Published November 7, 2013

The 11th Dahlia Greidinger Memorial Symposium:
Advanced Methods for Investigating Nutrient Dynamics in Soils and Ecosystems

The Potential of Metagenomic Approaches for Understanding Soil Microbial Processes

David D. Myrold*
Lydia H. Zeglin
Dep. of Crop and Soil Science
Oregon State Univ.
Corvallis, OR 97331

Janet K. Jansson
Earth Sciences Division
Lawrence Berkeley National Lab.
Berkeley, CA 94720

Technological advances in sequencing technologies and bioinformatics analysis tools now enable the generation of a metagenome from soil, although the ultimate goal of obtaining the entire complement of all genes of all organisms in a given sample of soil still lies in the future. The rich information obtained from a soil metagenome will undoubtedly provide new insights into the taxonomic and functional diversity of soil microorganisms; the question is whether it will also yield greater understanding of how C, N, and other nutrients cycle in soil. The purpose of this review is to describe the steps involved in producing a soil metagenome, including some of the potential pitfalls associated with its production and annotation. Possible solutions to some of these challenges are presented. Selected examples from published soil metagenomic studies are discussed, with an emphasis on clues that they have provided about biogeochemical cycling.

Abbreviations: bp, base pairs; DNA, deoxyribonucleic acid; mRNA, messenger ribonucleic acid; RNA, ribonucleic acid; rRNA, ribosomal ribonucleic acid.

Soil microbial communities are known to be incredibly diverse, harboring tens of thousands of species of bacteria and thousands of species of fungi in a gram of soil. This knowledge is based primarily on recent advances in DNA sequencing, which have made it possible to generate millions of sequence reads quickly and economically. The initial application of this high-throughput sequencing technology explored the taxonomic diversity and composition of soil microbial communities using a polymerase chain reaction (PCR)-based approach that focused on phylogenetically informative ribosomal genes (Buée et al., 2009; Roesch et al., 2007). This has become known as *pyrotagged sequencing* and, with the incorporation of barcoded primers or tags (Hamady et al., 2008), has rapidly became the standard approach for describing the taxonomic composition of soil microbial communities—the soil "microbiome". The pyrotagging approach has subsequently been extended to include targeted functional genes, such as *nif*H, *amo*A, etc. (Mao et al., 2011). Targeted pyrosequencing has the advantage of being able to focus on a gene, or a few genes, of specific interest; however, for some applications, such as the interaction among multiple community members or their collective response to environmental perturbations, a more comprehensive and complete inventory of microbial genes is desired—a "metagenome".

Initial soil metagenomic studies relied on constructing libraries (e.g., plasmid, bacterial artificial chromosome [BAC], cosmid, fosmid) that were then sequenced with the intent of finding genes that encoded for products of interest, such as antimicrobials or enzymes (Daniel, 2005). The first attempt to generate a comprehensive soil metagenome was reported by Tringe et al. (2005), who constructed it by Sanger sequencing of random lengths of DNA that were

Soil Sci. Soc. Am. J. 78:3–10
doi:10.2136/sssaj2013.07.0287dgs
Received 17 July 2013.
*Corresponding author (david.myrold@oregonstate.edu).

cloned into a virus (i.e., a phage library). Although the depth of sequencing was insufficient to achieve significant assembly of the DNA sequences into larger contiguous segments, or *contigs*, the individual reads were useful in comparing the gene content of the Minnesota farm soil metagenome with those obtained from other environments. Subsequent efforts to produce soil metagenomes have primarily used a shotgun sequencing approach that eliminates the need for making libraries and directly sequences the extracted DNA (Table 1).

Most soil shotgun metagenomes have been obtained using the 454 GS FLX platform with Titanium chemistry, which generates reads of 400 to 500 bp in length. These studies have generated 100-fold differences in the amount of sequence generated, with 0.1 to 1.0 million reads per soil sample being typical and a maximum metagenome size of about 0.5 Gbp. Assembly was attempted in about one-third of these metagenomic studies with some success, although most contigs were relatively short, <1000 bp (Delmont et al., 2012b; Kanokratana et al., 2011; Yergeau et al., 2010). A few studies have used the Illumina sequencing platform for generating shotgun metagenomes from soils (Fierer et al., 2012b; Mackelprang et al., 2011), and one has combined both sequencing platforms (Uroz et al., 2013). The soil metagenomes produced using the Illumina system generally produced many more sequences (0.4–29 million reads per sample) and a maximum metagenome size of about 4.0 Gbp. Although a relatively small fraction of the short reads could be assembled into larger contigs, sufficient assembly was obtained in a permafrost soil to produce a draft genome of the dominant methanogen (Mackelprang et al., 2011).

A major goal of soil metagenomic studies is to identify the functional potential of the complex microbial communities, whether using individual reads or assembled contigs. As might be expected, greater success in assigning functions has been obtained with the longer reads generated with 454 sequencing: 20 to 60% assignment depending on the databases used (Table 1). In contrast, only 10 to 25% of the shorter Illumina reads have been successfully assigned, although the 10-fold greater number of sequences resulted in more total functional gene assignments.

Whether using individual reads or assembled contigs, the studies to date have been effective in understanding the functional potential of microbial communities in soil and in distinguishing among soils and treatments. For practical reasons, such as cost and computational constraints, many studies, particularly the earlier ones, have not had true biological replication.

Of course, many more soil metagenomic projects are in process: as of May 2013, 48 soil metagenomes were registered at www.genomesonline.org. Most of these are not yet published and not all are shotgun metagenomes. As part of our research on Mollisol soils of the Great Plains, several metagenomes have been obtained, including some in excess of 500 million Illumina reads (unpublished data). Assembly of such large datasets is challenging, however.

METAGENOMIC INSIGHTS INTO BIOGEOCHEMICAL CYCLES

One goal of soil metagenomic studies is to gain insights into soil C, N, P, S, and other elemental cycles. Mackelprang et al. (2011) compared the response to thawing of the metagenome of microbial communities in the active (seasonally frozen) and permafrost layers of an Alaskan Gelisol. Thawing resulted in shifts in both microbial community structure and function. The increased rate of CH_4 consumption was positively correlated with an increase in particulate methane monooxygenase genes and the 16S rRNA genes of type II, but not type I, methanotrophs. No change in functional genes associated with methanogenesis was observed. Thawing of the permafrost core also brought about an increase in genes associated with denitrification. Another interesting observation is that thawing resulted in changed abundances of functional genes of the permafrost in the two replicate cores, but this was not matched by a convergence of 16S rRNA gene composition, which suggests that microbial communities of different structure may have similar functions.

Andreote et al. (2012) generated metagenomes from submerged sediment of four mangrove forests in Brazil. Based on this metagenomic information, they constructed metabolic pathways associated with C, N, and S cycling. Not surprisingly, pathways for anaerobic metabolic processes (e.g., NO_3 reduction and SO_4 reduction) were dominant.

Fierer et al. (2012a) performed a comparative metagenomic study to focus on responses of soil metagenomes to N fertilization using two long-term experiments in the Upper Midwest of the United States. Several consistent responses to N fertilization were observed at the two sites, including: increased relative abundance of genes associated with respiration, protein metabolism, and nucleic acid metabolism; and decreased abundances of genes associated with urea decomposition and tricarboxylate transporters. They did not specifically report on changes in genes associated with classic N cycling processes of N_2 fixation, nitrification, or denitrification. Based on assumptions that assign bacterial phyla into copiotrophic and oligotrophic life histories, they suggested that N fertilization causes a shift toward more copiotrophic phyla.

CHALLENGES OF METAGENOMICS

The initial call by Vogel et al. (2009b) to sequence the soil metagenome resulted in a number of commentaries, replies, and reviews about the efficacy of doing so (Baveye, 2009; Morales and Holben, 2011; Myrold and Nannipieri, 2013; Singh et al., 2009; Vogel et al., 2009a), but, as reported above, more than a dozen soil metagenomes have been published (Table 1) and more are in process. The challenges have not, however, disappeared (Thomas et al., 2012). Sequencing methods have improved and costs have diminished, but there at least three areas of opportunity for advancing metagenomics.

Table 1. Summary of soil metagenome studies.

Location	Site description	Soil type	Experimental design	Biological replication	Sequencing platform	Sequencing depth	Assembly	Functional assignment†	Comments	Reference
								%		
Nunavut, Canada	tundra	permafrost	two depths: active layer and permafrost	no	454 GS FLX Titanium	853 Mbp (0.35–0.99 million reads/sample)	Phrap assembler (140 Mbp, 134,000 contigs)	not given	multiple displacement amplification (MDA) was needed to generate sufficient DNA; detected genes for methanogenesis	Yergeau et al. (2010)
Rothamsted, UK	Park Grass experiment (permanent grassland)	silty clay loam (Chromic Luvisol)	comparison of direct and indirect DNA extraction methods	no	454 GS FLX Titanium	~1 million reads	none	not given	extraction methods gave similar functional gene profiles	Delmont et al. (2011)
Pru Toh Daeng, Thailand	swamp forest	peat	pooled sample	no	454 GS FLX	45.9 Mbp (0.18 million reads)	GS De Novo Assembler (13.4 Mbp, 54,000 contigs averaging 248 bp)	SEED: 53	searched mainly for polysaccharide degrading genes	Kanokratana et al. (2011)
Alaska	Hess Creek (black spruce forest)	permafrost	two depths: active layer and permafrost; before and 2 and 7 d after thawing	yes (n = 2)	IlluminaGAII (2 × 113 bp)	39.8 Gbp (176 million reads)	Velvet assembler (9.6 Mbp, 3758 contigs >1 kb)	KEGG: 11	used emulsion polymerase chain reaction to generate sufficient DNA; draft genome of dominant methanogen obtained	Mackelprang et al. (2011)
New Hampshire	Harvard Forest (northern hardwood forest)	sandy loam (Typic Dystrochrept)	single composite from two soil cores (0–10 cm)	no	454 GS FLX Titanium	748 Mbp (1.4 million reads)			also companion metatranscriptome	Stewart et al. (2011)
São Paulo, Brazil	mangrove forest	submerged sediment	samples from four sites, one of which was impacted by oil contamination	no	454 GS FLX Titanium	215 Mbp (~0.25 million reads/site)	none	COG: 60; KEGG: 30	metabolic reconstruction of C, N, and S cycles	Andreote et al. (2012)
Rothamsted, UK	Park Grass experiment (permanent grassland)	silty clay loam (Chromic Luvisol)	13 samples differing by date, depth, and DNA extraction method	no	454 GS FLX Titanium	4.8 Gbp (12.5 million reads)	Newbler assembler (15.2 Mbp, 267,000 contigs)	SEED: 56	extraction method showed more variation than date or depth of sampling	Delmont et al. (2012a)
Michigan and Minnesota	Kellogg Biological Station (cropped fields) and Cedar Creek Ecosystem Science Reserve (successional grasslands)	sandy loam (Hapludalf) and sand (Udipsamment)	low, medium, and high N addition at each site	yes (n = 3)	454 GS FLX Titanium	518 Mbp (1.35 million reads)	none	SEED: 25–35	shift in functional capacity with N input; significant correlation of metagenome functional and phylogenetic composition	Fierer et al. (2012a)
Worldwide	cold and hot deserts, selected nondesert biomes	various	16 total soils	no	Illumina GAIIx (2 × 100 bp)	6.2 Gbp total (0.39–1.9 million reads per soil)	none	SEED: 13–23	functional composition correlated with site characteristics	Fierer et al. (2012b)
Lucknow, India	dumpsite	not specified	three soils with a gradient of hexachlorocyclohexane contamination	no	454 GS FLX Titanium	1.2 Gbp (1.1–1.2 million reads per soil)	none	not given	draft genome of *Chromohalobacter salaxigenes*	Sangwan et al. (2012)
Nevada,	Nevada Desert FACE Facility (shrubland)	loamy sand (Aridisol)	four pooled samples (ambient and elevated CO_2 plots, with two locations [creosote bush and interspace])	no	454 GS FLX Titanium	724 Mbp total (0.31–0.68 million reads per sample)	none	SEED: 36–40	functional genes less discriminating than 16S rRNA genes for the effect of elevated CO_2	Steven et al. (2012)
Nunavut, Canada	tundra	permafrost	four pooled samples (control and three times for treated biopiles of oil-contaminated soil)	no	454 GS FLX Titanium	463 Mbp total (0.11–0.46 reads per sample)	none	not applicable	used BLAST to focus on hydrocarbon-degrading genes (found to be higher in treated than control)	Yergeau et al. (2012)
Breuil-Chenue, France	Norway spruce plantation	Alocrisol soil	closely spaced soil cores separated into organic and mineral horizons	yes (n = 3)	454 GS FLX Titanium Illumina HiSeq 2000 (1 × 75 bp)	1.9 Gbp total of 454 (0.41–0.62 million reads per sample); 11.9 Gbp total of Illumina (23–29 million reads per sample)	unspecified assembler of Illumina data (0.77 Mbp, 3492 contigs)	SEED: 36–60 (454 data); ~25 (Illumina data)	differences in functional subsystems between organic and mineral horizons	Uroz et al. (2013)

† COG, Clusters of Orthologous Groups of proteins; KEGG, Kyoto Encyclopedia of Genes and Genomes; SEED, SEED subsystem hierarchy.

Obtaining a Representative Sample

Collecting a sample that is representative of a soil is a long-standing challenge for any type of soil study because soils are by nature heterogeneous and vary both spatially and temporally. Often stratified sampling, taking composite samples, and the use of replicated designs can help in obtaining representative samples appropriate for addressing the objective of an experiment, at least when the cost of analyses is not a major limitation (Prosser, 2010; Wollum, 1994). Although some level of replication in soil metagenomic studies is now financially feasible, other issues remain (Knight et al., 2012; Lombard et al., 2011). Working with a soil from a single location, Delmont et al. (2012a) found that DNA extraction protocols displayed greater variation in metagenomic composition than either soil depth or season of sampling, which was perhaps a surprising result but is consistent with their earlier observation using other molecular methods (Delmont et al., 2011). Because current metagenomic sequencing typically requires a few hundred nanograms of DNA, soils that yield small amounts of DNA, such as permafrost soils, may require additional DNA amplification, a step that may introduce some bias (Mackelprang et al., 2011; Yergeau et al., 2010). The difficulty in adopting standard methods in soil microbiology is not unique to DNA extraction (Philippot et al., 2012)—and a perfect extraction method is unlikely to be developed. Nevertheless, it is likely that the community will select one as a common protocol, as has been done by the Earth Microbiome Project (http://press.igsb.anl.gov/earthmicrobiome/).

Metagenome Assembly

The lack of efficient and effective strategies for assembling and annotating shotgun metagenomic data is a major challenge facing bioinformaticians. The sheer magnitude of data bogs down even the fastest, largest computers using standard assembly algorithms. The result is that only a small fraction of soil metagenomic data can be assembled, usually into relatively small contigs (Table 1); however, a number of metagenome-specific assemblers have been developed in an attempt to overcome this hurdle (Scholz et al., 2012; Segata et al., 2013). Such new approaches to de novo assembly often take the approach of normalizing and partitioning the sequence reads, which leads to greater efficiency and generally more complete assembly (Dröge and McHardy, 2012; Howe et al., 2012; Jones et al., 2012). Although larger amounts of sequencing should increase the number of contigs that can be assembled, more reads did not increase contig length in one study (Delmont et al., 2012a).

It should be noted that even if significant assembly of sequences into contigs is not achieved, insights can be gained by comparing individual reads with gene sequences archived in annotated databases. The results of annotation tools, such as MG-RAST and IMG-ER (Markowitz et al., 2009; Meyer et al., 2008), are dependent on database matching parameters as well as the content and quality of the sequence databases. Current sequence databases have a number of limitations that are important to recognize. For example, gene sequences may not be annotated correctly and genes from soil microorganisms are often underrepresented in these databases. Attempts to address the latter shortcoming are the Genomic Encyclopedia of Bacteria and Archaea and 1000 Fungal Genomes projects being undertaken by the Department of Energy's Joint Genome Institute (http://genome.jgi.doe.gov/programs/bacteria-archaea/GEBA.jsf and http://genome.jgi.doe.gov/programs/fungi/1000fungalgenomes.jsf), which should greatly expand coverage of the microbial world.

Metagenome Interpretation

Even if well-assembled, unbiased soil metagenomes are obtained, there are challenges in how these data will be integrated and synthesized into a coherent representation of the potential functioning of the microbial community. A variety of analytical methods have been used, and are being developed, to provide insights into these complex systems and their emergent properties, such as clustering, ordination, and artificial neural networks, to name a few (Cloots and Marchal, 2011; Faust and Raes, 2012; Gianoulis et al., 2009; Muller et al., 2013; Segata et al., 2013). Two recent examples for soil are the use of random matrix theory to infer the interaction networks of microbial communities and how these networks changed in response to elevated CO_2 (Zhou et al., 2011) and the co-occurrence of microbial taxa in relation to their life-history strategies (Barberan et al., 2012). Another option would be to extend metabolic modeling of individual organisms (Henry et al., 2010) to soil microbial communities, as has been done for simpler systems (Stolyar et al., 2007; Zomorrodi and Maranas, 2012). Finally, it may be possible to incorporate metagenomic data into trait-based models of soil biogeochemical cycles (Allison, 2012; Bouskill et al., 2012).

EXTRAPOLATING FROM METAGENOMES TO FUNCTION

A common goal of metagenomic studies has been to link the structure, or composition, of the microbial community with its function. Metagenomic data provide insights into the metabolic, or functional, potential of the microbial community, but other approaches are needed to assess and strengthen the connections with actual microbial activity. This may be especially important in the soil environment, where many organisms detected on the basis of DNA may be quiescent or not growing (Bottomley et al., 2012; Buerger et al., 2012; Lennon and Jones, 2011). The development of additional "omics" methods along the canonical DNA-RNA-protein continuum (Fig. 1) is one approach to explore the functional activity of microbial communities (Muller et al., 2013).

Metatranscriptomics

It is generally assumed that gene expression (the transcription of DNA into RNA) is indicative of microbial activity and reflects the response of microorganisms to environmental cues. The cellular rRNA content is roughly proportional to the growth rate, whereas the amount of the rRNA gene is relatively constant,

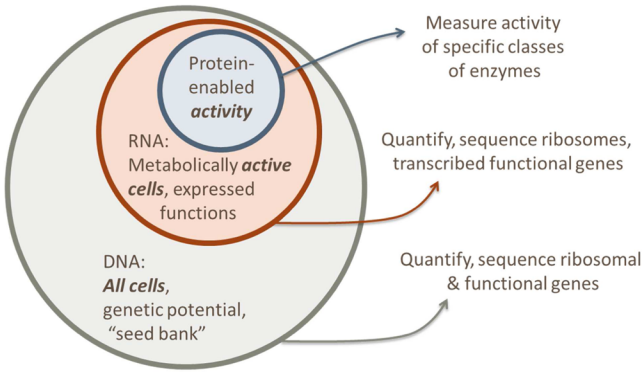

Fig. 1. The canonical continuum of transcription of DNA to RNA and translation of RNA to protein.

leading to the use of the rRNA/DNA ratio as an indicator of relative microbial activity (Blazewicz et al., 2013; DeAngelis et al., 2010; Muttray et al., 2001). Several studies have also found that the composition of microbial communities based on RNA differs from that based on DNA and concluded that the active microbial community is just a subset of the potentially active microbial community (Anderson and Parkin, 2007; Baldrian et al., 2012; Griffiths et al., 2000). Such differences between genes and transcripts have also been observed when metagenomes and metatranscriptomes from marine or soil environments have been compared (Frias-Lopez et al., 2008; Gilbert et al., 2008; Urich et al., 2008).

There are just a few published soil metatranscriptomic studies (Table 2). Two have taken advantage of the poly-A tails of eukaryotic mRNA to selectively isolate this fraction for sequencing (Bailly et al., 2007; Damon et al., 2012). Shotgun metatranscriptomics have been used in two studies, an approach sometimes known as RNA-Seq (Carvalhais et al., 2012; Croucher and Thomson, 2010). The first study (Urich et al., 2008) did not use any type of enrichment procedure to maximize the number of mRNA sequences, which made up about one-quarter of the total. A customized rRNA hybridization technique has been developed to enrich the fraction of mRNA sequences significantly (Stewart et al., 2011). Annotation of functional gene transcripts was about as successful as it is for gene annotation (Tables 1 and 2). The utility of soil metatranscriptomics has not yet been fully explored but has provided insights about microbial functions in other complex microbial systems (e.g., Poretsky et al., 2010).

Metaproteomics

Microbial functions are mediated by proteins, thus the measurement of the microbial proteins should be the most direct "omics" estimator of the potential activity of the microbial community (Fig. 1). As with metagenomics and metatranscriptomics, there has been a steady evolution of the methodology for the extraction and analysis of proteins from soils (Bastida et al., 2009; Hettich et al., 2012; Siggins et al., 2012). Nonetheless, the extraction of proteins lags behind that of nucleic acids, in part because of the strong interactions between proteins and other organic molecules and inorganic minerals in soils (Adamczyk et al.,

Table 2. Summary of soil metatranscriptomes.

Location	Site description	Soil type	Experimental design	Biological replication	RNA preparation	Sequencing platform	Sequencing depth	Functional assignment†	Comments	Reference
								%		
Southwestern France	*Pinus pinaster* plantation on coastal sand dune	nutrient poor sand	single composite of 27 samples (0–20 cm)	no	polyT capture of eukaryotic mRNA, cloned into plasmid library	Sanger	119 clones sequenced	COG: 68	first metatranscriptome for soil.	Bailly et al. (2007)
Am Rotböll, Germany	lawn	sandy	single composite of three samples (0–10 cm)	no	no mRNA enrichment	454 GS20	260,000 sequences (75% rRNA)	SEED: 33	first shotgun metatranscriptome from soil; noted database limitations for identifying transcripts based on sequenced genomes	Urich et al. (2008)
New Hampshire	Harvard Forest (northern hardwood forest)	sandy loam (Typic Dystrochrept)	single composite from two soil cores (0–10 cm)	no	mRNA enrichment with custom rRNA subtraction hybridization	454 GS FLX Titanium	1.2 million sequences, 458 Mbp (17% rRNA)	KEGG: 45	observed lower relative abundances of transcripts than genes for motility and transport; higher abundances of transcripts than genes for carbohydrate and nucleotide metabolism; also companion metagenome	Stewart et al. (2011)
Central France	Breuil-Chenue forest (*Picea abies* and *Fagus sylvatica* plantations)	sandy clay	single composites from 14–16 soil cores in each stand (3–7 cm)	no	polyT capture of eukaryotic mRNA, cloned into plasmid library	Sanger	8606 (*Pinus*), 7905 (*Fagus*)	Blast2GO: 32–39	transcripts for enzymes that degrade plant constituent comprised just 0.5–0.8% of all functional transcripts	Damon et al. (2012)

† COG, Clusters of Orthologous Groups of proteins; KEGG, Kyoto Encyclopedia of Genes and Genomes; SEED, SEED subsystem hierarchy.

2008; Kleber et al., 2007). Although advances in protein extraction methods have been made (Chourey et al., 2010; Kanerva et al., 2013; Keiblinger et al., 2012; Taylor and Williams, 2010), a standard method is not currently available, and it is possible that soil-specific modifications will always be needed.

To date, soil proteomic studies have typically identified a relatively small number of proteins based on gel isolation methods (Table 3). The development of shotgun proteomics using more sophisticated two-dimensional liquid chromatography separation methods and subsequent tandem mass spectrometry analysis have increased that number to hundreds to thousands (Chourey et al., 2010; Keiblinger et al., 2012). This is far fewer identified molecules than for metagenomic or metatranscriptomic methods and may be a limitation of metaproteomic analyses, which identify only the most abundant proteins, those that are often associated with housekeeping activities rather than enzymes associated with specific biogeochemical processes.

CONCLUSIONS

This is clearly an exciting time for soil microbial ecology as advances in sequencing technologies generate huge amounts of information about microbial communities. At present, the metagenomics of soils is transitioning from studies about how to do it to those that focus on what can be done with it. Soil metagenomics has already provided insight into the long-standing questions of "who's there?" and is making inroads into the question of "what are they doing?" Progress will require further developments related to metagenomics (e.g., data reduction and analysis, improvements in functional gene annotation, and database curation) combined with knowledge about the spatial and temporal variability of the soil habitat and its influence on microbial activities. This, as well as the integration with other "omics" methods, promises to enhance our understanding about the functioning of soil ecosystems.

ACKNOWLEDGMENTS

We thank Peter Bottomley for reviewing an early draft of this manuscript. This material is based on work supported by the National Science Foundation under Grant no. 1051481 and by the USDOE Contract no. DE-SC0004953. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the USDOE.

Table 3. Summary of soil metaproteomes.

Location	Site description	Soil type	Experimental design	Biological replication	Extraction method	Protein analysis approach†	Comments	Reference
Waldstein, Germany	spruce forest, 145 yr old	Podzol	mineral (heavy, >1.6 g/cm^3) fraction isolated	no	proteins released by dissolving minerals i n 10% HF	SDS-PAGE followed by LC-MS/MS of tryptic digest	identified eight enzymes involved in C cycling	Schulze et al. (2005)
Leipzig, Germany	laboratory study	compost soil	column with and without addition of 2,4-D-degrading bacteria	no	NaOH and phenol	SDS-PAGE followed by LC-MS/MS of tryptic digest	identified four proteins	Benndorf et al. (2007)
Washington and California	laboratory study		soils with and without addition of a known bacterial species; four extraction protocols	no	SDS and TCA (and others for comparison)	2D LC-MS/MS of tryptic digest	shotgun proteomics identified 145 to 925 proteins; for unamended soil, 333 nonredundant proteins were identified	Chourey et al. (2010)
Mississippi	agricultural soil	Paleudult	0–10 cm, control or with addition of toluene or glucose	no	indirect with phenol extraction	SDS-PAGE followed by MALDI-TOF/TOF	identified 16 proteins from toluene-amended soils, eight from glucose-amended soils	Williams et al. (2010)
Southeast, central, and southwest China	agricultural soils			no	SDS and phenol	2D PAGE followed by MALDI-TOF/TOF	just rice rhizosphere analyzed; 122 proteins identified	Wang et al. (2011)
Henan, China	*Rehmannia glutinosa* cropped field		fallow and 1- and 2-yr monoculture	no	SDS and phenol	2D PAGE followed by MALDI-TOF/TOF	identified 103 proteins, with 33 differentially expressed among the cultural treatments	Wu et al. (2011)
Schottenwald, Austria	beech forest	Dystric Cambisol	0–10 cm, four extraction procedures, also a potting soil mix	no	SDS, NaOH–phenol, SDS–phenol	2D LC-MS/MS of tryptic digest	shotgun proteomics identified 226–494 proteins in forest soil and 80–237 in potting soil, depending on extraction method, with just 59 or seven in common	Keiblinger et al. (2012)

† LC-MS/MS, liquid chromatography–tandem mass spectrometry; MALDI-TOF/TOF, matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate.

REFERENCES

Adamczyk, B., V. Kitunen, and A. Smolander. 2008. Protein precipitation by tannins in soil organic horizon and vegetation in relation to tree species. Biol. Fertil. Soils 45:55–64. doi:10.1007/s00374-008-0308-0

Allison, S.D. 2012. A trait-based approach for modelling microbial litter decomposition. Ecol. Lett. 15:1058–1070. doi:10.1111/j.1461-0248.2012.01807.x

Anderson, I.C., and P.I. Parkin. 2007. Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. J. Microbiol. Methods 68:248–253. doi:10.1016/j.mimet.2006.08.005

Andreote, F.D., D.J. Jimenez, D. Chaves, A.C.F. Dias, D.M. Luvizotto, F. Dini-Andreote, et al. 2012. The microbiome of Brazilian mangrove sediments as revealed by metagenomics. PLoS ONE 7:E38600. doi:10.1371/journal.pone.0038600

Bailly, J., L. Fraissinet-Tachet, M.C. Verner, J.C. Debaud, M. Lemaire, M. Wesolowski-Louvel, and R. Marmeisse. 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME J. 1:632–642. doi:10.1038/ismej.2007.68

Baldrian, P., M. Kolarik, M. Stursova, J. Kopecky, V. Valaskova, T. Vetrovsky, et al. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J. 6:248–258. doi:10.1038/ismej.2011.95

Barberan, A., S.T. Bates, E.O. Casamayor, and N. Fierer. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 6:343–351. doi:10.1038/ismej.2011.119

Bastida, F., J.L. Moreno, C. Nicolas, T. Hernandez, and C. Garcia. 2009. Soil metaproteomics: A review of an emerging environmental science. Significance, methodology and perspectives. Eur. J. Soil Sci. 60:845–859. doi:10.1111/j.1365-2389.2009.01184.x

Baveye, P.C. 2009. To sequence or not to sequence the whole-soil metagenome? Nat. Rev. Microbiol. 7:756. doi:10.1038/nrmicro2119-c2

Benndorf, D., G.U. Balcke, H. Harms, and M. von Bergen. 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME J. 1:224–234. doi:10.1038/ismej.2007.39

Blazewicz, S.J., R.L. Barnard, R.A. Daly, and M.K. Firestone. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: Limitations and uses. ISME J. (in press). doi:10.1038/ismej.2013.102

Bottomley, P., A.E. Taylor, and D.D. Myrold. 2012. A consideration of the relative contributions of different microbial subpopulations to the soil N cycle. Front. Terr. Microbiol. 3:373. doi:10.3389/fmicb.2012.00373

Bouskill, N., J. Tang, W.J. Riley, and E.L. Brodie. 2012. Trait-based representation of biological nitrification: Model development, testing, and predicted community composition. Front. Aquat. Microbiol. 3:364. doi:10.3389/fmicb.2012.00364

Buée, M., M. Reich, C. Murat, E. Morin, R.H. Nilsson, S. Uroz, and F. Martin. 2009. 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytol. 184:449–456. doi:10.1111/j.1469-8137.2009.03003.x

Buerger, S., A. Spoering, E. Gavrish, C. Leslin, L. Ling, and S.S. Epstein. 2012. Microbial scout hypothesis, stochastic exit from dormancy, and the nature of slow growers. Appl. Environ. Microbiol. 78:3221–3228. doi:10.1128/AEM.07307-11

Carvalhais, L.C., P.G. Dennis, G.W. Tyson, and P.M. Schenk. 2012. Application of metatranscriptomics to soil environments. J. Microbiol. Methods 91:246–251. doi:10.1016/j.mimet.2012.08.011

Chourey, K., J. Jansson, N. VerBerkmoes, M. Shah, K.L. Chavarria, L.M. Tom, et al. 2010. Direct cellular lysis/protein extraction protocol for soil metaproteomics. J. Proteome Res. 9:6615–6622. doi:10.1021/pr100787q

Cloots, L., and K. Marchal. 2011. Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. Curr. Opin. Microbiol. 14:599–607. doi:10.1016/j.mib.2011.09.003

Croucher, N.J., and N.R. Thomson. 2010. Studying bacterial transcriptomes using RNA-seq. Curr. Opin. Microbiol. 13:619–624. doi:10.1016/j.mib.2010.09.009

Damon, C., F. Lehembre, C. Oger-Desfeux, P. Luis, J. Ranger, L. Fraissinet-Tachet, and R. Marmeisse. 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoS ONE 7(1):e28967. doi:10.1371/journal.pone.0028967

Daniel, R. 2005. The metagenomics of soil. Nat. Rev. Microbiol. 3:470–478. doi:10.1038/nrmicro1160

DeAngelis, K.M., W.L. Silver, A.W. Thompson, and M.K. Firestone. 2010. Microbial communities acclimate to recurring changes in soil redox potential status. Environ. Microbiol. 12:3137–3149. doi:10.1111/j.1462-2920.2010.02286.x

Delmont, T.O., E. Prestat, K.P. Keegan, M. Faubladier, P. Robe, I.M. Clark, et al. 2012a. Structure, fluctuation and magnitude of a natural grassland soil metagenome. ISME J. 6:1677–1687. doi:10.1038/ismej.2011.197

Delmont, T.O., P. Robe, S. Cecillon, I.M. Clark, F. Constancias, P. Simonet, et al. 2011. Accessing the soil metagenome for studies of microbial diversity. Appl. Environ. Microbiol. 77:1315–1324. doi:10.1128/AEM.01526-10

Delmont, T.O., P. Simonet, and T.M. Vogel. 2012b. Describing microbial communities and performing global comparisons in the 'omic era. ISME J. 6:1625–1628. doi:10.1038/ismej.2012.55

Dröge, J., and A.C. McHardy. 2012. Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. Brief. Bioinform. 13:646–655. doi:10.1093/bib/bbs031

Faust, K., and J. Raes. 2012. Microbial interactions: From networks to models. Nat. Rev. Microbiol. 10:538–550. doi:10.1038/nrmicro2832

Fierer, N., C.L. Lauber, K.S. Ramirez, J. Zaneveld, M.A. Bradford, and R. Knight. 2012a. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J. 6:1007–1017. doi:10.1038/ismej.2011.159

Fierer, N., J.W. Leff, B.J. Adams, U.N. Nielsen, S.T. Bates, C.L. Lauber, et al. 2012b. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. 109:21390–21395. doi:10.1073/pnas.1215210110

<conf>Frias-Lopez, J., Y. Shi, G.W. Tyson, M.L. Coleman, S.C. Schuster, S.W. Chisholm, and E.F. DeLong. 2008. Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. 105:3805–3810. doi:10.1073/pnas.0708897105</conf>

Gianoulis, T.A., J. Raes, P.V. Patel, R. Bjornson, J.O. Korbel, I. Letunic, et al. 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. Proc. Natl. Acad. Sci. 106:1374–1379. doi:10.1073/pnas.0808022106

Gilbert, J.A., D. Field, Y. Huang, R. Edwards, W.Z. Li, P. Gilna, and I. Joint. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS ONE 3:E3042. doi:10.1371/journal.pone.0003042

Griffiths, R.I., A.S. Whiteley, A.G. O'Donnell, and M.J. Bailey. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl. Environ. Microbiol. 66:5488–5491. doi:10.1128/AEM.66.12.5488-5491.2000

Hamady, M., J.J. Walker, J.K. Harris, N.J. Gold, and R. Knight. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods 5:235–237. doi:10.1038/nmeth.1184

Henry, C.S., M. DeJongh, A.A. Best, P.M. Frybarger, B. Linsay, and R.L. Stevens. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28:977–982. doi:10.1038/nbt.1672

Hettich, R.L., R. Sharma, K. Chourey, and R.J. Giannone. 2012. Microbial metaproteomics: Identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. Curr. Opin. Microbiol. 15:373–380. doi:10.1016/j.mib.2012.04.008

Howe, A.C., J. Jansson, S.A. Malfatti, S.G. Tringe, J.M. Tiedje, and C.T. Brown. 2012. Assembling large, complex environmental metagenomes. arXiv:1212.2832v2 [q-bio.GN]. Cornell Univ., Ithaca, NY.

Jones, D.C., W.L. Ruzzo, X. Peng, and M.G. Katze. 2012. Compression of next-generation sequencing reads aided by highly efficient de novo assembly. Nucleic Acids Res. 40:e171. doi:10.1093/nar/gks754

Kanerva, S., A. Smolander, V. Kitunen, R.A. Ketola, and T. Kotiaho. 2013. Comparison of extractants and applicability of MALDI-TOF-MS in the analysis of soil proteinaceous material from different types of soil. Org. Geochem. 56:1–9. doi:10.1016/j.orggeochem.2012.12.003

Kanokratana, P., T. Uengwetwanit, U. Rattanachomsri, B. Bunterngsook, T. Nimchua, S. Tangphatsornruang, et al. 2011. Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. Microb. Ecol. 61:518–528. doi:10.1007/s00248-010-9766-7

Keiblinger, K.M., I.C. Wilhartitz, T. Schneider, B. Roschitzki, E. Schmid, L. Eberl, et al. 2012. Soil metaproteomics: Comparative evaluation of protein extraction protocols. Soil Biol. Biochem. 54:14–24. doi:10.1016/j.soilbio.2012.05.014

Kleber, M., P. Sollins, and R. Sutton. 2007. A conceptual model of organo-mineral interactions in soils: Self-assembly of organic molecular fragments into zonal structures on mineral surfaces. Biogeochemistry 85:9–24. doi:10.1007/s10533-007-9103-5

Knight, R., J. Jansson, D. Field, N. Fierer, N. Desai, J.A. Fuhrman, et al. 2012. Unlocking the potential of metagenomics through replicated experimental design. Nat. Biotechnol. 30:513–520. doi:10.1038/nbt.2235

Lennon, J.T., and S.E. Jones. 2011. Microbial seed banks: The ecological and evolutionary implications of dormancy. Nat. Rev. Microbiol. 9:119–130. doi:10.1038/nrmicro2504

Lombard, N., E. Prestat, J.D. van Elsas, and P. Simonet. 2011. Soil-specific

limitations for access and analysis of soil microbial communities by metagenomics. FEMS Microbiol. Ecol. 78:31–49. doi:10.1111/j.1574-6941.2011.01140.x

Mackelprang, R., M.P. Waldrop, K.M. DeAngelis, M.M. David, K.L. Chavarria, S.J. Blazewicz, et al. 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480:368–371. doi:10.1038/nature10576

Mao, Y., A.C. Yannarell, and R.I. Mackie. 2011. Changes in N-transforming archaea and bacteria in soil during the establishment of bioenergy crops. PLoS ONE 6(9):e24750. doi:10.1371/journal.pone.0024750

Markowitz, V.M., K. Mavromatis, N.N. Ivanova, I-M.A. Chen, K. Chu, and N.C. Kyrpides. 2009. IMG ER: A system for microbial genome annotation expert review and curation. Bioinformatics 25:2271–2278. doi:10.1093/bioinformatics/btp393

Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E.M. Glass, M. Kubal, et al. 2008. The metagenomics RAST server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinf. 9:386. doi:10.1186/1471-2105-9-386

Morales, S.E., and W.E. Holben. 2011. Linking bacterial identities and ecosystem processes: Can 'omic' analyses be more than the sum of their parts? FEMS Microbiol. Ecol. 75:2–16. doi:10.1111/j.1574-6941.2010.00938.x

Muller, E.E.L., E. Glaab, P. May, N. Vlassis, and P. Wilmes. 2013. Condensing the omics fog of microbial communities. Trends Microbiol. 21:325–333. doi:10.1016/j.tim.2013.04.009

Muttray, A.F., Z.T. Yu, and W.W. Mohn. 2001. Population dynamics and metabolic activity of *Pseudomonas abietaniphila* BKME-9 within pulp mill wastewater microbial communities assayed by competitive PCR and RT-PCR. FEMS Microbiol. Ecol. 38:21–31. doi:10.1111/j.1574-6941.2001.tb00878.x

Myrold, D.D., and P. Nannipieri. 2013. Classical techniques versus omics approaches. In: G. Pietramellara et al, editors, Omics in soil science. Horizon Sci. Press, Poole, UK. p. 179–187.

Philippot, L., K. Ritz, P. Pandard, S. Hallin, and F. Martin-Laurent. 2012. Standardisation of methods in soil microbiology: Progress and challenges. FEMS Microbiol. Ecol. 82:1–10. doi:10.1111/j.1574-6941.2012.01436.x

Poretsky, R.S., S.L. Sun, X.Z. Mou, and M.A. Moran. 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ. Microbiol. 12:616–627. doi:10.1111/j.1462-2920.2009.02102.x

Prosser, J.I. 2010. Replicate or lie. Environ. Microbiol. 12:1806–1810. doi:10.1111/j.1462-2920.2010.02201.x

Roesch, L.F., R.R. Fulthorpe, A. Riva, G. Casella, A.K.M. Hadwin, A.D. Kent, et al. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1:283–290. doi:10.1038/ismej.2007.53

Sangwan, N., P. Lata, V. Dwivedi, A. Singh, N. Niharika, J. Kaur, et al. 2012. Comparative metagenomic analysis of soil microbial communities across three hexachlorocyclohexane contamination levels. PLoS ONE 7(9):e46219. doi:10.1371/journal.pone.0046219

Scholz, M.B., C.C. Lo, and P.S.G. Chain. 2012. Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. Curr. Opin. Biotechnol. 23:9–15. doi:10.1016/j.copbio.2011.11.013

Schulze, W.X., G. Gleixner, K. Kaiser, G. Guggenberger, M. Mann, and E.D. Schulze. 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia 142:335–343. doi:10.1007/s00442-004-1698-9

Segata, N., D. Boernigen, T.L. Tickle, X.C. Morgan, W.S. Garrett, and C. Huttenhower. 2013. Computational meta'omics for microbial community studies. Mol. Syst. Biol. 9:666. doi:10.1038/msb.2013.22

Siggins, A., E. Gunnigle, and F. Abram. 2012. Exploring mixed microbial community functioning: Recent advances in metaproteomics. FEMS Microbiol. Ecol. 80:265–280. doi:10.1111/j.1574-6941.2011.01284.x

Singh, B.K., C.D. Campbell, S.J. Sorenson, and J.Z. Zhou. 2009. Soil genomics. Nat. Rev. Microbiol. 7:756. doi:10.1038/nrmicro2119-c1

Steven, B., L. Gallegos-Graves, S.R. Starkenburg, P.S. Chain, and C.R. Kuske. 2012. Targeted and shotgun metagenomic approaches provide different descriptions of dryland soil microbial communities in a manipulated field study. Environ. Microbiol. Rep. 4:248–256. doi:10.1111/j.1758-2229.2012.00328.x

Stewart, F.J., A.K. Sharma, J.A. Bryant, J.M. Eppley, and E.F. DeLong. 2011. Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities. Genome Biol. 12:R26. doi:10.1186/gb-2011-12-3-r26

Stolyar, S., S. Van Dien, K.L. Hillesland, N. Pinel, T.J. Lie, J.A. Leigh, and D.A. Stahl. 2007. Metabolic modeling of a mutualistic microbial community. Mol. Syst. Biol. 3:92. doi:10.1038/msb4100131

Taylor, E.B., and M.A. Williams. 2010. Microbial protein in soil: Influence of extraction method and C amendment on extraction and recovery. Microb. Ecol. 59:390–399. doi:10.1007/s00248-009-9593-x

Thomas, T., J. Gilbert, and F. Meyer. 2012. Metagenomics: A guide from sampling to data analysis. Microb. Inf. Exp. 2:3. doi:10.1186/2042-5783-2-3

Tringe, S.G., C. von Mering, A. Kobayashi, A.A. Salamov, K. Chen, H.W. Chang, et al. 2005. Comparative metagenomics of microbial communities. Science 308:554–557. doi:10.1126/science.1107851

Urich, T., A. Lanzen, J. Qu, D.H. Huson, C. Schleper, and S.C. Schuster. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS ONE 3(6):e2527. doi:10.1371/journal.pone.0002527

Uroz, S., P. Ioannidis, J. Lengelle, A. Cebron, E. Morin, M. Buée, and F. Martin. 2013. Functional assays and metagenomic analyses reveals differences between the microbial communities inhabiting the soil horizons of a Norway spruce plantation. PLoS ONE 8(2):e55929. doi:10.1371/journal.pone.0055929

Vogel, T.M., P.R. Hirsch, P. Simonet, J.K. Jansson, J.M. Tiedje, J.D. van Elsas, et al. 2009a. Advantages of the metagenomic approach for soil exploration: Reply from Vogel et al. Nat. Rev. Microbiol. 7:756–757. doi:10.1038/nrmicro2119-c3

<unknown>Vogel, T.M., P. Simonet, J.K. Jansson, P.R. Hirsch, J.M. Tiedje, J.D. van Elsas, et al. 2009b. TerraGenome: A consortium for the sequencing of a soil metagenome. Nat. Rev. Microbiol. 7:252. doi:10.1038/nrmicro2119</unknown>

Wang, H.-B., Z.-X. Zhang, H. Li, H.-B. He, C.-X. Fang, A.-J. Zhang, et al. 2011. Characterization of metaproteomics in crop rhizospheric soil. J. Proteome Res. 10:932–940. doi:10.1021/pr100981r

Williams, M.A., E.B. Taylor, and H.P. Mula. 2010. Metaproteomic characterization of a soil microbial community following carbon amendment. Soil Biol. Biochem. 42:1148–1156. doi:10.1016/j.soilbio.2010.03.021

Wollum, A.G. 1994. Soil sampling for microbiological analysis. In: R.W. Weaver et al., editors, Methods of soil analysis. Part 2. Microbiological and biochemical properties. SSSA Book Ser. 5. SSSA, Madison, WI. p. 1–14. doi:10.2136/sssabookser5.2.c1

Wu, L., H. Wang, Z. Zhang, R. Lin, Z. Zhang, and W. Lin. 2011. Comparative metaproteomic analysis on consecutively *Rehmannia glutinosa*-monocultured rhizosphere soil. PLoS ONE 6(5):e20611. doi:10.1371/journal.pone.0020611

Yergeau, E., H. Hogues, L.G. Whyte, and C.W. Greer. 2010. The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. ISME J. 4:1206–1214. doi:10.1038/ismej.2010.41

Yergeau, E., S. Sanschagrin, D. Beaumier, and C.W. Greer. 2012. Metagenomic analysis of the bioremediation of diesel-contaminated Canadian high Arctic soils. PLoS ONE 7(1):e30058. doi:10.1371/journal.pone.0030058

Zhou, J., Y. Deng, F. Luo, Z. He, and Y. Yang. 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2. mBio 2(4):e00122–11. doi:10.1128/mBio.00122-11

Zomorrodi, A.R., and C.D. Maranas. 2012. OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput. Biol. 8(2):e1002363. doi:10.1371/journal.pcbi.1002363