The Earth’s bounty: assessing and accessing soil microbial diversity

Michelle R. Rondon, Robert M. Goodman and Jo Handelsman

The study of microbial diversity represents a major opportunity for advances in biology and biotechnology. Recent progress in molecular microbial ecology shows that the extent of microbial diversity in nature is far greater than previously thought. Here, we discuss methods to analyse microorganisms from natural environments without culturing them and new approaches for gaining access to the genetic and chemical resources of these microorganisms.

For millennia, diverse natural microorganisms have yielded important biological materials useful to humans. Over the past 50 years, products derived from microbial secondary metabolites have been used to meet medical, industrial and agricultural needs (e.g. antibiotics, anticancer drugs, antifungal compounds, immunosuppressive agents, enzyme inhibitors, anti-parasitic agents, herbicides, insecticides and growth promoters). As soil microorganisms have been a natural products from soil microorganisms.

The Earth’s bounty: assessing and accessing soil microbial diversity

Michelle R. Rondon, Robert M. Goodman and Jo Handelsman

The study of microbial diversity represents a major opportunity for advances in biology and biotechnology. Recent progress in molecular microbial ecology shows that the extent of microbial diversity in nature is far greater than previously thought. Here, we discuss methods to analyse microorganisms from natural environments without culturing them and new approaches for gaining access to the genetic and chemical resources of these microorganisms.
constitute a significant fraction of the total community then they represent a novel and untapped resource for natural-product discovery. Below, we discuss some common methods for examining microorganisms in nature, with particular attention to the study of soil microorganisms.

Methods to analyse microbial diversity

Microorganisms are extremely difficult to study in nature, owing to their small size and morphological simplicity (Fig. 1). These challenges have led to the use of culturing to analyse microorganisms, with the result that some microorganisms have been extremely well studied but the great majority have not been studied at all. Current attempts to describe and understand microbial diversity are aimed at overcoming the culturing bias in an effort to provide a more accurate picture of microbial diversity and function in natural environments. First has been a cataloging stage, supported mainly by 16S ribosomal RNA (rRNA) gene-sequence studies, which strives to answer the question ‘What is out there?’ This is being followed by more multidimensional studies aimed at understanding the geographical distribution and functional roles of microorganisms in the environment. Further goals include the development of a cohesive ecological framework for microbial life and of improved molecular methods to access this diversity for basic and applied research goals.

Molecular analysis of microbial communities has provided evidence that unexploited microbial diversity exists in many environments. Investigations of phylogenetic diversity by 16S-rRNA-gene-sequence analysis have yielded similar results in numerous habitats; many new sequence types are found that do not correspond to sequences in the database. More sophisticated analyses of microbial diversity have been performed on simple communities such as microbial mats, where temporal and geographical variations can be monitored.

Many methods involving the use of single genes such as the 16S rRNA gene are used to examine phylogenetic diversity; others, such as the cloning approaches, are aimed at understanding and exploiting the functional diversity of microorganisms in nature (Table 1).

<table>
<thead>
<tr>
<th>Method</th>
<th>Comments</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culturing</td>
<td>Not representative</td>
<td>6,10</td>
</tr>
<tr>
<td>16S rRNA gene sequence analysis</td>
<td>Cloning required; provides identification of members of community</td>
<td>6,10</td>
</tr>
<tr>
<td>In situ hybridization</td>
<td>Labor intensive, can be used to identify metabolically active microorganisms</td>
<td>8,14</td>
</tr>
<tr>
<td>Substrate utilization</td>
<td>Measures metabolic diversity</td>
<td>15-17</td>
</tr>
<tr>
<td>DNA–DNA reassociation kinetics</td>
<td>Provides a global view of genetic complexity of sample</td>
<td>18</td>
</tr>
<tr>
<td>ARDRA</td>
<td>More useful for simple communities, useful for comparative analysis</td>
<td>19</td>
</tr>
<tr>
<td>PCR amplification or expression cloning</td>
<td>Functional diversity targeted</td>
<td>20.22</td>
</tr>
<tr>
<td>BAC libraries</td>
<td>Permanent archive of genetic information from sampled environment; phylogenetic and functional diversity</td>
<td>22</td>
</tr>
<tr>
<td>Flow cytometry</td>
<td>Enumeration of microorganisms</td>
<td>23</td>
</tr>
<tr>
<td>RNA dot or slot blot</td>
<td>Representation of metabolically active members of a community</td>
<td>24</td>
</tr>
<tr>
<td>SSCP</td>
<td>Comparative analysis</td>
<td>25</td>
</tr>
<tr>
<td>%GC content</td>
<td>Global view of community diversity</td>
<td>26,27</td>
</tr>
<tr>
<td>T-RFLP</td>
<td>Comparative analysis</td>
<td>28</td>
</tr>
<tr>
<td>DGGE or TGGE</td>
<td>Used to monitor enrichment, comparative analysis</td>
<td>29,30</td>
</tr>
</tbody>
</table>

Abbreviations: ARDRA, amplified ribosomal DNA restriction analysis; BAC, bacterial artificial chromosomes; DGGE, denaturing gradient gel electrophoresis; SSCP, single-strand conformational polymorphisms; TGGE, temperature-gradient gel electrophoresis; T-RFLP, terminal restriction fragment-length polymorphisms.
obtained by differential centrifugation. Recombining measurements revealed that the DNA isolated directly from soil was much more complex than expected and suggested that thousands of independent genomes were present in the sample. A similar analysis on 206 cultured bacteria from the same sample yielded much less diversity. Extrapolation of the data suggests that there may be thousands of microbial types in a gram of soil, many of which are assumed not to be culturable. This kind of analysis serves as a global measurement of the heterogeneity of environmental DNA and lends support to the idea that microbial diversity has not yet been adequately described.

**16S-rRNA-based methods**

Many methods for investigating microbial diversity utilize 16S-rRNA-gene sequence analysis. Commonly, researchers isolate DNA from environmental sources, amplify 16S-rRNA-gene sequences from the sample and analyse the amplified sequences by several methods, including cloning and sequencing, amplified ribosomal-DNA-restriction analysis, denaturing gradient gel electrophoresis (DGGE), temperature-gradient gel electrophoresis (TGGE), dot blots, single-strand conformational polymorphisms, and terminal-restriction-fragment-length polymorphism (T-RFLP) analysis (Table 1). Reverse-transcription PCR, TGGE or DGGE, T-RFLP and in situ hybridization are also used to identify metabolically active or numerically dominant populations. These methods can be adapted for use with any gene of interest that has enough sequence conservation to allow primers to be designed. The results of 16S-rRNA-gene-sequence studies have provided the strongest evidence that the microbial populations in nature contain many surprises.

**Limitations of current methods**

It is important to recognize the limitations of these molecular methods, which differ from the limitations imposed by culturing. Recent reviews have discussed the limitations of some of these approaches, particularly those based on PCR. Many common pitfalls are that many studies compare environmental sequences with the databases and conclude that a lack of a match is evidence that the organism has not been cultured. This ignores the fact that the 16S-rRNA in the database indicates that the organism has not been compared with the databases and conclude that a lack of a match does not mean that a lack of a match means that the organism has not been cultured. In addition, we need to move beyond a focus on 16S-rRNA-gene sequence analysis as the dominant measure of diversity (as summarized in Table 1). Natural-product production has, in many cases, been shown to be strain and not species specific, suggesting that 16S-rRNA-gene-based approaches underestimate the total functional diversity of microbial populations. Methods such as the bacterial artificial chromosome (BAC) approach are needed to access the physiological and biosynthetic diversity of microorganisms in nature.

**Soil microbial diversity: terra incognita**

Both culture-based and culture-independent approaches support the statement that soil represents one of the most diverse habitats for microorganisms. Along with marine and geothermal habitats, soil has been a major focus of molecular-ecological studies. Molecular investigations have confirmed soil as an environment particularly rich in diversity, as most studies obtain 16S rRNA sequences from several divergent bacterial divisions ('division' is used in the sense of Ref. 40). The α, β, γ and δ Proteobacteria are usually well represented, as are the Cytophagales, Actinobacteria and low-GC Gram positives. Other soil inhabitants appear to define the 'environmental subdivision' of the low-GC Gram-positive division as well as new groups in the Actinobacteria division. These discoveries are particularly relevant to natural-product discovery because cultured members of these two divisions are prolific producers of antibiotics.

Interestingly, researchers are also finding evidence of the existence of soil members of bacterial divisions not usually associated with soil habitats, such as the green non-sulfur bacteria, planctomycetes and spirochetes. Importantly, 16S-rRNA-gene sequences obtained from soil figure prominently in several newly proposed bacterial divisions and have contributed greatly to our expanding picture of bacterial diversity. Several newly prominent divisions merit further discussion.

- **16S-rRNA-gene sequences** from the Holophaga–Acidobacterium division and the low-GC Gram-positive division have been found in soil, as well as in other habitats. In fact, sequences of this group have been found in most, if not all, soil samples tested to date, including soils from at least four continents. They appear to be as common in soil as the well-known culturable members of the soil community are. These are few cultured species of this group. Up to eight subdivisions of this group have been defined and their phylogenetic depth approaches that of the well-known and physiologically diverse Proteobacteria.

- **4.10**
The Verrucomicrobia division is another example of the impact of culture-independent environmental studies on our understanding of bacterial phylogeny and diversity. This division is also poorly represented by cultured members but is richly represented in environmental sequences. Again, soil seems to be an especially rich source of sequences of this group, although the cultured members are mainly from aquatic sources. Sequences from soil are also well represented in other new candidate divisions. Some divisions, such as OP11, TM7, TM6 and WS1, do not contain any cultured members, suggesting that there might be major bacterial groups in soil about which we are almost completely ignorant.

Novel soil Archaea

In addition to many reports of novel soil bacterial 16S rRNA-gene sequences, there appear to be equally novel and unexpected members of the Archaea in soil. Several molecular-ecological studies have documented Crenarchaeota in soil\(^\text{49,52,58,59,64}\). The cultured Crenarchaeota are thermophiles, so the discovery of presumed mesophiles members of this clade in soil was unexpected. Additionally, novel methanogens and other ‘peculiar’ Archaea have been described in environmental samples\(^\text{52,65}\). These results further demonstrate the diversity of microorganisms in soil and underscore our lack of knowledge about them.

Accessing the unknown microorganisms

Our ability to detect potentially novel microbial types has been tantalizing. How can we learn more about these microorganisms? How can we gain access to their metabolic potential if we cannot culture them? Below, we describe two approaches to answering these questions, based on the direct extraction of community DNA from environmental samples.

The metagenome approach

One approach has its origin in eukaryotic genomics and involves the use of a BAC vector to clone large segments of DNA from environmental samples. BACs are becoming the vector of choice for eukaryotic genomics owing to their ability to maintain large fragments of foreign DNA stably in the \(\text{Escherichia coli}\) host. Eukaryotic BAC libraries are useful for genomic mapping, in vivo complementation and sequencing projects\(^\text{66,67}\). We have begun to use BACs as a surrogate expression system to study bacterial genomes, from both cultured bacteria\(^\text{68}\) and total microbial DNA extracted directly from soil (the metagenome). Our method is to clone large DNA fragments into the BAC vector and analyze the resulting libraries for novel phenotypic expression in the host \(\text{E. coli}\) strain\(^\text{22,68}\) (Fig. 2).

The adaptation of the BAC system to bacterial genomics has all of the advantages of BAC technology used in eukaryotic genomics plus the possibility that some gene expression from BAC clones will be obtained, because the insert DNA is prokaryotic. This could be very useful for the discovery of new natural products. The genes required for the biosynthesis of many antibiotics and other metabolites are usually clustered together, along with the genes for self-resistance\(^\text{43}\), and are often large and difficult to clone using traditional approaches.

Given the large size of BAC inserts, it is feasible to clone an entire pathway in one BAC plasmid, which provides a method for capturing, expressing and therefore detecting natural products produced from a BAC library made from environmental DNA. Furthermore, production in a heterologous and genetically defined system such as \(\text{E. coli}\) makes manipulating these pathways easier. Because each BAC clone will represent a subgenomic fragment and the host properties are well...
defined, there will be less chance for multiple activities per clone than in an entire organism. Screening isolated bacterial species can be complicated by the presence of multiple activities, as a given species often produces numerous biologically active metabolites.

The use of BACs for microbial-genomic analysis can be broadened by introducing the clones into other hosts such as Streptomyces or Bacillus and by cloning microbial DNA from other environments. Screening BAC libraries in these hosts can be achieved by developing shuttle BAC vectors that allow conjugation between E. coli and the alternative host or by constructing specialized BAC vectors specifically for use in other species.

BAC libraries of environmental DNA are also a resource for examining soil-microbial diversity by hybridization with specific probes, random sequencing or clone walking. BAC libraries provide a useful tool for examining the total genomic content of soil microflora. By cloning and analyzing large segments of soil microbial DNA, we can begin to ask more detailed questions about the physiology and functioning of microorganisms in nature. BACs thus offer a way to assess more completely the total diversity in a given environment (Table 1) by enabling us to examine the functional genomics of members of soil microflora, even if we lack the means to culture many of the organisms present. This approach provides a unique tool for expanding our knowledge of microbial diversity in nature, especially because, unlike most other methods for investigating environmental microbial diversity, it is not based on PCR. By combining studies of 16S-rRNA sequences with metagenomic analysis, we will take the first steps towards linking phylogeny and function of the total microflora in soil.

The expression-cloning approach

Another route to access the genomes of uncultured organisms is expression cloning. Developed for screening libraries of DNA from fungal isolates for enzyme production, this approach has been used with environmental samples presumably containing uncultured bacteria. DNA is isolated directly from an environmental sample, digested and cloned into a high-copy-number plasmid expression vector (Fig. 2), and the resulting clones screened for the production of industrial and biotechnological enzymes. This approach has been used to isolate novel enzymes from environmental sources without prior culturing of the producing organisms.

An adaptation of this approach involves PCR-based amplification of selected sequences from environmental DNA. It has been shown that novel polyketide-synthase-gene fragments can be isolated by this method, thus providing new sources of molecular diversity for combinatorial biosynthesis and module switching. The introduction of unknown DNA into a host can yield hybrid products whose synthesis is directed in part by the host genome and in part by the introduced clone.
Maintaining and protecting microbial biodiversity

- Our view of microbial diversity is expanding greatly, and techniques to measure, monitor and manipulate this diversity are being rapidly developed. We would like to suggest several broad questions to stimulate thought about future directions, answers to which we believe will certainly uncover more of the secrets of microbial life in soil.
- What fraction of soil microorganisms are "unculturable"? Represent new species?
- How can this be determined?
- What function do the unculturable microorganisms perform in soil?
- Is the phylogenetic diversity of the uncultured soil microflora reflected in its chemical diversity?

The vast microbial diversity of the natural world, combined with ingenious methods to access this diversity, can provide us with a bountiful source of new and useful natural products. The preservation of our valuable microbial resources is a major challenge, with the future benefit of humankind.

References

Rationalizing the design of polymeric biomaterials

Nela Angelova and David Hunkeler

Polymers are a promising class of biomaterials that can be engineered to meet specific end-use requirements. They can be selected according to key ‘device’ characteristics such as mechanical resistance, degradability, permeability, solubility and transparency, but the currently available polymers need to be improved by altering their surface and bulk properties. The design of macromolecules must therefore be carefully tailored in order to provide the combination of chemical, interfacial, mechanical and biological functions necessary for the manufacture of new and improved biomaterials.

Polymers remain the most versatile class of biomaterials, being extensively applied in medicine and biotechnology, as well as in the food and cosmetics industries. Applications include surgical devices, implants and supporting materials (e.g. artificial organs, prostheses and sutures), drug-delivery systems with different routes of administration and design, carriers of immobilized enzymes and cells, components of diagnostic assays, biosensors, ocular devices, and materials for orthopaedic applications.

Polymers used as biomaterials can be synthesized to have appropriate chemical, physical, interfacial and biomimetic characteristics, which permit various specific applications. Compared with other types of biomaterial, such as metals and ceramics, polymeric systems offer the advantage that they can be prepared in different compositions with a wide variety of structures and properties. Current research and development is focused on tissue engineering, for which such materials are considered to have a particularly significant potential.

After more than three decades of development, in which numerous polymers have been used to replace body parts or to help to restore vital functions, clinical success is still relatively rare. Therefore, in spite of the large number of existing biomaterials, additional studies in this field, concomitant with an often lengthy regulatory-approval process, are motivated by the need for more-durable implants and transplants. For example, hip replacement, which can be carried out only twice, has a mean duration of ten years. In addition, there is a growing interest in specific polymeric systems.

Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biocompatibility</td>
<td>The ability of a material to perform with an appropriate host response in a specific application.</td>
</tr>
<tr>
<td>Biomimetic material</td>
<td>Artificial material that resembles the original, biologically produced precursor in micro- and macrostructure.</td>
</tr>
<tr>
<td>Endotoxins</td>
<td>Lipo polysaccharides, a toxic component of Gram negative bacterial cell walls.</td>
</tr>
<tr>
<td>Inlet of Langerhans</td>
<td>Cluster of pancreatic cells that produces insulin. A family of cells, present in all tissues and arising from three germ layers, specialized for the establishment of the tissue structure.</td>
</tr>
<tr>
<td>Fibroblast cells</td>
<td>Cells belonging to the immune system. The tendency for a material to induce clot formation when in contact with blood.</td>
</tr>
<tr>
<td>Macrophages</td>
<td>Formation of a blood vessel system within a tissue.</td>
</tr>
<tr>
<td>Thrombogenicity</td>
<td></td>
</tr>
</tbody>
</table>

N. Angelova (nangelova@imperial.ac.uk) and D. Hunkeler (david.hunkeler@epfl.ch) are at the Laboratory of Polymers and Biomaterials, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland.