

FORUM: Microbiology

The life beneath our feet

Our planet's soils teem with microorganisms that regulate processes from crop productivity to carbon sequestration. Molecular analysis contributes hugely to the characterization of microbial communities, but how can we better understand their ecological functions? Two microbiologists discuss the advantages of data-mining approaches versus targeted experiments.

THE TOPIC IN BRIEF

- The diverse activities of mixed populations of soil microorganisms are fundamental to ecological processes (Fig. 1).
- These communities are also important indicators of responses to changed conditions.
- Rapid DNA sequencing and other molecular-analysis technologies can

provide large-scale data on the combined genomes of a microbial community, potentially revealing unanticipated community members or activities.

- But experiments that test hypotheses on microbe–environment associations may allow more direct identification and analysis of these processes.

Exploring Earth's dark matter

JANET K. JANSSON

Investigation of the crucial environmental parts played by microbes has accelerated with the advent of 'omics' technologies, which has been made possible by low-cost and high-throughput nucleic-acid sequencing, and advances in technologies for studying other biological macromolecules (proteins and metabolites). By studying the total composition of DNA (genomics), RNA (transcriptomics), proteins (proteomics) or metabolites (metabolomics) for multiple organisms, it is possible to produce 'meta-omic' data that encompass all microorganisms in a given habitat. So, can omics provide insight into microbial ecology that cannot be achieved using traditional methods? It is important to keep in mind that omics approaches are not themselves the science, but rather tools, and that the key difference between them and other methods is the amount of data they generate — and the potential that these 'big data' have to expand our knowledge.

Sequencing circumvented the need to generate pure cultures of microorganisms for many studies of microbial processes, but also revealed that most of the 'species' detected were unknown. This led to the concept of microbial life as Earth's 'dark matter', analogous to the unknown realms of the cosmos. In many ways, the challenges faced by astrophysicists and microbial ecologists are indeed

similar, and both fields rely on large quantities of data and supercomputing facilities. But, in microbial ecology, a lack of resources with which to analyse these data has led to a bottleneck in scientific discovery. So, should major investments be made in big-data infrastructure to support the analysis of Earth's microbial communities, similar to those afforded to astrophysics?

One critique against the use of omics in microbial ecology is that the resulting data are merely descriptive. The problem with this criticism is that we don't know what we don't know, similar to studies of the cosmos. There are an estimated 10^{24} stars in the Universe and 10^{30} bacteria

“The discovery of a microbial species with potentially novel functions would be just as interesting as the discovery of a star.”

on our planet, and I would argue that the discovery of a microbial species with potentially novel functions would be just as interesting as the discovery of a star. Until we dive in and use the best tools at our disposal to explore a given habitat, we often do not even know what questions to ask, or we may be asking the wrong questions.

There are numerous examples that illustrate how omics investigation of microbial samples from diverse environments can lead to discovery. Before metagenome sequencing of the Sargasso Sea¹, it was not known that Earth's oceans contain a predominance of bacteria with a previously unknown light-harvesting system called proteorhodopsin. Metagenomics

also led to the identification of ammonia-oxidizing archaea², which have subsequently been shown to be active in a variety of habitats. Furthermore, combining metagenomic studies with metatranscriptomic and/or metaproteomic studies can reveal which microbial functions are expressed under certain conditions. An example of this approach was the finding of active, yet uncultivated, alkane-degrading *Oceanospirillales* bacteria in the deep sea following the 2010 Deepwater Horizon oil spill in the Gulf of Mexico³.

Another promise of omics is that the data will themselves be hypothesis-generating. The Earth Microbiome Project⁴, which aims to systematically categorize the identities and functions of microbial communities across the planet, is illustrative of this point. One of the project's underlying hypotheses is that certain environmental features are correlated with specific combinations of microbial species, and that knowledge of these patterns can be used in a predictive capacity. For example, studies of the temporal variability of microorganisms in the English Channel, combined with environmental data, led to the proposal that such information could be used to predict the seasonal variability of specific microbes and their metabolic products, and this was shown to be the case⁵.

Although characterizing metagenomes from habitats with high microbial diversity, such as soil, remains challenging, the approach promises to provide a more comprehensive view of the composition and function of their microbial communities than has been possible to date. A recent example was the use of metagenomics to determine that permafrost-soil microbial communities are strongly affected by a short-term thaw⁶. A draft genome of a novel methane-producing bacterium was assembled from the permafrost soil, leading to the hypothesis that these bacteria have a major role in generating methane as the permafrost thaws and suggesting an avenue for further experimentation to test this hypothesis. As these tools are further refined and validated, they should lead to the discovery and understanding of more of our planet's microbial species and capabilities.

Janet K. Jansson is in the Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. e-mail: jrjansson@lbl.gov

Think before you sequence

JAMES I. PROSSER

Descriptions of soil microbial communities have been transformed by sequencing and, increasingly, by omics approaches. We must marvel at the technological advances that have made this possible, but they have not been matched by increases in our understanding of the ecology of such communities or links between community composition, diversity and ecosystem function. A potential reason, in my opinion, is an overemphasis on descriptive approaches to soil-microbial ecology compared with hypothesis-driven experiments.

Debates on the relative value of descriptive and data-mining approaches versus hypothesis-driven science are not new. The typical conclusion is that these methods are complementary, rather than mutually exclusive^{7,8}. Hypotheses are constructed to explain observed phenomena, and they may be influenced by existing knowledge, a situation even accepted by philosopher Karl Popper, who wrote⁹: “Some scientists find, or so it seems, that they get their best ideas when smoking; others by drinking coffee or whisky. Thus there is no reason why I should not admit that some may get their ideas by observing, or by repeating observations.”

Hypotheses lack value, however, if they are based solely on observations, or if they are relevant only to the data used to construct them. They are worthwhile if they incorporate novel ideas and flashes of inspiration; they can propose (ideally universal) explanations and mechanisms; and they generate predictions that can be tested by experimentation. It is this process, and not the initial observations, that truly increases understanding. Hypothesis-driven research can thus provide counter-observational, non-intuitive predictions and conceptual frameworks, and can indicate which techniques are, and are not, needed to test them.

Omics descriptions of microbial communities are therefore not analogous to attempts to determine the nature of dark matter in the Universe, which was discovered by, and is being investigated using, hypothesis-driven research. In practice, purely descriptive studies of microbial communities are rare. Most genomic studies compare the sequences present in different, or differently treated, soils, often alongside descriptions of the soils' properties and correlations between these sequences



Figure 1 | Data mining. The abundant microorganisms in Earth's soils perform myriad ecosystem services, many of which are still poorly understood or remain unrecognized. The best ways of identifying and studying these processes is a topic of debate in the ecology community.

and soil properties. Such studies are pointless unless they are used to generate hypotheses, but they are generally driven by a question, although this is often unstated or framed only vaguely. So these studies are, even if unconsciously, based on hypotheses — for example, that gene sequences provide useful information on microbial identity and function, and

“A hypothesis generated by a descriptive study lacks value, and does not increase understanding, unless it is tested by experiment.”

that soil properties influence the assembly and activity of the resident microbe communities. The ability of descriptive studies to increase understanding is obviously limited by the techniques used, which may or may not be relevant to underlying mechanisms. In addition, the non-explicit nature of the ideas that underlie such studies can compromise attempts to explain their findings, for example when assumptions are generated retrospectively or the experimental design is inadequate. It is also noteworthy that data used to generate a hypothesis cannot also be used to test the hypothesis or assess its value. Thus, a hypothesis generated by a descriptive study lacks value, and does not increase understanding, unless it is subsequently tested by experiment.

Unfortunately, the appetite for experimental testing seems to be smaller than that for further descriptive sequencing, even though the *raison d'être* for omic approaches is their greater potential explanatory power. This

may result, in part, from the reduced costs of sequencing. Generating large quantities of sequence data is inexpensive and relatively simple, and it is easier to describe and compare data than to construct and test hypotheses.

Of course, there are examples of hypothesis-driven omics research in microbial ecology. There is also no shortage of questions, ideas, concepts and ecological theory that omics can address, test and extend. Indeed, the enormous complexity and heterogeneity of the soil environment demands courage and intellectual effort in the construction and explicit stating of hypotheses, and critical and focused experimental testing of predictions. It seems to me that this approach offers a more efficient way to use limited resources than the relentless cataloguing and correlation of sequences from ever more soils. In my opinion, it is more challenging and requires greater thought — and is also more enjoyable and rewarding. ■

James I. Prosser is at the Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK. e-mail: j.prosser@abdn.ac.uk

1. Venter, J. C. *et al.* *Science* **304**, 66–74 (2004).
2. Treusch, A. H. *et al.* *Environ. Microbiol.* **7**, 1985–1995 (2005).
3. Mason, O. U. *et al.* *ISME J.* **6**, 1715–1727 (2012).
4. www.earthmicrobiome.org
5. Gilbert, J. A. *et al.* *ISME J.* **6**, 298–308 (2012).
6. Mackelprang, R. *et al.* *Nature* **480**, 368–371 (2011).
7. Kell, D. B. & Oliver, S. G. *BioEssays* **26**, 99–105 (2004).
8. Casadevall, A. & Fang, F. C. *Infect. Immun.* **76**, 3835–3836 (2008).
9. Popper, K. R. *Realism and the Aim of Science* (ed. Bartley, W. W. III) 36 (Routledge, 1983).