The Potential of Metagenomic Approaches for Understanding Soil Microbial Processes

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.

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 nifH, amoA, 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
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 metagenome 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 microbial community structure and function. The increased rate of CH4 consumption was positively correlated with an increase in particulate methane monoxygenase 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., NO3 reduction and SO4 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 N2 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.

<table>
<thead>
<tr>
<th>Location</th>
<th>Site description</th>
<th>Soil type</th>
<th>Experimental design</th>
<th>Biological replication</th>
<th>Sequencing platform</th>
<th>Sequencing depth</th>
<th>Assembly</th>
<th>Functional assignment†</th>
<th>Comments</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nunavut, Canada</td>
<td>tundra</td>
<td>permafrost</td>
<td>two depths: active layer and permafrost</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>853 Mbp (0.35–0.99 million reads/sample)</td>
<td>Phrap assembler (140 Mbp, 134,000 contigs)</td>
<td>not given</td>
<td>multiple displacement amplification (MDA) was needed to generate sufficient DNA; detected genes for methanogenesis</td>
<td>Yergeau et al. (2010)</td>
</tr>
<tr>
<td>Rothamsted, UK</td>
<td>Park Grass experiment (permanent grassland)</td>
<td>silty clay loam</td>
<td>comparison of direct and indirect DNA extraction methods</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>~1 million reads</td>
<td>none</td>
<td>not given</td>
<td>extraction methods gave similar functional gene profiles</td>
<td>Delmont et al. (2011)</td>
</tr>
<tr>
<td>Pro Toh Daieng, Thailand</td>
<td>swamp forest</td>
<td>peat</td>
<td>pooled sample</td>
<td>no</td>
<td>454 GS FLX</td>
<td>45.9 Mbp (0.18 million reads)</td>
<td>GS De Noso Assembler (13.4 Mbp, 54,000 contigs, averaging 248 bp)</td>
<td>SEED: 53</td>
<td>searched mainly for polysaccharide-degrading genes</td>
<td>Kanokratana et al. (2011)</td>
</tr>
<tr>
<td>Alaska</td>
<td>Hess Creek (black spruce forest)</td>
<td>permafrost</td>
<td>two depths: active layer and permafrost, before and 2 and 7 after thawing</td>
<td>yes (n = 2)</td>
<td>Illumina GAII (213 bp)</td>
<td>39.8 Gbp (176 million reads)</td>
<td>Velvet assembler (9.6 Mbp, 3758 contigs &gt;1 kb)</td>
<td>KEGG: 11</td>
<td>used emulsion polymerase chain reaction to generate sufficient DNA; draft genome of dominant methanogen obtained</td>
<td>Makielski et al. (2011)</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>Harvard Forest (northern hardwood forest)</td>
<td>sandy loam (Typic Hapludalf)</td>
<td>single composite from two soil cores (0–10 cm)</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>748 Mbp (3.4 million reads)</td>
<td>none</td>
<td>COG: 60</td>
<td>also companion metatranscriptome</td>
<td>Stewart et al. (2011)</td>
</tr>
<tr>
<td>São Paulo, Brazil</td>
<td>mangrove forest</td>
<td>submerged sediment</td>
<td>samples from four sites, one of which was impacted by oil contamination</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>215 Mbp (~0.25 million reads/site)</td>
<td>none</td>
<td>KEGG: 30</td>
<td>metabolic reconstruction of C, N, and S cycles</td>
<td>Andreote et al. (2012)</td>
</tr>
<tr>
<td>Rothamsted, UK</td>
<td>Park Grass experiment (permanent grassland)</td>
<td>silty clay loam</td>
<td>13 samples differing by date, depth, and DNA extraction method</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>4.8 Gbp (12.5 million reads)</td>
<td>Newbler assembler (152 Mbp, 267,000 contigs)</td>
<td>SEED: 56</td>
<td>extraction method showed more variation than date or depth of sampling</td>
<td>Delmont et al. (2012a)</td>
</tr>
<tr>
<td>Michigan and Minnesota</td>
<td>Kellogg Biological Station (cropped fields) and Cedar Creek Ecosystem Science Reserve (successional grasslands)</td>
<td>sandy loam (Hapludalf) and sand (Udipsamment)</td>
<td>low, medium, and high N addition at each site</td>
<td>yes (n = 3)</td>
<td>454 GS FLX Titanium</td>
<td>518 Mbp (1.35 million reads)</td>
<td>none</td>
<td>SEED: 25–35</td>
<td>shift in functional capacity with N input; significant correlation of metagenome functional and phylogenetic composition</td>
<td>Fierer et al. (2012a)</td>
</tr>
<tr>
<td>Worldwide</td>
<td>cold and hot deserts, selected nondesert biomes</td>
<td>various</td>
<td>16 total soils</td>
<td>no</td>
<td>Illumina GAIIx (2100 bp)</td>
<td>6.2 Gbp total (0.39–3.9 million reads per soil)</td>
<td>none</td>
<td>SEED: 13–23</td>
<td>functional composition correlated with site characteristics</td>
<td>Fierer et al. (2012b)</td>
</tr>
<tr>
<td>Lucens, India</td>
<td>dumpsite</td>
<td>not specified</td>
<td>three soils with a gradient of hexachlorocyclohexane contamination</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>1.2 Gbp (1.1–1.2 million reads per soil)</td>
<td>none</td>
<td>not given</td>
<td>draft genome of Chromohalobacter salinigenes</td>
<td>Sangwan et al. (2012)</td>
</tr>
<tr>
<td>Nevada</td>
<td>Nevada Desert FACE facility (shrubland)</td>
<td>loamy sand (Andosol)</td>
<td>four pooled samples (ambient and elevated CO₂ plots, with two locations [creosote bush and interspace])</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>724 Mbp total (0.31–0.66 million reads per sample)</td>
<td>none</td>
<td>SEED: 36–40</td>
<td>functional genes less discriminating than 16S rRNA genes for the effect of elevated CO₂</td>
<td>Steven et al. (2012)</td>
</tr>
<tr>
<td>Nunavut, Canada</td>
<td>tundra</td>
<td>permafrost</td>
<td>four pooled samples (control and three times for treated biopiles of oil-contaminated soil)</td>
<td>no</td>
<td>454 GS FLX Titanium</td>
<td>463 Mbp total (0.11–0.46 million reads per sample)</td>
<td>none</td>
<td>not applicable</td>
<td>used BLAST to focus on hydrocarbon-degrading genes (found to be higher in treated than control)</td>
<td>Yergeau et al. (2012)</td>
</tr>
<tr>
<td>Brazil-Châmere, France</td>
<td>Norway spruce plantation</td>
<td>Allocron soil</td>
<td>closely spaced soil cores separated into organic and mineral horizons</td>
<td>yes (n = 3)</td>
<td>454 GS FLX Titanium</td>
<td>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)</td>
<td>unspecified assembler of Illumina data (0.77 Mbp, 3492 contigs)</td>
<td>SEED: 36–60</td>
<td>differences in functional subsystems between organic and mineral horizons</td>
<td>Uroz et al. (2013)</td>
</tr>
</tbody>
</table>

† 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; Wöllum, 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.

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,
leading to the use of the rRNA/DNA ratio as an indicator of relative microbial activity (Blazewicz et al., 2013; DeAngelis et al., 2010; Mutteray 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 and inorganic molecules in soils (Adamczyk et al.,

---

**Table 2. Summary of soil metatranscriptomes.**

<table>
<thead>
<tr>
<th>Location</th>
<th>Site description</th>
<th>Soil type</th>
<th>Experimental design</th>
<th>Site description</th>
<th>Experimental design</th>
<th>Soil type</th>
<th>Sequencing platform</th>
<th>Sequencing depth</th>
<th>Functional assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern France</td>
<td>Pinus pinaster plantation on coastal sand dune</td>
<td>Sand</td>
<td>poly-A capture of total mRNA, cloned into plasmid library</td>
<td>119 clones sequenced</td>
<td>COG: 66</td>
<td>first metatranscriptome for soil</td>
<td>260,000 sequences</td>
<td>75% rRNA</td>
<td>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 analysis</td>
<td></td>
</tr>
<tr>
<td>Am Rotböll, Germany</td>
<td>lawn sandy single composite of three samples (0–10 cm)</td>
<td>Sandy loam</td>
<td>no mRNA enrichment</td>
<td>454 GS20</td>
<td>260,000 sequences</td>
<td>SEED: 33</td>
<td>first shotgun metatranscriptome from soil; no database limitations for identifying transcripts based on gene sequences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>Harvard Forest (northern hardwood forest)</td>
<td>Sandy loam</td>
<td>single composite of 14–16 soil cores (0–10 cm)</td>
<td>454 FLX Titanium</td>
<td>1.2 million sequences</td>
<td>KEGG: 45</td>
<td>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 analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central France</td>
<td>Breuil-Chenue forest (Picea abies and Fagus sylvatica plantations)</td>
<td>Sandy loam</td>
<td>single composite of 14–16 soil cores (0–10 cm)</td>
<td>Sanger</td>
<td>Blast2GO: 52</td>
<td>860 (Picea), 790 (Fagus)</td>
<td>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 analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

REFERENCES


Baveye, P. C. 2009. To sequence or not to sequence the whole-soil metagenome? Nat. Rev. Microbiol. 7:56–66. doi:10.1038/nrmicro2046


Mackelprang, R., M.P. Waldrop, K.M. DeAngelis, M.M. David, K.L. Chavarria, 
Mackelprang, R., M.P. Waldrop, K.M. DeAngelis, M.M. David, K.L. Chavarria, 

2012. Targeted and shotgun metagenomic approaches provide different 
descriptions of dryland soil microbial communities in a manipulated 
2299.2012.00328.x

Community transcriptomics reveals universal patterns of protein sequence 
doi:10.1186/gb-2011-12-3-r26

Stolarz, S., V. Van Dien, K.L. Hillisland, N. Pindel, T.J. Lie, J.A. Leigh, and D.A. 
Mol. Syst. Biol. 3:92. doi:10.1038/msb410131

Taylor, E.B., and M.A. Williams. 2010. Microbial protein in soil: Influence of 
extraction method and C amendment on extraction and recovery. Microb. 

Thomas, T.J. Gilbert, and F. Meyer. 2012. Metagenomics: A guide from sampling 

Tringe, S.G., C. von Mering, A. Kobayashi, A.A. Salamov, K. Chen, H.W. Chang, 
308:554–557. doi:10.1126/science.1107851

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. Ceborn, 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 
pone.0055929

al. 2009a. Advantages of the metagenomic approach for soil exploration: 
nrmicro2119-8

al. 2009b. TerraGenome: A consortium for the sequencing of 
a soil metagenome. Nat. Rev. Microbiol. 7:252. doi:10.1038/ 
nrnicro2119</unknown

Characterization of metaproteomes in crop rhizospheric soil. J. Proteome 
Res. 10:932–940. doi:10.1021/pr100981t

characterization of a soil microbial community following carbon 
soilbio.2010.03.021

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

metaproteomic analysis on consecutively Rehmannia glutinosa-
pone.0020611

potential of high Arctic permafrost revealed by metagenomic sequencing, 
sPCR and microarray analyses. ISME J. 4:1206–1214. doi:10.1038/ 
ismej.2010.41

analysis of the bioremediation of diesel-contaminated Canadian high 

ecological network of soil microbial communities in response to elevated 

framework for the metabolic modeling and analysis of microbial 
pcbi.1002363