Journal of Microbiological Methods 91 (2012) 246–251

Contents lists available at SciVerse ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Review

Application of metatranscriptomics to soil environments

Lilia C. Carvalhais [a,1], Paul G. Dennis [b,c,1], Gene W. Tyson [b,c], Peer M. Schenk [a,*]

[a] School of Agriculture and Food Sciences, The University of Queensland, Brisbane, QLD 4072, Australia
[b] Australian Centre for Ecogenomics, The University of Queensland, Brisbane, QLD 4072, Australia
[c] Advanced Water Management Centre, The University of Queensland, Brisbane, QLD 4072, Australia

ARTICLE INFO

Article history:
Received 23 April 2012
Received in revised form 10 August 2012
Accepted 22 August 2012
Available online 29 August 2012

Keywords:
Metatranscriptomics
Metagenomics
Soil
High-throughput sequencing

ABSTRACT

The activities of soil microbial communities are of critical importance to terrestrial ecosystem functioning. The mechanisms that determine the interactions between soil microorganisms, their environment and neighbouring organisms, however, are poorly understood. Due to advances in sequencing technologies, an increasing number of metagenomics studies are being conducted on samples from diverse environments including soils. This information has not only increased our awareness of the functional potential of soil microbial communities, but also constitutes powerful reference material for soil metatranscriptomics studies. Metatranscriptomics provides a snapshot of transcriptional profiles that correspond to discrete populations within a microbial community at the time of sampling. This information can indicate the potential activities of complex microbial communities and the mechanisms that regulate them. Here we summarise the technical challenges for metatranscriptomics applied to soil environments and discuss approaches for gaining biologically meaningful insight into these datasets.

Contents

1. Introduction

Soil microbial communities are involved in critical ecosystem functions such as decomposition and geochemical cycling (Carney and Matson, 2005; Nielsen et al., 2011) and strongly influence soil physical characteristics (Feeney et al., 2006; Rillig and Mummey, 2006) as well as plant health and nutrition (Dennis et al., 2010). Soils are complex and provide a vast diversity of habitats that result from structural aspects such as the size, shape and connectivity of pore networks, as well as other factors including the complexity of

* Corresponding author at: School of Agriculture and Food Sciences, The University of Queensland, St. Lucia QLD 4072, Australia. Tel./fax: +61 7 33658817.
E-mail addresses: l.carvalhais@uq.edu.au (L.C. Carvalhais), p.dennis@uq.edu.au (P.G. Dennis), g.tyson@awmc.uq.edu.au (G.W. Tyson), p.schenk@uq.edu.au (P.M. Schenk).
[1] Equal contribution.

0167-7012/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.mimet.2012.08.011

resources, physicochemical conditions and biological interactions. Microbial community structure may be influenced by a range of environmental parameters, including: pH (Dennis et al., 2009), temperature (Ward et al., 1998), moisture content (Zhou et al., 2002), nutrient status (Broughton and Gross, 2000), substrate availability and complexity (Dennis et al., 2012), exposure to the roots of different plant species (Kuske et al., 2002), contamination with pollutants (Muller et al., 2001), salinity (Nubel et al., 2000), predation (Jurgens and Matz, 2002), and other variables such as the architecture of their habitats (Sessitsch et al., 2001). This environmental heterogeneity is thought to contribute to the maintenance of soil microbial communities that typically represent the largest fraction of below-ground biomass (Hassink et al., 1993) and are estimated to constitute somewhere in the order of tens of thousands of microbial 'species' per gram of soil (Gans et al., 2005; Roesch et al., 2007). Nonetheless, the relative influence of these parameters on microbial activities is poorly understood.

Studies aiming to investigate the diversity and functioning of soil microbial communities were hampered for a long time by the inability of the vast majority of microorganisms to grow in standard culture media (Vartoukian et al., 2010). Consequently, the development of culture-independent approaches has significantly increased our understanding of soil microbial ecology. DNA, RNA, proteins and metabolites can be extracted directly from environmental samples and analysed via metagenomics, metatranscriptomics, metaproteomics and metabolomics, respectively. The advent of high-throughput sequencing technologies used in metagenomics and metatranscriptomics has made it possible to obtain datasets that are commensurate to the complexity of these microbial communities. Metagenomics offers novel insights into the functional potential of microbial communities and provides reference genes and genomes for metatranscriptomics (Shi et al., 2011). Metatranscriptomics facilitates insight into the potential expression of genes at the time of the sampling. While post-transcriptional and post-translational gene expression can regulate protein synthesis, transcriptional level control of gene expression enables organisms such as bacteria to rapidly adapt to changing environmental conditions (Moran, 2009). For this reason, immediate regulatory responses to environmental changes may be better reflected by the metatranscriptome than the metaproteome (the assemblage of proteins present in an environmental sample; Moran, 2009). In this review, we summarise the technical challenges relevant to metatranscriptomics applied to soil environments and the methodological and analytical solutions that can be used to circumvent them.

2. High-throughput sequencing as a key tool for soil metatranscriptomics

High-throughput sequencing generates large volume of data and facilitates characterisation of transcripts without any *a priori* knowledge of their nucleotide sequences. A key consideration before applying metatranscriptomics to soil-associated microbial communities is the depth of coverage that is required to address the focal research question. To some extent this will determine the choice of platforms used for a metatranscriptomics study. Currently, the most common high-throughput sequencing platforms used in metatranscriptomics studies are the 454 Genome Sequencer FLX systems (Roche) and the HiSeq 2000 (Illumina, Inc.). Despite some technical differences between platforms, both are based on miniaturised, individual sequencing-by-synthesis reactions and allow multiplexing of samples. The platforms are designed to optimise the spatial arrangement of each reaction, and facilitate large numbers of individual sequencing reactions to be performed in parallel. At present, the 454 GS FLX Titanium platform provides among the longest average read lengths (~700 bp); however, relative to the HiSeq 2000 (600 Gb) its total sequence output per run (0.45–0.75 Gb) is low. The 454 platform is prone to read errors in homo-polymer stretches (Metzker, 2010). Nonetheless, an advantage of the long reads is that repetitive regions can be mapped more effectively. The HiSeq 2000 currently produces reads of up to 150 bp in length and provides a throughput up to 600 Gb per run, although the run time is considerably longer (~11 days). The HiSeq 2000 is well suited to gene expression studies because of its ability to generate large volumes of sequence data, which provides sufficient coverage to overcome some of the problems associated with differences in transcript abundance and quality (Birzele et al., 2010; Camarena et al., 2010).

Other manufacturers are developing platforms that differ from the current fluorophore-based chemistries. Ion Torrent and Ion Proton (Life technologies), for example, use ion-sensitive field effect transistors (ISFETs) that measure changes in pH to detect nucleotide incorporation during sequencing-by-synthesis. This method of detection facilitates shorter run times than fluorescence-based detection systems. Nanopore detection systems measure differences in conductivity across a nanoscale pore, eliminating the need for optics and DNA amplification (Niedringhaus et al., 2011). This technology is used in the GridION system (Oxford Nanopore Technologies) which promises short run-times, massively high-throughput and up to 10 kb reads. The main obstacle for these systems at present is that bioinformatics tools are yet to be developed that correct for the specific sequencing errors generated by each platform.

3. Methodological challenges

3.1. RNA instability and extraction

A summary of the key steps in a metatranscriptomics experiment is presented in Table 1. Key limitations inherent to metatranscriptomics are that the average half-lives of mRNA molecules are in the range of seconds to minutes (Deutscher, 2006). mRNA stability also differs between microbial species (Bernstein et al., 2002; Selinger et al., 2003; Hambraeus et al., 2003) and can be influenced by the nutritional status of individual cells (Redon et al., 2005). Furthermore, genes that share biological functions are implicated to display similar mRNA degradation rates, with house-keeping genes having more stable mRNAs (Bernstein et al., 2002; Selinger et al., 2003; Hambraeus et al., 2003). To minimise changes in transcriptional profiles that may occur as a consequence of sampling, it is thus imperative to snap-freeze samples in liquid nitrogen or to transfer them to an RNA preservation solution (*e.g.* LifeGuard™ Soil Preservation Solution, MO BIO Laboratories, Inc, Carlsbad, CA) as soon after sampling as possible. Ideally, this delay should be in the range of seconds rather than minutes. By taking multiple samples over time, metatranscriptomics should highlight the relative stability of different transcripts and indicate which transcripts are associated with constitutively expressed *vs.* acutely responsive genes.

RNA isolation from soils is especially challenging due to ineffective cell lysis, adsorption of RNA to soil particles and the presence of RNases. In addition, adsorption to soil particles is increased by conditions that are typical for RNA extraction buffers, such as low pH, which is used to isolate RNA from DNA (Chomczynski and Sacchi, 1987), and high salt conditions, under which RNases are inactivated.

Most soil RNA extraction methods employ bead-beating as an initial step. RNA extraction methods involving microwave-based rupture (Orsini and Romano-Spica, 2001), liquid nitrogen grinding (Volossiouk et al., 1995), and enzymatic lysis (Zhou et al., 1996) have been shown to be less efficient than those involving bead-beating (Lakay et al., 2007). Currently, there are five commonly used commercially available extraction kits: 1) PowerSoil™ Total RNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA), 2) E.Z.N.A.® Soil RNA kit (Omega Bio-tek, Norcross, GA, USA), 3) FastRNA® Pro Soil-Direct kit (MP Biomedicals, Solon, OH, USA), 4) FastRNA® Pro Soil-Indirect kit (MP Biomedicals, Solon, OH, USA), and 5) IT 1-2-3 Platinum Path™ Sample Purification kit (Idaho Technology Inc. Salt Lake City, UT, USA). Currently the

Table 1
Generalised pipeline for metatranscriptomics approaches.

Sequential steps	Widely used methods/kits
Soil RNA extraction	PowerSoil™ Total RNA Isolation Kit (MoBio Laboratories, Carlsbad, CA USA), E.Z.N.A.® Soil RNA kit (Omega Bio-tek, Norcross, GA, USA), FastRNA® Pro Soil-Direct kit, FastRNA® Pro Soil-Indirect kit (MP Biomedicals, Solon, OH, USA)
mRNA enrichment	mRNA-ONLY™ Prokaryotic mRNA Isolation kit (Epicentre Biotechnologies, Madison, WI, USA), MICROBExpress™ Bacterial mRNA Enrichment Kit (Invitrogen, Carlsbad, CA, USA), size separation by gel electrophoresis (McGrath et al., 2008), sample-specific subtractive hybridisation (Stewart et al., 2010)
Reverse transcription	Superscript® III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA, USA), Omniscript RT Kit (Qiagen, Valencia, CA), MMLV Reverse Transcriptase cDNA Synthesis Kit (Epicentre Biotechnologies, Madison, WI, USA)
DNA fragmentation	Nebulization (AIR™ DNA Fragmentation Kit, Bioo Scientific Corporation, Austin, TX, USA), sonication, cavitation (Bioruptor®, Diagenode, Denville, NJ, USA), hydrodynamic breakage ((Joneja and Huang, 2009; Nesterova et al., 2012), HydroShear, Holliston, MA), treatment with enzymes
Size selection	Gel-based size selection, electrophoresis platforms (Pippin Prep and Blue Pippin, Sage Science, Beverly, MA, USA, automated size selection coupled with fractionation systems (*e.g.* LabChip® XT, Caliper Life Sciences Hopkinton, MA, USA), Bead-based AxyPrep™ FragmentSelect kit (Axygen, Union City, CA, USA)
Sequencing	454 Genome Sequencer FLX systems (Roche), HiSeq 2000 (Illumina, Inc.)

PowerSoil™ Total RNA Isolation Kit is the most commonly used kit (DeCoste et al., 2011; Di Gennaro et al., 2009).

Complex organic molecules, such as humic and fulvic acids, typically co-precipitate during nucleic acid extraction from soils. These compounds often inhibit PCR by limiting template availability by sequence-specific binding (Arbeli and Fuentes, 2007; Opel et al., 2010). For this reason, methods have been developed that aim to eliminate humic and fulvic acids during nucleic acid extraction. These include: 1) adsorption with powdered activated charcoal (Desai and Madamwar, 2007); 2) precipitation with aluminium sulphate prior to cell lysis (Persoh et al., 2008); 3) pre-treatment of soils with $CaCO_3$ (Sagova-Mareckova et al., 2008); 4) addition of polyvinyl polypyrolidone (PVPP; Rajendhran and Gunasekaran, 2008), 5) isolation of extracted nucleic acids by $CaCl_2$ (Sagova-Mareckova et al., 2008); and 6) extraction of RNA at pH 5.0 followed by purification using Q-Sepharose columns, supplemented with cetyl trimethylammonium bromide (CTAB) and vitamins (Mettel et al., 2010). The presence of genomic DNA (gDNA) in RNA extracts can lead to overestimation of RNA concentration when using UV spectrophotometry for quantification, as the absorption wavelengths of RNA and DNA overlap. In addition, DNA fragments can be mistaken for transcripts post-sequencing. Co-extracted gDNA can be minimised by treatment with DNaseI (Rio et al., in press; Marchetti et al., 2012).

3.2. mRNA enrichment

The total RNA pool in environmental microbial communities consists primarily of ribosomal RNA (rRNA) and transfer RNA (Karpinets et al., 2006), with approximately 1–5% mRNA (He et al., 2010). Isolation or enrichment of mRNA is, therefore, an important step in metatranscriptomic experiments. Several methods for mRNA recovery from environmental samples have been described, including: 1) subtractive hybridisation (MICROBExpress Bacterial mRNA Enrichment Kit, Ambion; Pang et al., 2004), 2) exonuclease treatment, which preferentially degrades rRNA (mRNA-ONLY Prokaryotic mRNA Isolation kit, EPICENTRE Biotechnologies, Madison; USA), 3) size separation by gel electrophoresis (McGrath et al., 2008), and 4) duplex specific nuclease (DSN) treatment (Yi et al., 2011). A comparation of two of these approaches (MICROBExpress and DSN) indicated that removal efficiencies vary according to RNA integrity and the environment from which the microbial communities are sampled (He et al., 2010). Subtractive hybridisation (rather than exonuclease treatment) was found to be more effective in preserving the relative abundance of different transcripts (He et al., 2010). A comparison between subtractive hybridisation using the MICROBExpress kit and the DSN treatment revealed that the latter was more efficient at enriching for mRNA (Yi et al., 2011). The most recently developed approach for mRNA enrichment uses a specific probe mix for subtractive hybridisation of rRNA using antisense rRNA probes generated by *in vitro* transcription of PCR products amplified from coupled DNA samples (Stewart et al., 2010).

Another issue that may arise from applying metatranscriptomics to soil samples is the presence of eukaryotic RNA, *e.g.* from fungi and plants. Bacterial and archaeal RNAs can be enriched using surfaces coated with poly(dT) probes, which capture eukaryotic RNAs that contain 3′-poly-A tails. Similarly, eukaryotic mRNAs can be isolated by mRNA-specific cDNA synthesis using anchored oligo dT primers, or by affinity capture using magnetic beads that are coated with poly-dT oligonucleotides, which bind to the 3′ polyadenylated (poly-A) tails associated with the eukaryotic mRNAs (Bailly et al., 2007). These methods exploit the fact that non-eukaryotic 3′-poly-A RNA molecules are rare and are rapidly degraded when present (Belasco, 2010; Dreyfus and Regnier, 2002).

3.3. Issues relating to cDNA synthesis and amplification

Soil RNA extraction typically yields small amounts of mRNA and an additional amplification step may be required to achieve sufficient starting material for downstream applications. This process is typically performed using linear amplification, which involves several steps. Firstly, *Escherichia coli* poly-A polymerase is used to polyadenylate the RNA prior to reverse transcription during which polyadenylated RNA is converted to cDNA using an oligo-dT primer containing a T7 RNA polymerase promoter sequence and a recognition site for a restriction enzyme (Frias-Lopez et al., 2008). This procedure can also be applied to eukaryotic RNA, although in this case the RNA does not need to be polyadenylated as it already contains poly-A tails. In addition, the oligo-dT primer does not need to include a recognition site for a restriction enzyme because deadenylases (or poly-A nucleases) can be used instead. After *in vitro* transcription, large quantities of single-stranded antisense RNA are generated. Double stranded cDNA can then be synthesised by reverse-transcription using random primers. Lastly, poly-A tails are removed by enzymatic digestion using the restriction sites built into the oligo-dT primers (Frias-Lopez et al., 2008; Stewart et al., 2010). Another method that has been used to amplify small amounts of DNA, and could be applied to cDNA, is multiple displacement amplification (MDA; Blanco et al., 1989; Gilbert et al., 2008). This method uses random hexamers as primers and Phi29 DNA polymerase, which has high fidelity and strand displacement activity at a constant temperature. MDA is known to compromise quantitative analysis of metagenomes due to DNA amplification biases (Yilmaz et al., 2010). Amplification biases are to be expected, therefore, when applying MDA to cDNAs. This limits the interpretation of such data to the presence/absence of transcripts.

In general, current high-throughput sequencing platforms require cDNA as template, which typically undergo reverse transcription, shearing, size selection, end polishing, and ligation of adapters. Reverse transcriptases can introduce errors during cDNA synthesis (Roberts et al., 1989). Furthermore, long transcripts appear to be reverse-transcribed less efficiently than short transcripts (Stewart et al., 2010). It has been suggested that spurious cDNA molecules can be generated by

primer-independent cDNA synthesis caused by non-target RNA molecules acting as primers (Haddad et al., 2007; Stahlberg et al., 2004). However, a higher specificity was found when reverse transcription was performed at higher temperatures, in the presence of RNase H^+ (Haddad et al., 2007). In addition, chimeric cDNA molecules can be generated through template switching by the reverse transcriptase in high homology regions of cDNA (Cocquet et al., 2006; Zeng and Wang, 2002). Direct sequencing of RNA may help to avoid these issues (Mamanova et al., 2010; Ozsolak et al., 2009); however, this approach presents additional challenges and is yet to be widely implemented (Ozsolak et al., 2010; Ozsolak and Milos, 2011a,b).

3.4. Targeting transcripts from fewer populations

Despite advances in sequencing technology, the fact remains that most microbial communities are so diverse that they are difficult to study. There is interest, therefore, in reducing the complexity of metagenomic and metatranscriptomic datasets by targeting single cells, or isolating specific populations and/or assemblages. The advantage of having transcripts from a smaller number of populations is that the sequence coverage for those populations would be greater. Populations/assemblages that are actively catabolising specific compounds within their environment can be isolated using stable isotope probing (SIP; Dumont and Murrell, 2005; Whiteley et al., 2007). In SIP experiments, specific or broad-range substrates are highly enriched with a stable isotope such as ^{13}C, ^{15}N or ^{18}O, and nucleic acids and other anabolic products from microbial cells utilising these substrates become isotopically labelled. After sampling, isotope labelled and unlabelled molecules such as DNA/RNA can be separated by buoyant density gradient centrifugation. The fractions can then be isolated and analysed using a wide range of molecular techniques, including metagenomics and metatranscriptomics. Chen et al. (2008) applied SIP to isolate DNA associated with methanotrophs in a peatland soil. They then used metagenomics to analyse the isolated DNA, but had to perform MDA before doing so in order to obtain sufficient DNA. Low RNA yields from SIP are thus likely to present a potential challenge for downstream metatranscriptomics analyses unless MDA is used. This approach deserves further investigation.

4. Data processing

4.1. Bioinformatics

Typically the first step in analysing metatranscriptomic data involves removal of short or poor quality sequences and error correction. Sequences should also be trimmed as sequencing errors become more frequent towards the ends of reads (Balzer et al., 2010). Error detection and removal/correction algorithms have been developed for 454 (Quince et al., 2009) and Illumina data (Dolan and Denver, 2008; Rougemont et al., 2008), but are not yet available for newer platforms (discussed above). Irrespective of whether methods are used to enrich for mRNA, sequence data can include considerable numbers of reads derived from rRNA (51–60% of total RNA (Stewart et al., 2010). These should be identified by comparison with a comprehensive rRNA gene sequence database and then removed. Likewise, if during the mRNA amplification step, transcripts were poly-adenylated, the artificial poly-A tails need to be removed prior to subsequent analyses (Frias-Lopez et al., 2008).

In the next step, sequences should be assigned a description by comparison with publically available databases, such as the National Centre for Biotechnological Information (NCBI) non-redundant (nr) database (http://www.ncbi.nlm.nih.gov/), the Integrated Microbial Genomes database (IMG/M; http://img.jgi.doe.gov/; Markowitz et al., 2008) and the Metagenomics Analysis Server (MG-RAST, http://metagenomics.anl.gov). Mapping reads to known sequences allows the relative frequencies of genes to be compared; however, to determine whether genes are up- or down-regulated, gene frequencies should be normalised by the gene abundances within a coupled genome/metagenome from the same nucleic acid extraction. For this reason, more useful information can be extracted from metatranscriptomics datasets if reads are mapped to custom databases generated using metagenomic data from the same or highly similar communities (Frias-Lopez et al., 2008; Shi et al., 2009). Mapping can be performed using a range of tools, including the Burrow-Wheeler Aligner (BWA, http://bio-bwa.sourceforge.net/) or the Blast-Like Alignment Tool (BLAT; Kent, 2002). Although not common practice, it may be useful to assemble cDNA reads into full gene transcripts, or polycistronic operons if a complementary metagenome is not available. This approach may simplify the assignment step, reduce the size of the dataset and provide coverage information that could be used to evaluate differences in transcript abundances. Examples of assemblers include: Velvet (Zerbino and Birney, 2008), Newbler (Chaisson and Pevzner, 2008) and Genovo (Laserson et al., 2010). Assembly is likely to be most effective for highly abundant transcripts from simple communities so this approach requires testing for complex microbial communities such as those found in soil environments.

Functional categorisation of transcripts can be obtained using the Kyoto Encyclopedia of Genes and Genomes (KEGG; (Kanehisa et al., 2004), the Clusters of Orthologous Groups (COGs; Tatusov et al., 2003), and the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG; Jensen et al., 2008) databases. These databases consist of groups of genes that have been assigned to different functional pathways (*e.g.* denitrification or nitrogen fixation) based on the similarity of protein orthologs from sequenced isolate microbial genomes. A BLAST against these databases assigns transcripts with significant similarity to functional pathways. This approach helps to determine whether whole pathways, rather than single genes, are differentially expressed between treatments.

Once sequences have passed quality control and been mapped to databases that identify genes and indicate whether whole functional pathways are represented, comparative analyses can be performed. By comparing metatranscriptomes from samples obtained under controlled conditions, different locations or time points, it is possible to determine whether genes and functional pathways are up- or down-regulated.

4.2. Statistical analyses

The output of various bioinformatics analyses can be tabulated at varying levels of organisation and complexity, which facilitates statistical analyses that address well defined research questions ranging from those concerning broad patterns of gene expression to those focusing on the expression of specific functional pathways. At present, knowledge of community level patterns of gene expression in soil environments is poor. Therefore, the theoretical content of most analyses is generally low, with experimental objectives being largely exploratory in nature. Exploratory multivariate statistical models can be used to identify differences in gene expression patterns between treatments, and along environmental gradients. Examples include: Between Group Analysis (Culhane et al., 2002), Redundancy Analysis (RDA; Joh et al., 2007), Canonical Correspondence Analysis (CCA; Liang et al., 2010), Permutational Multivariate Analysis of Variance (PERMANOVA; Anderson, 2001; Zapala and Schork, 2006), or Analysis of Similarities (ANOSIM; Huse et al., 2010). The advantage of RDA, CCA and BGA is that they allow the relationships between sites, species and treatments to be interpreted concomitantly. PERMANOVA, however, enables the inclusion of interaction terms in models. Comparisons can also be made based on the richness, equitability and phylogenetic distinctiveness of functional genes and/or pathways.

Exploratory analyses are effective for identifying sets of genes that correlate with treatments and environmental gradients; however, to determine whether these relationships are direct/indirect or causal/non-causal requires the use of confirmatory analyses with

greater theoretical content. The information provided by exploratory analyses facilitates development of testable multivariate hypotheses that can provide greater insight into the complexity of community level patterns of transcription. Structural equation modelling (SEM; Grace, 2006) is a multivariate statistical tool that enables multivariate hypothesis testing. SEM models are generally represented graphically and incorporate both empirical data and theoretical constructs. SEM facilitates significance testing of user-defined models and thereby gives an indication of the likely validity of system theories. Such analyses could lead to novel experiments designed to empirically test likely theories that may explain community level transcriptional patterns.

5. Concluding remarks

Despite the complexity of soil microbial communities a wide-range of existing methodological and analytical approaches should facilitate application of metatranscriptomics to soil environments. Combined with rigorously designed experiments, which perturb soils through the addition of substrates or modification of environmental conditions, metatranscriptomics will enhance our understanding of microbial responses and functionality. This should reveal mechanisms to enhance the abundance and activities of assemblages that perform desired ecosystem services such as: nutrient mobilisation, pathogen suppression and breakdown of organic pollutants. Improved understanding of short-term responses of microbial communities through metatranscriptomics should, therefore, aid the development of effective strategies to manage terrestrial ecosystems.

Acknowledgements

We thank Phil Hugenholtz for useful discussions. This work was supported by the Australian Research Council (DP1094749).

References

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46.
Arbeli, Z., Fuentes, C.L., 2007. Improved purification and PCR amplification of DNA from environmental samples. FEMS Microbiol. Lett. 272, 269–275.
Bailly, J., Fraissinet-Tachet, L., Verner, M.C., Debaud, J.C., Lemaire, M., Wesolowski-Louvel, M., Marmeisse, R., 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME J. 1, 632–642.
Balzer, S., Malde, K., Lanzen, A., Sharma, A., Jonassen, I., 2010. Characteristics of 454 pyrosequencing data-enabling realistic simulation with flowsim. Bioinformatics 26, 420–425.
Belasco, J.G., 2010. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. Nat. Rev. Mol. Cell Biol. 11, 467–478.
Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 99, 9697–9702.
Birzele, F., Schaub, J., Rust, W., Clemens, C., Baum, P., Kaufmann, H., Weith, A., Schulz, T.W., Hildebrandt, T., 2010. Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. Nucleic Acids Res. 38, 3999–4010.
Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C., Salas, M., 1989. Highly efficient DNA synthesis by the phage phi-29 DNA polymerase — symmetrical mode of DNA replication. J. Biol. Chem. 264, 8935–8940.
Broughton, L.C., Gross, K.L., 2000. Patterns of diversity in plant and soil microbial communities along a productivity gradient in a Michigan old-field. Oecologia 125, 420–427.
Camarena, L., Bruno, V., Euskirchen, G., Poggio, S., Snyder, M., 2010. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA sequencing. PLoS Pathog. 6, e1000834.
Carney, K.M., Matson, P.A., 2005. Plant communities, soil microorganisms, and soil carbon cycling: does altering the world belowground matter to ecosystem functioning? Ecosystems 8, 928–940.
Chaisson, M.J., Pevzner, P.A., 2008. Short read fragment assembly of bacterial genomes. Genome Res. 18, 324–330.
Chen, Y., Dumont, M.G., Neufeld, J.D., Bodrossy, L., Stralis-Pavese, N., McNamara, N., Ostle, N., Briones, M.J.I., Murrell, J.C., 2008. Revealing the uncultivated majority: combining DNA stable-isotope probing, mulitple displacement amplification and metagenomic analyses of uncultivated *Methylocystis* in acidic peatlands. Environ. Microbiol. 10, 2609–2622.
Chomczynski, P., Sacchi, N., 1987. Single step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. Anal. Biochem. 162, 156–159.
Cocquet, J., Chong, A., Zhang, G.L., Veitia, R.A., 2006. Reverse transcriptase template switching and false alternative transcripts. Genomics 88, 127–131.
Culhane, A.C., Perriere, G., Considine, E.C., Cotter, T.G., Higgins, D.G., 2002. Between-group analysis of microarray data. Bioinformatics 18, 1600–1608.
DeCoste, N.J., Gadkar, V.J., Filion, M., 2011. Relative and absolute quantitative Real-time PCR-based quantifications of *hcnC* and *phlD* gene transcripts in natural soil spiked with *Pseudomonas* sp. strain LBUM300. Appl. Environ. Microbiol. 77, 41–47.
Dennis, P.G., Hirsch, P.R., Smith, S.J., Taylor, R.G., Valsami-Jones, E., Miller, A.J., 2009. Linking rhizoplane pH and bacterial density at the microhabitat scale. J. Microbiol. Methods 76, 101–104.
Dennis, P.G., Miller, A.J., Hirsch, P.R., 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? FEMS Microbiol. Ecol. 72, 313–327.
Dennis, P.G., Rushton, S.P., Newsham, K.K., Lauducina, V.A., Ord, V.J., Daniell, T.J., O'Donnell, A.G., Hopkins, D.W., 2012. Soil fungal community composition does not alter along a latitudinal gradient through the maritime and sub-Antarctic. Fungal Ecol. 5, 403–408.
Desai, C., Madamwar, D., 2007. Extraction of inhibitor-free metagenomic DNA from polluted sediments, compatible with molecular diversity analysis using adsorption and ion-exchange treatments. Bioresour. Technol. 98, 761–768.
Deutscher, M.P., 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. Nucleic Acids Res. 34, 659–666.
Di Gennaro, P., Moreno, B., Annoni, E., Garcia-Rodriguez, S., Bestetti, G., Benitez, E., 2009. Dynamic changes in bacterial community structure and in naphthalene dioxygenase expression in vermicompost-amended PAH-contaminated soils. J. Hazard. Mater. 172, 1464–1469.
Dolan, P.C., Denver, D.R., 2008. TileQC: a system for tile-based quality control of Solexa data. BMC Bioinformatics 9, 250.
Dreyfus, M., Regnier, P., 2002. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. Cell 111, 611–613.
Dumont, M.G., Murrell, J.C., 2005. Stable isotope probing — linking microbial identity to function. Nat. Rev. Microbiol. 3, 499–504.
Feeney, D.S., Crawford, J.W., Daniell, T., Hallett, P.D., Nunan, N., Ritz, K., Rivers, M., Young, I.M., 2006. Three-dimensional microorganization of the soil–root–microbe system. Microb. Ecol. 52, 151–158.
Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., DeLong, E.F., 2008. Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. U. S. A. 105, 3805–3810.
Gans, J., Wolinsky, M., Dunbar, J., 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309, 1387–1390.
Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W.Z., Gilna, P., Joint, I., 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One 3, e3042.
Grace, J.B., 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press, Cambridge, UK.
Haddad, F., Qin, A.Q.X., Giger, J.M., Guo, H.Y., Baldwin, K.M., 2007. Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. BMC Biotechnol. 7, 21.
Hambraeus, G., von Wachenfeldt, C., Hederstedt, L., 2003. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. Mol. Genet. Genomics 269, 706–714.
Hassink, J., Bouwman, L.A., Zwart, K.B., Brussaard, L., 1993. Relationships between habitable pore-space, soil biota and mineralization rates in grassland soils. Soil Biol. Biochem. 25, 47–55.
He, S.M., Wurtzel, O., Singh, K., Froula, J.L., Yilmaz, S., Tringe, S.G., Wang, Z., Chen, F., Lindquist, E.A., Sorek, R., Hugenholtz, P., 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat. Methods 7, 807–812.
Huse, H.K., Kwon, T., Zlosnik, J.E., Speert, D.P., Marcotte, E.M., Whiteley, M., 2010. Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations *in vivo*. mBio 1, e00199-10.
Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., Bork, P., 2008. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. 36, D250–D254.
Joh, J.H., Lee, S.H., Lee, J.S., Kim, K.H., Jeong, S.J., Youn, W.H., Kim, N.K., Son, E.S., Cho, Y.S., Yoo, Y.B., Lee, C.S., Kim, B.G., 2007. Isolation of genes expressed during the developmental stages of the oyster mushroom, *Pleurotus ostreatus*, using expressed sequence tags. FEMS Microbiol. Lett. 276, 19–25.
Joneja, A., Huang, X.H., 2009. A device for automated hydrodynamic shearing of genomic DNA. Biotechniques 46, 553–556.
Jurgens, K., Matz, C., 2002. Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. Anton. Leeuw. Int. J. G. 81, 413–434.
Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M., 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. 32, D277–D280.
Karpinets, T.V., Greenwood, D.J., Sams, C.E., Ammons, J.T., 2006. RNA: protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. BMC Biol. 4, 30.
Kent, W.J., 2002. BLAT — the BLAST-like alignment tool. Genome Res. 12, 656–664.
Kuske, C.R., Ticknor, L.O., Miller, M.E., Dunbar, J.M., Davis, J.A., Barns, S.M., Belnap, J., 2002. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. Appl. Environ. Microbiol. 68, 1854–1863.

Lakay, F.M., Botha, A., Prior, B.A., 2007. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. J. Appl. Microbiol. 102, 265–273.
Laserson, J., Jojic, V., Koller, D., 2010. Genovo: *de novo* assembly for metagenomes. J. Comput. Biol. 6044, 341–356.
Liang, Y., Van Nostrand, J.D., Deng, Y., He, Z., Wu, L., Zhang, X., Li, G., Zhou, J., 2010. Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. ISME J. 5, 403–413.
Mamanova, L., Andrews, R.M., James, K.D., Sheridan, E.M., Ellis, P.D., Langford, C.F., Ost, T.W.B., Collins, J.E., Turner, D.J., 2010. FRT-seq: amplification-free, strand-specific transcriptome sequencing. Nat. Methods 7, 130–163.
Marchetti, A., Schruth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, A.E., Armbrust, E.V., 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proc. Natl. Acad. Sci. U. S. A. 109, E317–E325.
Markowitz, V.M., Ivanova, N.N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., Chen, I.M.A., Grechkin, Y., Dubchak, I., Anderson, I., Lykidis, A., Mavromatis, K., Hugenholtz, P., Kyrpides, N.C., 2008. IMG/M: a data management and analysis system for metagenomes. Nucleic Acids Res. 36, D534–D538.
McGrath, K.C., Thomas-Hall, S.R., Cheng, C.T., Leo, L., Alexa, A., Schmidt, S., Schenk, P.M., 2008. Isolation and analysis of mRNA from environmental microbial communities. J. Microbiol. Methods 75, 172–176.
Mettel, C., Kim, Y., Shrestha, P.M., Liesack, W., 2010. Extraction of mRNA from soil. Appl. Environ. Microbiol. 76, 5995–6000.
Metzker, M.L., 2010. Applications of next generation sequencing technologies — the next generation. Nat. Rev. Genet. 11, 31–46.
Moran, A.M., 2009. Metatranscriptomics: eavesdropping on complex microbial communities. Microbe 4, 329–335.
Muller, A.K., Westergaard, K., Christensen, S., Sorensen, S.J., 2001. The effect of long-term mercury pollution on the soil microbial community. FEMS Microbiol. Ecol. 36, 11–19.
Nesterova, I.V., Hupert, M.L., Witek, M.A., Soper, S.A., 2012. Hydrodynamic shearing of DNA in a polymeric microfluidic device. Lab Chip 12, 1044–1047.
Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P., Barron, A.E., 2011. Landscape of next-generation sequencing technologies. Anal. Chem. 83, 4327–4341.
Nielsen, U.N., Ayres, E., Wall, D.H., Bardgett, R.D., 2011. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity–function relationships. Eur. J. Soil Sci. 62, 105–116.
Nubel, U., Garcia-Pichel, F., Clavero, E., Muyzer, G., 2000. Matching molecular diversity and ecophysiology of benthic cyanobacteria and diatoms in communities along a salinity gradient. Environ. Microbiol. 2, 217–226.
Opel, K.L., Chung, D., McCord, B.R., 2010. A study of PCR inhibition mechanisms using real-time PCR. J. Forensic Sci. 55, 25–33.
Orsini, M., Romano-Spica, V., 2001. A microwave-based method for nucleic acid isolation from environmental samples. Lett. Appl. Microbiol. 33, 17–20.
Ozsolak, F., Milos, P.M., 2011a. RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. 12, 87–98.
Ozsolak, F., Milos, P.M., 2011b. Single-molecule direct RNA sequencing without cDNA synthesis. WIREs RNA 2, 565–570.
Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., Milos, P.M., 2009. Direct RNA sequencing. Nature 461, 814–818.
Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., Milos, P.M., 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell 143, 1018–1029.
Pang, X., Zhou, D.S., Song, Y.J., Pei, D.C., Wang, J., Guo, Z.B., Yang, R.F., 2004. Bacterial mRNA purification by magnetic capture-hybridization method. Microbiol. Immunol. 48, 91–96.
Persoh, D., Theuerl, S., Buscot, F., Rambold, G., 2008. Towards a universally adaptable method for quantitative extraction of high-purity nucleic acids from soil. J. Microbiol. Methods 75, 19–24.
Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., Sloan, W.T., 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat. Methods 6, 639–641.
Rajendhran, J., Gunasekaran, P., 2008. Strategies for accessing soil metagenome for desired applications. Biotechnol. Adv. 26, 576–590.
Redon, E., Loubière, P., Cocaign-Bousquet, M., 2005. Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. J. Biol. Chem. 280, 36380–36385.
Rillig, M.C., Mummey, D.L., 2006. Mycorrhizas and soil structure. New Phytol. 171, 41–53.
Rio, D.C., Ares Jr., M., Hannon, G.J., Nilsen, T.W., in press. Removal of DNA from RNA. Cold Spring Harb. Protoc. http://dx.doi.org/10.1101/pdb.prot5443.
Roberts, J.D., Preston, B.D., Johnston, L.A., Soni, A., Loeb, L.A., Kunkel, T.A., 1989. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis *in vitro*. Mol. Cell. Biol. 9, 469–476.
Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1, 283–290.
Rougemont, J., Amzallag, A., Iseli, C., Farinelli, L., Xenarios, I., Naef, F., 2008. Probabilistic base calling of Solexa sequencing data. BMC Bioinformatics 9, 431.
Sagova-Mareckova, M., Cermak, L., Novotna, J., Plhackova, K., Forstova, J., Kopecky, J., 2008. Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Appl. Environ. Microbiol. 74, 2902–2907.
Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., Rosenow, C., 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res. 13, 216–223.
Sessitsch, A., Weilharter, A., Gerzabek, M.H., Kirchmann, H., Kandeler, E., 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl. Environ. Microbiol. 67, 4215–4224.
Shi, Y.M., Tyson, G.W., DeLong, E.F., 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature 459, 266–269.
Shi, Y.M., Tyson, G.W., Eppley, J.M., DeLong, E.F., 2011. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J. 5, 999–1013.
Stahlberg, A., Hakansson, J., Xian, X.J., Semb, H., Kubista, M., 2004. Properties of the reverse transcription reaction in mRNA quantification. Clin. Chem. 50, 509–515.
Stewart, F.J., Ottesen, E.A., DeLong, E.F., 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME J. 4, 896–907.
Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41.
Vartoukian, S.R., Palmer, R.M., Wade, W.G., 2010. Strategies for culture of 'unculturable' bacteria. FEMS Microbiol. Lett. 309, 1–7.
Volossiouk, T., Robb, E.J., Nazar, R.N., 1995. Direct DNA extraction for PCR-mediated assays of soil organisms. Appl. Environ. Microbiol. 61, 3972–3976.
Ward, D.M., Ferris, M.J., Nold, S.C., Bateson, M.M., 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol. Mol. Biol. Rev. 62, 1353–1370.
Whiteley, A.S., Thomson, B., Lueders, T., Manefield, M., 2007. RNA stable-isotope probing. Nat. Protoc. 2, 838–844.
Yi, H., Cho, Y.J., Won, S., Lee, J.E., Yu, H.J., Kim, S., Schroth, G.P., Luo, S., Chun, J., 2011. Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. Nucleic Acids Res. 39.
Yilmaz, A., Allgaier, M., Hugenholtz, P., 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat. Methods 7, 943–944.
Zapala, M.A., Schork, N.J., 2006. Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. Proc. Natl. Acad. Sci. U. S. A. 103, 19430–19435.
Zeng, X.C., Wang, S.X., 2002. Evidence that BmTXK beta-BmKCT cDNA from Chinese scorpion *Buthus martensii* karsch is an artifact generated in the reverse transcription process. FEBS Lett. 520, 183–184.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhou, J.Z., Bruns, M.A., Tiedje, J.M., 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62, 316–322.
Zhou, J.Z., Xia, B.C., Treves, D.S., Wu, L.Y., Marsh, T.L., O'Neill, R.V., Palumbo, A.V., Tiedje, J.M., 2002. Spatial and resource factors influencing high microbial diversity in soil. Appl. Environ. Microbiol. 68, 326–334.