the active metabolic processes, including environmental sensing and phenotype plasticity, as well as the microbial responses to nutrient supply and environmental stimuli that trigger physiological processes (Xu et al., 2021; Law et al., 2022). With most microbes exhibiting transcriptional gene control, metatranscriptomics enables us to elucidate the rapid adaptation of microbial communities to environmental change. In conclusion, metatranscriptomics is a highly effective method for uncovering hidden environmental information in soil microbial ecology, particularly the complex structure-function changes in microbiomes, expanding the capabilities of culture-based strategies or other omics approaches.

Over the past two decades, there has been a substantial accumulation of research in soil metatranscriptomics, as evidenced by the body of literature published (Fig. 1). The first study of metatranscriptomics was focused on the transcriptional activity of tidal marsh communities, using Sanger sequencing of enriched mRNA (mRNA) (Poretsky et al., 2005). Later research has demonstrated the potential of metatranscriptomics in exploring small RNA (sRNA) (Box 1, see below) molecules in natural environments (Shi et al., 2009). Subsequently, metatranscriptomic sequencing evolved toward utilizing pyrosequencing and second/third generation high-throughput technologies such as the Illumina platform. The advent of quantitative metatranscriptomics was marked by the pioneering studies conducted by Gifford et al. (2011) and Moran et al. (2013), which involved the insertion of standardized control molecules before sequencing. These studies opened the door for accurate quantification using metatranscriptomics.

Metatranscriptomics has emerged as a highly desirable method for investigating diverse soil ecosystems, leading to the identification of active and novel microbial pathways. For instance, environmental genomics and transcriptomics revealed a previously unknown fungal pathway for xylose degradation in permafrost (Woodcroft et al., 2018), while another study uncovered a new methanogenic pathway involved in hydrocarbon degradation in an oilfield (Zhou et al., 2022). Metatranscriptomics plays a crucial role in ecological assessments by disentangling soil RNA profiles encompassing coding and **non-coding RNA** (Blazewicz et al., 2013; Yates et al., 2021). Moreover, it can address various health issues, which align with recent technological advancements and ongoing research efforts. More specifically, the application of metatranscriptomics would significantly enhance our understanding of the functional complexity of soil microbiota.

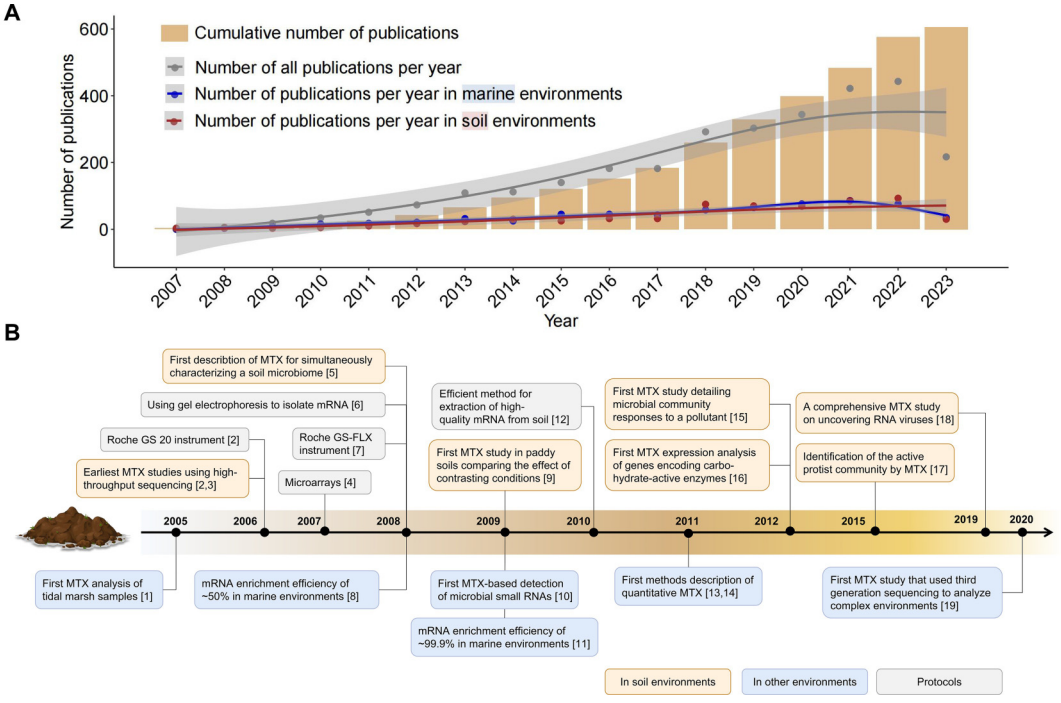


Fig. 1 A global perspective on the development and advancements in metatranscriptomic research. (A) The total number of publications indexed in the PubMed database up to July 2023 for metatranscriptomic research (black) and further divided into those with specific focus on marine systems (blue), and soils (red). The bar chart indicates the cumulative number of publications in the field of soil metatranscriptomics up to the respective year. (B) Timeline of methodological and scientific progress in metatranscriptomic research. Data relate to the following publications: [1], (Poretsky et al., 2005); [2], (Leininger et al., 2006); [3], (Grant et al., 2006); [4], (Parro et al., 2007); [5], (Urich et al., 2008); [6], (McGrath et al., 2008); [7], (Gilbert et al., 2008); [8], (Frias-Lopez et al., 2008); [9], (Shrestha et al., 2009); [10], (Shi et al., 2009); [11], (Gilbert et al., 2009); [12], (Mettel et al., 2010); [13], (Gifford et al., 2011); [14], (Moran et al., 2013); [15], (de Menezes et al., 2012); [16], (Damon et al., 2012); [17], (Geisen et al., 2015); [18], (Starr et al., 2019); [19], (Semmouri et al., 2020). Abbreviation: MTX represents metatranscriptomics.

2 Application of metatranscriptomics illuminating soil ecology

After two decades of advancements, metatranscriptomics has become a comparatively mature methodology. However, its application range still needs to be broadened to major soil ecology research areas such as the Rhizobiome, the Methanome, the Resistome, and the Virome. Metatranscriptomic research on these microbial components has been largely neglected but is nonetheless represented by an accessible RNA fraction that contains structural (taxonomic) and functional information of tremendous ecological importance. Hence, metatranscriptomics has the potential to significantly enhance our mechanistic understanding in these research areas.

2.1 Rhizobiome

Rhizobiome refers to the microbial community in the rhizosphere. In contrast, the term "holobiont" encompasses the entire host organism beneath and its associated microbial community, recognizing their integrated nature and functional interplay, which is beyond the scope of our current discussion (Berg et al., 2020). Metatranscriptomics represents an essential platform for the comprehensive exploration of functional dynamics and microbial interactions within highly heterogeneous and complex soils, providing valuable insights into the transcriptional patterns characterizing plant-microbe interactions. For example, metatranscriptomics discloses the activity of both beneficial and plant-pathogenic microorganisms in soils (Turner et al., 2013). Several discoveries have been made, such as a distinct transcriptional response of the interactome between willow roots and the rhizobiome to soil contamination (Yergeau et al., 2018), changes in the iron metabolism of the root-associated sorghum rhizobiome in response to drought (Xu et al., 2021), the pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome of sugar beet (Middleton et al., 2021), the discovery of auxin-synthesis expression in rhizosheath, as well as the prevalent replacement in host-coordination of ectomycorrhizal specialists with metabolically more versatile species in the boreal forest (Law et al., 2022). The rhizobiome holds promise as an indicator of disease susceptibility and a predictor of resistance to biotic and abiotic stressors, as revealed by metatranscriptomics (Gu et al., 2022). Metatranscriptomics also allowed for analyzing climate change effects on the rhizosphere microbiomes in European grassland (Bei et al., 2019; Bei et al., 2023). It unravels active plant-microbe communication networks, including niche differentiation and assembly process, shedding light on the signaling molecules, small RNAs, and other molecular cues that underlie these interactions (Nuccio et al., 2020). In particular, specific sRNAs, such as miRNAs, have been proposed to establish a cross-kingdom, bi-directional communication channel between the plant and its associated rhizospheric microbiome (Carrión et al., 2019), thereby regulating microbial composition and activity in the rhizosphere. By exploring the underlying mechanisms on species and strain level, metatranscriptomics allows to assess the efficiency of plant growth-promoting rhizobacteria as biofertilizers and enhancers of nutrient uptake for plants. Therefore, decoding rhizobiomes through metatranscriptomics may provide a roadmap to targeted interventions, enhancing crop productivity and sustainability and unveiling the underlying molecular interactions that shape plant–soil dynamics, guiding the development of sustainable agricultural practices.

2.2 Methanome

Metatranscriptomics plays a crucial role in unraveling the composition and activity of the **methanome** (Lackner et al., 2022). Significant methanome discoveries involved the identification of nontraditional *mcrA*-containing archaea (*Archaeoglobales* and *Candidatus Nezhaarchaeota*) as the key methane producers in geothermal springs (Wang et al., 2023), non-syntrophic methanogenic hydrocarbon degradation in oilfields (Zhou et al., 2022), and the predominance of a canonical acetoclastic pathway for methanogenesis in the oxic soil (Angle et al., 2017). Metatranscriptomic research also provided a system-level understanding of the temperature effect on the methanogenic food web in rice field soil (Peng et al., 2018) and revealed that conventional methanotrophs are responsible for the high-affinity oxidation of the tiny amount of atmospheric methane oxidation (1.9 ppm) in paddy soils (Cai et al., 2016). Another example is the co-occurrence of aerobic methane production (bacterio)chlorophyll metabolism, photosynthesis, and proteobacterial degradation of methylphosphonate in aquatic ecosystems (Perez-Coronel and Michael Beman, 2022). Moreover, metatranscriptomics has shed light on the characteristics of ANaerobic MEthanotrophs (ANME) archaea, revealing insights into the pleomorphic lifestyle of *Candidatus Methanoperedens nitroreducens* and its prophages (McIlroy et al., 2023), as well as the active transcription of large plasmids associated with ANME archaea, like *Candidatus Methanoperedens*, in anoxic environments (Schoelmerich et al., 2022). These notable findings pave the way for a deeper understanding of methane dynamics in soil ecosystems, facilitating future research and potential applications in mitigating greenhouse gas emissions.

2.3 Resistome

An emerging issue in soil metatranscriptomics research is

elucidating the environmental impact on **antimicrobial resistance (AMR)** transmission. Metatranscriptomics allows for examining environmental AMR spread by detecting the transcriptional activity of antibiotic resistance genes (ARGs), collectively termed the **resistome**, within a microbial community and revealing variations in the movement and expression of ARGs under specific environmental conditions. Indeed, ARGs act as a bridge between diverse ecosystems, enabling the exchange and dissemination of resistance determinants between soil, animal, and human wastes (Wang et al., 2021; Despotovic et al., 2023), which underlies the urgent need for exploring ARG issues in soil. For instance, Lawther et al. (2022) combined metatranscriptomics and metagenomics to identify the active ARG reservoirs in four microbiomes, highlighting the importance of AMR surveillance in soil and livestock microbiomes, in addition to human AMR surveillance. An *in-silico* analysis of the resistome in 453 ruminal microbial genomes revealed a high abundance of genes encoding tetracycline resistance (Sabino et al., 2019). Concurrently, metatranscriptomics confirmed the expression of the most abundant antibiotic-resistance genes, shedding light on the key role of mobile genetic elements in shaping the resistome of the rumen microbiome. Moreover, the overuse of antibiotics has been identified as the leading cause of ARG spread in wastewater treatment plants (WWTPs). Previous research has successfully quantified resistome genes, including ARGs, biocide resistance genes (BRGs), and metal resistance genes (MRG), in communal WWTPs, suggesting the importance of considering these factors when assessing the biological significance of the resistome in WWTPs (Ju et al., 2019). Notably, a genome-centric quantitative metatranscriptomic approach has been employed to achieve high-resolution qualitative and quantitative analyses of bacterial hosts of ARGs in wastewater treatment plants (Yuan et al., 2021). However, compromising the effectiveness of our targeted interventions, no or limited research has been conducted on the expression activity of ARGs in soil ecosystems, even though the emergence of AMRs in soils poses a grave threat to human, animal, and environmental health. By drawing upon the expertise gained in these analogous settings like WWTPs and rumen microbiomes, we can adapt and expand our understanding of soil ecosystems, shedding light on the prevalence and expression of ARGs in this important environmental context. By comprehensively analyzing the soil resistome, we would take a crucial step toward safeguarding global health, preserving the efficacy of antimicrobial treatments, and ensuring a sustainable future for humanity.

2.4 Virome

Soil ecosystems can function as environmental reservoirs for potentially virulent viruses (Box 2, see below) (Callanan et al., 2020), although RNA viruses have often escaped detection. Metatranscriptomics has emerged as a powerful method for discovering the active viral communities in soil at the RNA level (Shi et al., 2018). This approach has helped to link the abundant soil **virome** to their natural habitats and hosts (Chen et al., 2022) and to identify their activity (Emerson et al., 2018). A recent breakthrough in metatranscriptomics has been the ability to independently detect RNA viral sequences, utilizing a computational downstream analysis after direct sequencing of total sample RNA and classifying viruses based on a universally conserved **RNA-dependent RNA polymerase (RdRP)** gene possessed by most RNA viruses (Shi et al., 2016). This approach has dramatically increased the recovery and diversity of RNA viruses (Chen et al., 2022; Muscatt et al., 2022; Neri et al., 2022; Xia et al., 2023), including RNA **phages** (Neri et al., 2022). It has expanded our understanding of the role of RNA viromes in carbon cycling (Starr et al., 2019) and the impact of soilborne RNA viruses on their host organisms (Hillary et al., 2022). Although recovering virus particles from soil can be challenging, protocols have been developed to enrich RNA viruses in soil (Hillary et al., 2022). However, lessons from our lack of knowledge of soil viral ecology demand more detailed research and caution in interpreting and reporting uncultivated virus genomes, particularly in research not explicitly devoted to investigating viruses (Roux et al., 2019). Hence, increased awareness of bias risks for discovering RNA viruses will lay the basis to unambiguously uncover the true diversity of the 'villains' (Jansson and Wu, 2022). In the future, intensified virome research may also lead to the development of innovative biotechnology tools and bio-control agents. This may involve the increased use of phage therapies (Callanan et al., 2020) and discoveries related to CRISPR–Cas systems (Burstein et al., 2017; Esser et al., 2023).

3 Key procedural steps in metatranscriptomic research

Metatranscriptomics involves the extraction of total RNA, which may optionally be enriched for mRNA. This is followed by cDNA library preparation and deep sequencing (Fig. 2). These subsequent procedural steps finally generate large RNA data sets for taxonomic and functional analysis, quantification of transcripts, and genome mapping in biological samples. Extracting total RNA of high quality and yield is a prerequisite for success in metatranscriptomic research. Indeed, the extraction of total RNA from the soil is the most challenging experimental step in metatranscriptomic research. Given the low proportion of mRNA in total RNA,

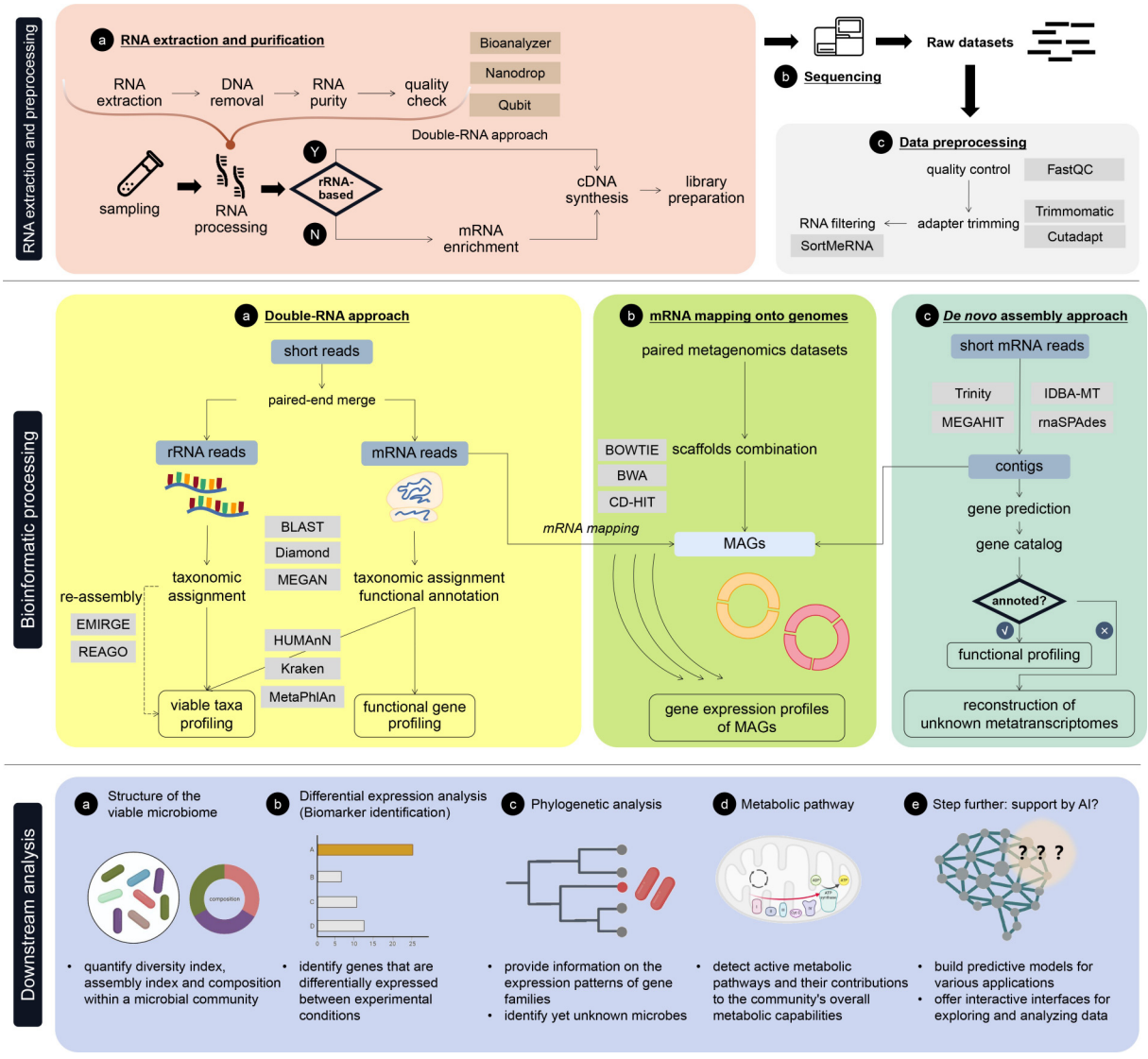


Fig. 2 Three-step workflow in metatranscriptomic research. (1) RNA extraction and preprocessing of raw RNA-seq data. (a) Extraction of total RNA and its quality check followed by either the double-RNA approach (involves immediate cDNA library preparation and sequencing for simultaneous analysis of both rRNA and mRNA) or mRNA enrichment by subtractive removal of the rRNA. The enriched mRNA will then be used for cDNA library preparation and subsequent sequencing. (b and c) Preprocessing of the raw reads, involving quality control using Trimmomatic, FastQC, and Cutadapt. The rRNA and mRNA reads will be distinguished by SortMeRNA or Blast search against the NCBI nt/nr database. (2) Bioinformatic processing of the quality-checked RNA-seq reads using "non-assembly-based" (a) or "assembly-based" (b and c) approaches, which will be chosen depending on the experimental objectives. (3) Further downstream analysis, including the analysis of diversity and composition of the viable microbiome (a), differential gene expression analysis (b), phylogenetic analysis (c), detection of the active metabolic pathways (d), as well as building models and interactive interfaces by AI support (e).

the quantitative removal of rRNA (rRNA) can significantly enhance the metatranscriptomic significance and the resolution of functional gene expression patterns. However, the short half-life of mRNA, the presence of high amounts of inhibitory compounds (e.g., humic acids and fulvic acids) co-extracted with the RNA, and the potential for methodological pitfalls during rRNA removal can significantly affect the outcome of metatranscriptomic research. Therefore, it may be advisable to avoid rRNA removal and directly sequence the total RNA, referred to as the "Double-RNA approach."

3.1 Extraction of total RNA

Compared to DNA, RNA yields from soil samples are low. In addition, RNA, in particular mRNA, is more prone to degradation into smaller fragments than DNA. Therefore, special attention must be paid to sample preservation before RNA extraction. When freezing preservation is unavailable in field sampling, kit-based solutions such as RNAlater® (Invitrogen) can help maintain the environmental transcript pool (Liao et al., 2023). Several commercial kits are available based on these principles, with the most commonly used being

RNeasy PowerSoil Total RNA Kit (Qiagen). Despite this, many scientists still favor manual methods, which are more economical and flexible for samples from different soils and specific quality criteria (Mettel et al., 2010; Levy-Booth et al., 2021).

3.2 Double-RNA approach

The double-RNA approach, which does not require mRNA enrichment, simultaneously enables taxonomic and functional profiling (Peng et al., 2018; Hempel et al., 2022). Using the two most common marker molecules (rRNA and mRNA), the simultaneous analysis of community-wide microbial activity (structure) and gene expression (function) was first described and validated by Urich et al. (2008). While the importance of rRNA as the standard phylogenetic biomarker needs to be re-emphasized, its RNA-seq reads represent the overwhelming majority (84.39%−90.74%) in metatranscriptomic data sets (Stewart et al., 2010). Their taxon-specific abundances generally reflect the overall structural activity of the viable microbial community (Blazewicz et al., 2013). More specifically, the double-RNA strategy involves a two-step analysis process to identify tags derived from rRNA and mRNA. Using rRNA allows for assessing taxonomic diversity across all three domains of life. Indeed, rRNA-centered metatranscriptomics is more accurate in taxonomic identifications than metagenomics, even at sequencing depths that are one order of magnitude lower (Hempel et al., 2022). Metatranscriptomic research collectively confirmed that mRNA is more responsive to environmental change than rRNA. The strength of the double-RNA approach lies in the ability to simultaneously analyze both the taxonomic composition of the viable community and the gene expression activity of particular functional guilds.

3.3 mRNA-centered metatranscriptomics

mRNA, accounting for less than 5% of total RNA, is recognized as a taxonomic and functional biomarker for microbial activity (Peng et al., 2018; Täumer et al., 2022). Various methods have been developed to enrich mRNA for specific downstream analysis. The most widely applied procedure involves the depletion of rRNA through subtractive hybridization between rRNA and oligonucleotides that target conserved regions of bacterial rRNA. After subtractive hybridization, the residual rRNA in the rRNA-depleted sequencing libraries may range from 0.5% to 20% (He et al., 2010). Exonuclease digestion is an alternative strategy, but its efficacy may be affected by residual humic acid impurities in soil samples (He et al., 2010; Mettel et al., 2010). Nonetheless, it is recommended to combine these two approaches if the research objective is to analyze unprocessed mRNA that is triphophorylated (Mettel et al., 2010). The duplex nuclease specific (DSN) normalization, a non-targeted rRNA removal approach with high removal efficiency, relies on low RNA input (Chappell, 2012). Some other methods, such as electrophoretic size selection, RNase H digestion, and sequence-specific blockage of reverse transcription, have been suggested but have limitations (Huang et al., 2020). Notably, some commercial kits, such as MICROBExpress™ Bacterial mRNA Enrichment Kit (Invitrogen), RiboMinus™ Transcriptome Isolation Kit (Bacteria) (Invitrogen), and RiboZero™ Magnetic Kit (Bacteria) (Epicenter), are available for rRNA removal, with the RiboZero shown to be the most efficient without introducing major bias (Mettel et al., 2010). Furthermore, it has been verified that riboPOOLs (siTOOLs Biotech) can be as efficient in rRNA depletion as the former RiboZero (Wahl et al., 2022). Other custom-made approaches can be applied, such as repeated rounds of subtractive hybridization (He et al., 2010; Stewart et al., 2010) or a combination of different methods, including customized sample-specific probes (Culviner et al., 2020). The analysis focus on enriched mRNA may be particularly advisable for assessing the transcriptional activity of genes or metabolic pathways that exhibit low expression levels. If the expression analysis of eukaryotic microorganisms is specifically aimed, the enrichment of their mRNA can be achieved by exploiting the poly-A tail, which selectively targets and captures the coding regions of eukaryotic genes rather than prokaryotic genes (Passmore and Coller, 2022).

4 Challenges in bioinformatic analysis

We refer to Table 1 for additional information on the software tools mentioned in the following text. Paired-end or single-end Illumina are the most commonly used platforms to generate up to billions of reads in RNA-seq studies. To eliminate interference during subsequent annotation and assembly, trimmed reads are sorted by comparing against the reference databases or using software, such as SortMeRNA (Kopylova et al., 2012) and BBDuk within the BBTools suite (e.g., the website doe.gov). The downstream analysis strategy of metatranscriptomics is usually divided into two categories: reference-based versus assembly-based. However, these strategies may differ between the analysis of rRNA and mRNA reads (Fig. 2): 1) Paired-end merging approach: this approach involves the direct BLAST of unassembled merged reads, in which the merged reads act as query sequences in a BLAST search against a comprehensive reference database (Söllinger et al., 2022); 2) *De novo* assembly approach: this approach focuses on the assembly of the reads into contigs, which are subsequently subjected to a BLAST search against a reference database (Bei et al.,

Table 1 Software tools for downstream analysis of soil metatranscriptomic data sets[a].

Software	Function	Prevalence (%)[b]	Reference
Preprocessing			
Trimmomatic	Trimming and filtering raw sequencing reads.	34	Bolger et al., 2014
FastQC	Providing modular analysis for quality control of raw sequencing reads.	34	-
Cutadapt	Removal of adapter and other unwanted sequences, as well as quality trimming of raw sequencing reads.	37	Martin, 2011
SortMeRNA	Fast and accurate mapping and filtering of rRNA sequences from metatranscriptomic data.	29	Kopylova et al., 2012
De novo assembly			
Trinity	*De novo* assembly transcripts to good-quality contigs. It excels at capturing isoforms and alternative splicing events.	24	Grabherr et al., 2011
MEGAHIT	Assembly of metagenomic and metatranscriptomic reads to contigs with low computational resources.	46	Li et al., 2015
rnaSPAdes	*De novo* assembler for RNA-Seq data with particular assembly accuracy in differentiating between isoforms.	55	Bushmanova et al., 2019
IDBA-MT	Extensions of IDBA for metatranscriptomic assembly, aiming to handle microbial communities with multiple strains or species.	18	Leung et al., 2013
Double-RNA approach			
BLAST	Compares nucleotide or protein sequences to reference databases and calculates the statistical significance.	20	Altschul et al., 1990
Diamond	Accelerated BLAST compatible local sequence aligner.	35	Buchfink et al., 2015
Re-assembly			
EMIRGE	Reconstruction of full-length ribosomal genes from short rRNA reads.	11	Miller et al., 2011
REAGO	An assembly tool for rRNA recovery from metagenomic and metatranscriptomic data.	11	Yuan et al., 2015
mRNA mapping onto genomes			
Bowtie/Bowtie2	Ultrafast and memory-efficient aligner of sequencing reads to reference genomes.	31	Langmead and Salzberg, 2012
BWA	Burrows-Wheeler Aligner for accurate alignment of sequencing reads to reference genomes.	27	Li and Durbin, 2009
CD-HIT	A fast program for clustering and comparing large data sets of protein or nucleotide sequences	30	Fu et al., 2012
Annotation			
MEGAN	Taxonomic assignment, functional annotation, and visualization of Blast output data.	18	Huson et al., 2007
HUMAnN	Taxonomic assignment and functional annotation of quality-controlled DNA/RNA reads.	26	Franzosa et al., 2018
Kraken/Kraken2	Fast taxonomic classification of quality-controlled DNA/RNA reads using exact *k-mer* matches.	40	Wood et al., 2019
MetaPhlAn	Taxonomic assignment of quality-controlled DNA/RNA reads.	29	Blanco-Míguez et al., 2023

[a]The incorporation of the software tools into the computational downstream analysis of the RNA-seq reads is shown in Fig. 2. [b]Prevalence (%) refers to the frequency with which the software tool has been mentioned in papers published over the past two years relative to the total number of publications that refer to this software tool. The search was conducted using Google Scholar and the following search terms: "[Software]" AND metatranscriptomics AND (soil OR wetland OR marsh OR bog OR swamp OR mangrove OR forest OR permafrost OR tundra OR grassland OR rhizosphere).

2023) or mapped to either known reference genomes or paired MAGs, with the latter also termed environmental transcriptome mapping (Levy-Booth et al., 2021).

4.1 Paired-end merging approach

The paired-end reads can be merged into fragments if the target fragment size is smaller than twice the length of the single-end reads, resulting in an overlap. In contrast, if the target fragment size exceeds the sum of the lengths of the individual reads, these reads may be too short to generate enough high-quality bases, highlighting the need to evaluate size selection during library preparation and sequencing depth. To improve merging efficiency, various paired-read merging tools, such as fast-join, PANDAseq, PEAR, and NGmerge, have been developed and successfully applied in the metatranscriptomic analysis (Ivanova et al., 2016; Täumer et al., 2022). Taxonomic assignments of rRNA and

mRNA can be classified against SILVA SSU and NCBI nr/nt databases (Quast et al., 2012). Some undesired sequences that do not provide functional information, such as noncoding regions, should be discarded in advance. Functional annotation of mRNA in the environmental samples can be achieved by mapping them to databases such as SEED subsystems, KEGG categories, COG database (Hayden et al., 2018), or specialized databases such as Carbohydrate-Active EnZymes (CAZyme) (Huang et al., 2018), Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2020), Virulence Factor Database (VFDB) (Liu et al., 2019), and IMG/VR (Camargo et al., 2023).

4.2 *De novo* assembly approach

mRNA reads can be analyzed using two distinct approaches, as demonstrated by Ivanova et al. (2016). In contrast to the conventional method of mapping mRNA-tags to a reference database, the *de novo* assembly approach offers the advantage of identifying novel expressed genes or noncoding RNAs and providing more reliable annotations, albeit at the expense of sensitivity, particularly for low-abundance transcripts. To ensure accurate research in complex environments, it is recommended to include unassembled reads in the final annotation of metatranscriptomics data (Toseland et al., 2014). However, a universally proven solution for optimizing run time, memory requirements, and overall accuracy in choosing an appropriate assembler for metatranscriptomic data sets is currently lacking. Previous studies highlighted Trinity as the optimal assembler for metatranscriptomics data (Celaj et al., 2014). However, the evaluation of eukaryotic assemblies indicated that MEGAHIT excels in a long-read generation, while rnaSPAdes exhibits superior mapping and annotation rates (Krinos et al., 2023). Acknowledging the quest for improved assembly efficiency, it is widely recognized that combining multiple assemblers yields better results (Ortiz et al., 2021; Krinos et al., 2023). A similar approach has been applied to reconstruct long SSU rRNA reads, enabling the recovery of 16S rRNA and 18S rRNA genes from metatranscriptomes, using tools such as EMIRGE, MetaRib, and RiboTaxa (Chakoory et al., 2022). This rRNA-based method has been applied to natural microbial communities and successfully yielded full-length sequences that can be classified to the species level (Nuccio et al., 2021). For instance, the phylogenetic distinctiveness of temperature-responding species-level populations within the hydrogenotrophic methanogen genus *Methanocella* could be differentiated in metatranscriptomic research on the methanogen dynamics in flooded rice field soil using EMIRGE for analysis of the archaeal 16S rRNA reads (Peng et al., 2018). Alternatively, long-read sequencing, which symbolizes the third generation sequencing, has emerged as a promising approach that circumvents the need for poor assemblies and captures complete or near full-length segments directly (Ciuffreda et al., 2021). The quality of long sequencing reads can be enhanced through error correction and assembly, resulting in more comprehensive transcripts that can address truncated reads (Ojala et al., 2023). It is expected to benefit significantly from short-read metatranscriptomics (Shakya et al., 2019) and facilitate discrimination among different isoforms, providing an accurate search, especially for low-abundance species (Van Goethem et al., 2021).

4.3 Environmental transcriptome mapping

Single metatranscriptomics makes it challenging to distinguish gene-loss events from reduced expression, while paired with metagenomics integrates transcript abundances with corresponding gene copy numbers from the community. The relative abundance and activity of **metagenome-assembled genomes (MAGs)** can be determined by assembling metagenomes separately, followed by mapping metatranscriptomic mRNA or assembled contigs to the MAGs. The fundamental step in this process involves mapping trimmed mRNA reads to all assembled genomes with alignment programs such as Bowtie2 (Langmead and Salzberg, 2012) or Burrows-Wheeler alignment tool (BWA) (Li and Durbin, 2009). Recent advances in the field of genome-centric metatranscriptomics have allowed for the exploration of active microbial populations and their metabolite pathways in complex environments, including soils (Woodcroft et al., 2018), re-vegetated mine wastelands (Tan et al., 2019), oily sludges (Zhou et al., 2022), and geothermal springs (Levy-Booth et al., 2021; Yin et al., 2022). Environmental transcriptome mapping can be accomplished using either unassembled mRNA reads or *de novo* assembled contigs. *De novo* assembled contigs have the advantage that they can be directly blasted against databases to achieve robust taxonomic assignment and functional annotation of the query sequences. Such a direct approach is often impossible with unassembled mRNA sequences, primarily due to the limited read length.

5 Concluding remarks

Here, we highlight the crucial role of metatranscriptomics in building a comprehensive environmental transcriptome that connects physiological and functional processes within natural and anthropogenically affected environments. Despite experimental challenges associated with effort and cost, the numerous advantages of this approach cannot be overlooked. Metatranscriptomics has allowed researchers to better understand how the active soil microbes contribute to local and global ecosystem functioning. The integration of

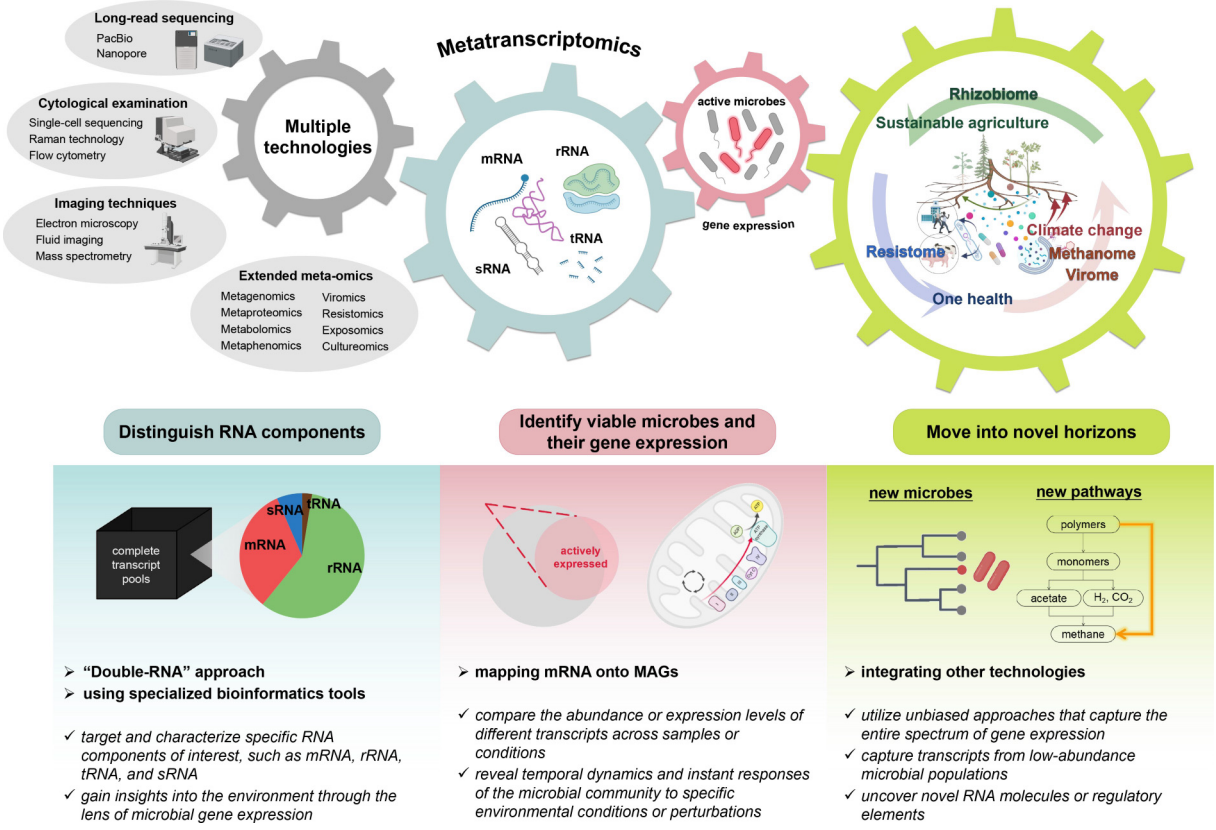


Fig. 3 Proposed framework for further metatranscriptomic research and for improving soil health to achieve sustainable agriculture. The application of conceptually well-chosen multi-methods will be crucial for system-level understanding of the functional roles and activities of microbes in terrestrial ecosystems. The sophisticated interplay between the experimental and computational approaches will help us cope with the threats of antibiotic pollutions and viruses for soil health, green agriculture and climate change.

metatranscriptomics and metagenomics, along with experimental methods like flow cytometry, fluid imaging, Raman technology, microscopy, and single-cell sequencing, will be essential for further in-depth elucidation of the actual functional activities of the soil microbiome (Fig. 3). Furthermore, sophisticated computational methods, such as multivariate statistics, niche modeling and deep learning, will also be required to facilitate predictions at local, regional, and global scales. These efforts will enable scientists to move forward from hindcasting to forecasting and to unlock how microbial diversity and activities that drive local interactions translate to broader system impacts see the Electronic Supplementary Material: Outstanding Questions. Ultimately, this review will help to better understand the crucial role of soil metatranscriptomics in enhancing our knowledge of the interactions between the soil microbiome and the terrestrial ecosystem.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (41977038 and 42277307) and the National Key Research and Development Program of China (2021YFD1900100).

Author contributions

J.J.P., X.Z., C.R., W.L., and Y.G.Z. wrote the manuscript. All authors read and approved the final manuscript.

Declaration of interests

No interests are declared.

Glossary

Antimicrobial resistance (AMR): the ability of microorganisms, such as bacteria, viruses, fungi, and parasites, to resist the effects of drugs.

Auxiliary metabolic genes (AMGs): while not essential for basic cellular functions, AMGs encode proteins involved in various host metabolic pathways. Originating from bacterial cells, they are frequently found in many bacteriophages.

Bacteriophage (phage): a virus that explicitly infects, and replicates within, bacteria or archaea.

Metagenome-assembled genomes (MAGs): genomes reconstructed from metagenomic data by binning or clustering contigs that are derived from the same organism or

closely related organisms. MAGs represent the genomic information of individual microorganisms present in a metagenomic sample.

Metagenomics: a technology that allows for direct sequencing of genetic material from environmental or clinical samples without the need to culture or isolate individual organisms.

Metatranscriptomics: a technology that allows for the sequencing analysis of total RNA or enriched mRNA extracted from a complex biological sample. It aims to understand and interpret the transcriptome of entire microbial communities within a specific environment.

Methanome: comprises the methane-producing (methanogenic) and methane-oxidizing (methanotrophic) microorganisms, which inhabit a wide range of natural ecosystems.

Non-coding RNA (ncRNA): a broad category of RNA molecules that do not code for proteins but have diverse functional roles in the cell. These molecules include small regulatory RNAs (sRNAs), rRNAs (rRNAs), and tRNAs (tRNAs).

Resistome: the collective genetic reservoir of resistance genes in a microbial community of a given environment. These genes encode proteins that enable microorganisms to withstand the effects of antibiotics and other antimicrobial agents.

Rhizobiome: the collection of microorganisms inhabiting the rhizosphere, with the latter being defined as the soil surrounding the plant roots while influenced by them. The rhizobiome encompasses bacteria, fungi, archaea, and other microorganisms that are thought to interact with plant roots. The term “rhizobiome” is derived from the word “rhizosphere” (the soil zone influenced by plant roots) and “biome” (a complex ecological community).

RNA-dependent RNA polymerase (RdRP): an enzyme responsible for the replication and transcription of RNA in certain viruses.

Virome: the collective genetic material of viruses presenting in a particular ecosystem or within a specific organism, and the term is indeed derived from “virus” and “biome” (a complex ecological community). It encompasses all the viruses, including bacteriophages, in a given environment or host.

Text boxes

Box 1 Hidden treasure uncovered: Non-coding small RNA

Recent scientific investigations have transcended the conventional realms of “three flavors” (mRNA, rRNA, and tRNA) and expanded their view to encompass the often overlooked class of non-coding RNAs, particularly the small RNAs (sRNAs) which are essential components in the global control of transcriptional regulation (Eddy, 2001). sRNAs are a crucial type of non-coding RNA molecules that typically vary in size from 50 to 500 bases. It has been a decade since the first discovery of sRNAs in the ocean through metatranscriptomics (Shi et al., 2009). sRNAs are involved in several biological processes, including translational control, RNA-RNA mediated interactions, nutrient cycling, stress response, and quorum sensing. Metatranscriptomics has enabled a more comprehensive study of sRNAs, allowing for detection of their expression in non-protein-coding regions and the community-level analysis of sRNA abundances (Lott et al., 2020). Through a comparison of cDNAs with metagenomic data retrieved form marine samples, Shi et al. (2009) found that over 28% of the unassigned cDNA reads are represented by sRNAs, indicating the importance of this class of RNA in environmental processes like carbon metabolism and nutrient acquisition. More recently, Gelsinger et al. (2020) employed metatranscriptomics to elucidate the abundance and diversity of non-coding sRNAs within an extremophilic microbial community. Their study highlighted the regulatory role of sRNAs in gene expression and response to environmental stress by demonstrating a correlation between sRNA expression levels and their corresponding target mRNAs (Gelsinger et al., 2020).

Box 2 Soil viruses under climate change

Soil viruses, abundant and diverse biological entities, are crucial in shaping soil microbiota and essential soil functions (Liao et al., 2022). As our planet experiences the effects of global warming and extreme weather events, the pivotal role of viruses in shaping soil microbiota and their functional activities has gained significant research attention. Based on their life cycle, phages can be categorized as virulent or temperate, which differ in their infective properties (Chevallereau et al., 2022). Virulent phages operate as predators, exerting control over bacterial communities by lysing cells, aligning with the “Kill-the-Winner” model (Muscatt et al., 2022). By liberating cellular matter through host lysis, they contribute to the soil organic carbon pool, thus facilitating what we refer to as the “viral shunt.” In marine ecosystems, phages lyse 20%−40% of prokaryotes daily, thus constituting a significant proportion (6%−26%) of the marine carbon exchange and exerting a considerable impact on the marine carbon cycle (Tong et al., 2023). Experimental evidence supports the existence of a “viral shuttle” in soils, where soil phages enhance the abundance of recalcitrant dissolved organic matter components, creating a delicate balance between net carbon sink and carbon source during the lysis process (Tong et al., 2023). In contrast, temperate phages integrate their genetic material into the host's chromosomes and replicate their nucleic acids alongside the host, in what is known as the lysogenic

cycle (Chen et al., 2018). Contrary to virulent phages whose dynamics are explained by the “Kill-the-Winner” model, temperate (or lysogenic) phages enable their hosts to endure harsh environments by integrating **auxiliary metabolic genes (AMGs)** into the host's genome, an intriguing concept known as the “Piggyback-the-Winner” model (Jansson and Wu, 2022). For instance, viruses have been found to encode glycoside hydrolases that can potentially influence complex carbon degradation (Emerson et al., 2018; Jansson and Wu, 2022). Furthermore, the changing climate, characterized by elevated temperatures, thawing permafrost, and altered precipitation patterns, significantly influences soil viral communities (Jansson and Wu, 2022). Increased soil temperatures, fostering microbial activity, may induce temperate phages to transition into the lytic life cycle, thereby triggering an escalating arms race between viruses and hosts and, in consequence, contributing to carbon cycling through the viral shunt (Chen et al., 2018; Jansson and Wu, 2022). These discoveries emphasize the substantial impact of viruses on soil services and functions in crucial terrestrial ecosystems influenced by climate change.

Electronic supplementary material

Supplementary material is available in the online version of this article at https://doi.org/10.1007/s42832-023-0217-z and is accessible for authorized users.

References

Alcock, B.P., Raphenya, A.R., Lau, T.T., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.L.V., Cheng, A.A., Liu, S., 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research 48, D517–D525.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Angle, J.C., Morin, T.H., Solden, L.M., Narrowe, A.B., Smith, G.J., Borton, M.A., Rey-Sanchez, C., Daly, R.A., Mirfenderesgi, G., Hoyt, D.W., Riley, W.J., Miller, C.S., Bohrer, G., Wrighton, K.C., 2017. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. Nature Communications 8, 1567.

Bei, Q., Moser, G., Wu, X., Müller, C., Liesack, W., 2019. Metatranscriptomics reveals climate change effects on the rhizosphere microbiomes in European grassland. Soil Biology & Biochemistry 138, 107604.

Bei, Q., Reitz, T., Schnabel, B., Eisenhauer, N., Schädler, M., Buscot, F., Heintz-Buschart, A., 2023. Extreme summers impact cropland and grassland soil microbiomes. ISME Journal 17, 1589–1600.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.C.C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome 8, 1–22.

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P., Dubois, L., Huang, K.D., Thomas, A. M., Nickols, W.A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., Berry, S.E., Spector, T.D., Franzosa, E.A., Pasolli, E., Asnicar, F., Huttenhower, C., Segata, N., 2023. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nature Biotechnology 41, 1–12.

Blazewicz, S.J., Barnard, R.L., Daly, R.A., Firestone, M.K., 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME Journal 7, 2061–2068.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30, 2114–2120.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12, 59–60.

Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A., Banfield, J.F., 2017. New CRISPR–Cas systems from uncultivated microbes. Nature 542, 237–241.

Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D., 2019. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. GigaScience 8, giz100.

Cai, Y., Zheng, Y., Bodelier, P.L., Conrad, R., Jia, Z., 2016. Conventional methanotrophs are responsible for atmospheric methane oxidation in paddy soils. Nature Communications 7, 11728.

Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R. P., Hill, C., 2020. Expansion of known ssRNA phage genomes: from tens to over a thousand. Science Advances 6, eaay5981.

Camargo, A.P., Nayfach, S., Chen, I.M.A., Palaniappan, K., Ratner, A., Chu, K., Ritter, S.J., Reddy, T., Mukherjee, S., Schulz, F., Call, L., Neches, R.Y., Woyke, T., Ivanova, N.N., Eloe-Fadrosh, E.A., Kyrpides, N.C., Roux, S., 2023. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. Nucleic Acids Research 51, D733–D743.

Carrión, V.J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., De Hollander, M., Ruiz-Buck, D., Mendes, L.W., van Ijcken, W.F., Gomez-Exposito, R., Elsayed, S.S., Mohanraju, P., Arifah, A., van der Oost, J., Paulson, J.N., Mendes, R., van Wezel, G.P., Medema, M.H., Raaijmakers, J.M., 2019. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science 366, 606–612.

Celaj, A., Markle, J., Danska, J., Parkinson, J., 2014. Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. Microbiome 2, 1–13.

Chakoory, O., Comtet-Marre, S., Peyret, P., 2022. RiboTaxa: combined approaches for rRNA genes taxonomic resolution down to the species level from metagenomics data revealing novelties. NAR Genomics and Bioinformatics 4, lqac070.

Chappell, L., 2012. Finding a needle in a haystack. Nature Reviews Microbiology 10, 446–446.

Chen, J., Quiles-Puchalt, N., Chiang, Y.N., Bacigalupe, R., Fillol-

Salom, A., Chee, M.S.J., Fitzgerald, J.R., Penadés, J.R., 2018. Genome hypermobility by lateral transduction. Science 362, 207–212.

Chen, Y.M., Sadiq, S., Tian, J.H., Chen, X., Lin, X.D., Shen, J.J., Chen, H., Hao, Z.Y., Wille, M., Zhou, Z.C., Wu, J., Li, F., Wang, H.W., Yang, W.D., Xu, Q.Y., Wang, W., Gao, W.H., Holmes, E. C., Zhang, Y.Z., 2022. RNA viromes from terrestrial sites across China expand environmental viral diversity. Nature Microbiology 7, 1312–1323.

Chevallereau, A., Pons, B.J., van Houte, S., Westra, E.R., 2022. Interactions between bacterial and phage communities in natural environments. Nature Reviews Microbiology 20, 49–62.

Ciuffreda, L., Rodríguez-Pérez, H., Flores, C., 2021. Nanopore sequencing and its application to the study of microbial communities. Computational and Structural Biotechnology Journal 19, 1497–1511.

Culviner, P.H., Guegler, C.K., Laub, M.T., 2020. A simple, cost-effective, and robust method for rRNA depletion in RNA-sequencing studies. mBio 11, e00010–e00020.

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., Marmeisse, R., 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoS One 7, e28967.

de Menezes, A., Clipson, N., Doyle, E., 2012. Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. Environmental Microbiology 14, 2577–2588.

Despotovic, M., de Nies, L., Busi, S.B., Wilmes, P., 2023. Reservoirs of antimicrobial resistance in the context of One Health. Current Opinion in Microbiology 73, 102291.

Eddy, S.R., 2001. Non-coding RNA genes and the modern RNA world. Nature Reviews Genetics 2, 919–929.

Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H.B., Singleton, C.M., Solden, L.M., Naas, A.E., Boyd, J.A., Hodgkins, S.B., Wilson, R.M., Trubl, G., Li, C., Frolking, S., Pope, P.B., Wrighton, K.C., Crill, P.M., Chanton, J.P., Saleska, S.R., Tyson, G.W., Rich, V.I., Sullivan, M.B., 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nature Microbiology 3, 870–880.

Esser, S.P., Rahlff, J., Zhao, W., Predl, M., Plewka, J., Sures, K., Wimmer, F., Lee, J., Adam, P.S., McGonigle, J., Turzynski, V., Banas, I., Schwank, K., Krupovic, M., Bornemann, T.L.V., Figueroa-Gonzalez, P.A., Jarett, J., Rattei, T., Amano, Y., Blaby, I.K., Cheng, J.F., Brazelton, W.J., Beisel, C.L., Woyke, T., Zhang, Y., Probst, A.J., 2023. A predicted CRISPR-mediated symbiosis between uncultivated archaea. Nature Microbiology 8, 1619–1633.

Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J. G., Segata, N., Huttenhower, C., 2018. Species-level functional profiling of metagenomes and metatranscriptomes. Nature Methods 15, 962–968.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S. C., Chisholm, S.W., DeLong, E.F., 2008. Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences of the United States of America 105, 3805–3810.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England) 28, 3150–3152.

Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M., Urich, T., 2015. Metatranscriptomic census of active protists in soils. ISME Journal 9, 2178–2190.

Gelsinger, D.R., Uritskiy, G., Reddy, R., Munn, A., Farney, K., DiRuggiero, J., 2020. Regulatory noncoding small RNAs are diverse and abundant in an extremophilic microbial community. mSystems 5, e00584–e00519.

Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., Moran, M.A., 2011. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. ISME Journal 5, 461–472.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., Joint, I., 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One 3, e3042.

Gilbert, J.A., Thomas, S., Cooley, N.A., Kulakova, A., Field, D., Booth, T., McGrath, J.W., Quinn, J.P., Joint, I., 2009. Potential for phosphonoacetate utilization by marine bacteria in temperate coastal waters. Environmental Microbiology 11, 111–125.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29, 644–652.

Grant, S., Grant, W.D., Cowan, D.A., Jones, B.E., Ma, Y., Ventosa, A., Heaphy, S., 2006. Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. Applied and Environmental Microbiology 72, 135–143.

Gu, Y., Banerjee, S., Dini-Andreote, F., Xu, Y., Shen, Q., Jousset, A., Wei, Z., 2022. Small changes in rhizosphere microbiome composition predict disease outcomes earlier than pathogen density variations. ISME Journal 16, 2448–2456.

Hayden, H.L., Savin, K.W., Wadeson, J., Gupta, V.V., Mele, P.M., 2018. Comparative metatranscriptomics of wheat rhizosphere microbiomes in disease suppressive and non-suppressive soils for *Rhizoctonia solani* AG8. Frontiers in Microbiology 9, 859.

He, S., Wurtzel, O., Singh, K., Froula, J.L., Yilmaz, S., Tringe, S.G., Wang, Z., Chen, F., Lindquist, E.A., Sorek, R., Hugenholtz, P., 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nature Methods 7, 807–812.

Hempel, C.A., Wright, N., Harvie, J., Hleap, J.S., Adamowicz, S.J., Steinke, D., 2022. Metagenomics versus total RNA sequencing: most accurate data-processing tools, microbial identification accuracy and perspectives for ecological assessments. Nucleic Acids Research 50, 9279–9293.

Hillary, L.S., Adriaenssens, E.M., Jones, D.L., McDonald, J.E., 2022. RNA-viromics reveals diverse communities of soil RNA viruses with the potential to affect grassland ecosystems across multiple trophic levels. ISME Communications 2, 1–10.

Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., Yi, H., Yang, Z., Yin, Y., 2018. dbCAN-seq: a database of carbohydrate-

active enzyme (CAZyme) sequence and annotation. Nucleic Acids Research 46, D516–D521.

Huang, Y., Sheth, R.U., Kaufman, A., Wang, H.H., 2020. Scalable and cost-effective ribonuclease-based rRNA depletion for transcriptomics. Nucleic Acids Research 48, e20.

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Research 17, 377–386.

Ivanova, A.A., Wegner, C.E., Kim, Y., Liesack, W., Dedysh, S.N., 2016. Identification of microbial populations driving biopolymer degradation in acidic peatlands by metatranscriptomic analysis. Molecular Ecology 25, 4818–4835.

Jansson, J.K., Wu, R., 2022. Soil viral diversity, ecology and climate change. Nature Reviews Microbiology 21, 296–311.

Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H. P., Johnson, D.R., Zhang, T., Bürgmann, H., 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME Journal 13, 346–360.

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics (Oxford, England) 28, 3211–3217.

Krinos, A.I., Cohen, N.R., Follows, M.J., Alexander, H., 2023. Reverse engineering environmental metatranscriptomes clarifies best practices for eukaryotic assembly. BMC Bioinformatics 24, 1–36.

Lackner, M., Drew, D., Bychkova, V., Mustakhimov, I., 2022. Value-Added Products from Natural Gas Using Fermentation Processes: Fermentation of Natural Gas as Valorization Route, Part 1. In: Ravanchi, M.T., ed. Natural Gas—New Perspectives and Future Developments. IntechOpen Limited, London. pp. 23–46.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.

Law, S.R., Serrano, A.R., Daguerre, Y., Sundh, J., Schneider, A.N., Stangl, Z.R., Castro, D., Grabherr, M., Näsholm, T., Street, N.R., Hurry, V., 2022. Metatranscriptomics captures dynamic shifts in mycorrhizal coordination in boreal forests. Proceedings of the National Academy of Sciences of the United States of America 119, e2118852119.

Lawther, K., Santos, F.G., Oyama, L.B., Rubino, F., Morrison, S., Creevey, C.J., McGrath, J.W., Huws, S.A., 2022. Resistome analysis of global livestock and soil microbiomes. Frontiers in Microbiology 13, 897905.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J.I., Schuster, S., Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442, 806–809.

Leung, H.C., Yiu, S.M., Parkinson, J., Chin, F.Y., 2013. IDBA-MT: *de novo* assembler for metatranscriptomic data generated from next-generation sequencing technology. Journal of Computational Biology 20, 540–550.

Levy-Booth, D.J., Hashimi, A., Roccor, R., Liu, L.Y., Renneckar, S., Eltis, L.D., Mohn, W.W., 2021. Genomics and metatranscriptomics of biogeochemical cycling and degradation of lignin-derived aromatic compounds in thermal swamp sediment. ISME Journal 15, 879–893.

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 (Oxford, England) 31, 1674–1676.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics (Oxford, England) 25, 1754–1760.

Liao, H., Li, H., Duan, C.S., Zhou, X.Y., Luo, Q.P., An, X.L., Zhu, Y. G., Su, J.Q., 2022. Response of soil viral communities to land use changes. Nature Communications 13, 6027.

Liao, H., Liu, C., Ai, C., Gao, T., Yang, Q.E., Yu, Z., Gao, S., Zhou, S., Friman, V.P., 2023. Mesophilic and thermophilic viruses are associated with nutrient cycling during hyperthermophilic composting. ISME Journal 17, 916–930.

Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J., 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Research 47, D687–D692.

Lott, S.C., Voigt, K., Lambrecht, S.J., Hess, W.R., Steglich, C., 2020. A framework for the computational prediction and analysis of non-coding RNAs in microbial environmental populations and their experimental validation. ISME Journal 14, 1955–1965.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17, 10–12.

McGrath, K.C., Thomas-Hall, S.R., Cheng, C.T., Leo, L., Alexa, A., Schmidt, S., Schenk, P.M., 2008. Isolation and analysis of mRNA from environmental microbial communities. Journal of Microbiological Methods 75, 172–176.

McIlroy, S.J., Leu, A.O., Zhang, X., Newell, R., Woodcroft, B.J., Yuan, Z., Hu, S., Tyson, G.W., 2023. Anaerobic methanotroph '*Candidatus* Methanoperedens nitroreducens' has a pleomorphic life cycle. Nature Microbiology 8, 321–331.

Mettel, C., Kim, Y., Shrestha, P.M., Liesack, W., 2010. Extraction of mRNA from soil. Applied and Environmental Microbiology 76, 5995–6000.

Middleton, H., Yergeau, É., Monard, C., Combier, J.P., El Amrani, A., 2021. Rhizospheric plant–microbe interactions: miRNAs as a key mediator. Trends in Plant Science 26, 132–141.

Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., Banfield, J.F., 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biology 12, 1–14.

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.K., Meng, J., Durham, B.P., Shen, C., Varaljay, V.A., Smith, C. B., Yager, P.L., Hopkinson, B.M., 2013. Sizing up metatranscriptomics. ISME Journal 7, 237–243.

Muscatt, G., Hilton, S., Raguideau, S., Teakle, G., Lidbury, I.D., Wellington, E.M., Quince, C., Millard, A., Bending, G.D., Jameson, E., 2022. Crop management shapes the diversity and activity of DNA and RNA viruses in the rhizosphere. Microbiome 10, 1–16.

Neri, U., Wolf, Y.I., Roux, S., Camargo, A.P., Lee, B., Kazlauskas, D., Chen, I.M., Ivanova, N., Allen, L.Z., Paez-Espino, D., 2022. Expansion of the global RNA virome reveals diverse clades of bacteriophages. Cell 185, 4023–4037.e18.

Nuccio, E.E., Nguyen, N.H., Nunes da Rocha, U., Mayali, X., Bougoure, J., Weber, P.K., Brodie, E., Firestone, M., Pett-Ridge, J., 2021. Community RNA-Seq: multi-kingdom responses to living

versus decaying roots in soil. ISME Communications 1, 1–10.

Nuccio, E.E., Starr, E., Karaoz, U., Brodie, E.L., Zhou, J., Tringe, S. G., Malmstrom, R.R., Woyke, T., Banfield, J.F., Firestone, M.K., Pett-Ridge, J., 2020. Niche differentiation is spatially and temporally regulated in the rhizosphere. ISME Journal 14, 999–1014.

Ojala, T., Häkkinen, A.E., Kankuri, E., Kankainen, M., 2023. Current concepts, advances, and challenges in deciphering the human microbiota with metatranscriptomics. Trends in Genetics 39, 686–702.

Ortiz, R., Gera, P., Rivera, C., Santos, J.C., 2021. Pincho: a modular approach to high quality *de novo* transcriptomics. Genes 12, 953.

Parro, V., Moreno-Paz, M., González-Toril, E., 2007. Analysis of environmental transcriptomes by DNA microarrays. Environmental Microbiology 9, 453–464.

Passmore, L.A., Coller, J., 2022. Roles of mRNA poly (A) tails in regulation of eukaryotic gene expression. Nature Reviews Molecular Cell Biology 23, 93–106.

Peng, J., Wegner, C.E., Bei, Q., Liu, P., Liesack, W., 2018. Metatranscriptomics reveals a differential temperature effect on the structural and functional organization of the anaerobic food web in rice field soil. Microbiome 6, 1–16.

Perez-Coronel, E., Michael Beman, J., 2022. Multiple sources of aerobic methane production in aquatic ecosystems include bacterial photosynthesis. Nature Communications 13, 6454.

Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A., Hollibaugh, J.T., 2005. Analysis of microbial gene transcripts in environmental samples. Applied and Environmental Microbiology 71, 4121–4126.

Prosser, J.I., 2015. Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. Nature Reviews Microbiology 13, 439–446.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41, D590–D596.

Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A., Amid, C., Aziz, R.K., Bordenstein, S.R., Bork, P., Breitbart, M., Cochrane, G.R., Daly, R.A., Desnues, C., Duhaime, M.B., Emerson, J.B., Enault, F., Fuhrman, J.A., Hingamp, P., Hugenholtz, P., Hurwitz, B.L., Ivanova, N.N., Labonté, J.M., Lee, K.B., Malmstrom, R.R., Martinez-Garcia, M., Mizrachi, I.K., Ogata, H., Páez-Espino, D., Petit, M.A., Putonti, C., Rattei, T., Reyes, A., Rodriguez-Valera, F., Rosario, K., Schriml, L., Schulz, F., Steward, G.F., Sullivan, M.B., Sunagawa, S., Suttle, C.A., Temperton, B., Tringe, S.G., Thurber, R.V., Webster, N.S., Whiteson, K.L., Wilhelm, S.W., Wommack, K.E., Woyke, T., Wrighton, K.C., Yilmaz, P., Yoshida, T., Young, M.J., Yutin, N., Allen, L.Z., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2019. Minimum information about an uncultivated virus genome (MIUViG). Nature Biotechnology 37, 29–37.

Sabino, Y.N.V., Santana, M.F., Oyama, L.B., Santos, F.G., Moreira, A.J.S., Huws, S.A., Mantovani, H.C., 2019. Characterization of antibiotic resistance genes in the species of the rumen microbiota. Nature Communications 10, 5252.

Schoelmerich, M.C., Ouboter, H.T., Sachdeva, R., Penev, P.I., Amano, Y., West-Roberts, J., Welte, C.U., Banfield, J.F., 2022. A widespread group of large plasmids in methanotrophic *Methanoperedens* archaea. Nature Communications 13, 7085.

Semmouri, I., De Schamphelaere, K.A., Mees, J., Janssen, C.R., Asselman, J., 2020. Evaluating the potential of direct RNA nanopore sequencing: Metatranscriptomics highlights possible seasonal differences in a marine pelagic crustacean zooplankton community. Marine Environmental Research 153, 104836.

Shakya, M., Lo, C.C., Chain, P.S., 2019. Advances and challenges in metatranscriptomic analysis. Frontiers in Genetics 10, 904.

Shi, M., Lin, X.D., Chen, X., Tian, J.H., Chen, L.J., Li, K., Wang, W., Eden, J.S., Shen, J.J., Liu, L., Holmes, E.C., Zhang, Y.Z., 2018. The evolutionary history of vertebrate RNA viruses. Nature 556, 197–202.

Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X. C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.Z., 2016. Redefining the invertebrate RNA virosphere. Nature 540, 539–543.

Shi, Y., Tyson, G.W., DeLong, E.F., 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature 459, 266–269.

Shrestha, P.M., Kube, M., Reinhardt, R., Liesack, W., 2009. Transcriptional activity of paddy soil bacterial communities. Environmental Microbiology 11, 960–970.

Söllinger, A., Séneca, J., Borg Dahl, M., Motleleng, L.L., Prommer, J., Verbruggen, E., Sigurdsson, B.D., Janssens, I., Peñuelas, J., Urich, T., Richter, A., Tveit, A.T., 2022. Down-regulation of the bacterial protein biosynthesis machinery in response to weeks, years, and decades of soil warming. Science Advances 8, eabm3230.

Starr, E.P., Nuccio, E.E., Pett-Ridge, J., Banfield, J.F., Firestone, M. K., 2019. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. Proceedings of the National Academy of Sciences of the United States of America 116, 25900–25908.

Stewart, F.J., Ottesen, E.A., DeLong, E.F., 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME Journal 4, 896–907.

Tan, S., Liu, J., Fang, Y., Hedlund, B.P., Lian, Z.H., Huang, L.Y., Li, J.T., Huang, L.N., Li, W.J., Jiang, H.C., Dong, H.L., Shu, W.S., 2019. Insights into ecological role of a new deltaproteobacterial order *Candidatus* Acidulodesulfobacterales by metagenomics and metatranscriptomics. ISME Journal 13, 2044–2057.

Täumer, J., Marhan, S., Groß, V., Jensen, C., Kuss, A.W., Kolb, S., Urich, T., 2022. Linking transcriptional dynamics of CH_4-cycling grassland soil microbiomes to seasonal gas fluxes. ISME Journal 16, 1788–1797.

Tong, D., Wang, Y., Yu, H., Shen, H., Dahlgren, R.A., Xu, J., 2023. Viral lysing can alleviate microbial nutrient limitations and accumulate recalcitrant dissolved organic matter components in soil. ISME Journal 17, 1247–1256.

Toseland, A., Moxon, S., Mock, T., Moulton, V., 2014. Metatranscriptomes from diverse microbial communities: assessment of data reduction techniques for rigorous annotation. BMC Genomics 15, 1–7.

Turner, T.R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Osbourn, A., Grant, A., Poole, P.S., 2013. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. ISME Journal 7, 2248–2258.

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S. C., 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS One 3, e2527.

Van Goethem, M.W., Osborn, A.R., Bowen, B.P., Andeer, P.F., Swenson, T.L., Clum, A., Riley, R., He, G., Koriabine, M., Sandor, L., Yan, M., Daum, C.G., Yoshinaga, Y., Makhalanyane, T. P., Garcia-Pichel, F., Visel, A., Pennacchio, L.A., O'Malley, R.C., Northen, T.R., 2021. Long-read metagenomics of soil communities reveals phylum-specific secondary metabolite dynamics. Communications Biology 4, 1–10.

Wahl, A., Huptas, C., Neuhaus, K., 2022. Comparison of rRNA depletion methods for efficient bacterial mRNA sequencing. Scientific Reports 12, 1–11.

Wang, F., Fu, Y.H., Sheng, H.J., Topp, E., Jiang, X., Zhu, Y.G., Tiedje, J.M., 2021. Antibiotic resistance in the soil ecosystem: A One Health perspective. Current Opinion in Environmental Science & Health 20, 100230.

Wang, J., Qu, Y.N., Evans, P.N., Guo, Q., Zhou, F., Nie, M., Jin, Q., Zhang, Y., Zhai, X., Zhou, M., Yu, Z., Fu, Q.L., Xie, Y.G., Hedlund, B.P., Li, W.J., Hua, Z.S., Wang, Z., Wang, Y., 2023. Evidence for nontraditional *mcr*-containing archaea contributing to biological methanogenesis in geothermal springs. Science Advances 9, eadg6004.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biology 20, 1–13.

Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A., Hoelzle, R.D., Lamberton, T.O., McCalley, C. K., Hodgkins, S.B., Wilson, R.M., Purvine, S.O., Nicora, C.D., Li, C., Frolking, S., Chanton, J.P., Crill, P.M., Saleska, S.R., Rich, V. I., Tyson, G.W., 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560, 49–54.

Xia, R., Sun, M., Balcázar, J.L., Yu, P., Hu, F., Alvarez, P.J., 2023. Benzo[a]pyrene stress impacts adaptive strategies and ecological functions of earthworm intestinal viromes. ISME Journal 17, 1004–1014.

Xu, L., Dong, Z., Chiniquy, D., Pierroz, G., Deng, S., Gao, C., Diamond, S., Simmons, T., Wipf, H.M.L., Caddell, D., Varoquaux, N., Madera, M.A., Hutmacher, R., Deutschbauer, A., Dahlberg, J.A., Guerinot, M.L., Purdom, E., Banfield, J.F., Taylor, J.W., Lemaux, P.G., Coleman-Derr, D., 2021. Genome-resolved metagenomics reveals role of iron metabolism in drought-induced rhizosphere microbiome dynamics. Nature Communications 12, 3209.

Yates, M.C., Derry, A.M., Cristescu, M.E., 2021. Environmental RNA: a revolution in ecological resolution? Trends in Ecology & Evolution 36, 601–609.

Yergeau, E., Tremblay, J., Joly, S., Labrecque, M., Maynard, C., Pitre, F.E., St-Arnaud, M., Greer, C.W., 2018. Soil contamination alters the willow root and rhizosphere metatranscriptome and the root–rhizosphere interactome. ISME Journal 12, 869–884.

Yin, Z., Ye, L., Jing, C., 2022. Genome-resolved metagenomics and metatranscriptomics reveal that Aquificae dominates arsenate reduction in Tengchong geothermal springs. Environmental Science & Technology 56, 16473–16482.

Yuan, C., Lei, J., Cole, J., Sun, Y., 2015. Reconstructing 16S rRNA genes in metagenomic data. Bioinformatics (Oxford, England) 31, i35–i43.

Yuan, L., Wang, Y., Zhang, L., Palomo, A., Zhou, J., Smets, B.F., Bürgmann, H., Ju, F., 2021. Pathogenic and indigenous denitrifying bacteria are transcriptionally active and key multi-antibiotic-resistant players in wastewater treatment plants. Environmental Science & Technology 55, 10862–10874.

Zhou, Z., Zhang, C., Liu, P., Fu, L., Laso-Pérez, R., Yang, L., Bai, L., Li, J., Yang, M., Lin, J., Wang, W., Wegener, G., Li, M., Cheng, L., 2022. Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species. Nature 601, 257–262.