ELSEVIER

INSTITUT PASTEUR

Research in Microbiology xx (2010) 1–9

www.elsevier.com/locate/resmic

Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology

Stefano Mocali*, Anna Benedetti

CRA- Centro di Ricerca per lo Studio delle relazioni tra Pianta e Suolo, Via della Navicella, 2/4, 00184 Roma, Italy

Received 19 November 2009; accepted 13 April 2010

Abstract

Soil is one of the most complex and challenging environments for microbiologists. In fact, although it contains the largest microbial diversity on the planet, the majority of these microbes are still uncharacterized and represent an enormous unexplored reservoir of genetic and metabolic diversity. Metagenomics, the study of the entire genome of soil biota, currently represents a powerful tool for assessing the diversity of complex microbial communities, providing access to a number of new species, genes or novel molecules that are relevant for biotechnology and agricultural applications. In this paper, the onset of new high-throughput metagenomic approaches and new perspectives in soil microbial ecology and data handling are discussed.

Keywords: Metagenomics; Microbial diversity; Soil conservation; Microbial ecology; Genetic resources

1. Introduction

It is well known that soil microorganisms are fundamental for terrestrial processes, playing important roles in various biogeochemical cycles by contributing to plant nutrition and soil health (Arias et al., 2005; Nannipieri et al., 2003; Wardle et al., 2004), even in agricultural and extreme environments (Mäder et al., 2002; Wall and Virginia, 1999). Soil represents the most immense source of microbial diversity on the entire planet, a "hidden" biodiversity which could be a great resource of natural products for agriculture and biotechnological applications. In fact, the total number of prokaryotic cells on earth has been estimated at $4-6 \times 10^{30}$ including $10^{6}-10^{8}$ individual genomes belonging to different species (Sleator et al., 2008). For this reason, accessing and preserving the diversity of soil microorganisms is crucial, for they contain a large pool of unknown genes that encode novel enzymes and proteins.

Direct culture or molecular methods can be used to assess soil microbial diversity. Nevertheless, traditional microbiological approaches present severe limitations, as only a small fraction of the soil bacteria is cultivable using standard methods (Torsvik et al., 1996). Therefore, in the last two decades, several molecular approaches have been proposed (Bloem et al., 2006; Kirk et al., 2004; Kowalchuk et al., 2004; Sorensen et al., 2009; van Elsas et al., 2007) and recently, the exploration of entire genomes present in a soil sample, metagenomics, has provided a new approach for detailed assessment (Daniel, 2005; Handelsman, 2004; Lorenz and Eck, 2005; Langer et al., 2006; Schloss and Handelman, 2005). It is not our goal in this paper to evaluate the state of work in this emerging area of research. Rather, we seek to highlight and discuss both the limitations of traditional microbiological approaches and recent advances in the application of metagenomic and bioinformatic tools to the soil environment. In particular, new research possibilities and novel database developments that hold promise for advancing functional knowledge obtained from metagenomics will be briefly discussed.

2. The difficulty in studying soil microorganisms

Given the importance of soil functions for most aspects of our lives, surprisingly little is known about the subsurface

* Corresponding author. Tel.: +39 (0) 6 7008721; fax: +39 (0) 6 7005711.
E-mail addresses: stefano.mocali@entecra.it (S. Mocali), anna.benedetti@entecra.it (A. Benedetti).

0923-2508/$ - see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2010.04.010

living world. In fact, soil is known to be one of the most challenging environments for microbiologists with respect to microbial community functions and diversity. This is because soil is a heterogeneous environment and we do not fully understand its biogeochemistry or spatial complexity, nor how soil properties and processes such as carbon cycling interact with microbial life. It is known that soil mineral and organic particles strongly interact with soil biota through formation and stabilization of microaggregates, micropores and clay-organic matter complexes which are the dominant structural characteristics of soil matrix and among the most important parameters that affect microbial composition (Girvan et al., 2003). Moreover, as mentioned above, not only is the soil matrix extremely complex and heterogeneous but microbial communities are randomly spread out, following nutrient gradients, moisture content, etc., determining the so-called "hot-spot" distribution (Nunan et al., 2002). Thus, as investigators do not use a standard soil sample size, this may lead to misinterpretations when community structures from different soils are compared. Franklin and Mills (2003) used multiple spatial scales, with sampling intervals ranging from 2.5 cm to 11 m, to study the spatial heterogeneity of soil microbial communities in an agricultural soil. They confirmed that microbial communities may have several nested levels of organization and that they could be dependent on different soil properties or groups of properties. Therefore, since within a few millimeters it is possible to find a variety of microsites with different conditions, operating by traditional soil sampling microbial diversity and functions could be grossly underestimated, resulting in high variability between replicates and low statistical power. For instance, Parkin found that almost all bacterial denitrification in a 98 g sample was situated within a 0.08 g subsample containing plant debris (Parkin, 1987). Therefore, the procedures for soil sampling and sample size are extremely critical when assessing microbial community diversity or functions, and they have to be carefully selected according to the research goals. For example, some authors considered large soil samples (1.0 and 10.0 g) to be appropriate for assessment of bacterial structure, whereas small soil samples (0.01 and 0.1 g) are recommended if new strains of a bacterial community are to be discovered (Ellingsøe and Johnsen, 2002). Klironomos and co-workers suggested using a combination of geostatistical analyses to describe spatial distribution of subsurface microorganisms together with power analyses to assess the required sample size. This approach should reduce variability in samplings and provide a more representative sampling regime (Klironomos et al., 1999).

Our poor knowledge of soil microbial diversity is also limited by our inability to properly study soil microorganisms. In fact, it was shown that 1 g of soil contains more than 10^7 prokaryotic cells (Gans et al., 2005) and it has been assumed that less than 1% of soil prokaryotes are culturable by traditional cultivation and isolation methods (Torsvik et al., 1990a,b).

There are several reasons for microbial unculturability under laboratory conditions, for example, extremely high substrate concentrations or the lack of specific nutrients required for growth. Some authors suggested that "unculturable" bacteria should be called "not-yet-culturable", as novel isolation and culturing methods could successfully culture diverse microbes which may simply be in a physiological state that eludes our actual ability to culture them (Rondon et al., 1999). However, it is also likely that unculturable microorganisms are simply phenotypically and genetically different from the 1% which is amenable to laboratory cultivation and that represents the minority of the population (Rondon et al., 2000).

A broad-scale analysis has evidenced these limitations by estimating that the community genome size in 30–100 cm^3 samples equals the size of 6000–10,000 *Escherichia coli* genomes in unperturbed organic soils and 350–1500 genomes in arable or heavy metal-polluted soils whereas, in contrast, the genomic complexity recovered by culturing methods was less than 40 genomes (Torsvik and Øvreås, 2002; Torsvik et al., 2002). However, these numbers might be underestimated, because they represent just few dominant species rather than large numbers of rare species which might have been excluded from these analyses. In fact, genetic diversity has been determined as a result of the use of an analytical approach that implicitly assumes that all bacterial species in a sample are equally abundant, whereas accurate computational analysis of microbial diversity revealed that rare organisms comprise most of this diversity (Gans et al., 2005; Podar et al., 2007). Nevertheless, the mathematical models still seems to be limited by technical difficulties and biases, as the survey size required for accurate inference analysis of soil microbial diversity is impractically large (Curtis and Sloan, 2005). For instance, for a typical gram of soil containing a billion bacterial cells, accurately estimating diversity in a community with a log-normal species abundance distribution requires sampling at least 10^6 16S rRNA gene sequences, three orders of magnitude larger than current survey efforts, to sample 80% of diversity in a community with 10,000 species (Curtis et al., 2002).

These issues raise the question: what is the significance of research on microbial diversity if it is limited to culturable organisms? In order to overcome such problems, various culture-independent methods have been developed, including phospholipid fatty acid analysis (PLFA) and numerous DNA- and RNA-based molecular approaches (Kowalchuk et al., 2004). Microbial genetic diversity is commonly studied through a number of PCR-based methods which analyze the variability of genes encoding 16S rRNA (18S rRNA for eukaryotes) or ITS regions, considered as "molecular clocks" because they occur in all microorganisms and have well defined and conserved regions for taxonomic classification, but some specific primer systems have also been designed to specifically amplify genes involved in particular metabolic pathways (i.e. *amoA* or *nifH* genes). Since cloning and sequencing are too labor-intensive and time-consuming to be routinely used for large sample sets, the development of molecular fingerprinting techniques to analyze PCR products represented a rapid and powerful tool for understanding the dynamics and diversity of soil microbial communities

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

(van Elsas et al., 2007). However, these approaches are not without their own limitations and biases, mainly related to the characteristics of target gene and PCR amplification efficiency, which have always limited knowledge to a restricted part of the microbial communities (Kirk et al., 2004; Prosser, 2002; Wardle et al., 2004; Wintzingerode et al., 1997).

For these reasons, novel approaches to the exploration of the vast majority of soil microbial diversity are required. Some interesting improvements in culture methods have recently been made (Balestra and Misaghi, 1997; Kaeberlein et al., 2002). For example, Zengler and co-workers presented an impressive method based on the combination of single-cell encapsulation and flow cytometry (or fluorescence-activated cell sorting, FACS) that enabled cells to grow with nutrients that are present at environmental concentrations, and detected microdroplets containing microcolonies of previously uncultured microorganisms (Zengler et al., 2002). Such efforts may help to reduce the need for indirect molecular approaches in the near future, but the current challenge for soil molecular ecology is "metagenomics".

3. The metagenomic approach

Given the enormous utility and importance of soil microorganisms for all biological systems, methods are needed to access huge quantities of information within the whole microbial DNA isolated from a soil sample. This DNA represents the "collective DNA of all the indigenous soil biota" and is referred to as the *metagenome* (Handelsman et al., 1998). In recent years, sequencing of the soil metagenome provided new insights into the ecology of soil microorganisms and proved to be a powerful tool for recovery of novel genes and biomolecules (Daniel, 2005) (Table 1). However, the technical process is not so simple and there are still a lot of challenges.

Conceptually, metagenomics is a simple culture-independent approach and usually consists of cloning and analyzing the microbial DNA extracted directly from an environmental sample. It involves the following main steps: i) isolation of soil DNA; ii) fragmentation and insertion of DNA into appropriate vectors; iii) DNA cloning and transformation of suitable host cells; iv) delivering a metagenomic library; and v) screening of the clone library (Fig. 1).

First of all, the construction of a soil-based library requires sufficient amounts of high-quality DNA which is representative of the soil microbial community (Bertrand et al., 2005). Therefore, appropriate DNA extraction methods and cloning strategies are required in order that entire gene clusters encoding biosynthetic pathway for secondary metabolites can be cloned. In fact, the contaminant substances of soil matrix could interfere with restriction-enzyme digestion and PCR amplification and reduce cloning and transformation efficiency. For these reasons, this step is still considered to be one of the most critical and challenging (Daniel, 2005). Moreover, it was shown that the microbial physiological status in the soil is also an important determinant of soil DNA extraction. For instance, DNA extracted from bacteria that live in a dormant or inactive state appeared to be smaller in size than DNA fragments extracted from active cells (Bertrand et al., 2005). Therefore, a number of enrichment strategies could lead to increasing the proportion of certain bacterial taxa or of bacteria with specific catabolic ability, for example, by using stable isotope probing (Dumont and Murrell, 2005). For more detailed information about isolating, purifying, and cloning DNA methods from diverse soil microbiota, a number of excellent manuals and articles are available, such as the paper presented by Liles et al. (2008).

The second step is the construction of soil libraries by cloning the extracted DNA into appropriate host cells using different vectors. Cloning strategies and the choice of vectors depend strongly on the final objective of the experiment. For instance, if the DNA is used to construct gene banks, large fragments are required in order to minimize the number of clones that need to be cloned, whereas if the DNA is to be used in PCR, DNA yield could be more important than its size. Thus, the following vectors are commonly cloned into host

Table 1
Some examples of screening for novel biomolecules from soil metagenomics libraries.

Biomolecule	Library type	Average insert size (kb)	Number of clones screened	Library size (Mb)	Number of hits	Hit rate (hit per Mb)	Ref.
Esterase/lipase	Plasmid	6	286,000	1, 716	3	1/572	Henne et al., 2000
Esterase/lipase	Plasmid	6	730,000	4380	1	1/4380	Henne et al., 2000
Esterase/lipase	BAC	27	3648	100	2	1/50	Rondon et al., 2000
Oxidation of polyols	Plasmid	3	900,000	2700	15	1/180	Knietsch et al., 2003a
Oxidative coupling enzymes	Cosmid	25	10,000,000	400,000	25	1/16,000	Banik and Brady, 2008
Alcohol oxidoreductase	Plasmid	4	400,000	1600	10	1/160	Knietsch et al., 2003b
Amidase	Plasmid	5	193,000	965	7	1/138	Gabor et al., 2004a
Amylase	BAC	27	3648	100	8	1/12	Rondon et al., 2000
Biotin production	Cosmid	35	50,000	1750	7	1/250	Entcheva et al., 2001
Protease	Plasmid	10	100,000	1000	1	1/1000	Gupta et al., 2002
β-Lactamase	Plasmid	5	80,000	400	4	1/100	Gabor, 2004
Cellulase	Cosmid	22	1700	37	8	1/5	Voget et al., 2006
Antibiotic	Fosmid	35,6	100,000	3560	13	1/274	van Elsas et al., 2008
Antibiotic	BAC	63	12,000	756	4	1/189	MacNeil et al., 2001
Antibiotic	BAC	44. 5	24,546	1092	3	1/364	Gillespie et al., 2002

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

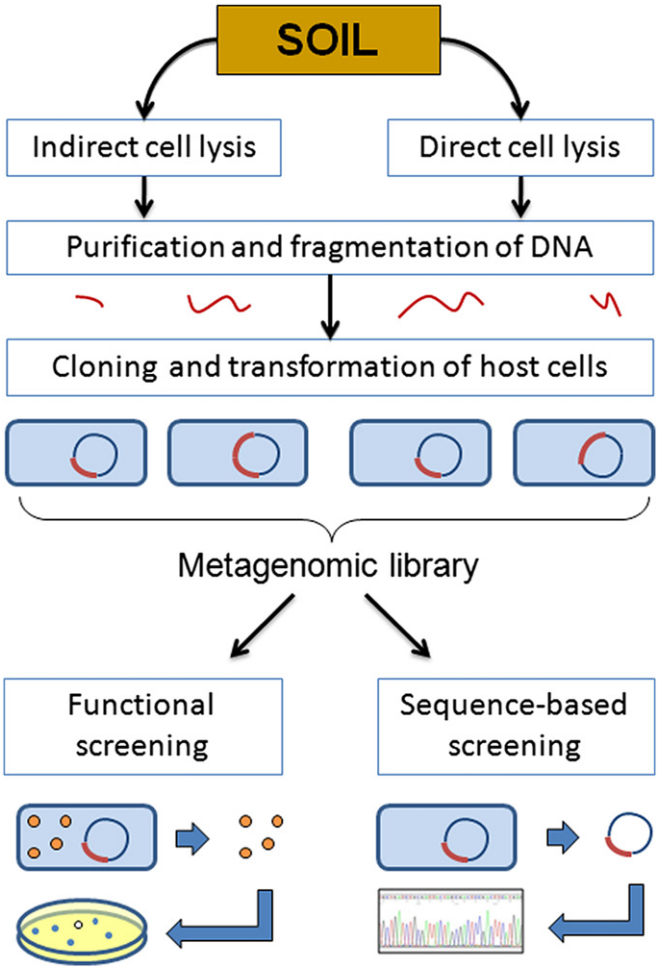

Fig. 1. General scheme of the metagenomic approach. Soil DNA is recovered by direct or indirect lysis of cells from soil particles. Recovered DNA is purified, fragmented and ligated into the linearized cloning vector which is introduced into a suitable bacterial host cell. The delivered library can be screened for specific functions or sequences.

cells (usually E. coli) for construction of metagenomic libraries: plasmids (if the insert is less than 15 kb), cosmids and fosmids (if the insert is less than 40 kb) or BAC vectors (if the insert is more than 40 kb). However, a number of other host strains such as *Steptomyces lividans*, *Rhizobium leguminosarum* and *Pseudomonas aeruginosa* have also been used for the detection of bioactive compounds (Singh et al., 2009a).

Once the library is obtained, it is necessary to proceed with screening of clones. There are two different metagenomic approaches: functional and sequence-based screenings. The first is based on the metabolic activity of clones and the second on nucleotide sequences. In general, a functional approach is used to mine for genes encoding novel enzymes or drugs (Courtois et al., 2003; Fierer et al., 2007; Lorenz and Eck, 2005; Langer et al., 2006; Rondon et al., 2000) and does not require sequence analysis. However, one of the main limitations is that many genes are not or are only poorly expressed in a specific host cell such as *E. coli*. Gabor and co-workers indicated that one possible reason is that the level of metagenomics expression that could function in an *E. coli* host cell is only 40% (Gabor et al., 2004b). Therefore, although novel host strains and novel intracellular types of screens have been developed, such as METREX (Williamson et al., 2005) and SIGEX (Uchiyama et al., 2005), further developments in high-throughput screenings are needed for function-based metagenomics.

On the other hand, sequence-driven strategies involve complete sequencing of clones and are used to explore microbial diversity by analysis of conserved rRNA gene sequences (Riesenfeld et al., 2004) or direct evaluation of shotgun sequencing-derived datasets (Edwards et al., 2006; Manichanh et al., 2008). Microarray technology could be used to analyze metagenomic libraries, providing information on the composition and activity of complex microbial communities (Sebat et al., 2003), but showed 100−10,000-fold lower sensitivity than PCR for gene detection (Zhou and Thompson, 2002). The main problem is that, although novel high-throughput sequencing technologies gave access to the enormous resources of microbial diversity, our ability to assemble sequences recovered from shotgun libraries from complex soil communities decreases dramatically with increased complexity of the community, as reported by Tringe et al. (2005). In this work, about 140 Mb of sequence from farmland soil revealed less than 1% of sequences showing any overlap and produced no contigs, indicating that complete sequencing of such habitats is still not practically feasible. In fact, to date, the main challenges for soil metagenomics were represented by clone numbers and economic costs. Since it was estimated that soil contains approximately 1000 Gbp of microbial genome sequences per gram of soil, more than 10^6 BAC clones (100 kb inserts) are required to represent the entire soil metagenome (Handelsman et al., 1998). Furthermore, as mentioned above, these estimates are based on the assumption that all species are equally abundant, underestimating the true data. Compared with the Human Genome project in which 3 Gbp were sequenced (IHGSC, 2004) and sequencing projects that target microbial habitats, such as the Sargasso Sea (Venter et al., 2004), for which 6 Gbp were sequenced, metagenomic sequencing of soil remains rudimentary and continues to constitute a huge and ambitious challenge.

All metagenomic output is collected and shared across public databases and bioinformatics tools which deal with the enormous amount of data continuously being generated (Table 2). As a consequence, given the quantity of genomic data increasing at an exponential rate, it is imperative that these data be captured electronically in a comprehensive standard format, through open-access and international working bodies. This demand led to the establishment of the Genomic Standards Consortium (GSC) in late 2005 (http://gensc.org/gc_wiki/index.php/GSC) and the first step of this international community was to define the "minimum information about a genome sequence" (MIGS) and the minimum information about metagenome sequence" (MIMS) specifications, with the intent of promoting and discussing the resources that will be required to develop improved mechanisms of metadata capture and exchange (Field et al., 2008).

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

Table 2
Some bioinformatic tools and databases used in metagenomics.

Name	Web site	Short description
AMPHORA	http://bobcat.genomecenter.ucdavis.edu/AMPHORA	Software for phylogenetic analysis of single gene or whole genomes
ARB	http://www.arb-home.de	Interacting software tools for sequence database, maintenance and analysis
CAMERA	http://camera.calit2.net	Metagenomic database of marine and oceanic sequences
CARMA	http://www.cebitec.uni-bielefeld.de/brf/carma/carma.html	Algorithm characterizing the genetic diversity of short-read metagenomes
COG	http://www.ncbi.nlm.nih.gov/COG	Database for phylogenetic classification of proteins encoded in complete genomes
DDBJ	http://www.ddbj.nig.ac.jp	Database tools for collection and analysis of nucleotide sequences
DOTUR	http://www.plantpath.wisc.edu/fac/joh/dotur.html	Sequence assignment to OTUs and richness estimation
EMBL	http://www.ebi.ac.uk/embl	Database tools for collection and analysis of nucleotide sequences
INSDC	http://www.insdc.org	International synchronized collaboration between DDBJ (Japan),GenBank(USA) and the EMBL (Europe) databases
KEGG	http://www.genome.jp/kegg	Database tools for computational prediction of cellular metabolic processes
GenBank	http://ncbi.nlm.nih.gov/Genbank/metagenome.html	An annotated collection database of all publicly available DNA sequences
GOLD	http://genomesonline.org	It provides metadata information related to genome and metagenome projects worldwide
GSC	http://gensc.org/gc_wiki/index.php/Main_Page	A consortium which provides genomic standards and methods for harmonization of metadata collections
MEGAN	http://www-ab.informatik.uni-tuebingen.de/software/megan	Algorithm which supports comparison of multiple and large datasets
Megx.net	http://www.megx.net	Database tools for analysis of Marine Metagenomics
MetaGene	http://metagene.cb.k.u-tokyo.ac.jp	Algorithm for prokaryotic gene-finding from environmental genome shotgun sequences
NAST	http://greengenes.lbl.gov/NAST	Algorithm for creating multiple sequence alignments
MG-RAST	http://metagenomics.nmpdr.org	Bioinformatic tool for phylogenetic and functional analysis of metagenomes
PHACCS	http://biome.sdsu.edu/phaccs	Bioinformatic tool for the analysis of viral metagenomic data
RefSeq	http://ncbi.nlm.nih.gov/RefSeq	Collection of sequences representing genomic data and proteins from 2400 organisms
SILVA	http://www.arb-silva.de	Database for analysis and alignment of high-quality ribosomal RNA sequence data
SINA	http://www.arb-silva.de/aligner	Bioinformatic tool for sequence alignment based on SEED
StrainInfo.net	http://www.straininfo.net	A bioportal of information integration services for the microbial community
TETRA	http://www.megx.net/tetra	A tetranucleotide-based tool correlating large DNA sequences
UniFrac	http://bmf2.colorado.edu/unifrac/index.psp	Comparison of microbial communities using phylogenetic information
XplorSeq	http://vent.colorado.edu/phloware	Compilation, management and phylogenetic analysis of DNA sequences

4. Recent studies and new perspectives in soil metagenomics

Recently, the advent of next-generation sequencing tools has increased the potential of metagenomics and environmental microbiology. In fact, actual shotgun studies using a mass genome sequence followed by scaffold reconstruction and gene annotation has given notable results (Edwards et al., 2006). For example, to date, there are more than 1000 complete or nearly complete genome sequences of microbes available and several impressive sequencing projects have been founded, such as Global Ocean Survey, GOS (http://www.jcvi.org/cms/research/projects/gos), or Genomic Encyclopedia of Bacteria and Archaea, GEBA (http://www.jgi.doe.gov/programs/GEBA).

However, soil metagenomic projects still require great scientific and economic effort because, although the current pyrosequencing technology reads up to 400 bp providing data within a single run of more than 100 million reads and 10 Gb sequence data, the most critical challenges lie in the construction of scaffolds from a huge number of mixed short sequences (Margulies et al., 2005; Roesch et al., 2007).

Recently, an ambitious international consortium was presented by Vogel and co-workers, called TerraGenome (www.terragenome.org) (Vogel et al., 2009a). The project proposed using a metagenomic approach to a reference soil from the Park Grass located in the Rothamsted Institute (UK), which is considered the temple of soil science, as agronomic field experiments have been running for more than 150 years. However the discussion about the advantages and the risks of such initiative has grown (Singh et al., 2009b; Vogel et al., 2009b). In fact, although metagenomics have led to the discovery of new useful molecules (Table 1), poor knowledge of soil complexity, methodological approach biases and heterogeneous microbial spatial distribution make questionable the value of such efforts in linking microbial diversity with soil functions.

To fully realize the potential of soil metagenomics, a number of obstacles currently need to be overcome. Perhaps among the most significant of these are both microbial complexity in most communities and the huge difficulty in managing such amounts of genomic data.

In the first case, as mentioned above, “simple” systems (e.g. extreme environments) are needed to obtain complete metagenomic information from soil (Tyson et al., 2004; Tringe et al., 2005). This strategy was recently used in a METACONTROL project (van Elsas et al., 2008) in which four disease-suppressive soils were considered as naturally enriched in antiphytopathogenic microorganisms. One of the main purposes of this work was the exploitation of microbial strains involved in the production of specific antibiotics. However, despite the fact that these soils had been selected for

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

their anti-phytopathogen properties, functional screening of the metagenomics library led to the detection of on a few positive clones (<0.05%), confirming the difficulties in obtaining high yields of expression signals, as previously reported (Chung et al., 2008; Henne et al., 1999).

Nevertheless, sequence-based screening technologies have always been limited by the huge amounts of sequence data. Assembling so many sequence reads, while simultaneously accounting for heterogeneities between genomes, introduces unique challenges for each study. To date, the major problem in annotating environmental samples is identifying the 99% organisms which have never been cultured and sequenced before, including previously unknown species. Although new databases and informatics tools have been recently presented (Huson et al., 2009; Meyer et al., 2008), further developments in computational analysis are required for interpretation of metagenomic data.

New highly sensitive approaches have been reported for accessing the genetic and functional diversity of uncultured microorganisms, such as the GeoChips technology (He et al., 2007), a comprehensive microarray for studying biogeochemical processes and functional activities of microbial communities. It is particularly useful for providing direct linkages of microbial genes/populations to ecosystem processes and functions. Another approach, multiple displacement amplification (MDA), has been successfully applied to soil metagenomics, leading to increased access to genomic DNA, in particular in low-biomass environments (Abulencia et al., 2008; Binga et al., 2008). Recently, a new sensitive metabolite array, called the "*reactome array*", has been presented (Beloqui et al., 2009): the system represents a tool for genome sequence-independent functional analysis of metabolic phenotypes and networks, the *reactomes*, of cell populations and communities. In that work, a library of 2483 identified metabolites that collectively serve as substrates for all possible reactions described in KEGG and PubMed databases were used to compare the metabolism of three different communities. Each substrate-metabolite was coupled to a quenched dye and to a linker molecule, a Co(II) complex containing a poly(A) tail linked to the array, and the catalytic reaction of an enzyme on the substrate resulted in the release of the product and of unquenched dye which gives a fluorescent signal (Fig. 2). This impressive approach provides a link between metabolome and genome which could lead to the discovery of unknown metabolic activities and enzyme production in soil microbial communities, representing one of the most innovative tools in current microbial ecology.

Another promising approach is based on single-cell metagenomics. For example Marcy et al., (2007) presented an amazing microfluid system to isolate uncultivated TM7 cells from a mixed bacterial community. The device coupled FACS with MDA strategy to shuffle and sort individual bacterial cells and carry out subsequent MDA reactions on isolated bacteria. After selecting TM7 species by FACS, they carried out pyrosequencing on MDA-amplified DNA of the selected isolates that were identified by their 16S rRNA sequences. This approach enables single-cell genetic analysis of any uncultivated minority member of a microbial community and, in our opinion, it could finally overcome our need to culture organisms to gain access to their full genomic and functional potential.

In this scenario, the Genomic Standards Consortium (GSC) is currently leading a community effort to establish a richer, more standardized description of the world's collection of genomes and metagenomes. For example, it is developing the Genomic Contextual Data Markup Language (GCDML), which provides official implementation of MIGS/MIMS describing aspects of genomes and metagenomes, such as geographic location and habitat type from which the sample was taken, as well as details of the processing of a sample from the time of sampling up to sequencing and subsequent analysis. Moreover, GSC is creating a mapping of identifiers describing complete genomes across as many genomic databases as possible so that information about genomes and the organism from which they derive can be more easily integrated. This mapping is named "Genomic Rosetta Stone"

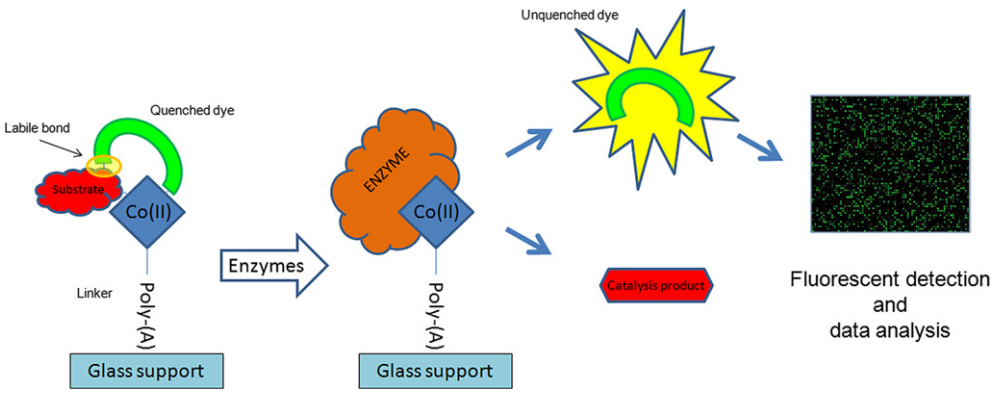

Fig. 2. The reactome strategy. The process initially involved three linked molecules: the enzyme substrate-metabolite, the quenched dye (Cy3) and the linker. The substrate-metabolite is linked to the quencher through a labile nitrogen bond and both are anchored to the Co(II)-containing poly(A) linker by hystidine tags. Enzymatic recognition of the substrate induces a reaction which causes a chemical change and rupture of the labile bond with the release of quenched dye. The consequent release of the reaction product exposes the active cobalt cation that ligates and immobilizes the enzyme on the glass support (array spot). The released dye is no longer quenched and gives a fluorescent signal.

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

(GRS) and it aims to enrich our ever growing data collection of genomic and metagenomic sequences (van Brabant et al., 2008).

As a consequence, the role of international archives and databases, such as the International Nucleotide Sequence Database Collaboration (INSDC, DDBJ/EMBL/GenBank), which provide public and comprehensive access to nucleotide sequencing information, from raw sequencing machine output through to functional annotation, is also progressively changing and moving forward, integrating recent community-focused initiatives and structures for more specific information, for example, the emerging strategy of incorporation of MIGS compliance data into INSDC records.

These trends could represent the foundation for the world's bioinformatics common data infrastructure which is needed to make metagenomic data something more than a simple "megagenomic encyclopedia" without any ecological sense.

5. Conclusions

The enormous potential of metagenomics to access soil microbial diversity, the widest biodiversity on the entire planet, could be a great resource in agriculture and biotechnology. In fact, although the key role of microbial diversity is maintenance of soil quality, productivity and health, traditional culture-dependent and culture-independent microbiological approaches have failed to exploit the genetic resources of this "hidden diversity".

At present, high-throughput sequencing technologies coupled with gene arrays, proteomics, expression-based analyses and traditional methods, can provide insight into both agricultural problems, such as soil fertility and sustainability, plant health and biotechnological processes, and microbial strains or biomolecules of particular interest currently hindered by our inability to culture most microorganisms in pure culture. However, further large-scale metagenomic sequencing efforts will be necessary to resolve the complexity of the soil microbiome and to provide sufficient data to understand soil microbial community diversity and functions. The success of soil metagenomics depends on a combination of appropriate sample selection, efficient DNA extraction methods, (eventual) cloning, screening strategies and sequencing approaches, together with improved open system data management and bioinformatics.

The growing need for integration of massive datasets of metagenomic information with biological information and resources across the scientific community is inducing international genomic databases, such as INSDC, to provide standardized data organization enabling simple retrieval of and computation based on small and large datasets of interest. In this scenario, new open-source bioinformatic tools and internet applications, such as LinkOut resources (http://www.ncbi.nlm.nih.gov/projects/linkout) or StrainInfo (http://www.straininfo.net), will represent opportunities for facilitating access to relevant online resources, including full-text publications, biological databases, Biological Resource Centers (BRC), consumer health information, research tools and more. These efforts could represent an implementation of the "ecosystem approach" established during the seventh ordinary meeting of the Conference of the Parties to the Convention on Biological Diversity (CBD) held in Malaysia in 2004 as an integrating framework for implementation of objectives of the Convention on Biological Diversity (Decision VII/11). This trend in managing metagenomic data and the application of the ecosystem approach will help to attain the three objectives of the CBD: i) conservation; ii) sustainable use; and iii) fair and equitable sharing of benefits arising from the utilization of genetic resources. However, in our opinion, a combination of scientific, technical, legal, institutional and normative efforts will be required to both design an integrated infrastructure in microbial research, such as "Microbial Commons" (http://www.microbialcommons.ugent.be) and implement the "International Treaty on Plant Genetic Resources for Food and Agriculture" (http://www.planttreaty.org) or the activities of Bioversity International (http://www.bioversityinternational.org) with microbial genetic resources.

In fact, environmental sustainability depends on soil health, which is regulated by soil microorganisms. Crop production depends on soil fertility, which is regulated by soil microorganisms. However, much about these microorganisms remains mysterious. Thus, the aim of metagenomics is not simply to provide a collection of billions of biological items, but to offer a unique opportunity to explore how microbial communities interact with both soil and crops and to eventually exploit the power of soil microbial communities to produce healthier and more robust crops or novel biomolecules. Therefore, as soil environment represents the most incredible and unexplored biodiversity source for several biotechnological applications and ecological purposes, the conservation of this "invisible biological treasure" must be pursued no matter what.

References

Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W., Chang, S. H., Chang, H.W., Watson, D., et al., 2008. Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl. Environ. Microbiol. 72 (5), 3291–3301.

Arias, M.E., Gonzalez-Perez, J.A., Gonzalez-Vila, F.J., Ball, A.S., 2005. Soil health-a new challenge for microbiologists and chemists. Int. Microbiol. 8, 13–21.

Balestra, G.M., Misaghi, I.J., 1997. Increasing the efficency of the plate counting method for estimating bacterial diversity. J. Microbiol. Methods 30, 111–117.

Banik, J.J., Brady, S.F., 2008. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. PNAS 105, 17273–17277.

Beloqui, A., Guazzaroni, M.E., Pazos, F., Vieites, J.V., Godoy, M., Golyshina, O.V., Chernikova, T.N., Waliczek, A., et al., 2009. Reactome array: forging a link between metabolome and genome. Science 326 (5950), 252–257.

Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M., Simonet, P., 2005. High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. J. Microbiol. Methods 62 (1), 1–11.

Binga, E.K., Lasken, R.S., Neufeld, J.D., 2008. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. ISME J. 2, 233–241.

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

Bloem, J., Benedetti, A., Hopkins, D., 2006. Microbial Methods Assessing Soil Quality. CABI Publishing, CAB International, Wallingford Oxfordshire, UK, ISBN 0-85199-098-3.

Chung, E.J., Lim, H.K., Kim, J.-C., Choi, G.J., Park, E.J., Lee, M.H., Chung, Y.R., Lee S-, W., 2008. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. Appl. Environ. Microbiol. 74, 723–730.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., et al., 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. 69, 49–55.

Curtis, T.P., Sloan, W.T., Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. PNAS 99 (16), 10494–10499.

Curtis, T.P., Sloan, W.T., 2005. Exploring microbial diversity–a vast below. Science 309, 1331–1333.

Daniel, R., 2005. The metagenomics of soil. Nat. Rev. Microbiol. 3, 470–478.

Dumont, M.G., Murrell, J.C., 2005. Stable isotope probing — linking microbial identity to function. Nat. Rev. Microbiol. 3, 499–504.

Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M., Saar, M.O., Alexander, S., et al., 2006. Using pyrosequencing to shed light on deep mine microbial ecology. BMC. Genomics 7, 57.

Ellingsøe, P., Johnsen, K., 2002. Influence of soil sample sizes on the assessment of bacterial community structure. Soil. Biol. Biochem. 34, 1701–1707.

Entcheva, P., Liebl, W., Johann, A., Hartsch, T., Streit, W.R., 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl. Environ. Microbiol. 67, 89–99.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., 2008. The minimum information about a genome sequence (MIGS) specification. Nat. Biotechnol. 26 (5), 541–547.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.E., et al., 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of Bacteria, Archaea, Fungi, and Viruses in soil. Appl. Environ. Microbiol. 73 (21), 7059–7066.

Franklin, R.B., Mills, A.L., 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS Microbiol. Ecol. 44, 335–346.

Gabor, E.M., de Vries, E.J., Janssen, D.B., 2004a. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. Environ. Microbiol. 6, 948–958.

Gabor, E.M., 2004. Harvesting novel biocatalysts from the metagenome. thesis, Univ. Groningen.

Gabor, E.M., Alkema, W.B.L., Janssen, D.B., 2004b. Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ. Microbiol. 6, 879–886.

Gans, J., Wolinsky, M., Dunbar, J., 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309, 1387–1390.

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M., et al., 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. 68, 4301–4306.

Girvan, M.S., Bullimore, J., Pretty, J.N., Osborn, A.M., Ball, A.S., 2003. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Appl. Environ. Microbiol. 69 (3), 1800–1809.

Gupta, R., Beg, Q.K., Lorenz, P., 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. Appl. Microbiol. Biotechnol. 59, 15–32.

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 (10), 245–249.

Handelsman, J., 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68, 669–685.

He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., Wu, W., et al., 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. 1 (1), 67–77.

Henne, A., Daniel, R., Schmitz, R.A., Gottschalk, G., 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. 65, 3901–3907.

Henne, A., Schmitz, R.A., Bömeke, M., Gottschalk, G., Daniel, R., 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl. Environ. Microbiol. 66, 3113–3116.

Huson, D.H., Richter, D.C., Mitra, S., Auch, A.F., Schuster, S.C., 2009. Methods for comparative metagenomics. BMC Bioinformatics 10 (1), S12. doi:10.1186/1471-2105-10-S1-S12.

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945.

Kaeberlein, T., Lewis, K., Epstein, S., 2002. Isolating "uncultivable" microorganisms in a simulated natural environment. Science 296, 1127–1129.

Kirk, J.L., Beaudette, L.A., Hart, M., Moutoglis, P., Klironomos, J.N., Lee, H., Trevors, J.T., 2004. Methods of studying soil microbial diversity. J. Microbiol. Methods 58, 169–188.

Klironomos, J.N., Rillig, M.C., Allen, M.F., 1999. Designing belowground field experiments with the help of semi-variance and power analyses. Appl. Soil Ecol. 12, 227–238.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., Daniel, R., 2003a. Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. J. Mol. Microbiol. Biotechnol. 5, 46–56.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., Daniel, R., 2003b. Construction and screening of metagenomics libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. Appl. Environ. Microbiol. 69, 1408–1416.

Kowalchuk, G.A., de Bruijn, F.J., Head, I.M., Akkermans, A.D., van Elsas, J. D., 2004. Molecular Microbial Ecology Manual, Second ed. Kluwer Academic Publisher.

Liles, M.R., Williamson, L.L., Rodbumrer, J., Torsvik, V., Goodman, R.M., Handelsman, J., 2008. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. Appl. Environ. Microbiol. 74 (10), 3302–3305.

Lorenz, P., Eck, J., 2005. Metagenomics and industrial applications. Nat Rev Microbiol 3, 510–516.

Langer, M., Gabor, E.M., Liebeton, K., Meurer, G., Niehaus, F., Schulze, R., Eck, J., Lorenz, P., 2006. Metagenomics: an inexhaustible access to nature's diversity. Biotechnol. J. 1, 815–821.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Phillips, T., et al., 2001. Expression and isolation of antimicrobial small molecules from soil DNA libraries. J. Mol. Microbiol. Biotechnol. 3, 301–308.

Mäder, P., Fliebach, A., Dubois, D., Gunst, L., Fried, P., Niggli, U., 2002. Soil fertility and biodiversity in organic farming. Science 296 (5573), 1694–1697.

Manichanh, C., Chapple, C.E., Frangeul, L., Gloux, K., Guigo, R., Dore, J., 2008. A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. Nucleic. Acids. Res. 36, 5180–5188.

Marcy, Y., Ouverney, C., Bik, E.M., Lösekann, T., Ivanova, N., Garcia Martin, H., Szeto, E., Platt, D., et al., 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. PNAS 104, 11889–11894.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L. A., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., et al., 2008. The Metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386.

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G., 2003. Microbial diversity and soil functions. Eur. J. Soil Sci. 54, 655–670.

Nunan, N., Wu, K., Young, I.M., Crawford, J.W., Ritz, K., 2002. In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. Microb. Ecol. 44, 296–305.

Parkin, T.B., 1987. Soil microsites as a source of denitrification variability. Soil Sci. Am. J 51, 1194–1199.

Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L., Keller, M., 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl. Environ. Microbiol. 73 (10), 3205–3214.

Prosser, J.I., 2002. Molecular and functional diversity in soil microorganisms. Plant. Soil. 244, 9–17.

Riesenfeld, C.S., Goodman, R.M., Handelsman, J., 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ. Microbiol. 6, 981–989.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W. , 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1, 283–290.

Rondon, M.R., Goodman, R.M., Handelsman, J., 1999. The earth's bounty: assessing and accessing soil microbial diversity. Trends Biotechnol. 17, 403–409.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., et al., 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66, 2541–2547.

Schloss, P.D., Handelman, J., 2005. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome. Biol. 6 (8), 229.

Sebat, A.L., Colwell, F.S., Crawford, R.L., 2003. Metagenomic profiling: microarray analysis of an environmental genomic library. Appl. Environ. Microbiol. 69 (8), 4927–4934.

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., Bali, V., Batra, N., 2009a. Metagenomics: concept, methodology, ecological inference and recent advances. Biotechnol. J. 4, 480–494.

Singh, B.K., Campbell, C.D., Sorenson, S.J., Zhou, J., 2009b. Soil genomics. Nat. Rev. Microbiol. 7 (10), 756–757. doi:10.1038.nrmicro2119-c1.

Sleator, R.D., Shortall, C., Hill, C., 2008. Metagenomics. Lett. Appl. Microbiol. 47, 361–366.

Sorensen, J., Nicolaisen, M.H., Ron, E., Simonet, P., 2009. Molecular tools in rhizosphere microbiology–from single-cell to whole-community analysis. Plant Soil. 321, 483–512.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., et al., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43.

Torsvik, V., Goksoyr, J., Daae, F.L., 1990a. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56, 782–787.

Torsvik, V., Salte, K., Soerheim, R., Goksoeyr, J., 1990b. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl. Environ. Microbiol. 56, 776–781.

Torsvik, V., Sorheim, R., Goksoyr, J., 1996. Total bacterial diversity in soil and sediment communities–a review. J. Ind. Microbiol. 17, 170–178.

Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5, 240–245.

Torsvik, V., Øvreås, L., Thingstad, T.F., 2002. Prokaryotic diversity-magnitude, dynamics and controlling factors. Science 296, 1064–1066.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., et al., 2005. Comparative metagenomics of microbial communities. Science 308, 554–557.

Uchiyama, T., Abe, T., Ikemura, T., Watanabe, K., 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat. Biotechnol. 23, 88–93.

van Brabant, B., Gray, T., Verslyppe, B., Kyrpides, N., Dietrich, K., Glöckner, F.O., Cole, J., Farris, R., et al., 2008. Laying the foundation for a genomic rosetta stone: creating information hubs through the use of consensus identifiers. Jun. OMICS 12 (2), 123–127.

van Elsas, J.D., Jansson, J.K., Trevors, J.T., 2007. Modern Soil Microbiology. CRC Press. Taylor & Francis Group.

van Elsas, J.D., Costa, R., Jansson, J., Sjöling, S., Bailey, M., Nalin, R., Vogel, T.M., van Overbeek, L., 2008. The metagenomics of disease-suppressive soils - experiences from the METACONTROL project. Trends Biotechnol. 26 (11), 591–601.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., et al., 2004. Environmental shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M., van Elsas, J.D., Bailey, M.J., Nalin, R., et al., 2009a. TerraGenome: a consortium for the sequencing of a soil metagenome. Nat. Rev. Microbiol. 7 (4), 252.

Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M., van Elsas, J. D., Bailey, M.J., Nalin, R., et al., 2009b. Advantages of the metagenomic approach for soil exploration: reply from Vogel, et al. Nat. Rev. Microbiol. 7, 756. doi:10.1038/nrmicro2119-c3.

Voget, S., Steele, H.L., Streit, W.R., 2006. Characterization of a metagenome-derived halotolerant cellulase. J. Biotechnol. 126, 26–36.

Wall, D.H., Virginia, R.A., 1999. Controls on soil biodiversity: insights from extreme environments. Appl. Soil Ecol. 13, 137–150.

Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., van der Putten, W. H., Wall, D.H., 2004. Ecological linkages between aboveground and belowground biota. Science 304 (5677), 1629–1633.

Williamson, L.L., Borlee, B.R., Schloss, P.D., Guan, C., Allen, H.K., Handelsman, J., 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl. Environ. Microbiol. 71 (10), 6335–6344.

Wintzingerode, F.V., Gobel, U.B., Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS. Microbiol. Rev. 21, 213–229.

Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M., Keller, M., 2002. Cultivating the uncultured. PNAS 99 (24), 15681–15686.

Zhou, J., Thompson, D.K., 2002. Challenges in applying microarrays to environmental studies. Curr. Opin. Biotechnol. 13, 204–207.

Please cite this article in press as: Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology, Research in Microbiology (2010), doi:10.1016/j.resmic.2010.04.010