TerraGenome: An International Public Consortium for the Complete Sequencing of a Reference Soil Metagenome.

Pascal Simonet
Environmental Microbial Genomics Group, Laboratoire AMPERE
Ecole Centrale de Lyon, Université de Lyon, 36 avenue Guy de Collongue, 69134 Ecully, France.
Pascal.simonet@ec-lyon.fr

Abstract: The purpose of this presentation is to raise awareness in the wider international scientific community and funding bodies about the third global metagenomics sequencing initiative on our planet; i.e. to sequence the soil metagenome, following in the footsteps of similar initiatives for the marine environment and the human microbiome. The living soil is fundamental to all higher life on planet Earth and is of vital importance for terrestrial processes that determine our present life quality, including soil fertility, carbon and nutrient cycling and desertification. Given the importance of soil functions to most aspects of our lives, surprisingly little is understood of its vulnerability to perturbations or its functional resilience, for example, to land use changes or to changes in climate. Although microorganisms are responsible for the key functions in soils, the soil microbial community is one of the most diverse known and is largely unexplored. Soil microbes contain a largely hidden reservoir of genes that may be exploited for novel drugs, for bioremediation of pollutants and for improving the growth of plants or suppressing plant pathogens. Recent developments in high-throughput sequencing methods provide the essential tools for dissecting the information contained in the soil microbial gene pool. Large-scale metagenomic sequencing efforts will provide the data necessary to decipher the intricacies of the soil microbiome, including necessary information to understand microbial community diversity, interaction and function. Due to the magnitude of this task, we propose the establishment of a coordinated international effort to bring the global skills of the scientific community together to focus on sequencing and annotation of the soil metagenome. In this article, we highlight the key developments and challenges that need to be addressed for a large scale soil metagenomic sequencing effort and we call for the establishment of a global forum for debate and knowledge exchange.
Keywords: Soil, metagenome, sequencing, pyrosequencing, soil DNA metagenomic libraries

1. Introduction

Microbes are the foundation of the biosphere and have dominated life on Earth for most of its 4.5 billion year history. Life, as we know it, is completely dependent upon these microscopic life forms and humans, for instance, cannot survive without the rich diversity of microbes inhabiting their own bodies [1] and their surrounding environment. When examining the importance of different ecosystems, soil stands out as the habitat on Earth that harbours by far the largest microbial diversity per unit mass or volume [2, 3]. Soil microbial communities are known to drive major geochemical cycles, to support healthy plant growth, and to degrade organic matter and pollutants. However, little is known about the vulnerability of their key functions and how they respond to (and buffer) human-induced environmental perturbations, such as climate change and land use. The genetic resources in only a single gram of soil are vast, with 1,000 times more sequence information being present than in the human genome. Cultivation-based approaches show that soil harbours diverse antibiosis-related functions, pollutant-degrading bacteria, plant growth promoting bacteria, bacteria resistant to heavy metals and antibiotics, bacteria that can resist extreme conditions and microorganism directly involved in biogeochemical cycling, energy and green house gas (GHG) emissions [4-6]. Although soils and sediments are probably the greatest reservoirs of microbial diversity on the planet, only a fraction (less than 0.5%) has been grown in the laboratory and the complete genome sequence has only been determined for a select few. We have much to learn.

In order to push back the frontiers of our understanding, an intensive investigation of the basic units of microbial organization and their interactions is required. Descriptions of the temporal and spatial dimensions of microbial community structure and the complex gene expression patterns that underlie trophic interactions are fundamental to a more complete understanding of the biosphere. These biosphere descriptions will rely heavily on knowledge of the basic mechanisms contributing to genetic variation and speciation [7-9]. Knowledge of DNA sequence information from soil will be invaluable in this respect. Indeed, genome sequencing has revealed totally unexpected genetic plasticity within and among known microbial species, and horizontal DNA exchange is now recognized to be a major force in the shaping of their genomes and fostering biochemical innovation, both in isolated strains and in functional microbial communities [10, 11].

2. Soil metagenomics and sequencing.

In light of the difficulty to culture most microorganisms in soil, researchers have developed increasingly sophisticated molecular techniques that can explore a wider range of soil biodiversity. These techniques derive their strength from the direct analysis and exploration of soil-extracted microbial community DNA. Soil metagenomics is

the study and exploration of the collective genomes of all organisms present in a particular soil [12]. The basis of soil metagenomics relies on sequencing of DNA extracted from the soil microbial community for subsequent study. Recent advances in high-throughput cloning and new generation sequencing approaches make the prospect of completely sequencing a soil metagenome realistic. To date, there have been only superficial, although ambitious, attempts to explore the soil metagenome and we are still only exploring a fraction of what we need in order to completely understand soil microbial diversity and function. Genomic studies stand at the vanguard of science as technological advances are providing access to the functioning of different biological systems. After the Human Genome Project (3 Gbp), plant genome sequences (3-5 Gbp) and the highly publicized Venter Sargasso Sea marine sequencing effort (6 Gbp) [13], exploration of the most biodiverse environment, soil at about 10 Tbp [14], still remains rudimentary and constitutes a new and ambitious challenge. Just as knowledge of the human genome promises to revolutionize medical science, the application of genomic technologies to microbial evolution and environmental biology promises to revolutionize microbial biology.

3. Fundamental and applied interest of the DNA metagenomic approach for soil.

To date, limited resources have been directed towards the sequencing of soil metagenomes, when compared to those devoted to the human microbiome [15] or marine environments [16]. A number of basic questions can be addressed by deep exploration of soil using metagenomics approaches. These include the following: What is the extent of soil microbial diversity? What is the extent of strain and species variation in soil? What fraction of the soil community is active under a given condition? What are the key functions of the community? Who are the dominant and rare community members and what are their relative contributions to ecosystem functioning? What are the effects of perturbations, such as climate change and anthropogenic inputs, (e.g., pollutants and fertilizers) on the composition, activity and function of the soil community? What is the extent of horizontal gene transfer in soil and how is it influenced by different soil conditions? What is the relative abundance and functional significance of different domains of life in soil; eg. bacteria, archaea, fungi, viruses and protozoa? How do the composition of soil and the microbial community vary at a micro- and macro-scale and over time?

Sequencing the soil metagenome will bring considerable economic and environmental value. The soil microbial community is established as a key source, or “goldmine”, for those genes and pathways that encode novel biocatalysts involved in either biosynthetic or biodegradation processes, including development of original approaches for drug discovery [17-19], the degradation of human-made polluting compounds [20-22], the improvement of indispensable bioprocesses in the biotransformation industry, the production of biofuels [6] and the fundamentals of biodiversity and spatial complexity [23]. Sequencing of the soil metagenome will also provide insights into the ecology of microorganisms that are beneficial to, or threaten, crop production, that enhance food security through the development of sustainable agricultural practices and that ensure the quality and provision of ecosystem services.

4. Terragenome: The international consortium to achieve sequencing of a reference soil metagenome.

The complete sequencing of the soil metagenome, i.e. the collective genomes of all microorganisms inhabiting the soil environment is now within reach. Soil microbiologists, microbial ecologists, geneticists, molecular biologists and bioinformaticians will constitute an International Soil Metagenome Consortium, TerraGenome, with the goal of providing the first complete sequence of a soil metagenome over a ten-year period. These metagenome sequence data will constitute the “reference” sequence to which other soils around the world could be compared. Thus, other metagenomic projects devoted to sequencing (parts of) different soil metagenomes throughout the world will be able to use this complete soil metagenome as a scaffold for annotation of “core” genes representing common soil microorganisms in other soils and as a basis for estimating differences in diversity, completeness and richness between soils. The consortium will use the combination of metagenomics approaches and broad-scale sequencing to open a totally new era in soil microbiology with advances ranging from detection of climatic indicators and greenhouse gas production through green chemistry to drug discovery; from correlating biodiversity and function to predicting the biosphere’s resilience to human-induced perturbation. The soil system chosen for investigation, Park Grass, Rothamsted (UK), is a charismatic, internationally-recognized resource [24]. This unique long-term ecological site (LTER) includes ongoing experiments that have been running for over 140 years. The research center at Rothamsted provides a history of soil biology and chemistry, as well as an archive of soil samples from detailed studies of different plot treatments. Metadata are available concerning climate, soil use and inputs [25, 26].

The success of soil metagenomics depends to a large extent on intelligent decision-making concerning sample selection, DNA extraction methods, cloning strategies, screening methods, technological advances in sequencing approaches and data management and sharing. Recent advances in methods to capture the vast scale of genetic diversity within soil microbial communities will enable deep metagenomic sequencing. These advances include

methods to dissect the community using DNA or cell extraction fractionation methods that will be optimized to detect both the abundant and the rare members of the community and to distinguish between microorganisms in different physiological states, such as those which are active compared to those which are dormant. The technologies for library construction, functional screening and high-throughput sequence analyses are well established and the aligned developments in bioinformatics are advancing in parallel as are microarray-based tools that can interrogate the diversity and relative abundance of phylogenetic and functional genes. These approaches, coupled with novel bioinformatics methods that contrast metadata (e.g. SEED (http://www.theseed.org)) [16] provide an in-depth analysis of the large amount of sequence and array data that will be generated.

Putting into perspective the challenges we have posed in this article, we need international effort. Each set of a million clones with backup copies of the metagenomic library will require 5,210 386-well plates with a weight of half of tonne and ten -80°C freezers, and it will result in over 30 Gbp of DNA sequence. To meet even this apparently trivial physical storage problem, significant resource and infrastructure is needed to underpin the research and analysis of many research scientists, students and stakeholders. The hope of the authors of this perspectives paper is that our initiative will spur other researchers to collaborate on a global basis and provide input based on the activities of their teams, bringing insight, novel tools and expertise to specifically dissect and explore the microbiology of the Rothamsted field site. This will require an international effort in many fields, including soil ecology, microbial ecology, use of sequencing centres and bioinformatics.

Our global understanding of soil community diversity and function will benefit greatly from a concerted effort to obtain full genomic information from this reference site, as:

- this major endeavour will provide a forum for collaboration, exchange and deposit of publicly available soil metagenome sequences and

- the consortium will provide a benchmark metagenomic resource – clone libraries and sequence data with microbial (bacterial, archaeal, fungal and viral) sequences for current and future study.

Consortium members will participate in the sequencing of clones, sequencing of different target gene families and sequencing of the biota in different micro-scale soil niches. This information will serve as a starting block or platform for other investigators to more deeply explore this particular site and to add increasing levels of information about the composition and function of this reference soil as well as other soils. This initiative will also spur complementary efforts in other “omics” approaches, such as transcriptomics, proteomics and metabolomics of the soil to add different layers of information about gene expression, activity and function. In addition, there will be increasing needs for cultivation of members of the dominantly uncultured microbiota in order to perform biochemical and physiological studies necessary for assigning functions to the vast number of unknown or hypothetical genes that will be found.

This effort cannot be undertaken by a single laboratory or even a single country, and therefore, an international effort is required. This is the goal of the international public consortium “TerraGenome” which will build on the official launch meeting called “Metasted” held on December 13-14, 2008 in Lyon, France.

Acknowledgement

The author would like to thank all those who participated in the MicroEnGen III meeting in Lyon in December 2007 and the Metasted meeting in December 2008 and mainly Timothy M. Vogel, Janet K. Jansson, Penny R. Hirsch, James M. Tiedje, Jan Dirk van Elsas, Mark J. Bailey, Renaud Nalin, Laurent Philippot as members of the Terragenome steering committee.

References

[1] Turnbaugh, P.J. *et al.* 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature.* 444: 1027-1031.

[2] Torsvik, V., Ovreas, L. & Thingstad, T.F. 2002. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* 296: 1064-1066.

[3] Whitman, W.B., Coleman, D.C. & Wiebe, W.J. 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* 95: 6578-6583.

[4] Demanèche, S. *et al.*. 2009. Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. *Appl. Environ. Microbiol.* 75: 534-537.

[5] Dcosta, V.M., Griffiths, E. & Wright, G.D. 2007. Expanding the soil antibiotic resistome: exploring environmental diversity. *Curr. Opin. Microbiol.* 10: 481-489.

[6] Tiedje, J. & Donohue, T. 2008. Microbes in the energy grid. *Science.* 320: 985.

[7] Dopfer, D. *et al.*. 2008. Assessing genetic heterogeneity within bacterial species isolated from gastrointestinal and environmental samples: how many isolates does it take? *Appl. Environ. Microbiol.* 74: 3490-3496.

[8] Bohannon, J. 2008.Microbial ecology. Confusing kinships. *Science.* 320: 1031-1033.

[9] Achtman, M. & Wagner, M. 2008. Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* 6: 431-440.

[10] Sorek, R. *et al.* 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*. 318: 1449-1452.
[11] Bordenstein, S.R. 2007. Evolutionary genomics: transdomain gene transfers. *Curr. Biol.* 17: R935-R936.
[12.] Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5: R245-R249.
[13] Rusch, D.B. *et al.* 2007.The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *Plos. Biol.* 5: e77.
[14] Daniel, R. 2005.The metagenomics of soil. *Nat. Rev. Microbiol.* 3: 470-478.
[15] Blow, N. 2008. Metagenomics: exploring unseen communities. *Nature.* 453: 687-690.
[16] Dinsdale, E.A. *et al.* 2008. Functional metagenomic profiling of nine biomes. *Nature.* 452: 629-632.
[17] Ginolhac, A. *et al.* 2005. Type I polyketide synthases may have evolved through horizontal gene transfer. *J. Mol. Evol.* 60: 716-725.
[18] Lefevre, F. *et al.*. 2008. Drugs from hidden bugs: their discovery via untapped resources. *Res. Microbiol.* 159: 153-161.
[19] Van Elsas, J.D. *et al.*. 2009. The metagenomics of disease-suppressive soils - Experiences from the METACONTROL project. *Trends. Biotechnol.* 26: 591-601.
[20] Handelsman, J. & Wackett, L.P. 2002. Ecology and industrial microbiology: Microbial diversity - sustaining the Earth and industry. *Curr. Opin. Microbiol.* 5: 237-239.
[21] Galvao, T.C., Mohn, W.W. & de Lorenzo, V. 2005. Exploring the microbial biodegradation and biotransformation gene pool. *Trends. Biotechnol.* 23: 497-506.
[22.] Boubakri, H., Beuf, M., Simonet, P. & Vogel, T.M. 2006. Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. *Gene.* 375: 87-94.
[23] Roesch, L.F. *et al.* 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1: 283-290.
[24] Silvertown, J. *et al.* 2006. The Park Grass Experiment 1856-2006: Its contribution to ecology. *J. Ecol.* 94: 801-814.
[25] Tilman, D. 1998. The greening of the green revolution. *Nature* 396: 211-212.
[26] Rasmussen, P.E. *et al.* 1998. Long-term agroecosystem experiments: assessing agricultural sustainability and global change. *Science* 282 : 893-896.