# THE METAGENOMICS OF SOIL

*Rolf Daniel*

Abstract | Phylogenetic surveys of soil ecosystems have shown that the number of prokaryotic species found in a single sample exceeds that of known cultured prokaryotes. Soil metagenomics, which comprises isolation of soil DNA and the production and screening of clone libraries, can provide a cultivation-independent assessment of the largely untapped genetic reservoir of soil microbial communities. This approach has already led to the identification of novel biomolecules. However, owing to the complexity and heterogeneity of the biotic and abiotic components of soil ecosystems, the construction and screening of soil-based libraries is difficult and challenging. This review describes how to construct complex libraries from soil samples, and how to use these libraries to unravel functions of soil microbial communities.

BIOTA
The organisms that occupy an ecosystem.

Soil is probably the most challenging of all natural environments for microbiologists, with respect to the microbial community size and the diversity of species present. One gram of forest soil contains an estimated $4 \times 10^7$ prokaryotic cells[1], whereas one gram of cultivated soils and grasslands contains an estimated $2 \times 10^9$ prokaryotic cells[2]. Based on the reassociation kinetics of DNA isolated from various soil samples, the number of distinct prokaryotic genomes has been estimated to range from 2,000 to 18,000 genomes per gram of soil[3–6]. These numbers might be an underestimate because genomes representing rare and unrecovered species might have been excluded from these analyses[3]. Therefore, the prokaryotic diversity present in just one gram of soil might exceed that of the known catalogue of prokaryotes (16,177 species were listed in the statistics of the taxonomy browser of the National Center for Biotechnology Information on January 25th 2005). The extreme spatial heterogeneity, multiphase nature (including gases, water and solid material) and the complex chemical and biological properties of soil environments are thought to contribute to the microbial diversity present in soil samples.

## Soil as a microbial habitat

Soil comprises mineral particles of different sizes, shapes and chemical characteristics, together with the soil BIOTA and organic compounds in various stages of decomposition. The formation of clay–organic matter complexes and the stabilization of clay, sand and silt particles through the formation of aggregates are the dominant structural characteristics of the soil matrix. Soil-matrix-component aggregates range from approximately 2 mm or more (macroaggregates) to fractions of a micrometer for bacteria and colloidal particles (for models see REF. 2). Prokaryotes are the most abundant organisms in soil and can form the largest component of the soil biomass[7]. Soil microorganisms often strongly adhere or adsorb onto soil particles such as sand grains or clay–organic matter complexes. Microhabitats for soil microorganisms include the surfaces of the soil aggregates, and the complex pore spaces between and inside the aggregates[7,8]. Some pore spaces are inaccessible for microorganisms owing to size restrictions. The metabolism and the survival of soil microorganisms are strongly influenced by the availability of water and nutrients. In contrast to aquatic habitats, surfaces of soil environments undergo dramatic cyclic changes in water content, ranging from water saturation to extreme aridity. A fraction of the microbial community dies during each drying-and-wetting cycle[9]. As a consequence, the composition of soil microbial communities fluctuates. However, how microbial populations are altered depending on changes in the water content and other environmental factors such as pH, availability of oxygen or temperature has not been studied intensively.

*Abteilung Angewandte Mikrobiologie, Institut für Mikrobiologie und Genetik der Georg-August-Universität, Grisebachstrasse 8, 37077 Göttingen, Germany. e-mail: rdaniel@gwdg.de*
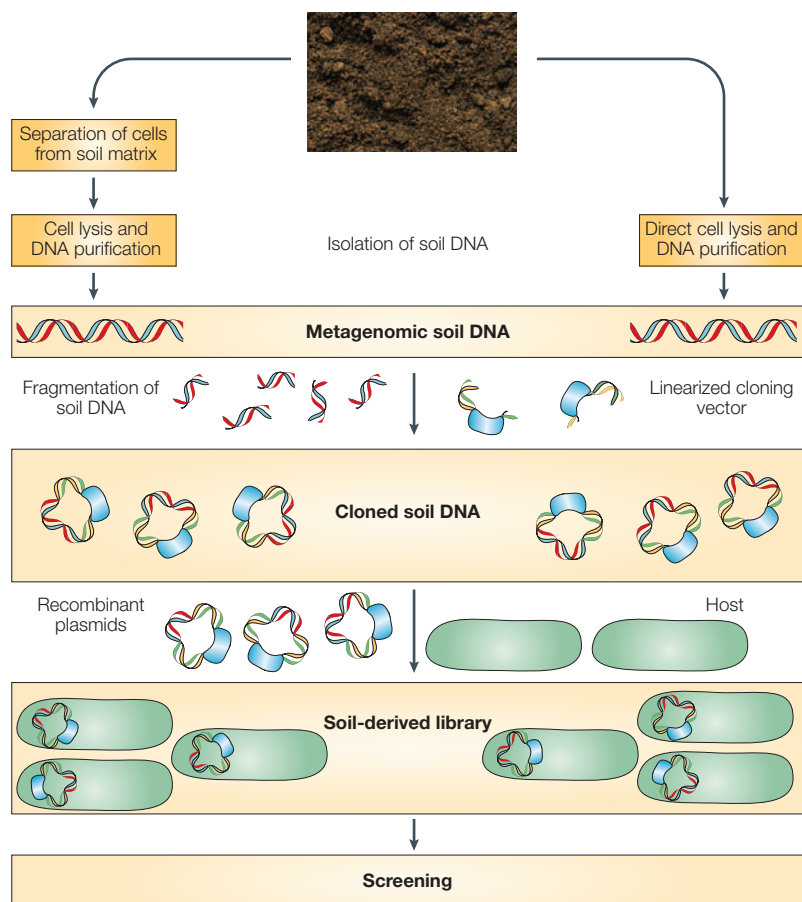
Figure 1 | **Essential steps to explore and exploit the genomic diversity of soil microbial communities by metagenomics.** Shown is a flow diagram of the main steps in the construction of a metagenomic DNA library from a soil sample. Soil DNA is recovered through separation of cells from soil particles followed by cell lysis and DNA recovery, or through direct lysis of cells contained within soil and recovery of DNA. Recovered soil DNA is fragmented and ligated into the linearized cloning vector of choice which might be a plasmid, cosmid, fosmid or BAC (bacterial artificial chromosome). Following the introduction of the recombinant vectors into a suitable bacterial cloning host, screening strategies can be designed to identify those clones which might contain new and useful genes.

Soil is an important reservoir for organic carbon, and prokaryotes are an essential component of the soil decomposition system[10]. Despite the high concentration of organic matter in most soil types, only low concentrations of organic carbon are readily available to microorganisms. Reasons for this include the transformation of most of the organic matter that is derived from plants, animals and microorganisms into HUMUS by a combination of microbiological and ABIOTIC processes, and the uneven distribution of microorganisms and organic compounds in the soil matrix. Humic substances are stable and recalcitrant to microbial decomposition processes — the half-life of these stable organic matter complexes with respect to biological degradation is approximately 2,000 years[2].

To adequately document the microbial diversity and the corresponding gene pool, the scale of soil surveys must be large. The versatility of soil microorganisms is also important for industry, as soil organisms have been the main sources of new natural products, including antibiotics[11].

HUMUS
A complex of heteropolymeric substances, including humic acids, humin and fulvic acids.

ABIOTIC
Non-living objects, substances or processes.

## Accessing the diversity of soil microorganisms

Direct cultivation or indirect molecular approaches can be used to explore and exploit the microbial diversity present in soil. Cultivation and isolation of microorganisms is the traditional method but, as only 0.1% to 1.0% of the soil bacteria are culturable using standard cultivation methods[3,12,13], the diversity of soil microbial communities has been mainly unexplored. Only a tiny portion of the gene pool has been characterized using cultivation and isolation. Recently, new approaches have been developed for the cultivation of soil bacteria[14–16], but these are not discussed in this review.

*Cultivation-independent techniques.* To circumvent some of the limitations of cultivation approaches, indirect molecular methods based on the isolation and analysis of nucleic acids (mainly DNA) from soil samples without cultivation of microorganisms have been developed. Theoretically, the microbial DNA isolated from a soil sample represents the collective DNA of all the indigenous soil microorganisms, and is named the soil metagenome[17,18]. Many protocols for the isolation of soil-derived microbial DNA have been published[19–27].

Considering the diversity of microbial species, the large populations of soil microorganisms and the complex soil matrix, which contains many compounds (such as humic acids) that bind to DNA and interfere with the enzymatic modification of DNA, recovery of microbial soil DNA that represents the resident microbial community and is suitable for cloning or PCR is still an important challenge. Phylogenetic surveys can be carried out by PCR amplification of 16S rRNA genes from soil DNA, using universal primers for bacteria and archaea. These surveys allow cataloguing and comparison of the microbial diversity in different soil habitats, and the comparative analysis of changes in community structure owing to altered environmental factors[27–32]. Other marker genes that are used to monitor microbial diversity include *dnaK*[33] (HSP-70-type molecular chaperone) and *amoA*[34] (ammonia monooxygenase). However, few soil environments have been surveyed, and the cataloguing of microbial diversity in soil is still in its infancy.

*Construction of soil DNA libraries.* Constructing soil-based libraries involves the same methods as the cloning of genomic DNA of individual microorganisms; that is, fragmentation of the soil DNA by restriction-enzyme digestion or mechanical shearing, insertion of DNA fragments into an appropriate vector system, and transformation of the recombinant vectors into a suitable host. Although the generation of soil libraries is conceptually simple, the size of the soil metagenome and the large number of clones that are required for full coverage make this a daunting task. The major breakthrough in soil metagenomics was the construction of libraries from soil DNA (FIG. 1) and screening of these libraries by functional and sequence-based approaches (TABLES 1,2). This

Table 1 | **Soil-based libraries constructed without enrichment steps before DNA isolation**

| Origin | Vector type | Number of clones | Average insert size (kb) | Total DNA (Gb) | Genes of interest | Year of construction | Refs |
|---|---|---|---|---|---|---|---|
| Meadow, sugar beet field, river valley | Plasmid | ~1,500,000 | 5–8 | 7.8 | 4-hydroxybutyrate utilization, lipolytic enzymes, antiporter | 1999 | 35,38,51 |
| Uncultivated soil | BAC | 3,648; 24,576 | 27; 44.5 | 1.19 | Antimicrobials, antibiotic resistance; 16S rRNA, various biocatalysts | 2000 | 37,48,50,69 |
| Soil type not specified | Not specified | Not specified | Not specified | - | Antimicrobials | 2000 | 46 |
| Soil type not specified | Cosmid | 700,000 | Not specified | 24.5* | Antimicrobials | 2000 | 36 |
| Soil type not specified | Cosmid | Not specified | Not specified | - | Pigments | 2001 | 64 |
| Uncultivated soil | BAC | 12,000 | 37 | 0.42 | Antimicrobials | 2001 | 47 |
| Soil type not specified | Cosmid | Not specified | Not specified | - | Fatty acid enol esters | 2002 | 73 |
| Alkaline loessian soil | Plasmid | 100,000 | 8–12 | 1.0 | Protease | 2002 | 41 |
| Calcerous grassland (sandy) | Fosmid | 25,278 | 32.5–43.5 | 0.90 | 16S rRNA genes | 2002 | 65 |
| Calcerous grassland (sandy) | Fosmid | 55,680 | 32.5–43.5 | 2.12 | Acidobacterial 16S rRNA genes | 2003 | 66 |
| Arable field | Cosmid | 5,000 | Not specified | 0.18* | Polyketide synthases, various other activities | 2003 | 49 |
| Meadow, sugar beet field, cropland | Plasmid | 583,000; 360,000; 324,000 | 4.4; 3.8; 3.5 | 4.05 | Carbonyl formation | 2003 | 43 |
| Sandy soil, sandy soil, mixed woodland soil | Fosmid | 25,344; 30,366; 19,978 | 33–45 | 3.03 | Taxonomic marker genes | 2004 | 58 |
| Clay loam sandy type | Fosmid | 100,000 | 30–40 | 3.50 | Polyketide synthase | 2004 | 67 |
| Forest soil | Fosmid | 33,700 | 35 | 1.18 | Lipolytic enzymes | 2004 | 39 |
| Soil type not specified | Cosmid | Not specified | Not specified | - | Long-chain *N*-acyltyrosines | 2004 | 61 |
| Plano silt loam soil | Plasmid | 200,000; 58,000; 250,000; 650,000 | 4.1; 2.7; 3.5; 3.5 | 4.2 | Antibiotic resistance | 2004 | 50 |
| Soil (surface covered with moss) | Plasmid | 30,000 | 3.5 | 0.11 | Amylolytic enzymes | 2004 | 45 |
| Agricultural field | Plasmid | 80,000 | 5.2 | 0.42 | Amidases | 2004 | 63 |

*An average insert size of 35 kb was assumed for cosmid libraries. BAC, bacterial artificial chromosome.

technology paved the way for elucidating the functions of organisms in soil communities, for genomic analyses of uncultured soil microorganisms and for the recovery of entirely novel natural products from soil microbial communities. In landmark studies, novel genes that encoded useful enzymes and antibiotics were recovered by direct cloning of soil DNA into plasmid, cosmid or BAC (bacterial artificial chromosome) vectors and screening of the generated libraries[35–37] (for the industrial impact of soil metagenomics see the article by P. Lorenz and J. Eck in this issue). The genes were identified using functional screens and had little homology to known genes, which illustrates the enormous potential of soil-based metagenomic libraries. The same approach has been used to clone genes from soil communities that code for lipases[38–40], proteases[41,42], oxidoreductases[43], amylases[44,45], antibiotics[46–49], antibiotic resistance enzymes[50] and membrane proteins[51]. The success of projects to generate and screen soil-derived metagenomic libraries depends on several factors: composition of the soil sample; collection and storage of the soil sample; the DNA extraction method used for high-quality DNA recovery; how representative the isolated DNA is of the microbial community present in the original sample; the host–vector systems used for cloning, maintenance and screening; and the screening strategy.

Table 2 | **Soil-based libraries constructed with enrichment steps before DNA isolation**

| Origin | Vector type | Number of clones | Average insert size (kb) | Total DNA (Gb) | Genes of interest | Year of construction | Ref. |
|---|---|---|---|---|---|---|---|
| Agricultural field, forest soil | Cosmid | Not specified | 30–40 | Not specified | Biotin synthesis | 2001 | 86 |
| Soil (agarolytic consortium) | Cosmid | Not specified | Not specified | Not specified | Novel biocatalysts | 2003 | 87 |
| Sugar beet field, river Grone, Solar lake, Gulf of Eilat | Plasmid | 100,000; 100,000; 100,000; 100,000 | 5.4, 3.3, 3.0, 5.6 | 1.31 | Alcohol oxidoreductase | 2003 | 85 |
| Sugar beet field, Solar lake, river Grone | Plasmid | 305,000; 301,000; 112,000 | 5.0, 3.4, 3.3 | 2.19 | Dehydratase | 2003 | 78 |
| Goose pond shore, agricultural field (loamy), lakeshore (sandy) | Plasmid | 25,000; 35,000; 30,000 | 5.2 | 0.47 | Amidases | 2004 | 63 |

*Isolation of high-quality DNA from soil.* Construction of a soil metagenomic library begins with sample collection (FIG. 1). As soil samples are heterogeneous, details of physical, chemical and biotic factors such as particle size, soil type, water content, pH, temperature and plant cover are useful for evaluation and comparison of the outcomes of soil-based studies[2]. Sampling is easier for surface soils compared with other environments such as SUBSURFACES. As microbial populations are large, sample volumes can be small (≤500 g in most studies)[25,35,37,47,52]. Disturbing soil during sampling might alter the composition of soil microbial communities, so the time that a sample is stored and transported should be kept to a minimum. A stored sample might not be representative of the undisturbed field soil[2].

Library construction requires sufficient amounts of high-quality DNA which is representative of the microbial community present. Because of the heterogeneity of soils, the extent of microbial diversity and the adherence of microorganisms to soil particles, DNA extraction is particularly challenging[53]. Also, extraction of soil DNA often results in coextraction of humic substances, which interfere with restriction-enzyme digestion and PCR amplification and reduce cloning efficiency, transformation efficiency and the specificity of DNA hybridization[21,54,55].

Many soil DNA extraction protocols have been published, and commercial soil DNA extraction kits are available[19–27]. The DNA extraction methods can be divided into two categories: direct lysis of cells contained in the sample matrix followed by separation of the DNA from the matrix and cell debris (pioneered by Ogram *et al.*[19]); or separation of the cells from the soil matrix followed by cell lysis (pioneered by Holben *et al.*[20]) (FIG. 1). The crude DNA recovered by both methods is purified by standard procedures. The amounts of DNA isolated from different soil types using a selection of protocols range from less than 1 μg to approximately 500 μg of DNA per gram of soil[24–26,35,56,57]. More DNA is recovered using the direct lysis approaches, for example, Gabor *et al.*[57] recorded a 10 to 100-fold reduction in the DNA yield using the cell separation approach compared with the direct lysis approach.

To achieve direct cell lysis, combinations of enzymatic treatment, high temperatures and detergent treatments have been used. In addition, several methods use mechanical disruption steps such as bead-beating, freeze–thawing or grinding of samples to lyse cells[19,24–27,57]. In addition to the DNA that is recovered from lysed prokaryotes, extracellular DNA and eukaryotic DNA are also recovered[27,57,58]. An excellent starting point for researchers is the direct lysis method of Hurt *et al.*[26], which allows simultaneous recovery of DNA and RNA from soils of different composition.

DNA extraction methods based on cell separation, although less efficient in terms of the amount of DNA recovered, are less harsh than direct lysis methods. The separation of microorganisms from the soil matrix is achieved by mild mechanical forces or chemical procedures such as blending, rotating pestle homogenization or the addition of cation-exchange resins, followed by density gradient or differential centrifugation[22,23,56,57]. The DNA obtained is almost entirely prokaryotic. Plus, DNA recovered by this method seems to be less contaminated with matrix compounds, including humic substances. In addition, the average size of the isolated DNA is larger than that typically obtained by the direct lysis approach[56] and is therefore more suitable for the generation of large-insert libraries.

*Library bias and DNA extraction.* As different soil microorganisms have different susceptibilities to cell lysis methods, the sequences present in the isolated DNA and the libraries is dependent on the extraction method[56,57,59].

How much bias in libraries is due to extraction methods has not been studied intensively. It is usually presumed that the DNA isolated by the direct lysis approach better represents the microbial diversity of a soil sample because this method does not include a cell separation step, so microorganisms that adhere to particles are also lysed[21,60]. However, Courtois *et al.*[56] found no significant difference in the spectrum of bacterial diversity during a comparison of DNA extracted directly from soil with DNA that was isolated from cells that were separated from the soil matrix. More studies comparing extraction methods and soil types would be helpful to determine the importance of this. Direct lysis approaches have been used more frequently than the separation techniques to isolate soil DNA for the construction of libraries[35–37,39,42,43,45,47,61].

SUBSURFACE
The geological zone below the surface of the Earth. It is not exposed to the Earth's surface.

Table 3 | **Pros and cons of small-insert and large-insert soil libraries**

| Advantages | Disadvantages |
|---|---|
| *Small-insert library (plasmids)* | |
| High copy number allows detection of weakly-expressed foreign genes | Small insert size |
| Expression of foreign genes from vector promoters is feasible | Large numbers of clones must be screened to obtain positives |
| Cloning of sheared DNA or soil DNA contaminated with matrix substances is possible | Not suitable for cloning of activities and pathways that are encoded by large gene clusters |
| Technically simple | |
| *Large-insert library (cosmids, fosmids, BACs)* | |
| Large insert size | Low copy-number might prevent detection of weakly-expressed foreign genes |
| Small numbers of clones can be screened to obtain positives | Limited expression of foreign genes by vector promoters |
| Suitable for cloning of enzyme activities and pathways that are encoded by large gene clusters | Requires high-molecular soil DNA of high purity for library construction |
| Suitable for partial genomic characterization of uncultured soil microorganisms | Technically difficult |

BACs, bacterial artificial chromosomes.

*Library size.* Libraries can be classified into two groups with respect to average insert size: small-insert libraries in plasmid vectors (less than 15 kb) and large-insert libraries in cosmid, fosmid (both up to 40 kb) or BAC vectors (more than 40 kb) (TABLE 3). The host for the initial construction and maintenance of almost all published libraries is *Escherichia coli*. Shuttle cosmid or BAC vectors can be used to transfer libraries that are produced in *E. coli* to other hosts such as *Streptomyces* or *Pseudomonas* species[49,62]. The choice of a vector system depends on the quality of the isolated soil DNA, the desired average insert size of the library, the vector copy number required, the host and the screening strategy that will be used, all of which depend on the aim of the study. Soil DNA that is contaminated with humic or matrix substances after purification or DNA sheared during purification might only be suitable for production of plasmid libraries. Small-insert soil-based libraries are useful for the isolation of single genes or small operons encoding new metabolic functions[35,38,43,45,50,51,63]. Large-insert libraries are more appropriate to recover complex pathways that are encoded by large gene clusters or large DNA fragments for the characterization of genomes of uncultured soil microorganisms[36,37,47–49,58,61,64–67]. It has been estimated that more than 10⁷ plasmid clones (5 kb inserts) or 10⁶ BAC clones (100 kb inserts) are required to represent the genomes of all the different prokaryotic species present in one gram of soil[17]. These estimates are based on the assumption that all species are equally abundant. To achieve substantial representation of the genomes from rare members (less than 1%) of the soil community, it has been calculated that libraries containing 10,000 Gb of soil DNA (10¹¹ BAC clones) might be required[68]. If these estimates are correct, the genetic contents stored in the soil-derived libraries already published (TABLES 1,2) do not come close to covering the entire soil metagenome. In addition, a comparison of the 16S rRNA genes in a BAC library with a collection

of DNA fragments that were generated by direct PCR amplification and cloning of the 16S rRNA genes from the same soil sample indicated that the representation of certain bacterial groups in the library was different from that present in the soil sample[69]. Despite these limitations, analysing and screening of libraries has yielded several novel biomolecules[35–51,61,63,70–73] and provided insights into the genomes of uncultured prokaryotic soil organisms and the ecology of the soil ecosystem[58,65,66,69].

## Functional screening of soil libraries

Several techniques have been used to identify and retrieve genes from soil-based libraries. Because of the complexity of the soil metagenome, high-throughput and sensitive screening methods are required. In principle, screens of soil-based libraries can be based either on metabolic activity (function-driven approach) or on nucleotide sequence (sequence-driven approach) (TABLE 4). PCR is most commonly used for sequence-driven screening of soil-based libraries or soil DNA[28,31,49,56,65–67,69,74–77]. Hybridization using target-specific probes has also been used to screen soil-based libraries[78]. Both approaches require suitable primers and probes that are derived from conserved regions of known genes and gene products, so applicability is limited to the identification of new members of known gene families. This approach has been used to identify phylogenetic anchors such as 16S rRNA genes[65,66,69] and genes encoding enzymes with highly conserved domains such as polyketide synthases[49,67,75], gluconic acid reductases[76] and nitrile hydratases[77]. To merely retrieve conserved genes from soil habitats by PCR, the construction of libraries is not a prerequisite. This approach often results in the amplification of partial genes, but the subsequent recovery of full-length genes from isolated complex soil DNA is difficult, whereas an insert from a clone that contained the gene of interest might harbour the

Table 4 | **Function-driven versus sequence-driven screening strategies**

| Advantages | Disadvantages |
| --- | --- |
| *Function-driven screening method* | |
| Completely novel genes can be recovered | Dependent on expression of the cloned genes by the bacterial host |
| Selects for full-length genes | Requires production of a functional gene product by the bacterial host |
| Selects for functional gene products | Dependent on the design of a simple activity-based screening strategy |
| *Sequence-driven screening method* | |
| Independent of expression of the cloned genes by the bacterial host used | Recovered genes are related to known genes |
| Similar screening strategies can be used for different targets, for example, colony hybridization and PCR | Partial genes can be cloned |
| | Not selective for functional gene products |

full-length gene. Stokes *et al.*[79] described a different PCR-based approach that uses primers that target a 59-bp recombination site. This site is present in different bacterial groups and flanks gene cassettes that are associated with integrons. Analysis of the gene cassettes isolated directly from soil DNA revealed that they contained full-length genes, most of which were not related to known genes.

The advantage of the identification of clones harbouring phylogenetic anchor genes on large inserts is that sequencing of the DNA surrounding these genes is feasible. This enables the partial genomic characterization of uncultivated soil microorganisms and yields insights into the physiology, ecological role and evolution of the organisms. This approach has been successfully used in the characterization of uncultivated members of the Acidobacteria phylum, which are abundant in soil but about which little is known[66,69], and to access the genomes of uncultivated Archaea in soil[58,65]. Theoretically, random sequencing of soil-derived libraries is another approach to characterize the soil ecosystem on a genomic level, but the species-richness of soil habitats would require enormous sequencing and assembly efforts.

Microarray technology could be useful for analysing the soil metagenome and profiling metagenomic libraries[80–84]. For example, genes encoding key reactions in the nitrogen cycle were detected using microarrays from samples that were collected from soil, and provided information on the composition and activity of the complex soil microbial community[80]. However, microarray methods for gene detection show a 100 to 10,000-fold lower sensitivity than PCR[81]. This difference might prevent the analysis of sequences from low-abundance soil microorganisms. The improvement of sensitivity and specificity are among the challenges of using complex soil DNA or RNA with microarray technology.

Most of the screening methods to isolate genes or gene clusters for novel biocatalysts or small molecules are based on detecting activity from library-containing clones[35–43,45–51,61,63,64,72,73]. As sequence information is not required, this is the only strategy that has the potential to identify new classes of genes that encode either known or new functions. This strategy has been validated by the isolation of novel genes that encode degradative enzymes[35,37–39,41,43,45,63], antibiotic resistance[50] and antibiotics[36,46–49]. Most of the biomolecules recovered by function-driven screens of complex soil libraries are either weakly related or entirely unrelated to known genes, and rediscovery of genes has not been reported. This confirmed that the amount of soil DNA that has been cloned and screened only represents the tip of the iceberg with respect to discovery of new natural products from the soil metagenome. Simple activity-based strategies are favoured, as the frequency of soil-derived metagenomic clones that express a specific activity is usually low, so large numbers of clones have to be tested. For example, the screening of 1,186,200 clones containing soil DNA resulted in the identification of 10 unique clones that confer antibiotic resistance[50]. Function-driven approaches can include the direct testing of colonies for a specific function. For example, chemical dyes and insoluble or chromophore-bearing derivatives of enzyme substrates can be incorporated into the growth medium solidified with agar to monitor enzymatic functions of individual clones. The sensitivity of these screens makes it possible to detect rare clones. An example is the screening of soil-based libraries for genes conferring polyol oxidoreductase activity[43,85], which was based on the ability of the recombinant *E. coli* strains to form carbonyls from polyols (FIG. 2a). Another example is the detection of *E. coli* clones with proteolytic activity on agar plates containing skimmed milk[41,42] (FIG. 2b). Another approach that allows detection of functional clones is the use of host strains or mutants of host strains that require heterologous complementation for growth under selective conditions. An example is complementation of a Na$^+$/H$^+$ antiporter-deficient *E. coli* strain with soil-derived libraries, which led to the identification of two new genes that encode Na$^+$/H$^+$ antiporters from a soil library consisting of 1,480,000 clones[51]. Although function-driven screens usually result in identification of full-length genes (and therefore functional gene products), one limitation of this approach is its reliance on the expression of the cloned gene(s) and the functioning of the encoded protein in a foreign host (TABLE 4).
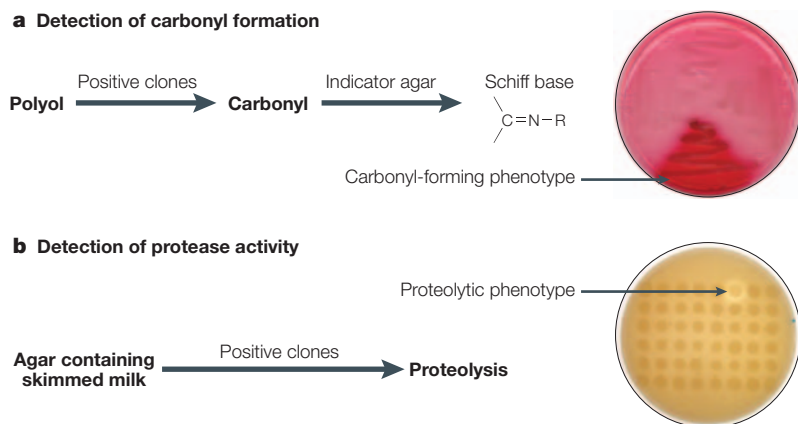
**a** Detection of carbonyl formation

Polyol $\xrightarrow{\text{Positive clones}}$ Carbonyl $\xrightarrow{\text{Indicator agar}}$ Schiff base
$$C=N-R$$

Carbonyl-forming phenotype

**b** Detection of protease activity

Proteolytic phenotype

Agar containing skimmed milk $\xrightarrow{\text{Positive clones}}$ Proteolysis

Figure 2 | **Examples of activity-based screens. a** | Detection of clones harbouring genes that confer carbonyl formation. Screening is based on the ability of the library-containing *Escherichia coli* clones to form carbonyls from test substrates, that is, polyols[43,85], during growth on indicator agar. The test substrates are included in the indicator agar, which contains a mixture of pararosaniline and sodium bisulphite (Schiff reagent). The production of carbonyls from test substrates on indicator plates by clones results in formation of a dark red Schiff base. The carbonyl-forming colonies are red and are surrounded by a red zone, whereas colonies failing to form carbonyls from the test substrate remain uncoloured. **b** | Detection of proteolytic activity. Proteolytic *E. coli* clones are detected on agar media containing skimmed milk by zones of clearance around the colonies.

Therefore, the low gene-detection frequencies or the inability to recover functional proteins encoded by metagenomic DNA during function-driven screening might also be due to the fact that many genes and gene products are not expressed and are inactive in the host strain. In most studies, *E. coli* has successfully been used as the host for functional screens. Recently, other bacterial hosts such as *Streptomyces* or *Pseudomonas* strains have been used to expand the range of soil-derived genes which can be detected during functional screens[46,49,62]. As expression in bacterial hosts is usually limited to prokaryotic genes and soil DNA can, depending on the isolation method, contain an important amount of eukaryotic DNA[57], using eukaryotic hosts could also be useful for function-driven screens of soil-based libraries.

### Enhancement of gene detection frequencies
The number of clones and, correspondingly, the size of cloned soil DNA that has to be screened to recover the genes of interest, is determined by the frequency of soil organisms that contain the desired genes in the soil sample used for DNA isolation and library construction. To increase this frequency, enrichment steps for microorganisms harbouring the desired traits have been used prior to library construction[63,78,85–87]. In most studies, carbon or nitrogen sources that are selective for microbial species containing the desired genes were used as growth substrates. A drawback of enrichment steps is the loss of microbial diversity, as fast-growing and culturable members of microbial CONSORTIA are usually selected. Nevertheless, a combination of traditional enrichment and metagenomic technologies is an efficient tool to increase the amount of positive clones in a screen and to isolate novel biomolecules

CONSORTIUM
Physical association between cells of two or more types of microorganisms. Such an association might be advantageous to at least one of the microorganisms.

when samples from complex habitats such as soil are used as starting material and non-vigorous enrichment steps are carried out[85,88]. In addition, using complex laboratory enrichments simplifies the isolation of high-quality DNA, which is required for the rapid construction of high-quality libraries. This strategy has been successfully used to isolate biotechnologically useful gene products, including alcohol oxidoreductases[85], coenzyme $B_{12}$-dependent dehydratases[78], amidases[63], agarases[87] and genes involved in biotin synthesis[86]. Other potential methods that could be used to enrich genomes from metabolically active members of the soil microbial community prior to library construction are stable isotope probing[89,90] (see also the article by M. G. Dumont and and J. C. Murrell in this issue) and enrichment with bromodeoxyuridine in the presence of selective substrates[91]. These techniques have not been used in soil library-based gene discovery to date. To improve the representation of rare genomes in a library, normalization procedures such as separating soil DNA based on its AT content might also be used for enrichment[40].

### Optimizing soil metagenomics
Bioinformatic methods that allow statistical comparisons of constructed libraries are necessary to determine whether differences in libraries are either artefacts of sampling and library construction or are caused by changes in the community composition. Programs such as ∫–LIBSHUFF[92], which has been employed for comparison of 16S rRNA gene libraries, might be useful for this purpose after further development. Functional and sequence-based screening of soil-based libraries have provided insights into soil microbial communities and have led to the identification of novel biomolecules, but these approaches have strengths and limitations (TABLE 4). To take full advantage of the enormous diversity of soil microorganisms, a combination of sequence-based and functional approaches and of different types of libraries should be used to probe the soil metagenome. Recently, a third high-throughput screening strategy, which is termed substrate-induced gene expression cloning (SIGEX) has been introduced for the identification and recovery of genes that encode catabolic pathways[93]. This method is based on the finding that genes encoding catabolic pathways are usually organized in operons that are induced by a relevant substrate, and are often controlled by regulatory elements located in the proximity of the catabolic genes. An operon-trap *gfp* (green fluorescent protein)-expression vector was constructed, which allowed shotgun cloning of metagenomic DNA upstream of the *gfp* gene, thereby placing the expression of this gene under the control of promoters that were present in the metagenomic DNA. Clones influencing the expression of *gfp* on addition of the substrate of interest can be isolated by fluorescence-activated cell sorting. SIGEX has the potential to sort through large-scale libraries that represent complex soil microbial communities but it has not yet been used for this purpose. Eukaryotes such as fungi are also an important component of the soil ecosystem, but their

genetic potential has not been fully integrated into soil-based gene discovery. In addition to the possibility of using eukaryotic hosts for activity-based screening, another option is the construction of soil libraries from cDNA, which would allow the retrieval of biocatalysts derived from eukaryotes. Therefore, efficient methods for the lysis of eukaryotic microorganisms in soil samples and the conservation of intact poly(A) mRNA are needed. The isolated poly(A) mRNA can then be used for cDNA synthesis and for library construction in expression vectors using standard procedures.

### Conclusions

Soil habitats probably contain the greatest microbial diversity of all the environments on earth. So far, metagenomic approaches have only scratched the surface of the genomic, metabolic and phylogenetic diversity stored in the soil metagenome. One of the major challenges for soil metagenomics is to develop methods to capture the heterogeneity and dynamics of complex soil microbial communities, both over time and spatially. Although considerable progress has been made in the characterization of microbial communities by random sequencing[94], a further improvement

of sequencing technologies and bioinformatic tools for analysing the enormous amount of data produced, combined with a reduction in sequencing costs, is required to apply this technique to the soil metagenome. The potential of microarrays for detecting and monitoring gene expression in soil microbial communities has already been proven, and monitoring microbial activities through protein arrays and proteomics will probably have an important role in the future. Soil microorganisms will continue to be the main source of novel natural products through the use of metagenomic technology. Taking the small fraction of the soil metagenome that has been accessed in screens of soil-based libraries so far and the relative wealth of new biomolecules that have been discovered, together with the limitations of library construction and screening methods (TABLES 3,4), soil microbial communities might be an almost unlimited resource of new genes encoding useful products.

Strategies to improve heterologous gene expression and production of functional recombinant proteins as well as new approaches for efficient screening of large soil libraries will further accelerate the speed of discovery and the diversity of the recovered biomolecules.

1. Richter, D. D. & Markewitz, D. How deep is soil? *Bioscience* **45**, 600–609 (1995).
2. Paul, E. A. & Clark, F. E. *Soil microbiology and biochemistry* (Academic Press, San Diego, 1989).
3. Torsvik, V., Daae, F. L., Sandaa, R.-A. & Øvreås, L. Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* **5**, 240–245 (2002).
4. Torsvik, V., Daae, F. L., Sandaa, R.-A. & Øvreås, L. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotech.* **64**, 53–62 (1998).
5. Torsvik, V., Sorheim, R. & Goksoyr, J. Total bacterial diversity in soil and sediment communities. *J. Ind. Microbiol.* **17**, 170–178 (1996).
6. Doolittle, W. Phylogenic classification and the universal tree. *Science* **284**, 2124–2128 (1999).
7. Hassink, J., Bouwman, L. A., Zwart, K. B. & Brussaard, L. Relationship between habitable pore space, soil biota, and mineralization rates in grassland soils. *Soil. Biol. Biochem.* **25**, 47–55 (1993).
8. Foster, R. C. Microenvironments of soil microorganisms. *Biol. Fertil. Soils* **6**, 189–203 (1988).
9. Kieft, T. L., Soroker, E. & Firestone, M. R. Microbial biomass response to a rapid change increase in water potential when dry soil is wetted. *Soil Biol. Biochem.* **19**, 119–126 (1987).
10. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
11. Osburne, M. S., Grossmann, T. H., August, P. R. & MacNeil, I. A. Tapping into microbial diversity for natural products drug discovery. *ASM News* **66**, 411–417 (2000).
12. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
13. Torsvik, V., Goksoyr, J. & Daae, F. L. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**, 782–787 (1990).
14. Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A. & Janssen, P. H. Laboratory cultivation of widespread and previously uncultured soil bacteria. *Appl. Environ. Microbiol.* **69**, 7210–7215 (2003).
15. Kaeberlein, T., Lewis, K. & Epstein, S. Isolating "uncultivable" microorganisms in a simulated natural environment. *Science* **296**, 1127–1129 (2002).
16. Zengler, K. *et al.* Cultivating the uncultured. *Proc. Natl Acad. Sci. USA* **99**, 15684–15686 (2002).
17. Handelsman, J., Rondon, M. R., Brady, S. F, Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998).
18. Rondon, M. R., Goodman, R. M. & Handelsman, J. The earth's bounty: assessing and accessing soil microbial diversity. *TIBTECH* **17**, 403–409 (1999).
19. Ogram, A., Sayler, G. S. & Barkay, T. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* **7**, 57–66 (1987).
20. Holben, W. E., Jansson, J. K., Chelm, B. K. & Tiedje, J. M. DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Appl. Environ. Microbiol.* **54**, 703–711 (1988).
21. Steffan, R. J., Goksoyr, J., Bej, A. K. & Atlas, R. M. Recovery of DNA from soils and sediments. *Appl. Environ. Microbiol.* **54**, 2908–2915 (1988).
22. Jacobsen, C. S. & Rasmussen, O. F. Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Appl. Environ. Microbiol.* **58**, 2458–2462 (1992).
23. Lindahl, V. & Bakken, L. R. Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiol. Ecol.* **16**, 135–142 (1995).
24. Zhou, J. M., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
25. Miller, D. N., Bryant, J. E., Madsen, E. L. & Ghiorse, W. C. Evaluation and optimization of DNA extraction for soil and sediment samples. *Appl. Environ. Microbiol.* **65**, 4715–4724 (1999).
26. Hurt, R. A. *et al.* Simultaneous recovery of RNA and DNA from soils and sediments. *Appl. Environ. Microbiol.* **67**, 4495–4503 (2001).
    **Development of a method for direct nucleic acid isolation from soils of various compositions.**
27. Lloyd-Jones, G. & Hunter, D. W. F. Comparison of rapid DNA extraction methods applied to contrasting New Zealand soils. *Soil Biol. Biochem.* **33**, 2053–2059 (2001).
28. Dunbar, J., Takala, S., Barns, S. M., Davis, J. A. & Kuske, C. R. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl. Environ. Microbiol.* **65**, 1662–1669 (1999).
29. Øvreås, L. Population and community level approaches for analysing microbial diversity in natural environments. *Ecol. Letts.* **3**, 236–251 (2000).
30. Dunbar, J., Barns, S. M., Ticknor, L. O. & Kuske, C. R. Empirical and theoretical bacterial diversity in four Arizona soils. *Appl. Environ. Microbiol.* **68**, 3035–3045 (2002).
31. Zhou, J. *et al.* Spatial and resource factors influencing high microbial diversity in soil. *Appl. Environ. Microbiol.* **68**, 326–334 (2002).
32. Yeager, C. M. *et al.* Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado plateau and Chihuahuan desert. *Appl. Environ. Microbiol.* **70**, 973–983 (2004).
33. Yap, W. H., Li, X., Soong, T. W. & Davies, J. E. Genetic diversity of soil microorganisms assessed by analysis of *hsp70* (*dnaK*) sequences. *J. Industr. Microbiol.* **17**, 179–184 (1996).
34. Webster, G., Embley, T. M. & Posser, J. L. Grassland management regimens reduce small-scale heterogeneity and species diversity of β-proteobacterial ammonia oxidisers. *Appl. Environ. Microbiol.* **68**, 20–30 (2002).
35. Henne, A., Daniel, R., Schmitz, R. A. & Gottschalk, G. 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 (1999).
    **First report on the isolation of biocatalysts from soil-derived plasmid libraries.**
36. Brady, S. F. & Clardy, J. Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J. Am. Chem. Soc.* **122**, 12903–12904 (2000).
    **First isolation of antibiotics from a soil-based cosmid library.**
37. Rondon, M. R. *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).
    **Generation of a large-insert soil-derived BAC library and illustration of the potential of this type of library for soil-based gene discovery.**
38. Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G. & Daniel, R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli. Appl. Environ. Microbiol.* **66**, 3113–3116 (2000).
39. Lee, S.-W. *et al.* Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl. Microbiol. Biotechnol.* **65**, 720–726 (2004).
40. Lorenz, P. & Schleper, C. Metagenome — a challenging source of enzyme discovery. *J. Mol. Catal. B Enzym.* **19**, 13–19 (2002).
41. Gupta, R., Berg, Q. K. & Lorenz, P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* **59**, 15–32 (2002).
42. Santosa, D. A. Rapid extraction and purification of environmental DNA for molecular cloning applications and molecular diversity studies. *Mol. Biotechnol.* **17**, 59–64 (2001).
43. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. & Daniel, R. 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 (2003).

44. Richardson, T. H. *et al.* A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable α-amylase. *J. Biol. Chem.* **277**, 26501–26507 (2002).

45. Yun, J. *et al.* Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl. Environ. Microbiol.* **70**, 7229–7235 (2004).

46. Wang, G. Y. *et al.* Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* **2**, 2401–2404 (2000).
    **First report on the use of a non-*E. coli* host for screening of metagenomic libraries.**

47. MacNeil, I. A. *et al.* Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.* **3**, 301–308 (2001).

48. Gillespie, D. E. *et al.* Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* **68**, 4310–4306 (2002).

49. Courtois, S. *et al.* Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* **69**, 49–55 (2003).
    **Strategy for increasing the screening efficiency of large-insert metagenomic libraries by using a cosmid shuttle vector.**

50. Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ. Microbiol.* **6**, 981–989 (2004).
    **References 50 and 51 illustrate the immense power of activity-based screening strategies using host strains or mutants of host strains that require heterologous complementation for growth under selective conditions.**

51. Majernik, A., Gottschalk, G. & Daniel, R. Screening of environmental DNA libraries for the presence of genes conferring Na⁺(Li⁺)/H⁺ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J. Bacteriol.* **183**, 6645–6653 (2001).

52. Liesack, W. & Stackebrandt, E. Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* **174**, 5072–5078 (1992).

53. Martin-Laurent, F. *et al.* DNA extractions from soils: old bias for new microbial diversity analysis methods. *Appl. Environ. Microbiol.* **67**, 2354–2359 (2001).

54. Tebbe, C. C. & Vahjen, W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and yeast. *Appl. Environ. Microbiol.* **59**, 2657–2665 (1993).

55. Tsai, Y.-L., & Olson, B. H. Detection of low numbers of bacterial cells in soils and sediments. *Appl. Environ. Microbiol.* **58**, 754–757 (1992).

56. Courtois, S. *et al.* Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ. Microbiol.* **3**, 431–439 (2001).

57. Gabor, E. M., de Vries, E. J. & Janssen, D. B. Efficient recovery of environmental DNA for expression cloning by indirect methods. *FEMS Microbiol. Ecol.* **44**, 153–163 (2003).
    **References 57 and 58 compare direct cell lysis with cell separation approaches with respect to species representation and content of eukaryotic DNA in the isolated soil DNA.**

58. Treusch, A. H. *et al.* Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ. Microbiol.* 6, 970–980 (2004).

59. Stach, J. E. M., Bathe, S., Clapp, J. P. & Burns, R. G. PCR–SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. *FEMS Microbiol. Ecol.* **36**, 139–151 (2001).

60. Leff, L. G., Dana, J. R., McArthur, J. V. & Shimkets, L. J. Comparison of methods of DNA extraction from stream sediments. *Appl. Environ. Microbiol.* **61**, 1141–1143 (1995).

61. Brady, S. F., Chao, C. J. & Clardy, J. Long-chain *N*-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* **70**, 46865–46870 (2004).

62. Martinez, A. *et al.* Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* **70**, 2452–2463 (2004).

63. Gabor, E. M., de Vries, E. J. & Janssen, D. B. 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 (2004).

64. Brady, S. F., Chao, C. J., Handelsman, J. & Clardy, J. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* **3**, 1981–1984 (2001).

65. Quaiser, A. *et al.* First insight into the genome of an uncultivated crenarchaeota from soil. *Environ. Microbiol.* **4**, 603–611 (2002).

66. Quaiser, A. *et al.* Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* **50**, 563–575 (2003).
    **Partial genomic characterization of uncultivated Acidobacteria.**

67. 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).

68. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**, 525–552 (2004).

69. Liles, M. R., Manske, B. F., Bintrim, S. B., Handelsman, J. & Goodman, R. M. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl. Environ. Microbiol.* **69**, 2684–2691 (2003).

70. Robertson, D. E. *et al.* Exploring nitrilase sequence for enantioselective catalysis. *Appl. Environ. Microbiol.* **70**, 2429–2436 (2004).

71. Gray, K. A., Richardson, T. H., Robertson, D. E., Swanson, P. E. & Subramanian, M. V. Soil-based gene discovery: a new technology to accelerate and broaden biocatalytic applications. *Adv. Appl. Microbiol.* **52**, 1–27 (2003).

72. Brady, S. F. & Clardy, J. Synthesis of long-chain fatty acid enol esters isolated from an environmental DNA clone. *Org. Lett.* **5**, 121–124 (2003).

73. Brady, S. F., Chao, C. J. & Clardy, J. New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* **124**, 9968–9969 (2002).

74. Borneman, J. & Triplett, E. W. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl. Environ. Microbiol.* **63**, 2647–2653 (1997).

75. Seow, K.-T. *et al.* A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J. Bacteriol.* **179**, 7360–7368 (1997).

76. Eschenfeldt, W. H. *et al.* DNA from uncultured organisms as a source of 2,5-diketo-ᴅ-gluconic acid reductases. *Appl. Environ. Microbiol.* **67**, 4206–4214 (2001).

77. Precigou, S., Goulas, P. & Duran, R. Rapid and specific identification of nitrile hydratase (Nhase)-encoding genes. *FEMS Microbiol. Lett.* **204**, 155–161 (2001).

78. Knietsch, A., Bowien, S., Whited, G., Gottschalk, G. & Daniel, R. Identification and characterization of genes encoding coenzyme B₁₂-dependent glycerol and diol dehydratases from metagenomic DNA libraries derived from enrichment cultures. *Appl. Environ. Microbiol.* **69**, 3048–3060 (2003).

79. Stokes, H. W. *et al.* Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl. Environ. Microbiol.* **67**, 5240–5246 (2001).
    **Library-independent approach for recovery of novel genes from environmental samples.**

80. Wu, L. *et al.* Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl. Environ. Microbiol.* **67**, 5780–5790 (2001).
    **One of the first reports on using DNA microarray technology for assessing functional gene diversity in soil microbial communities.**

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

82. Cho, J.-C. & Tiedje, J. M. Quantitative detection of microbial genes by using DNA microarrays. *Appl. Environ. Microbiol.* **68**, 1425–1430 (2002).

83. Denef, V. J. *et al.* Validation of a more sensitive method for using spotted oligonucleotide DNA microarrays for functional genomics studies on bacterial communities. *Environ. Microbiol.* **5**, 933–943 (2003).

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

85. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. & Daniel, R. Construction and screening of metagenomic 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 (2003).

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

87. Voget, S. *et al.* Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* **69**, 6235–6242 (2003).
    **The library derived from an agarolytic microbial consortium enriched from soil harbours a large number of genes encoding industrially relevant biocatalysts.**

88. Daniel, R. The soil metagenome — a rich resource for the discovery of novel natural products. *Curr. Opin. Microbiol.* **15**, 199–204 (2004).

89. Radajewski, S., McDonald, I. R. & Murrell, J. C. Stable-isotope probing of nucleic acids: a window to the function of uncultured microorganisms. *Curr. Opin. Biotechnol.* **14**, 296–302 (2003).

90. Wellington, E. M. H., Berry, A. & Krsek, M. Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr. Opin. Microbiol.* **6**, 295–301 (2003).

91. Yin, B., Crowley, D., Sparovek, G., De Melo, W. J. & Borneman, J. Bacterial functional redundancy along a soil reclamation gradient. *Appl. Environ. Microbiol.* **66**, 4361–4365 (2000).

92. Schloss, P. D., Larget, B. R. & Handelsman, J. Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl. Environ. Microbiol.* **70**, 5485–5492 (2004).
    **A statistical approach for comparing gene libraries is described.**

93. Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnol.* **23**, 88–93 (2005).
    **This strategy for screening of metagenomic libraries has enormous potential for soil-based gene discovery.**

94. Venter, J. C. *et al.* Environmental shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).

🌐 **Online links**

**FURTHER INFORMATION**
**Rolf Daniel's laboratory:**
http://wwwuser.gwdg.de/~genmibio/lab201.html
∫–**LIBSHUFF:**
http://www.plantpath.wisc.edu/fac/joh/S-LIBSHUFF.html
**Access to this interactive links box is free online.**