MICROBIOME — OPINION

Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology

James I. Prosser

Abstract | Technological advances are enabling the sequencing of environmental DNA and RNA at increasing depth and with decreasing costs. Metagenomic and transcriptomic analysis of soil microbial communities and the assembly of 'population genomes' from soil DNA are therefore now feasible. Although the value of such 'omic' approaches is limited by the associated technical and bioinformatic difficulties, even if these obstacles were eliminated and 'perfect' metagenomes and metatranscriptomes were available, important conceptual challenges remain. This Opinion article considers these conceptual challenges in the context of the current use of omics in soil microbiology, but the main arguments presented are also relevant to the application of omics to marine, freshwater, gut or other environments.

Microorganisms are major determinants of the physical, chemical and biological characteristics of soil, biogeochemical cycling, other terrestrial ecosystem functions and the sustainability of soil ecosystems. Understanding the ecology of soil microorganisms (the forces shaping soil microbial communities, their response to environmental change and their physicochemical and biological interactions) is difficult owing to the high degree of biological diversity and the inherent spatial heterogeneity that is present at scales <1 mm in soils[1,2]. Molecular studies (those that describe microbial taxonomy on the basis of the 16S rRNA gene) indicate that up to 1 million different bacterial and archaeal species are present in 10 g of soil, in the context of approximately 1 billion microbial cells[3–5]. This richness is no longer surprising, but its origins and causes are poorly understood. In addition, our understanding of functional redundancy and of the links between microbial richness, community composition and soil-ecosystem function is far from complete, despite the promises of 16S rRNA gene-based techniques.

Continuing advances in sequencing technology now make it possible to consider tackling these issues using metagenomics and metatranscriptomics, which are defined as the characterization of all genes and transcripts, respectively, in a soil sample (FIG. 1). In this Opinion article, I do not discuss targeted or single-gene high-throughput sequencing studies, which are sometimes described as 'metagenomics' but include data for only one gene and, therefore, do not encompass the holistic element of the 'omics' epithet. A single gram of soil contains in the order of 10^{12} prokaryotic genes and 10^{9} genomes[6,7]. Complete coverage of all of these genes by metagenomic sequencing is not yet achievable, but it might be within a few years; it is now becoming possible to reconstruct 'population genomes' from subsoil, where microbial abundance, biomass and diversity are lower[8], although genome closure in samples from complex topsoil remains a challenge[9,10]. It is also unlikely that genomes constructed from metagenomes will be derived from a single cell; rather, they will be a chimaera of genomes with undefined similarity. This problem could be solved by single-cell genome sequencing[11,12], but this currently yields only partial genomes, which may be sufficient for some applications (for example, for the study of a single metabolic pathway) but not others (for example, population genetics).

History has taught us that emerging techniques, which are often heralded as approaches that will revolutionize our understanding of soil communities, rarely accomplish what is wishfully anticipated. With the continuing rapid development and application of omics, it is therefore instructive to pause and consider whether and how these technologies can advance our understanding of the complex microbial communities in soil. In this Opinion article, I do not consider technical issues, which have been discussed elsewhere[13–18], or the issues associated with bioinformatics and annotation[19–23]. Major advances have been made in these areas and, although no technique is perfect, imperfections will become less important with time and are also expected to become easier to circumvent. Here, I discuss only the unavoidable conceptual issues and therefore assume the 'perfect' metagenome or metatranscriptome, including its characteristics and potential value. That is, I assume that nucleic acids can be extracted from every cell in an environmental sample of interest, without loss or bias, and that every gene or transcript can be sequenced and annotated accurately and its function correctly identified. Crucially, I distinguish between 'gene-centric' metagenomes, consisting of inventories of individual annotated genes, and 'genome-centric' metagenomes, consisting of completely sequenced genomes (FIG. 2). I then consider what these perfect 'omes' can and cannot tell us and explore their value in addressing specific, important questions in microbial ecology. The application of metagenomics in gene discovery[24–27] is not discussed as the conceptual issues are of less importance and relevance.

Metagenomics and soil function

Metagenomic approaches characterize communities on the basis of the relative abundance of genes and aim to provide a comprehensive, holistic view of these communities, although the focus is often on

© 2015 Macmillan Publishers Limited. All rights reserved

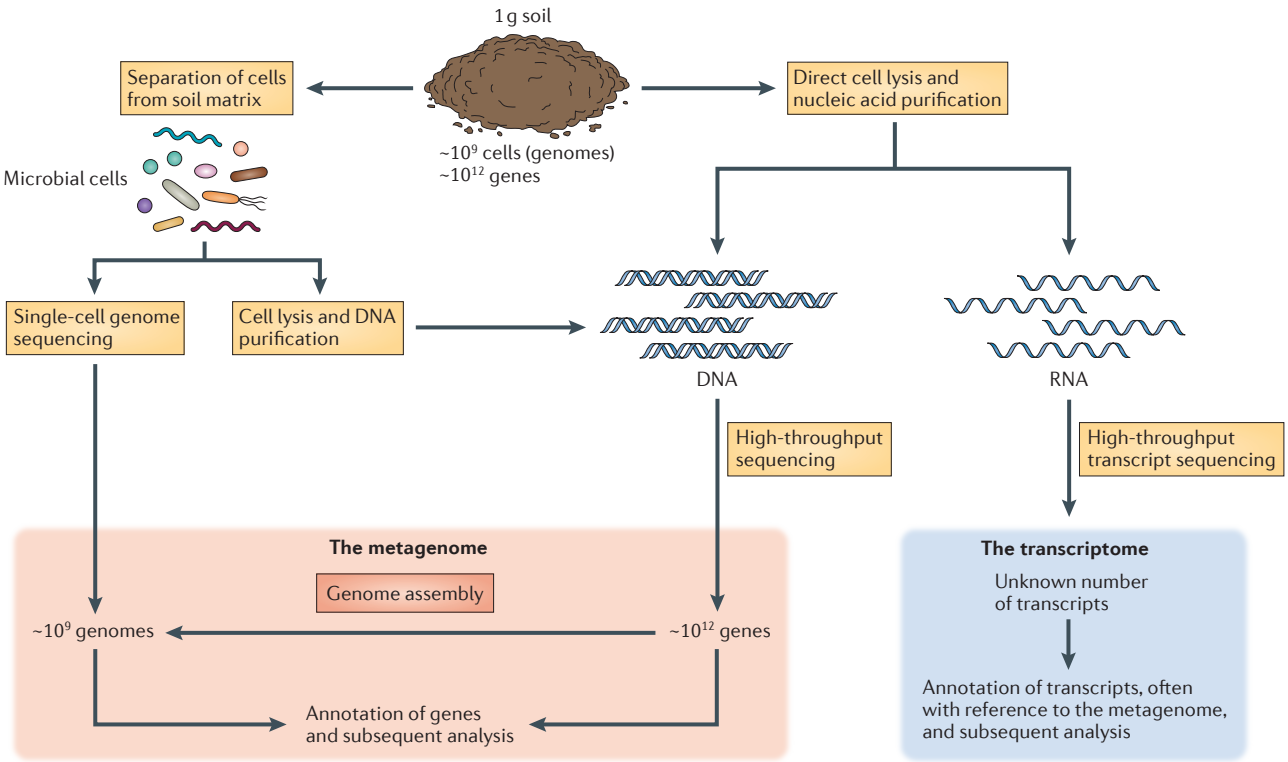

Figure 1 | **Metagenomic and metatranscriptomic analyses of soil samples.** Schematic representation of the main stages involved in generating metagenomic and metatranscriptomic libraries from 1 g of soil. The abundances of microbial cells and genes vary considerably in soil, but they are typically in the order of the amounts indicated here. Notably, the total number of transcripts in 1 g of soil at any particular time is difficult to estimate.

only a fraction of the physiological functions represented. Soil metagenomics is susceptible to limitations and biases that are common to all molecular techniques (such as those associated with cell lysis, nucleic acid extraction, stability and extractability of extracellular nucleic acids, and sequencing errors[28]), and there are major limitations in reliable annotation and quantification of sequenced genes. These limitations are not trivial, so complete coverage of a soil metagenome is not usually obtained; for example, in one study, it was shown that deep coverage of the majority of a soil community was not achieved, even with 300 Gbp of sequence data[29].

Qualitatively, the presence of a functional gene is not evidence of its activity. The host organism may be dormant, inactive or only active when alternative metabolic pathways that do not require the function of the detected gene are expressed. The gene may be cryptic, the gene transcript may not be translated, and the prevailing environmental conditions (including substrate concentration, temperature, pH and water availability) may inhibit activity. Soil microbial communities contain many genes that encode mutually exclusive processes; for example, nitrification and denitrification are favoured in aerobic and anaerobic conditions, respectively, and predicting process rates from the abundances of functional genes associated with each pathway is implausible without information on oxygen concentration and substrate availability. In addition, a single enzyme may participate in several distinct physiological pathways and contribute to different ecosystem functions or may itself have more than one function. For example, the enzyme nitrite reductase (encoded by *nirK*), which reduces nitrite to nitric oxide, is involved in ammonia oxidation, nitrite oxidation, denitrification and anaerobic ammonia oxidation. Ammonia monooxygenase (encoded by *amoA*) can oxidize ammonia, as well as methane and a range of organic compounds[30]. The presence of *nirK* or *amoA* genes in a metagenome therefore provides only limited information on the microbial functional groups that are present and on the processes operating within the soil from which they were obtained. This severely compromises any attempt to link gene-centric metagenomes to ecosystem function or soil characteristics. However, genome-centric metagenomic analyses may be of some value, as they provide information about the pathways that are associated with specific functional genes within a single cell.

Quantitatively, metagenomic data are usually presented as the relative abundance of functional genes, and we might be expected to infer relative activity — that is, the relative activity of one enzyme compared with that of another; for example, the activity of ammonia monooxygenase compared with that of DNA gyrase or citrate lyase. Indeed, the rationale for applying metagenomics is generally to assess whether differences in functional gene abundance are associated with, or predict, differences in the rates or potential rates of associated processes (for example, the differences in the functional characteristics of soil following fire[31], chitin amendment[32] or the presence of *Sphagnum* spp.[33]). It is also the basis for correlation-based studies and studies investigating niche specialization (see below). Thus, caution is required when assuming links between the relative abundance of genes and microbial activity. For example, the addition of ammonium to soil may stimulate nitrifier growth and lead to an increase in the abundance of associated functional genes, thereby reducing the relative abundance of other functional genes (for example, cellulase genes) without necessarily reducing the associated process rates (for example, cellulose degradation). Even if absolute gene abundance

© 2015 Macmillan Publishers Limited. All rights reserved

is known, there is little evidence that this is a reliable, quantitative measure of physiological activity or ecosystem function as it assumes constant enzyme activity per gene, which is unlikely if the same functional gene (and its encoded enzyme) is present in cells with different properties. For example, in cultivated soil, ammonia monooxygenase-specific activity is at least tenfold greater in bacterial ammonia oxidizers than in archaeal ammonia oxidizers because bacterial cells are approximately tenfold larger in size[34]. Furthermore, gene copy per genome varies between different phylotypes, and gene copy per cell will vary (sometimes more than fourfold) with specific growth rate[35] within a single organism. Even if gene abundance accurately reflects the amount of a particular enzyme in the soil, it may give little information about the process rate. Within a cell, flux of material through a pathway depends on the amounts of each enzyme in that pathway[36]. Even in a simple pathway, sensitivity to a particular enzyme may be low and, as a consequence, metabolic flux through the pathway will be insensitive to changes in the abundance of the encoding gene.

Thus, in my opinion, there is no evidence to suggest that gene abundance or relative gene abundance are reliable indicators of the rate of a physiological process or of soil functions, such as denitrification, ammonia oxidation or cellulose degradation. If they were, there would be no pressure to use metatranscriptomics or metaproteomics for this purpose.

A further conceptual issue is the degree to which genes provide information on ecologically important activities. Attention is often focused on genes that are involved in the conversion of substrates into products, but metagenomics provides little information on quantitative physiological characteristics such as maximum specific growth rate; saturation constant; optimum, minimum or maximum pH or temperature for growth; susceptibility to predation; and the speed of recovery after starvation. Attempts to determine the genetic and/or genomic basis of these characteristics are in their infancy and are rare, but attempts have been made to address some characteristics, such as maximum specific growth rate[37] and oligotrophy[38]. All of these characteristics, and many more, are important for the ecology of soil microorganisms and are not currently considered when analysing soil metagenomic data.

To illustrate the importance of these issues, it would not be possible with currently available approaches to use genome sequences to accurately predict, both qualitatively and quantitatively, the composition or activities of a community of microorganisms such as *Bacillus subtilis*, *Streptomyces coelicolor*, *Pseudomonas aeruginosa*, *Escherichia coli* and *Paracoccus denitrificans* from their known genome sequences following repeated subculturing in a defined laboratory medium. Although these are all well-characterized species, the presence of specific genes merely indicates the potential of these species to carry out corresponding functional activities (assuming we know the pathways that the genes are involved in), with little indication of growth dynamics or interactions. So, it is possible to speculate about activities, but there is no information on actual activities or growth rates and, consequently, there is little reliable information on which organisms will persist or be outcompeted. Thus, it is dangerous to assume that a similar approach can provide useful ecological information on 1 million species and 10 billion cells in a 10 g soil sample.

It is also difficult to draw meaningful information from correlations between the physicochemical characteristics of soil and metagenomic data. For example, temporal heterogeneity invalidates any potential links between the abundance of a functional gene and the environmental characteristics that influence both the associated process and the composition of the community containing this gene, such that the gene may be present in previously active cells that are now dormant. Spatial heterogeneity will physically separate substrates from cells that contain a functional gene involved in the metabolism of those substrates[39,40]. Descriptive and correlational studies are commonly described as reflecting an aggregated suite of environmental conditions; however, these studies are often based on measurements taken at a single time point and the methodology

Glossary

cDNA
Double-stranded DNA with a sequence that is complementary to that of the specific mRNA template from which it is synthesized in a reaction catalysed by reverse transcriptase.

Denaturing gradient gel electrophoresis
(DGGE). A gel electrophoresis-based method in which gradients in denaturing conditions (such as temperature, urea concentration or formamide concentration) are used to separate DNA fragments with different mobilities.

Extracellular polymeric material
Material secreted by microbial cells that mainly consists of polysaccharides and proteins; it is also known as extracellular polymeric substance (EPS). This may be released into the growth medium but it often remains attached to the cells and contributes to the formation and function of biofilms.

Genome closure
This process (also known as genome finishing) creates a complete genome by sequencing the gaps that remain when sequenced genome fragments are assembled into overlapping sequences known as contigs.

Maximum specific growth rate
The highest specific growth rate that is attainable under the prevailing environmental conditions with non-limiting substrate concentration and no inhibition of growth.

Neutral theory
A theory that assumes that all phylotypes within a microbial community follow the same rules, regardless of differences in phenotypic properties. The relative abundances of phylotypes and community composition are then determined by random birth, death, speciation and migration.

Nitrifier
Microorganisms that perform the process of nitrification, which is the sequential oxidation of reduced forms of nitrogen (usually in ammonia) to nitrite and then nitrate. In the soil, this process is carried out mainly by archaeal and bacterial ammonia oxidizers and by bacterial nitrite oxidizers.

Oligotrophy
Oligotrophic microorganisms are those that are adapted to exploit substrates present in low concentrations. The term is used to describe organisms that dominate natural habitats in which nutrients are scarce.

Phylotypes
A bacterial or archaeal phylotype is an evolutionarily related group of organisms. When communities are characterized in terms of 16S rRNA or functional gene sequences, phylotypes are defined as those sharing a particular level of sequence identity. Typically, for the 16S rRNA gene, this is 97% or 99% sequence identity, which reflects traditional classification criteria for microbial species.

Saturation constant
The concentration of a growth-limiting substrate at which the specific growth rate of a microbial population is half the maximum specific growth rate.

Seed bank
A term taken from plant ecology, which refers to the total number and diversity of seeds in the soil, including all ungerminated, viable seeds. By analogy, it can also be used to refer to all viable microorganisms within the soil, including those that are rare, inactive or dormant, and therefore constitutes the total richness of the soil microbial community.

Specific growth rate
In the context of a microbial population, this is the rate of increase in biomass relative to, or specific to, the current biomass. It has units of reciprocal time and is constant during exponential growth.

Tortuosity
A measure of the ability of fluid to flow through porous media. A soil with high tortuosity is one in which there is greater resistance to fluid flow owing to a more complex porous structure and greater path lengths.

© 2015 Macmillan Publishers Limited. All rights reserved

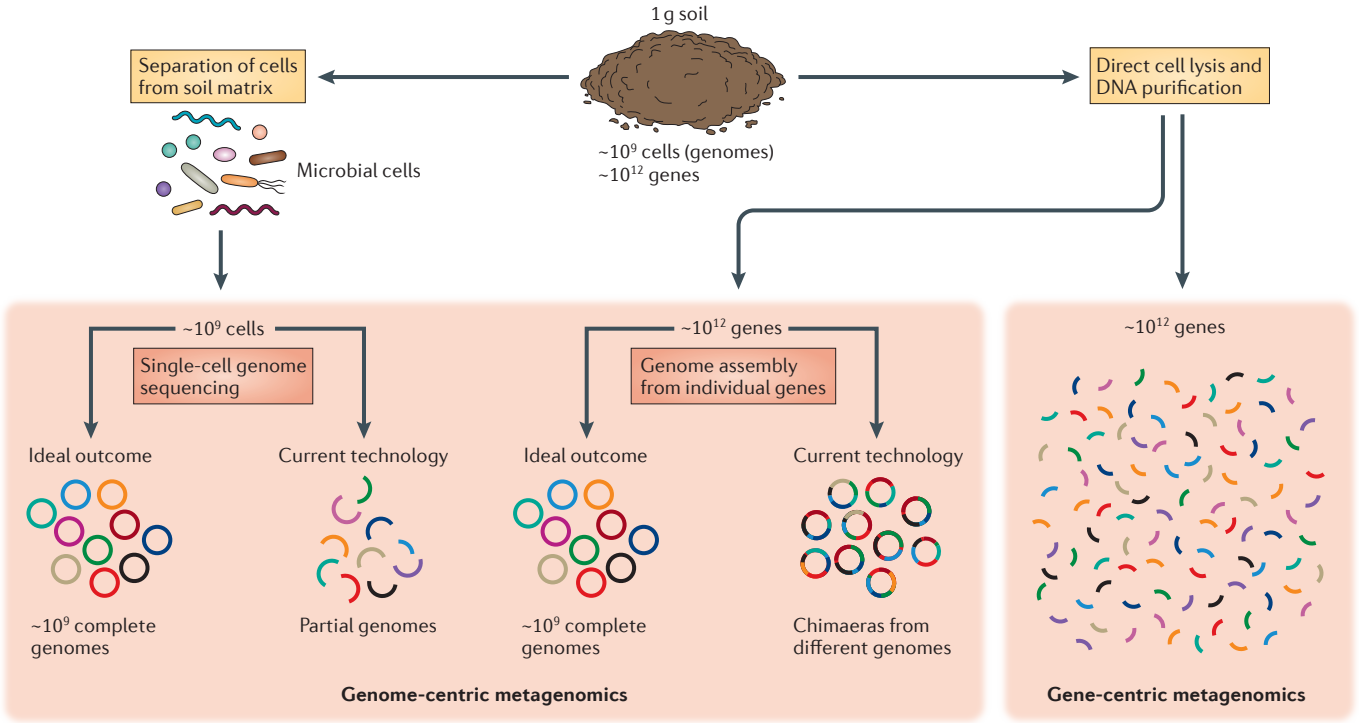

Figure 2 | **Distinction between 'gene-centric' and 'genome-centric' metagenomics.** 'Gene-centric' metagenomics involves the analysis of individual genes that have been isolated from genomes. It is therefore not possible to determine which genes originated from which genome or cell; thus, it is not possible to link phylogeny (based on the 16S rRNA gene or other taxonomic genes) to function or to reconstruct the metabolic pathways that operate in individual cells. By contrast, 'genome-centric' metagenomics aims to obtain complete genomes either through the assembly of individual genes or through single-cell genome sequencing. During genome assembly, there is a risk that chimaeras could form, in which segments of different genomes are assembled, leading to some of the disadvantages of gene-centric metagenomics. A drawback of single-cell genome sequencing is the difficulty in obtaining a complete genome assembly. These disadvantages may be reduced by future technical and bioinformatic developments. Genomes of different colours represent genomes from different cells.

that is used usually destroys spatial structure within the sample, which means that temporal and spatial homogeneity is an implicit assumption rather than a true reflection of the ecosystem. This is a concern if the caveats inherent in this assumption are not accounted for. The consequent problems of interpretation are not restricted to metagenomics, but are highlighted by the dominance of such approaches in metagenomic studies.

Extracting value from metagenomics

Soil metagenomics is currently dominated by descriptive studies that analyse metagenomes in soil samples taken at a single time point and location (snapshot measurements), possibly combined with analysis of soil characteristics and/or compared with snapshot measurements from different locations or treatments (for example, fertilizer addition, pH and soil type). Meaningful comparisons and correlations using such data require the assumption of temporal and spatial homogeneity, as discussed above, and are unlikely to provide useful information on current activity, ecosystem function (for example, process rates of biogeochemical cycling) or the drivers of community structure. Descriptive metagenomics are often justified as providing information on potential activity but, in my opinion, this is of little value unless it is combined with experimental studies that assess this potential in a rational way (see below).

However, the metagenome that is present at any point in time may provide historical information on microbial functions, processes or pathways that have operated in the past and on the environmental conditions that existed at that time. That said, we have virtually no basis on which to interpret the historical information provided by such metagenomic information, but consideration of the question alone can generate interesting avenues for research. For example, for how long are genes stable in soil, which soil characteristics enhance the retention and stability of genes and which characteristics lead to the rapid turnover of genes? If correctly employed, these questions provide avenues for useful, but underexploited applications of omics. It therefore exemplifies an important advantage of the use of omics, in that it has the potential to stimulate thought, leading to the consideration of phenomena and questions about previously ignored or unseen data and to open our minds to novel applications of these techniques.

To provide another example, although it is difficult to relate metagenomic data to soil functions because of temporal and spatial heterogeneity, metagenomic data may be used to characterize the heterogeneity of soil. If there is a link between phylogeny and function, the diversity of a particular functional gene might reflect the heterogeneity of microenvironments containing members of this functional group. This information on heterogeneity has enormous potential value if combined with theoretical analyses of the role of physicochemical and other factors, such as soil structure and tortuosity, in creating and destroying niches.

Metagenomics has much greater potential in experimental laboratory and field studies in which spatiotemporal dynamics and the influence of environmental (and

© 2015 Macmillan Publishers Limited. All rights reserved

other) factors are investigated. For example, if, following the addition of a substrate, a functional gene associated with the use of that substrate increases in abundance, this provides strong evidence of growth of the organisms encoding this gene, which can be used to infer process rate. Temporal changes in metagenome patterns following environmental change may explain why and how particular phylotypes or functional groups respond. Such extrapolation is feasible when sequenced genomes, or at least long reads, are available, and it could lead to the discovery of new pathways associated with the experimental conditions being applied. Genome-centric metagenomic data also enable fine-scale studies of population (rather than community) dynamics and population genetics, although population genetics requires sequencing technology and assembly methods for the construction of genomes from single cells. In other words, there is considerable potential value afforded by metagenomics in experimental studies in which temporal and/or spatial dynamics are investigated in a defined and controlled manner to test specific concepts, hypotheses or theories, or to address specific questions.

Metatranscriptomics and soil function

Metagenomics can provide information about which microorganisms are present (it identifies the seed bank) and what they are capable of doing, whereas metatranscriptomics has the potential to describe what those microorganisms are actually doing and how fast they are doing it. Metatranscriptomics generally involves the extraction of RNA from a microbial community, followed by the sequencing of mRNA (or the synthesis of cDNA, amplification and then sequencing), which generates data that estimate the relative abundance of gene transcripts, often after comparison with metagenomic data from the same sample or environment[17,41].

Technical issues associated with metatranscriptomics are similar to those associated with metagenomics, with further potential bias introduced during the construction of cDNA libraries and a requirement for rapid inactivation of samples to prevent mRNA turnover. In addition, metatranscriptomics generally requires a metagenome 'scaffold', and the identification of transcripts benefits from the availability of genome sequences obtained from pure cultures of microorganisms that are present (preferably in abundance) in the soil sample.

The conceptual issues associated with metatranscriptomics are also similar to those for metagenomics. In addition, a link between a transcript and the activity of its associated enzyme requires assumptions to be made about transcript and protein stability and turnover, and about the existence of environmental conditions that are permissive for metabolic activity. We know little about transcript dynamics, but marine studies indicate that transcripts are degraded within a few minutes following their transcription[42], which has enormous consequences for the links between transcript presence or abundance and metabolic activity. In addition, post-translational processes may be required for enzyme activity, or a stable enzyme may be active long after transcription has been completed and the transcript has been degraded. Similarly, the abundance of transcripts encoding constitutive versus inducible enzymes needs to be considered for correct interpretation of activity.

Extracting value from metatranscriptomics

The presence and abundance of a gene transcript is generally taken as evidence of current or future qualitative and quantitative activity of the enzyme encoded by the gene. More realistically, it provides information on the influence of the environment on a community. Thus, measurement of a change in the abundance of a gene transcript indicates upregulation or downregulation of that gene and potentially the process that it controls, and this change in transcript abundance may thereby provide information on current process rates and/or potential future process rates. The reliability of this information is limited by our lack of knowledge of the dynamics of the production and decay of both transcripts and the products they encode. The abundance of 16S rRNA transcripts provides information on the potential future activity of the taxa identified by this transcript, but it has significant limitations and caveats[43]. For example, the link between rRNA and both growth and metabolic activity is variable between organisms and is complex, such that rRNA levels can be high in dormant organisms, and there is limited information on links between non-growth related processes and rRNA levels.

Similarly to metagenomics, metatranscriptomic analysis of snapshot samples is uninformative. However, it is interesting to consider the potential value that this technique could bring if applied to soil ecosystems. For example, does an increase in transcripts for a functional gene indicate the upregulation of that function? Studies of temporal changes in metatranscriptomes are likely to provide valuable information on the response of a community to environmental change, the metabolic pathways involved in such a response and the potential consequences for ecosystem function. Assessment of the value of metatranscriptomics requires knowledge of transcript and protein turnover and is, obviously, limited to responses to environmental change that are associated with specific genes.

Transcriptional responses to environmental change can lead to the preliminary identification of novel sequences (based on similarities to known sequences or coordinate transcription) that encode products involved in potentially important functional processes. Metatranscriptomic analyses may also provide information on the heterogeneity of niches occupied by an existing community. Is a 'steady state' soil composed of a myriad of niches, many of which contribute to biogeochemical changes? Comparing transcriptomes with corresponding genomes may provide an index of pathway upregulation in response to changing niche conditions. Furthermore, metatranscriptomic data provide information on the interactions between microorganisms and their environment. This includes not only responses associated with nutrient supply and metabolic pathways, but also environmental triggers and responses associated with physiological processes such as the production of extracellular polymeric material and exoenzymes, stress responses and resuscitation following periods of dormancy[44].

Outstanding questions in soil microbiology

Above, I argue that, although descriptive omic studies are limited in advancing our understanding of soil microbial ecology, genome-centric omics and experimental approaches have greater potential to improve our understanding. In this context, I consider experimental approaches as those in which specific questions, concepts or theories of mechanisms controlling observed phenomena are tested in specifically designed laboratory or field experiments. To illustrate this, below I outline some hypothetical scenarios that I consider to be important and for which genome-centric omic analyses could be used to address questions regarding the response of the soil microbiota. I consider the benefits of perfect gene-centric and genome-centric metagenomic data, both with and without perfect metatranscriptomic data.

Impact of environmental change. Both the influence of environmental change on a microbial community and its contribution

© 2015 Macmillan Publishers Limited. All rights reserved

to soil biogeochemistry depend on the extent of the change and on the physiological characteristics of the community (FIG. 3). A small change in temperature (FIG. 3a), pH or a small pulse of substrate is unlikely to change community composition substantially. Physiological plasticity and flexibility within the existing community leads to subtle changes in metabolism and activity, with no perceptible change in community composition. Such a response would not be detectable in metagenomes; however, physiological changes might be detectable in metatranscriptomes. In the absence of complete genomes, metagenomic information would be descriptive only. Separating genes from genomes and from cells makes it virtually impossible to relate changes in gene expression to changes in metabolic pathways or to the phylotype containing the gene. The former can be achieved by combining metatranscriptomic data with genome-centric metagenomes; this facilitates the identification of the members of the community that have generated the response (thereby giving information on functional diversity and redundancy) and the response strategies involved (assuming that these are the result of changes in gene expression).

A moderate change in temperature or another environmental parameter will lead to the selection of organisms that are better adapted to the new conditions (FIG. 3b). If such adaptations and the consequent physiological diversity are linked to phylogenetic diversity, this will be reflected in changes in the relative abundance of different phylotypes. However, such changes can be detected without omics by determining changes in the relative abundance of 16S rRNA or functional genes using 'traditional' molecular techniques ranging from fingerprinting techniques, such as denaturing gradient gel electrophoresis (DGGE), to high-throughput sequencing. Coordinate increases in, and a correlation between, phylogenetic (such as the 16S rRNA gene) and functional genes may indicate mechanisms leading to adaptation and selection but, again, this is made difficult by the separation of genes from genomes and from cells. A combination of genome-centric metagenomics and metatranscriptomics can, however, indicate the metabolic basis for adaptation of particular phylotypes, including the mechanisms involved, and potentially provide a rational approach to discover novel adaptive responses to change.

A large change in environmental conditions, such as a large increase in temperature, is likely to reduce or prevent the activity of large proportions of the community and can lead to activation of dormant organisms or promote colonization by invader species that are adapted to the more 'extreme' conditions (FIG. 3c). Here, gene-centric metagenomes will provide little useful information but genome-centric metagenomes combined with metatranscriptomics could explain community changes and the metabolic characteristics that have enabled invasion. An alternative, or additional, response is adaptation and evolution of the members of the existing soil community. Again, gene-centric metagenomics is of little value for the analysis of mutational events and speciation, but single-cell genome sequencing or reliable (but currently unattainable) fine-scale genome reconstruction open up considerable opportunities for investigating the population genetics of soil organisms through the assessment of mutation rates, lateral gene transfer (including the transfer of plasmids), adaptive mechanisms and speciation.

Understanding soil biogeochemistry. Ultimately, changes in community composition are of little relevance to questions that relate specifically to the soil ecosystem if they do not influence soil ecosystem functions such as biogeochemical cycles, maintenance of soil structure and pest control[45]. As discussed above, genome-centric metagenomic data are a useful indicator of potential activity and the potential contribution of microorganisms to soil biogeochemistry; however, this approach does have some caveats, which have already been considered. Whether measurement of potential (as opposed to actual) activity is valuable depends on the questions being addressed, but experimental studies addressing specific questions or testing theory are required to determine whether, and under which conditions, such potential can be realized.

When combined with metatranscriptomics, genome-centric metagenomics has the potential to provide more accurate information about process rates, but snapshot data remain of little value without knowledge of transcript and protein stability in soil. Experimental studies designed to test, for example, proposed links between phylogeny and function have greater potential to assess the influence of communities on soil biogeochemistry following environmental change, the phylotypes that may be involved in such a response, the extent of functional diversity, and the importance of community stability, resilience and resistance to change (see, for example, REFS 44,46, which describe experiments designed to assess changes in communities and gene expression following the wetting of soil that was subjected to several months of drought).

An alternative approach is to predict the functional characteristics of a soil metagenome from the inputs (such as nutrients and energy) and outputs (such as nutrient transformations, biomass production and carbon dioxide production) from soil and the prevailing environmental conditions. For example, if nitrogen input to and output from a soil sample are characterized, is it possible to predict and quantify the nitrogen-cycle genes required for the nitrogen transformations involved? The approach could be extended to compare different soils, and it could be combined with metatranscriptomics to determine whether changes in input and environmental factors correctly predict changes in functional genes and (for genome-centric metagenomics) in metabolic pathways. In addition to increasing mechanistic understanding, this holistic approach has potential predictive power. It would also enable assessment of the importance of the assumptions underlying current omic studies, such as links between activity and gene abundance, the spatial and temporal homogeneity of ecosystems and whether separating genes from genomes and genomes from cells considerably reduces the value of gene-centric metagenome approaches.

Niche specialization and differentiation. The concept of niche specialization and differentiation is, intuitively, central to the investigation of links between soil heterogeneity and community structure and activity. It links evolutionary and ecological theory and provides the most attractive and widely adopted explanation for the composition of the soil microbial community. Niche specialization is a widely held assumption in soil microbial ecology, with most studies assuming a link between phylogeny and function (see REF. 39 for a discussion of the basis for this assumption). As such, niche specialization is also implicit in the other questions mentioned above; it forms the basis for investigating correlations between metagenomes and soil characteristics and is invoked to explain the response of communities to environmental change and their relationship to ecosystem function. All of these questions require knowledge of mutation and/or recombination, speciation, phenotypic diversity, links between phylogeny and function, environmental selection, dispersal and invasion. For those studies that consider the

© 2015 Macmillan Publishers Limited. All rights reserved

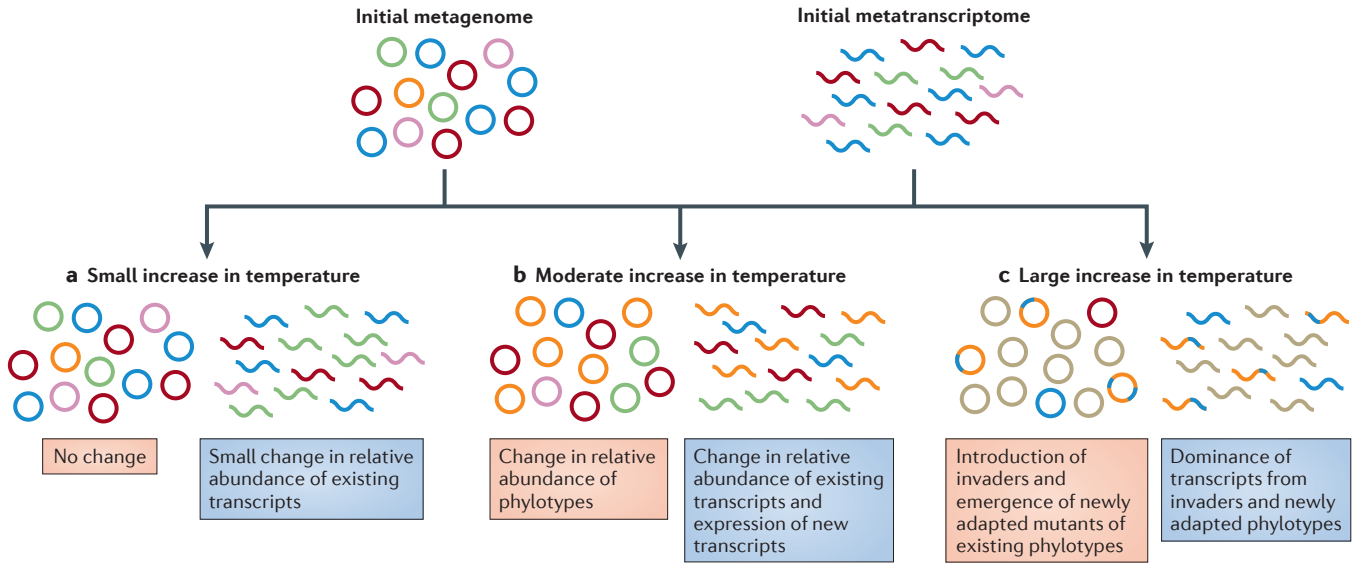

Figure 3 | **Changes in microbial metagenomes and metatranscriptomes following a change in temperature.** Metagenomes and metatranscriptomes are simplified as the most abundant genes (circles) or transcripts (wavy lines), respectively. Different genes and transcripts are represented by different colours. **a** | Following a small increase in temperature, no change occurs in the metagenome, whereas small changes in relative abundance occur at the level of the metatranscriptome. **b** | A moderate increase in temperature leads to a change in the relative abundance of existing genes owing to changes in the abundance of the different phylotypes in the community. At the level of the metatranscriptome, changes in the relative abundance of existing transcripts occur and, in addition, new genes belonging to phylotypes adapted to the higher temperature are expressed. **c** | A large increase in temperature may lead to the invasion of completely new phylotypes that also express new transcripts, and newly adapted mutants of existing phylotypes can emerge and contribute to the metatranscriptome. Similar changes are expected to arise following other forms of environmental change, such as pH change.

effect of the environment on communities, it is often assumed that members of distinct phylogenetic groups have evolved to respond functionally in the same manner to the same environmental change, leading to similar communities in similar environments. For studies that consider the effect of community composition on ecosystem function, a link between phylogeny and function can predict that soils with similar communities will have similar function. Specialization and differentiation result from the generation of genetic diversity followed by selection, which is based on activity rather than potential activity. Similarly, ecosystem function is associated with activity, not potential activity. Therefore, the use of omics to determine potential function and, indeed, measurement of potential activity itself, requires justification.

Neutral theory provides an alternative, or complementary, approach for the study of community assembly[47]. The inability of gene-centric metagenomic data to link phylogeny and function prevents its use in testing these two major theories and also questions whether it constitutes useful evidence for niche specialization. However, genome-centric data does have enormous potential for providing indications of which microorganisms are responsible for which functions and, when used in carefully designed experimental studies, in critically testing these concepts.

Conclusions

The term omics merely encompasses a suite of techniques. Like any other technique, their value is not determined by their engineering or technical brilliance or their cost but rather by their ability to help us test scientific theory and address scientific questions. Typically, but not inevitably, the emergence of a new technique heralds methodological studies that hopefully alleviate and solve current limitations, and these methodological studies are only justified if the technique can ultimately address specific scientific questions. In this article, I have assumed that the technical limitations of omics can be overcome and have focused instead on the conceptual issues associated with their application to soil microbial ecology. I have indicated the degree to which 'perfect' omics technology can address what I believe are important outstanding questions in this field of research. I argue that gene-centric omic analyses, in which phylogenetic genes and functional genes are separated from each other and from their host cells, are of little value in addressing these questions. Their value is reduced further by the current emphasis on snapshot analyses and the lack of identification of meaningful questions and hypotheses.

The real potential of metagenomics and metatranscriptomics in studying the ecology of microorganisms in soil, and in other environments (including marine and freshwater ecosystems and the human gut), will only be realized when genome-centric approaches are used in combination with other relevant techniques in experiments designed to test specific questions and hypotheses. This potential is beginning to be realized in other 'simpler' natural environments where genome-centric approaches are feasible. For example, metagenome assembly has been facilitated by the relatively low diversity of communities in extreme environments, such as acid mine drainage systems[48], hypersaline ecosystems[49] and a biofilm on a hospital sink drain[50]. Similarly, the lower spatial heterogeneity of marine ecosystems has enabled single-cell genome sequencing[51,52] and detailed metatranscriptomic studies[53,54].

The challenges faced by microbial ecologists in general, but especially soil microbiologists, are twofold. The first challenge is to resist the temptation to waste valuable resources on gene-centric metagenomics, particularly on unstructured studies involving snapshot and correlation-based

© 2015 Macmillan Publishers Limited. All rights reserved

approaches that are unlikely to provide any major advances in understanding. Rather, the focus should be placed on more meaningful and valuable genome-centric studies that investigate temporal changes, which can, in turn, address important ecological questions. The second is to explore the potential for omics in identifying new questions and phenomena and in generating new questions, ways of thinking, concepts and theories — all of which improve fundamental understanding and quantitative prediction of the activity and interactions of microorganisms in soil and other ecosystems.

James I. Prosser is at the Institute of Biological and Environmental Sciences, University of Aberdeen, Cruickshank Building, St. Machar Drive, Aberdeen, AB24 3UU, UK.

e-mail: j.prosser@abdn.ac.uk

doi:10.1038/nrmicro3468
Published online 8 June 2015

1. Young, I. M., Crawford, J. W., Nunan, N., Otten, W. & Spiers, A. *Advances in Agronomy* Vol. 100 (ed. Sparks, D. L.) 81–121 (Elsevier, 2008).
2. Raynaud, X. & Nunan, N. Spatial ecology of bacteria at the microscale in soil. *PLoS ONE* **9**, e87217 (2014).
3. Gans, J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**, 1387–1390 (2005).
4. Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput. Biol.* **2**, e92 (2006).
5. Roesch, L. *et al.* Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **1**, 283–290 (2007).
6. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
7. Trevors, J. T. One gram of soil: a microbial biochemical gene library. *Antonie Van Leeuwenhoek* **97**, 99–106 (2010).
8. Will, C. *et al.* Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. *Appl. Env. Microbiol.* **76**, 6751–6759 (2010).
9. Mackelprang, R. *et al.* Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**, 368–371 (2011).
10. Sharon, I. & Banfield, J. F. Genomes from metagenomics. *Science* **342**, 1057–1058 (2013).
11. Stepanauskas, R. Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* **15**, 613–620 (2012).
12. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
13. van Elsas, J. D. & Boersma, F. G. H. A review of molecular methods to study the microbiota of soil and the mycosphere. *Eur. J. Soil Biol.* **47**, 77–87 (2011).
14. Temperton, B. & Giovannoni, S. J. Metagenomics: microbial diversity through a scratched lens. *Curr. Opin. Microbiol.* **15**, 605–612 (2012).
15. Hirsch, P. R., Mauchline, T. H. & Clark, I. M. Culture-independent molecular techniques for soil microbial ecology. *Soil Biol. Biochem.* **42**, 878–887 (2010).
16. Morales, S. E. & Holben, W. E. Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? *FEMS Microbiol. Ecol.* **75**, 2–16 (2011).
17. Carvalhais, L. C., Dennis, P. G., Tyson, G. W. & Schenk, P. M. Application of metatranscriptomics to soil environments. *J. Microbiol. Meth.* **91**, 246–251 (2012).
18. Delmont, T. O., Simonet, P. & Vogel, T. M. Mastering methodological pitfalls for surviving the metagenomic jungle. *Bioessays* **35**, 744–754 (2013).
19. Scholz, M. B., Lo, C.-C. & Chain, P. S. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.* **23**, 9–15 (2012).
20. Wooley, J. C., Godzik, A. & Friedberg, I. A primer on metagenomics. *PLoS Comput. Biol.* **6**, 1–13 (2010).
21. Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. & Hugenholtz, P. A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* **72**, 557–578 (2008).
22. Prakash, T. & Taylor, T. D. Functional assignment of metagenomic data: challenges and applications. *Brief. Bioinform.* **13**, 711–727 (2012).
23. Thomas, T., Gilbert, J. & Meyer, F. Metagenomics — a guide from sampling to data analysis. *Microb. Inform. Exp.* **2**, 3 (2012).
24. Monier, J.-M. *et al.* Metagenomic exploration of antibiotic resistance in soil. *Curr. Opin. Microbiol.* **14**, 229–235 (2011).
25. Kakirde, K. S., Parsley, L. C. & Liles, M. R. Size does matter: application-driven approaches for soil metagenomics. *Soil Biol. Biochem.* **42**, 1911–1923 (2010).
26. Banik, J. J. & Brady, S. F. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opin. Microbiol.* **13**, 603–609 (2010).
27. Schloss, P. D. & Handelsman, J. Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* **14**, 303–310 (2003).
28. Lombard, N., Prestat, E. & van Elsas, J. D. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol. Ecol.* **78**, 31–49 (2011).
29. Howe, A. C. *et al.* Tackling soil diversity with the assembly of large, complex metagenomes. *Proc. Natl Acad. Sci. USA* **111**, 4904–4909 (2014).
30. Arp, D. J. & Stein, L. Y. Metabolism of inorganic N compounds by ammonia-oxidizing bacteria. *Crit. Rev. Biochem. Mol. Biol.* **38**, 471–495 (2003).
31. Tas, N. *et al.* Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *ISME J.* **8**, 1904–1919 (2014).
32. Bragina, A. *et al.* The *Sphagnum* microbiome supports bog ecosystem functioning under extreme conditions. *Mol. Ecol.* **23**, 4498–4510 (2014).
33. Jacquiod, S., Franqueville, L., Cécillon, S., Vogel, T. M. & Simonet, P. Soil bacterial community shifts after chitin enrichment: an integrative metagenomic approach. *PLoS ONE* **8**, e79699 (2013).
34. Prosser, J. I. & Nicol, G. W. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol.* **20**, 523–531 (2012).
35. Wang, J. D. & Levin, P. A. Metabolism, cell growth and the bacterial cell cycle. *Nat. Rev. Microbiol.* **7**, 822–827 (2009).
36. Kacser, H. The control of enzyme systems *in vivo*: elasticity analysis of the steady state. *Biochem. Soc. Trans.* **11**, 35–40 (1983).
37. Vieira-Silva, S. & Rocha, E. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* **6**, e1000808 (2010).
38. Lauro, F. M. *et al.* The genomic basis of trophic strategy in marine bacteria. *Proc. Natl Acad. Sci. USA* **106**, 15527–15533 (2009).
39. Prosser, J. I. Ecosystem processes and interactions in a morass of diversity. *FEMS Microbiol. Ecol.* **81**, 507–519 (2012).
40. Schimel, J. P. & Schaeffer, S. M. Microbial control over carbon cycling in soil. *Front. Microbiol.* **3**, 348 (2012).
41. Tveit, A., Schwacke, R., Svenning, M. M. & Urich, T. Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *ISME J.* **7**, 299–311 (2012).
42. Moran, M. A. *et al.* Sizing up metatranscriptomics. *ISME J.* **7**, 237–243 (2012).
43. Blazewicz, S. J., Barnard, R. L., Daly, R. A. & Firestone, M. K. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* **7**, 2061–2068 (2013).
44. Placella, S. A. & Firestone, M. K. Transcriptional response of nitrifying communities to wetting of dry soil. *Appl. Env. Microbiol.* **79**, 3294–3302 (2013).
45. Barrios, E. Soil biota, ecosystem services and land productivity. *Ecol. Econ.* **64**, 269–285 (2007).
46. Placella, S. A., Brodie, E. L. & Firestone, M. K. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. *Proc. Natl Acad. Sci. USA* **27**, 10931–10936 (2012).
47. Ofiteru, I. D. *et al.* Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl Acad. Sci. USA* **107**, 15345–15350 (2010).
48. Yelton, A. P. *et al.* Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. *BMC Genomics* **14**, 485 (2013).
49. Narasingarao, P. *et al. De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6**, 81–93 (2011).
50. McLean, J. S. *et al.* Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc. Natl Acad. Sci. USA* **110**, E2390–E2399 (2013).
51. Thrash, J. C. *et al.* Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J.* **8**, 1440–1451 (2014).
52. Giovannoni, S. J., Thrash, J. C. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).
53. Satinsky, B. M. *et al.* Microspatial gene expression patterns in the Amazon River plume. *Proc. Natl Acad. Sci. USA* **111**, 11085–11090 (2014).
54. Gifford, S. M., Sharma, S., Booth, M. & Moran, M. A. Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J.* **7**, 281–298 (2012).

Acknowledgements
The author is indebted to M. Firestone for invaluable constructive criticism of this article and to the Leverhulme Trust for the award of a Leverhulme Trust Research Fellowship.

Competing interests statement
The author declares no competing interests.

© 2015 Macmillan Publishers Limited. All rights reserved