

# Soil genomics

Brajesh K. Singh, Colin D. Campbell, Soren J. Sorenson and Jizhong Zhou

The recent editorial on the TerraGenome consortium by Vogel *et al.* (TerraGenome: a consortium for the sequencing of a whole-soil metagenome. *Nature Rev. Microbiol.* 7, 252 (2009))<sup>1</sup> elegantly highlighted the needs, requirements and challenges for sequencing the genomes of soil microorganisms in line with human genome and marine metagenomic projects. Microorganisms are the most diverse and abundant group of organisms and constitute 60% of the Earth's biomass. A current global estimate suggests that soil contains  $4\text{--}5 \times 10^{30}$  microbial cells (excluding viruses), 10 times more than the seas. In addition to reservoirs of industrial products worth UK£100s of billions, microorganisms play vital parts in biogeochemical cycling and sustainability<sup>2</sup>. Therefore, understanding microbial community structure, diversity, functions and stability is essential to our understanding of evolution, community formation and sustainability of life on the Earth. However, obtaining this information has been difficult, owing to our inability to grow microorganisms in laboratory conditions. It is estimated that >99% of microorganisms are currently unculturable under laboratory conditions<sup>3</sup>. Recent developments in technologies such as metagenomics offer a real opportunity for discoveries in the fundamental science of evolution and community formation.

Although large-scale genome sequencing has become cheaper, metagenomics is still an expensive and time-consuming approach. Based on current estimates of the diversity of soil microbial communities, it is estimated that at least 2 million metagenomic clones will need to be sequenced before we can consider statistically that all of the initial genomes have been included in the metagenomic library<sup>4</sup>. A large and interdisciplinary consortium is therefore needed to accomplish such an objective, and TerraGenome provides an important start for such an initiative if it is able to secure funding. However, for all practical purposes, this is just a first step forward, and many more concerted approaches will be needed in the future before we can even start to understand soil microbial diversity and community structure. The site (Park Grass) will provide a well-studied and highly relevant reference site. However, there is huge variability in the soil microbial community owing to space, time and management, and

therefore many more soils need to undergo metagenomic analysis. Soils in different locations are vastly different in their properties at the landscape scale, but heterogeneity in habitat also occurs at the micro-scale. This creates habitats within habitats at a range of scales from microenvironment (rhizospheres) through aggregates, soil profiles, profile sequences and management units (fields) to landscapes. However, because of cost and logistical reasons, metagenomics for many types of soils is not possible yet.

What is the way forward then? Several countries (for example, the United Kingdom, France, Ireland and the Netherlands) have recently included soil microbial community analysis in their national soil surveys. These surveys also include information on land use, and soil chemical, physical, biological and geological information. The use of microbial data produced by low-resolution DNA fingerprinting (for example, TRFLP (terminal restriction fragment length polymorphism) and ARISA (automated ribosomal intergenic spacer analysis) is an imperfect but good way to sort soils into different groups or microbial habitat types. We already know that some soil types with distinct properties determine rare above-ground biodiversity because they have constrained conditions that suit particular plants. But what about the below-ground biodiversity? Are rare (low extent) soil types with extreme or non-normal soil properties more likely to have highly different microbial communities than the more common, highly managed soils?

There are two strategies to tackle this issue. One (the strategy undertaken by TerraGenome) is to characterize a single soil metagenome in detail; such soil can then be used as a reference soil for future metagenomic analyses. However, a parallel strategy might also help us to characterize and understand soil biodiversity, and could be used to identify the greatest dissimilarities. With the invention of new large-scale sequencing technology (for example, 454 sequencing), it is now possible to use sequencing to estimate the microbial diversity and community structure for a substantial number of soil samples. Data from the sequencing can then be used to group soils further and, on this basis, a manageable number of soils can be selected

for further metagenomic analysis. At the Macaulay Institute we are responsible for the National Soil Inventory of Scotland (NSIS). The different geological parent materials, the cool, wet climate and the different land use and vegetation mean that soil resources in Scotland differ from those in many parts of Europe but still represent a broad range of geodiversity. We are profiling samples collected on a  $20 \times 20$  km grid for bacterial, fungal and archaeal communities, and for several functional genes involved in carbon and nitrogen cycling, mainly using TRFLP and limited cloning and sequencing. On the basis of this result we can select contrasting soils for large-scale sequencing (for structural diversity) and analysis by Geochip (for functional diversity)<sup>5</sup> to unravel the community structure of soil microorganisms. For this we have established and launched an international collaboration that includes groups from the Macaulay Institute (United Kingdom), the University of Copenhagen (Denmark), the University of Oklahoma (United States) and the University of Florida (United States) to work on the genomics of diverse soils. Several such small consortia will be needed to group soils from different geographical regions of the world. Such small consortia complement the work carried out by large consortia such as TerraGenome by feeding new information on community composition and its relationship with a wider range of soil chemical, physical and biological characteristics.

Similarly to our approach, systematic soil survey and community analysis of DNA is being carried out on a national scale in many countries, including France, Ireland, the Netherlands and the United Kingdom. This will provide an emerging opportunity to classify soils on the basis of habitat characteristics or microbial community structure at a regional to global scale. A major challenge will be to harmonize the methodologies and approaches for comparing global soil biodiversity. However, there is some flexibility in this approach because, in most cases, soil samples have been archived and this facilitates additional or repeat measures using consistent and agreed methods. Consequently, systematic surveys do present an opportunity to achieve harmonization and continuity of approaches if we wish to carry out a globally coordinated investigation.

Other aspects also need to be addressed. For example, there is an urgent need to sensitize funding agencies to the importance, advantage and necessity of such an

approach. Most of the funding agencies are hesitant to support soil metagenomics work owing to the high costs and implicit nature of the identified benefits. Sustained support for this activity is vital for the advancement of national and international science and practical outcomes that can benefit society. There is arguably a crucial need for a global inventory which contains metagenomics data that are freely available to all who need access. The Human Genome Project, for example, is a good model for such a database. Involvement of government agencies, policy makers and the general public from the onset and recognition that this approach would provide common societal benefits would go a long way to sustaining the extensive funding needed for soil metagenomic studies. TerraGenome is a great idea and deserves full support, but we also need to think about the diverse nature of soils and about future in-depth studies. There is, therefore, a great opportunity here to link both approaches together with forward planning and activity. Let's join up and think even bigger!

*Brajesh K. Singh and Colin D. Campbell are at the Macaulay Land Use Research Institute, Aberdeen, AB15 8QH, UK.*

*Soren J. Sorenson is at the Institute of Biology, University of Copenhagen, DK-1307 Copenhagen, Denmark.*

*Jizhong Zhou is at the Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma 73072, USA.*

*Correspondence to B.K.S.  
e-mail: [b.singh@macaulay.ac.uk](mailto:b.singh@macaulay.ac.uk)*

1. Vogel, T. M. *et al.* TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature Rev. Microbiol.* **7**, 252 (2009).
2. Curtis, T. P. & Sloan, W. T. Exploring microbial diversity — a vast below. *Science* **309**, 1331–1333 (2006).
3. Singh, B. K., Millard, P., Whiteley, A. S. & Murrell, J. C. Unravelling rhizosphere–microbial interactions: opportunities and limitations. *Trends Microbiol.* **12**, 386–393 (2004).
4. Ginolhac, A. *et al.* Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* **70**, 5522–5527 (2004).
5. He, Z. L. *et al.* Geochip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* **1**, 67–77 (2007).