BMC Methods

METHODOLOGY

Open Access

Assessment of rhizosphere microbial activity using optimized RNA extraction coupled with universal ribosomal RNA depletion techniques

Kipa Tamrakar[1,2], P. Winston Miller[1,2], Maureen C. Dolan[1,2,3] and Asela J. Wijeratne[1,2,3*]

Abstract

Background Metatranscriptomics, which analyzes gene expression patterns in a heterogeneous community, is a powerful tool to evaluate microbial functional activity. A key challenge in this process is obtaining high-quality RNA, which is complicated by the composition and ecosystem complexity of the soil matrix. Another crucial step involves the removal of highly abundant ribosomal RNA (rRNA), as its presence can dominate sequencing results and obscure the detection of messenger RNA (mRNA) expression. Conventional library preparation methods often struggle to efficiently remove rRNA from a complex mix of prokaryotic and eukaryotic organisms, further complicating mRNA isolation.

Methods To overcome these limitations, we have developed an optimized method for extracting RNA from soybean rhizosphere microbes, followed by universal rRNA depletion to create rRNA-free samples for sequencing. These samples were sequenced using an Illumina high-throughput sequencer.

Results Our optimized cetyltrimethylammonium bromide (CTAB) phenol–chloroform extraction protocol significantly improved RNA yield and quality from clay-rich soils, outperforming previously published methods and commercial kits. Illumina sequencing data revealed minimal rRNA contamination, and the resulting short reads could be successfully assembled into transcripts. These findings also demonstrate that using the Zymo-Seq RiboFree Total RNA Library Kit effectively enabled library preparation from complex environmental samples for Illumina sequencing.

Discussion This RNA sample preparation method, combined with our optimized extraction technique, provides a valuable approach for studying rhizosphere microbes that hold exciting potential for advancing soil health assessments and understanding plant–microbe-pathogen interactions.

Keywords RNA-extraction, Metatranscriptome, Microbiome, Rhizosphere

Introduction

The region in the soil influenced by plant roots, known as the rhizosphere, contains microbes that play a pivotal role in plant growth, health, and productivity [1]. The rhizosphere's plant-microbiome-soil interface regulates nutrient uptake and provides a primary defense against biotic and abiotic stresses, as extensively reviewed in previous literature [2, 3]. A shift in the composition or functional

*Correspondence:
Asela J. Wijeratne
awijeratne@astate.edu
[1] Molecular Biosciences Graduate Program, Arkansas State University, Jonesboro, AR, USA
[2] Arkansas Biosciences Institute, Jonesboro, AR, USA
[3] Department of Biological Sciences, Beck College of Sciences and Mathematics, Arkansas State University, Jonesboro, AR, USA

behavior of the rhizosphere microbial community (also known as rhizobiome) can drastically impact the host plant [4–6]. Therefore, understanding microbial functional dynamics within the rhizosphere is paramount.

Traditionally, researchers have depended on sequencing specific DNA markers, such as 16S ribosomal RNA (rRNA) or 18S rRNA genes, for identification, quantification, and functional characterization of soil microbes [7]. Recently, whole-genome shotgun sequencing (WGSS), which can capture both marker and functional genes, has gained traction [8]. These methods have yielded significant insights into soil microbiomes by identifying the presence and abundance of genes or organisms. However, prior studies showed that RNA-based sequencing outperforms DNA based sequencing in detecting microorganisms, directly correlating with their active involvement within the community [9, 10]. Metatranscriptomics addresses this limitation by capturing actively transcribed genes [11]. This approach allows for an analysis of functional changes within microbial communities in response to varying conditions and provides specific insight into microbial dynamics [12].

There are several challenges that impact successful metatranscriptome analysis. Extraction of good-quality RNA from rhizosphere soil is complicated due to copurification of phenolics, such as humic acid, during the extraction process. These compounds can hamper downstream molecular work [13]. Second, RNA extraction from soil requires a highly curated environment due to the ubiquity of robust RNA nucleases in soil [14]. While several commercial kits exist for soil RNA extraction, they can be cost-prohibitive when processing large sample numbers. Moreover, these kits often lack flexibility for optimization, making it challenging to adapt them to different soil types. Additionally, many of these kits involve complex, multi-step procedures, making them labor-intensive and time-consuming. Recently, Poursalavati et al. (2023) reported an optimized method for extracting RNA from rhizosphere soil samples. This method successfully produced RNA samples from both mineral and organic soils, free from contaminants like DNA, proteins, and humic acids. As a result, the samples are suitable for more accurate and reliable metatranscriptome analysis. However, further optimization is required to adapt this method for varying agricultural soils with higher clay content.

It is also necessary to remove highly abundant rRNA from total RNA samples for efficient mRNA sequencing. However, the removal of rRNA from heterogeneous multi-species samples is challenging due to differences in prokaryote and eukaryote RNA dynamics [15]. While there are various commercially available RNA library preparation kits for microorganisms, these platforms are designed to generate libraries for prokaryotic organisms independent of eukaryotic organisms. Classic prokaryotic library preparation kits use rRNA depletion during library preparation whereas poly(A) tail selection is a common eukaryotic method [16–18]. Approaches generating dual libraries are quite challenging on two fronts. It is physically and economically exhausting to create two libraries that in turn can lead to batch effects that further complicate the downstream analyses.

This project's main aim was to develop an optimized method for extracting mRNA from the plant rhizosphere, followed by universal rRNA depletion to obtain rRNA-free samples suitable for Illumina sequencing. By implementing our optimized mRNA extraction protocol, we successfully obtained a substantial yield of high-quality mRNA. Sequencing libraries were then prepared through rRNA depletion using Zymo-Seq RiboFree Total RNA Library Kit to facilitate high-throughput sequencing [19]. We confirmed effective rRNA removal *in-silico* and successfully assembled microbial transcript contigs, enabling the assessment of microbial activity. Our approach offers a valuable method for advancing the understanding of the functional roles of rhizosphere microbes, particularly in studies related to soil health and plant-pathogen interactions.

Methods

Soybean rhizosphere soil collection

Soil was collected from Arkansas State University Agricultural Teaching and Research Center (35° 50′ 16″ N, 90° 40′ 00″ W), comprised of a Collins slit loam [20]. After removing plant matter, soil was distributed to three autoclaved pots (top diameter 9″ and height 8 ½"). A single seed was sown into each of the three pots. Soybeans were grown to Reproductive Stage 4 with trifoliate leaves and pods [21]. Plants were carefully uprooted and shaken gently to remove the loose soil. The roots from the base of each plant stem were excised and collected in 50 ml conical tubes with 35 ml 1X PBS buffer [22]. The tubes were vortexed for 2 min to separate rhizosphere soil from the root tissue. The tubes were centrifuged at 3000 g for 5 min, supernatant was discarded, and the remaining pellet of rhizosphere soil was flash-frozen and stored at − 80 °C [23].

RNA Isolation and Purification

A cetyltrimethylammonium bromide (CTAB)-based phenol: chloroform extraction procedure described in Poursalavati et al. (2023) was optimized to develop the current RNA isolation protocol (Fig. 1A). The detailed protocol is published in protocol.io (https://doi.org/

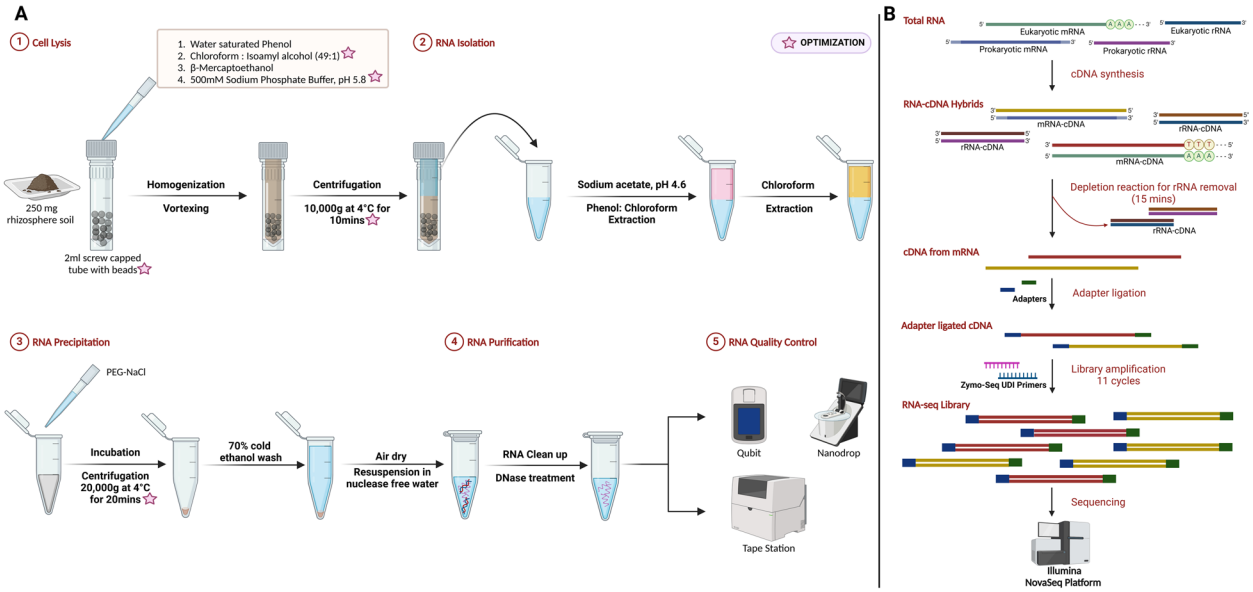

Fig. 1 **A** The optimized CTAB based Phenol chloroform method to extract RNA from the soybean rhizosphere. Steps indicated by stars are the adjustment implemented to the protocol described in Poursalavati et al. (2023); (**B**) The graphic representation of the steps involved in preparation of RNA sequencing library using Zymo-Seq RiboFree® Total RNA Library Kit. Total RNA is synthesized into RNA-cDNA hybrids. The rRNA-cDNA hybrids are depleted, and the remaining cDNA are amplified to produce sequencing library. This figure was created using Biorender

10.17504/protocols.io.6qpvr8weolmk/v1). Briefly, 250 mg rhizosphere soil samples were homogenized with a ratio of 19:1 of 0.1 mm and 0.5 mm silica beads in CTAB extraction buffer, water-saturated phenol, 49:1 Chloroform: Isoamyl alcohol [24], 500 mM sodium phosphate (NaP) buffer (pH 5.8), and 2-Mercaptoethanol. Samples were centrifuged at 10,000 g for 10 min at 4 °C [25]. Subsequent steps included Phenol: Chloroform: Isoamyl alcohol extraction and a second Chloroform: Isoamyl alcohol extraction. The aqueous phase recovered from this organic extraction was precipitated with one volume of PEG-NaCl precipitation solution, incubated on an ice bath in a refrigerator (4 °C) for 20 min and centrifuged at 20,000 g for 20 min at 4 °C [13, 26]. The recovered RNA-containing pellet was washed with 70% ice-cold ethanol, air-dried, and resuspended in nuclease-free water. Crude RNA was further purified using Zymo RNA Clean & Concentrator kits (Cat #R1015) supplemented with 1 U/µl DNase I (Zymo Research, Cat #E1010). RNA quality was assessed through a combination of quantification, purity, and integrity analyses. Quantification was performed using a Qubit 4 fluorometer (Thermo Fisher Scientific). Purity was established by determining the A260/A280 and A260/A230 ratios using a NanoDrop Microvolume spectrophotometer (Thermo Fisher Scientific). RNA integrity for prokaryote and eukaryote were measured using the Agilent 4150 TapeStation system (Agilent Technologies) and an RNA Integrity Number (RIN^e) for each sample was recorded.

RNA-sequencing library construction

Figure 1B illustrates the process used for rhizosphere soil RNA library preparation using Zymo-Seq RiboFree Total RNA Library Kit (Zymo Research, Cat #R3000). A step-by-step procedure for the library preparation can be found in protocol.io (https://doi.org/10.17504/protocols.io.6qpvr8weolmk/v1). Briefly, 250 ng of total RNA was subjected to cDNA synthesis. cDNA was treated with RiboFree® Universal Depletion reagents to remove rRNA-cDNA hybrids. The remaining cDNA from each sample was ligated with adapters and amplified using Zymo-Seq UDI Primers in accordance with the manufactures protocol. cDNA libraries were quantified using a Qubit 4 fluorometer (Thermo Fisher Scientific) and sequenced on the Illumina NovaSeq platform (20 million 150 PE reads per sample) at Novogene, Sacramento, CA.

Illumina short read analysis

Quality of sequencing reads was first assessed using *FastQC (v0.11.5)* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and quality trimming were performed using *BBMAP (v39.01)* (http://www.sourceforge.net/projects/bbmap/). To remove rRNA sequences, *SortMeRNA (v4.3.6)* was used with recommended settings and custom-built references from the Silva 138.1 small and large subunit references (retrieved December 20 th, 2023) [27, 28]. Following rRNA depletion, reads were aligned to the soybean genome (v4.0, retrieved from NCBI April 03, 2024) using *STAR (v2.7.11b)* aligner with

default parameters [29]. The aligned reads were removed from further analyses and remaining reads were assembled using *rnaSPAdes (v3.15.5)* [30]. The resulting assemblies were annotated (Taxonomy and Gene Ontology) using *DIAMOND (v2.1.9)* against the complete UniProt database (v2023_05, retrieved on March 1, 2024). To ensure accurate quantification, reads were aligned back to the assemblies using *Bowtie2 (v2.5.3)*, and contigs were then filtered to retain only those with 10 or more aligned reads [31].

Results

Good-quality RNA samples were obtained using optimized extraction protocol from rhizosphere soil

The nucleic acid concentrations of rhizosphere RNA samples (R1, R2, and R3), measured using the Qubit RNA HS Assay, ranged from 41.6 ng/µl to 132 ng/µl (Table 1). Spectrophotometer assessment ratios (Abs 260/280) of RNA samples were all approximately 2.0 (Table 1), suggesting pure, high-quality RNA [32]. While the R3 soil sample yielded an Abs 260/230 value in the desired RNA purity range (2.0–2.2), this metric varied and was lower in the rhizosphere total RNA assessed for the other two samples in this study. The lower Abs 260/230 ratios for R1 and R2 may be due to phenolics, salts, or other contaminants co-extracted from the soil [25]. For a more comprehensive assessment of RNA integrity of this optimized soil rhizosphere RNA extraction method, samples were evaluated using an Agilent TapeStation. In heterogenous samples both prokaryotic and eukaryotic RNA is expected to be present in soil rhizosphere with predicted RNA fragment sizes. Specifically, the largest 28S rRNA band (~ 4000 to 5000 nucleotides (nts)) sourced from eukaryotes, a slightly smaller 23S rRNA band (2900 nts) from prokaryotes, a smaller 18S rRNA (~ 1800 nts) band from eukaryotes, and the smallest 16S rRNA (1500 nts) prokaryotes ribosomal sub-units [33–35]. Prokaryotic RNA 23S and 16S rRNA bands were clearly detected in all samples and each species appeared as single bands indicative of intact rRNA (Fig. 2A). Conversely, the 28S rRNA band for eukaryotic RNA was less distinct compared to 18S rRNA band (Fig. 2B). All samples recorded RIN^e score greater than 8, that indicates good quality and is within the required integrity range for downstream analyses.

To confirm the suitability of these samples for Illumina library preparation and RNA sequencing, we used the Zymo-Seq RiboFree® Total RNA Library Kit to prepare samples for sequencing. The concentrations of the prepared libraries were 68.80 pg/µl, 36 pg/µl, and 71 pg/µl for R1, R2, and R3, respectively. The size distribution of the DNA fragments was assessed using TapeStation, and fragment sizes for all three samples ranged between 250–600 bp, with an average size of 350 bp (Supplementary Fig. 1). All three libraries passed quality control testing by the transcriptome core facility (Novogene Inc., Sacramento, CA).

Table 1 Qubit and Nanodrop results of the total RNA extracted from three rhizosphere samples

Samples	Qubit Concentration (ng/µL)	Nanodrop Spectrophotometer		
		Concentration (ng/µL)	A260/A280	A260/A230
R1	61	68.856	2.043	1.806
R2	41.6	68.725	1.867	1.523
R3	132	149.681	2.062	2.103

RiboFree Universal Depletion eliminates over 90% of rRNA

Consistent sequencing depth across all samples (between 20 and 22 million reads per sample) was successfully obtained. The quality-checked and trimmed sequences were then input into SortMeRNA, which identified and removed rRNA in-silico. The percentage of reads identified as rRNA varies across the samples; two samples (R1, R3) resulted in similar levels of 7.88% and 5.86% rRNA while a third sample (R2) displayed 0.17% rRNA content (Fig. 2C). However, this depletion step efficiently reduced the rRNA levels from over 90% to below 10%, which is sufficient to analyze soil rhizosphere complex transcriptome data.

Next, preprocessed short reads were assembled into contigs using the rnaSPAdes assembler [30]. The assembly resulted in 1.9 million total contigs. Furthermore, the longest contig was 57,272 nts long, and the E90 N50 metric indicated median length of contigs covering 90% of expression to be 266 nts. Alignment percentages were assessed utilizing Bowtie2 short-read aligner to map reads back to the assembled contigs. Our three sample alignment percentages ranged between 30 and 51% (Fig. 2D).

Transcriptomic analysis reveals active microbial populations in the soybean rhizosphere

A DIAMOND search of the UniProt database using assembled transcripts revealed 87% of them (1,656,097 out of 1,910,614) matched to least one sequence in the database with a majority sequence (52%) showing more than 90% sequence similarity to existing database sequences (Supplementary Fig. 2). Taxonomic annotation further indicated that 86% of the contigs (1,640,900 out of 1,910,614) received a classification (Fig. 3A). Among the contigs with taxonomic assignment, 98.79% (1,618,824 out of 1,640,900) of them were identified as bacterial origin, with the majority belonging to the Actinomycetota phylum (517,448), followed by Pseudomonadota

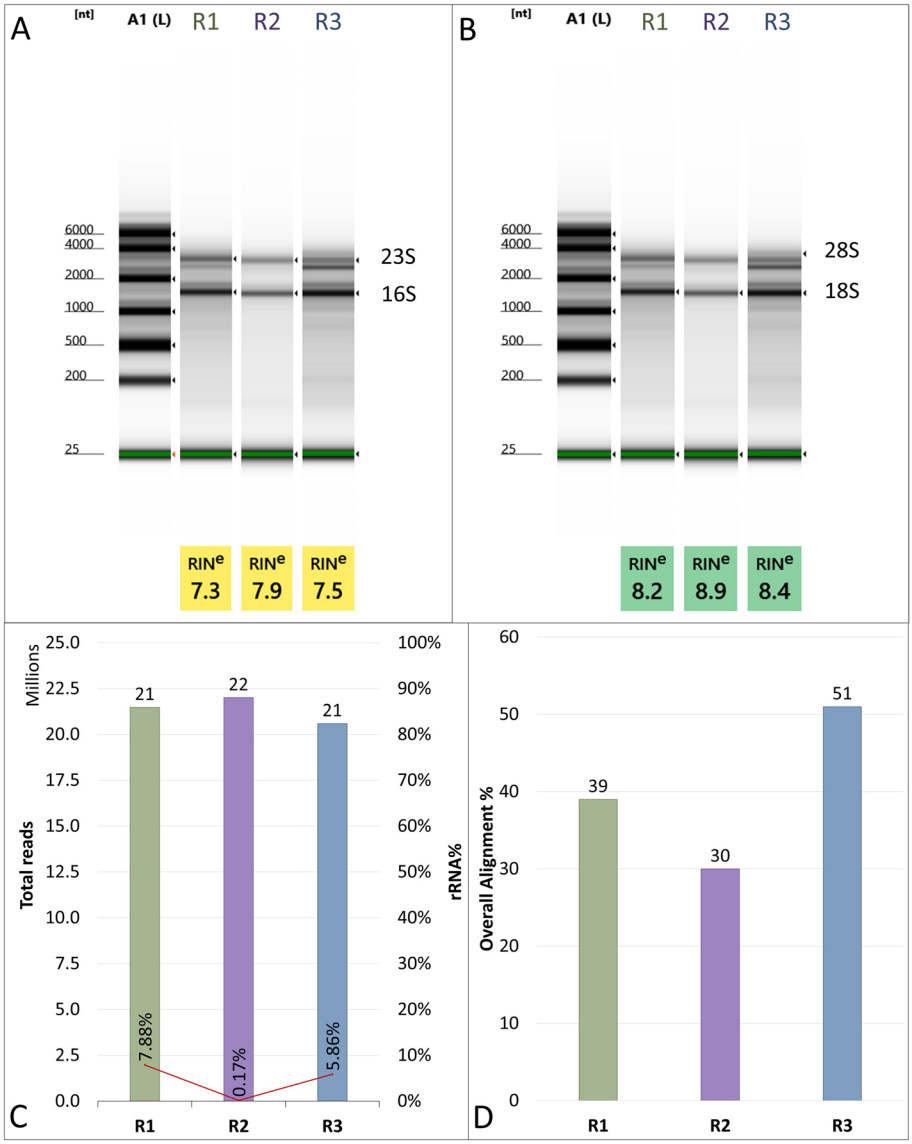

Fig. 2 TapeStation gel image of two rRNA bands obtained from (**A**) Prokaryote RNA analysis with RIN^e scores (yellow box) and (**B**) Eukaryotic RNA analysis with RIN.^{e} scores (green box). Lane A1(L) of both gel image is the RNA size marker; (**C**) Total reads (in millions) in y-axis at left, represented by bars generated after the sequencing using Novaseq platform and red line indicating the rRNA percentage (y-axis at right) detected in each sample; (**D**) Percentage of preprocessed reads, which were aligned to reference sequences using Bowtie2

(350,981) and Acidobacteriota (253,752). In contrast, Eukaryota (12, 935, 0.8%) and Archaea (6760, 0.4%) have low proportions, while Viruses contribute the least to the total count (2381, 0.1%). Moreover, 67% of the contigs (1,288,778 out of 1,910,614) were successfully assigned at least one Gene Ontology (GO) term, providing functional insights into the biological roles of these sequences (Fig. 3B). Then, we looked at the functional annotation of bacterial transcripts, given that they are the most overrepresented taxonomic group in this dataset. The most abundant biological processes identified in the bacterial metatranscripts are transmembrane transporter (359, 985), regulation of DNA-templated transcription (295,476), and proteolysis (250,197). In terms of molecular functions, our results indicated that ATP binding (872, 428), metal ion binding (540, 077), and DNA binding (460, 358) are the most dominant activities.

A four-step filtering process was implemented to assess further the quality of the assembled contigs (Supplementary Fig. 3). The first step retained contigs with at least ten short reads aligned, reducing the dataset to 101,102 contigs (5.29%). This threshold was set to exclude lowly

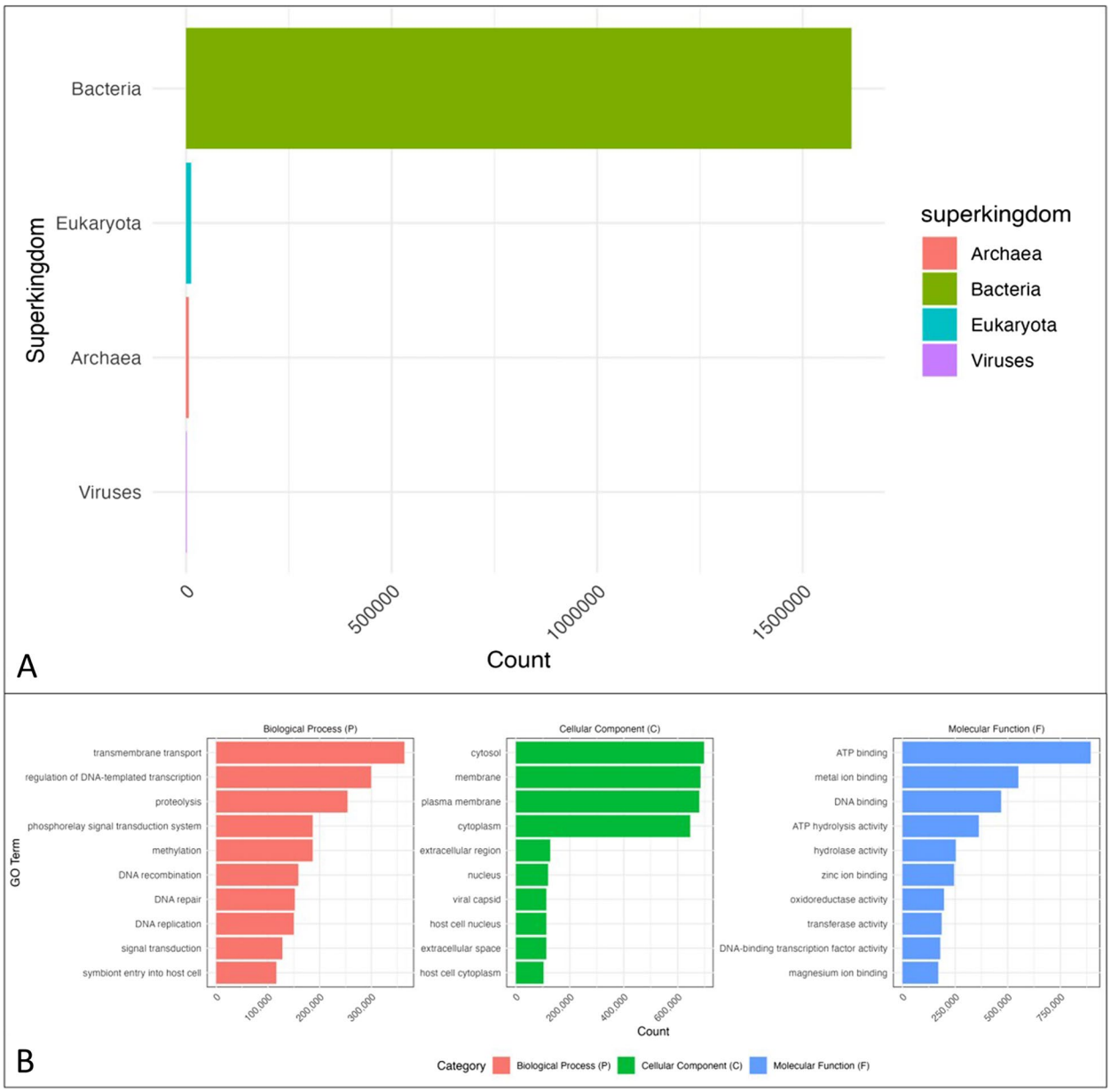

Fig. 3 (**A**) Taxonomy assignment of the genes with counts into three different superkingdoms: Archea, Bacteria, Eukaryotes, and Viruses; (**B**) Gene Ontology (GO) term analysis of transcript identified from three soybean rhizosphere samples including their respective counts, categorized into three groups: Biological Process (P), Cellular Component (**C**), and Molecular Function (**F**)

expressed contigs, which constituted almost 95% of the dataset. In the second step, a filter for valid start or stop codons within open reading frames (ORFs) intending to capture genes was applied, reducing the contig number to ~60,000. The third step selected contigs with at least 90% ORF coverage thus narrowing the dataset further to approximately 10,000 contigs. Finally, contigs with at least 90% sequence identity to reference sequences were retained, reducing the number to approximately 3,000 contigs. This bioinformatics pipeline ensured a set of high-quality, non-redundant sequences in the assembly.

Discussion

The study of the rhizosphere microbiome through metatranscriptomics remains relatively unexplored due to several technical challenges. One significant obstacle is the difficulty of extracting high-quality RNA from the rhizosphere, compounded by the need for separate

library preparation kits for prokaryotic and eukaryotic RNA. In this study, we optimized the RNA extraction protocol developed by Poursalavati et al. (2023), tailoring it to our clay-rich soil composition to improve RNA yield. We then employed universal rRNA depletion to create rRNA-free RNA-seq libraries. Finally, we present a streamlined bioinformatics pipeline to efficiently process the sequencing data, offering a more accessible approach for rhizosphere metatranscriptomic analysis. Together, these advancements provide a more reliable and scalable framework for studying soil microbial function.

A crucial factor limiting the study of microbial activity from the rhizosphere is inadequate RNA recovery from soil. Initially, we used commercially available kits that yielded either undetectable or very low levels of RNA that generated an insufficient amount of RNA for downstream analyses (Supplementary Table 1). These preliminary findings indicated that RNA extraction from soil requires careful consideration of soil characteristics and anticipated microbial content. A well-documented obstacle in soil-based RNA recovery is the strong adsorption of nucleic acids to the cations associated with the clay particles. These cations can bind nucleic acid molecules tightly, significantly reducing extraction efficiency [36, 37]. Considering the high clay content (18%) of Arkansas Delta soil [38], we tested an organic solvent-based approach using the CTAB Phenol: Chloroform method as outlined by Poursalavati et al. (2023) [39]. Although this method successfully produced detectable RNA, the yields were still low (Supplementary Table 2). In addition, all soil rhizosphere RNA samples analyzed had Nanodrop concentrations 2–3 folds higher than Qubit readings. As the latter measure is more specific for RNA, these results indicate the presence of contaminants and lower quality RNA that is problematic for downstream analyses.

To develop a more effective RNA extraction method for complex soils with higher clay content, we refined and optimized the Poursalavati et al. (2023) protocol. The key enhancement in our workflow was the introduction of a high ionic strength, low pH buffer (500 mM NaP buffer, pH 5.8) to improve RNA yield and quality. The elevated ionic strength of the NaP buffer likely minimized nucleic acid adsorption by clay minerals, thereby enhancing RNA concentration in the cell lysate [40, 41]. We further increased recovery by using a mixture of bead sizes for lysis (0.1 mm and 0.5 mm beads). This step likely increased cell disruption of varied sizes and compositions across kingdoms, improving RNA recovery [42]. Furthermore, the addition of isoamyl alcohol to chloroform, along with longer centrifugation, helped polysaccharides, proteins, and other cellular components in the debris to pellet, further improving RNA yield [43, 44].

We successfully applied this method to rhizosphere soil collected from field sites in Iowa with high-quality RNA recovered (data not shown). Central Iowa cropland consists of Haig and Grundy silt loam with even higher clay content, thus demonstrating the applicability of this method across different soil types and environmental conditions [45–47]. However, future studies should explore additional strategies to enhance soil rhizosphere RNA purity. In particular, while incorporating 500 mM NaP buffer improved RNA yield, it may not be sufficient to completely remove humic acid from soil during extraction, as we have observed less than ideal 260/230 values for some samples [39]. Despite this, all samples had RINe scores above 7.0, consistent with Prawer et al. (2023) confirming RNA quality suitable for library preparation and sequencing.

We utilized the Zymo-Seq RiboFree Total RNA Library Kit (Zymo-Seq kit), which effectively depletes rRNA in a domain-agnostic manner [48–50]. This approach uses double-stranded nuclease (DSN) to selectively degrade rRNA-derived cDNA by targeting early formed DNA-RNA hybrids and double-stranded DNA, allowing efficient probe-free rRNA depletion through precise timing [51]. Our data indicates that this method is highly effective at removing rRNA in complex samples, such as soil rhizosphere, containing RNA from multiple species without bias towards specific organisms [13].

While we observed variation in rRNA present among our three RNA library samples, residual rRNA were below the 10% threshold that is commonly used for transcriptome analyses [52]. Considering that RiboFree® Universal Depletion procedure was optimized for single organism samples, further optimizing this step for environmental samples that were the focus of this study is warranted. Variables such as the amount of RNA input into the depletion reaction have previously been shown to correlate strongly with rRNA presence in RNA library samples [53]. However, it is worth noting that the RIN^e values, the standard RNA quality metric, were very comparable between the three samples (R1-R3), and these values are considered well within the acceptable range for proceeding with RNA library applications [54].

Previously, a common approach to studying community mRNA involved using two independent commercial kits: one for prokaryotic transcripts and one for eukaryotic transcripts. While eukaryotic kits rely on poly(A) tail selection, prokaryotic kits typically use rRNA depletion due to the absence of poly(A) modifications in bacterial and archaeal transcripts [55]. However, this parallel processing increases labor, cost, and the risk of batch effects, which can complicate downstream analyses and data integration. Therefore, using the Zymo-Seq kit can streamline the process and increase reproducibility.

Notably, this is the first report to use the Zymo-Seq kit to examine microbial activities in the rhizosphere soil samples.

Data analysis of short transcript reads, as was used in our study, remains quite challenging. Historically, short reads can either be aligned to a pre-determined reference, or they can be assembled de novo and annotated [11]. The rhizosphere and soil samples are generally highly variable and contain many understudied taxa. To date, there is a lack of databases that are specific to plant host species, or to various soil types in which host plants are grown. Due to this paucity of information for soybean and its rhizosphere, we opted to use a de novo assembly-based approach. In this study, we aimed to generate a comprehensive assembly of community mRNA to understand microbial activity in the rhizosphere better. While a high proportion of assembled contigs matched known database sequences, incomplete transcripts remain a challenge, as shorter contigs may limit gene-level resolution. This incomplete assembly could contribute to the lower percentage of sequence alignments. We believe an increase in sequencing depth would improve all assembly metrics. To test this hypothesis, we subsampled our short reads and assembled each subsample (Supplementary Fig. 4). As anticipated, the number of contigs increased as the number of reads grew, supporting the need for deeper sequencing to enhance completeness and accuracy.

To better understand microbial activity in the soybean rhizosphere, we analyzed transcript annotations based on Gene Ontology (GO) terms. The functional profiles suggest a highly active bacterial community, with enriched genes associated with ATP binding, metal ion binding, and DNA binding, indicating elevated energy transfer, enzymatic activity, and gene regulation [56–58]. Additionally, genes involved in transmembrane transport, transcriptional regulation, and proteolysis highlight bacterial roles in nutrient uptake, gene expression control, and protein degradation, all of which are essential for maintaining metabolic balance in the soil ecosystem. [57, 59–61].

Taxonomic classification of assembled sequences provides further insight into the microbial dynamics of the rhizosphere. A significant proportion of assembled contigs were assigned to the genus or species level, with Actinomycetota, Pseudomonadota, and Acidobacteriota emerging as the most active microbial taxa. These groups play key roles in soil fertility, biogeochemical cycles, and plant–microbe interactions. Notably, Actinobacteria contribute to plant health by producing phytohormones, enzymes, and antibiotics, while Pseudomonadota and Acidobacteriota function as plant growth-promoting rhizobacteria (PGPRs), aiding in nutrient mobilization and stress resilience [62–64]. Their combined activities underscore the importance of microbial communities in shaping rhizosphere dynamics and supporting sustainable agricultural practices [65].

Approximately 98% of annotated contigs were identified as bacterial in origin. While it is possible that our extraction method may have disproportionately captured transcripts from prokaryotes, TapeStation data (Fig. 2A and 2B) suggests otherwise. Eukaryotic 18S rRNA bands appeared intense compared to 28S rRNA indicating RNA recovered from rhizosphere might have been affected by partial degradation. This could contribute to the incomplete assembly of eukaryotic transcripts. Also, we cannot rule out the possibility that the library preparation method may have eliminated eukaryotic transcripts during the rRNA depletion step. However, given that the probes used have been validated on several eukaryotic species, this also seems less probable [66, 67]. It is possible the longer eukaryotic transcripts may not have been fully assembled, or the annotation process is biased toward prokaryotic sequences. As of October 2024, the UniProt database, which we used for taxonomic annotation, contains significantly more bacterial sequences (153 million) compared to eukaryotes and fungal sequences (19 million), representing only about one-fifth of the total eukaryotic sequences. Such disparity in domain-specific sequence representation highlights a potential bias in the database toward bacterial sequence annotation, resulting in overrepresentation of bacterial contigs [68]. Overall, our results suggest that mRNA extraction from soil and de novo assembly of the resulting short reads into a metatranscriptome is a viable approach.

Conclusions

In this study, we successfully optimized a CTAB phenol–chloroform-based extraction method to enhance RNA yield and quality from clay-rich soils. Our approach leads to significantly enhanced RNA yield over other previously published methods and commercial kits. In addition, using Zymo-Seq RiboFree Total RNA Library Kit, we were able to fine-tune the metatranscriptomic library preparation methods for Illumina sequencing. Notably, the universal rRNA depletion step used in this kit eliminated the need for separate kits for prokaryotic and eukaryotic sample preparation. Our study also demonstrated that cost-effective, short-read sequencing could effectively assemble transcripts from diverse microbial communities. These findings also underscore the importance of increased sequencing depth to capture the full diversity and complexity of microbial transcripts in soil samples. Furthermore, the study highlights the necessity of selecting appropriate databases to improve the annotation accuracies of assembled transcripts.

This study presents an optimized RNA extraction protocol for clay-rich soils and a cost-effective preparation method for high-throughput sequencing, enabling transcriptomic analysis in complex agricultural soils. The development and adoption of robust and reliable experimental workflows are important to facilitating future research on active rhizosphere microbes, supporting comparative studies across various environmental and agricultural conditions. In turn, these studies will contribute to promoting sustainable agricultural practices and reducing reliance on chemical inputs through microbial-based solutions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44330-025-00027-6.

Supplementary Figure 1.
Supplementary Figure 2.
Supplementary Figure 3.
Supplementary Figure 4.
Supplementary Table 1.
Supplementary Table 2.

Acknowledgements
The authors acknowledge and are appreciative of the Arkansas High-Performance Computing Center.

Authors' contributions
K.T., W.P.M., M.D., and A.J.W. designed the study. K.T. performed all the laboratory experiments and collected the data. W.P.M. and A.J.W. conducted the data analysis. K.T., W.P.M., and A.J.W. wrote the manuscript. All authors have edited, read, and approved the final manuscript.

Funding
This work is supported by the Agricultural Microbiomes in Plant Systems and Natural Resources program's grant (project award no. 2022–67014 - 36674) from the U.S. Department of Agriculture's National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

Data availability
Sequence data that support the findings of this study have been deposited in The National Center for Biotechnology Information under the primary accession code BioProject: PRJNA1199510.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 13 November 2024 Accepted: 4 April 2025
Published online: 22 April 2025

References

1. Thepbandit W, Athinuwat D. Rhizosphere Microorganisms Supply Availability of Soil Nutrients and Induce Plant Defense. Microorganisms. 2024;12:558.
2. Pantigoso HA, Newberger D, Vivanco JM. The rhizosphere microbiome: Plant–microbial interactions for resource acquisition. J Appl Microbiol. 2022;133:2864–76.
3. Mohanram S, Kumar P. Rhizosphere microbiome: revisiting the synergy of plant-microbe interactions. Ann Microbiol. 2019;69:307–20.
4. Mendes R, Garbeva P, Raaijmakers JM. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol Rev. 2013;37:634–63.
5. Kong Z, Liu H. Modification of Rhizosphere Microbial Communities: A Possible Mechanism of Plant Growth Promoting Rhizobacteria Enhancing Plant Growth and Fitness. Front Plant Sci. 2022;13:920813.
6. Xun W, Shao J, Shen Q, Zhang R. Rhizosphere microbiome: Functional compensatory assembly for plant fitness. Comput Struct Biotechnol J. 2021;19:5487–93.
7. Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, et al. The ecologist's field guide to sequence-based identification of biodiversity. Methods Ecol Evol. 2016;7:1008–18.
8. Pérez-Cobas AE, Gomez-Valero L, Buchrieser C. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. Microb Genomics. 2020;6:mgen000409.
9. Li R, Tun HM, Jahan M, Zhang Z, Kumar A, Dilantha Fernando WG, et al. Comparison of DNA-, PMA-, and RNA-based 16S rRNA Illumina sequencing for detection of live bacteria in water. Sci Rep. 2017;7:5752.
10. Salgar-Chaparro SJ, Machuca LL. Complementary DNA/RNA-Based Profiling: Characterization of Corrosive Microbial Communities and Their Functional Profiles in an Oil Production Facility. Front Microbiol. 2019;10:2587.
11. Shakya M, Lo C-C, Chain PSG. Advances and Challenges in Metatranscriptomic Analysis. Front Genet. 2019;10:904.
12. Bashiardes S, Zilberman-Schapira G, Elinav E. Use of Metatranscriptomics in Microbiome Research. Bioinforma Biol Insights. 2016;10:19–25.
13. Mettel C, Kim Y, Shrestha PM, Liesack W. Extraction of mRNA from Soil. Appl Environ Microbiol. 2010;76:5995–6000.
14. Hurt RA, Qiu X, Wu L, Roh Y, Palumbo AV, Tiedje JM, et al. Simultaneous Recovery of RNA and DNA from Soils and Sediments. Appl Environ Microbiol. 2001;67:4495–503.
15. Xiang X, Poli D, Degnan BM, Degnan SM. Ribosomal RNA-Depletion Provides an Efficient Method for Successful Dual RNA-Seq Expression Profiling of a Marine Sponge Holobiont. Mar Biotechnol. 2022;24:722–32.
16. Anderson AJ, Culver HR, Prieto TR, Martinez PJ, Sinha J, Bryant SJ, et al. Messenger RNA enrichment using synthetic oligo(T) click nucleic acids. Chem Commun. 2020;56:13987–90.
17. Chen L, Yang R, Kwan T, Tang C, Watt S, Zhang Y, et al. Paired rRNA-depleted and polyA-selected RNA sequencing data and supporting multi-omics data from human T cells. Sci Data. 2020;7:376.
18. Peano C, Pietrelli A, Consolandi C, Rossi E, Petiti L, Tagliabue L, et al. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. Microb Inform Exp. 2013;3:1.
19. Zymo Research. Zymo-Seq RiboFree Total RNA Library Kit. https://www.zymoresearch.com/products/zymo-seq-ribofree-total-rna-library-kit.
20. Hale B, Watts C, Conatser M, Brown E, Wijeratne AJ. Fine-scale characterization of the soybean rhizosphere microbiome via synthetic long reads and avidity sequencing. Environ Microbiome. 2024;19:46.
21. Nleye T, Sexton P, Gustafson K, Miller JM. Soybean Growth Stages. In: iGrow Soybeans: Best Management Practices for Soybean Production; 2013. p. 24-35
22. Lu T, Ke M, Lavoie M, Jin Y, Fan X, Zhang Z, et al. Rhizosphere microorganisms can influence the timing of plant flowering. Microbiome. 2018;6:231.
23. McPherson MR, Wang P, Marsh EL, Mitchell RB, Schachtman DP. Isolation and Analysis of Microbial Communities in Soil, Rhizosphere, and Roots in Perennial Grass Experiments. J Vis Exp. 2018;137:57932.
24. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. Nat Protoc. 2006;1:581–5.
25. Thorn CE, Bergesch C, Joyce A, Sambrano G, McDonnell K, Brennan F, et al. A robust, cost-effective method for DNA, RNA and protein co-extraction from soil, other complex microbiomes and pure cultures. Mol Ecol Resour. 2019;19:439–55.

26. Holmes A, Birse L, Jackson RW, Holden NJ. An optimized method for the extraction of bacterial mRNA from plant roots infected with Escherichia coli O157:H7. Front Microbiol. 2014;5:286.
27. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–7.
28. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012;41:D590–6.
29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
30. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience. 2019;8:giz100.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
32. Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and Ionic Strength on the Spectrophotometric Assessment of Nucleic Acid Purity. Biotechniques. 1997;22:474–81.
33. Joseph S. Ribosomal RNAs and Protein Synthesis. In: Encyclopedia of Cell Biology; 2016. p. 384–95.
34. Aranda PS, LaJoie DM, Jorcyk CL. Bleach gel: A simple agarose gel for analyzing RNA quality. Electrophoresis. 2012;33:366–9.
35. Natsidis P, Schiffer PH, Salvador-Martínez I, Telford MJ. Computational discovery of hidden breaks in 28S ribosomal RNAs across eukaryotes and consequences for RNA Integrity Numbers. Sci Rep. 2019;9:19477.
36. Goring CAI, Bartholomew WV. Adsorption of Mononucleotides, Nucleic Acids, and Nucleoproteins by Clays. Soil Sci. 1952;74:149–64.
37. Yu WH, Li N, Tong DS, Zhou CH, Lin CX, Xu CY. Adsorption of proteins and nucleic acids on clay minerals and their interactions: A review. Appl Clay Sci. 2013;80–81:443–52.
38. USDA NRCS. Official Series Description - COLLINS Series. 2018. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/COLLINS.html.
39. Poursalavati A, Javaran VJ, Laforest-Lapointe I, Fall ML. Soil Metatranscriptomics: An Improved RNA Extraction Method Toward Functional Analysis Using Nanopore Direct RNA Sequencing. Phytobiomes J. 2023;7:42–54.
40. Novinscak A, Filion M. Effect of Soil Clay Content on RNA Isolation and on Detection and Quantification of Bacterial Gene Transcripts in Soil by Quantitative Reverse Transcription-PCR. Appl Environ Microbiol. 2011;77:6249–52.
41. Wang Y, Shimodaira J, Miyasaka T, Morimoto S, Oomori T, Ogawa N, et al. Detection of bphAa Gene Expression of Rhodococcus sp Strain RHA1 in Soil Using a New Method of RNA Preparation from Soil. Biosci Biotechnol Biochem. 2008;72:694–701.
42. Hopkins TR. Physical and Chemical Cell Disruption for the Recovery of Intracellular Proteins. In: Purification and Analysis of Recombinant Proteins. Marcel Dekker, Inc.; 1991. p. 57–83.
43. Greco M, Sáez CA, Brown MT, Bitonti MB. A Simple and Effective Method for High Quality Co-Extraction of Genomic DNA and Total RNA from Low Biomass Ectocarpus siliculosus, the Model Brown Alga. PLoS ONE. 2014;9:e96470.
44. Mulhardt C, Beese EW. Fundamental Methods. In: Molecular Biology and Genomics. Academic Press; 2007. p. 11–36.
45. Iowa State University. McNay Memorial Research and Demonstration Farm. Research and Demonstration Farms. https://farms.cals.iastate.edu/project/mcnay-memorial-research-and-demonstration-farm.
46. USDA NRCS. Official Series Description - GRUNDY Series. 2009. https://soilseries.sc.egov.usda.gov/OSD_Docs/G/Grundy.html.
47. USDA NRCS. Official Series Description - HAIG Series. 2015. https://soilseries.sc.egov.usda.gov/OSD_Docs/H/HAIG.html.
48. Gray J, Chandry PS, Kaur M, Kocharunchitt C, Fanning S, Bowman JP, et al. Colonisation dynamics of Listeria monocytogenes strains isolated from food production environments. Sci Rep. 2021;11:12195.
49. Overbey EG, Ng TT, Catini P, Griggs LM, Stewart P, Tkalcic S, et al. Transcriptomes of an Array of Chicken Ovary, Intestinal, and Immune Cells and Tissues. Front Genet. 2021;12:664424.
50. Willemse D, Moodley C, Mehra S, Kaushal D. Transcriptional Response of Mycobacterium tuberculosis to Cigarette Smoke Condensate. Front Microbiol. 2021;12:744800.
51. Nature Research, Zymo Research. A probe-free rRNA depletion method expands RNA-seq use. https://www.nature.com/articles/d42473-020-00248-z.
52. Herbert ZT, Kershner JP, Butty VL, Thimmapuram J, Choudhari S, Alekseyev YO, et al. Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics. 2018;19:199.
53. Haile S, Corbett RD, Bilobram S, Mungall K, Grande BM, Kirk H, et al. Evaluation of protocols for rRNA depletion-based RNA sequencing of nanogram inputs of mammalian total RNA. PLoS ONE. 2019;14: e0224578.
54. Prawer YDJ, Gleeson J, De Paoli-Iseppi R, Clark MB. Pervasive effects of RNA degradation on Nanopore direct RNA sequencing. NAR Genomics Bioinforma. 2023;5:lqad060.
55. New England Biolabs. RNA Depletion & mRNA Enrichment. https://www.neb.com/en-us/products/next-generation-sequencing-library-preparation/rna-depletion-and-mrna-enrichment/rna-depletion-and-mrna-enrichment.
56. Diep P, Mahadevan R, Yakunin AF. A microplate screen to estimate metal-binding affinities of metalloproteins. Anal Biochem. 2020;609:113836.
57. Akhtar AA, Turner DPj. The role of bacterial ATP-binding cassette (ABC) transporters in pathogenesis and virulence: Therapeutic and vaccine potential. Microb Pathog. 2022;171:105734.
58. Rodionov DA. Comparative Genomic Reconstruction of Transcriptional Regulatory Networks in Bacteria. Chem Rev. 2007;107:3467–97.
59. Bai X, Li Y, Jing X, Zhao X, Zhao P. Response mechanisms of bacterial communities and nitrogen cycle functional genes in millet rhizosphere soil to chromium stress. Front Microbiol. 2023;14:1116535.
60. Yeom J, Groisman EA. Reduced ATP-dependent proteolysis of functional proteins during nutrient limitation speeds the return of microbes to a growth state. Sci Signal. 2021;14:eabc4235.
61. Michalik S, Liebeke M, Zühlke D, Lalk M, Bernhardt J, Gerth U, et al. Proteolysis during long-term glucose starvation in *Staphylococcus aureus* COL. Proteomics. 2009;9:4468–77.
62. Silva GDC, Kitano IT, Ribeiro IADF, Lacava PT. The Potential Use of Actinomycetes as Microbial Inoculants and Biopesticides in Agriculture. Front Soil Sci. 2022;2:833181.
63. Sah S, Krishnani S, Singh R. Pseudomonas mediated nutritional and growth promotional activities for sustainable food security. Curr Res Microb Sci. 2021;2:100084.
64. Gonçalves OS, Fernandes AS, Tupy SM, Ferreira TG, Almeida LN, Creevey CJ, et al. Insights into plant interactions and the biogeochemical role of the globally widespread Acidobacteriota phylum. Soil Biol Biochem. 2024;192:109369.
65. Su Y, Wang J, Gao W, Wang R, Yang W, Zhang H, et al. Dynamic metabolites: A bridge between plants and microbes. Sci Total Environ. 2023;899:165612.
66. Whiteley LE, Rieckh G, Diggle FL, Alaga ZM, Nachbaur EH, Nachbaur WT, et al. Use of commercial or indigenous yeast impacts the S. cerevisiae transcriptome during wine fermentation. Microbiol Spectr. 2024;12:e0119424.
67. Ting MKY, Gao Y, Barahimipour R, Ghandour R, Liu J, Martinez-Seidel F, et al. Optimization of ribosome profiling in plants including structural analysis of rRNA fragments. Plant Methods. 2024;20:143.
68. UniProtKB statistics. https://www.uniprot.org/uniprotkb/statistics. Accessed 8 Oct 2024.

Publisher's Note