Discussion

Soil molecular microbial ecology at age 20: methodological challenges for the future

Andrew Ogram*

Soil and Water Science Department and Microbiology and Cell Science Department, University of Florida, 2169 McCarty Hall, Gainesville, FL 32611-0290, USA

Accepted 29 April 2000

Abstract

The year 2000 marks the twentieth anniversary of the publication in Soil Biology & Biochemistry by Vigdis L. Torsvik (University of Bergen) of the first procedure for isolation of bacterial DNA from soil (Torsvik, 1980), arguably initiating the subdiscipline of soil molecular microbial ecology. Since 1980, great strides have been made in the development of methods and in the application of genetic tools to analysis of soil microbial communities, and many soil microbiology laboratories routinely incorporate these tools in their research. It is likely that the concept of soil molecular ecology will soon disappear as a subdiscipline of microbial ecology, and that these tools will become as routine and indispensable as are genetic tools in microbial physiology. However, even though increasing numbers of soil microbiologists use molecular biology in their research, some fundamental obstacles must be overcome before these tools become as routine as are, for example, many soil chemical methods. This anniversary provides an opportunity for retrospection on the applicability of genetic tools to soil microbial ecology, and of methodological needs for the immediate future.

Keywords: Methods; Molecular ecology; Soil DNA

1. Purification of nucleic acids from soils

The first requirement of most molecular genetic procedures, and the first obstacle to overcome, is the purification of nucleic acids extracted from soil. The efficiency of cell lysis, efficiency and degree of nucleic acid purification, and size of the isolated nucleic acids are crucial to the success of the subsequent applications. The composition of the isolated DNA is dependent on the efficiency of lysis (some bacterial groups are far more difficult to lyse than others, Moré et al., 1994), and PCR-based methods are dependent on the purity of the samples (PCR is inhibited by humic contaminants, Tebbe and Vahjen, 1993). Most cloning applications require relatively large fragments of DNA, particularly if expression of entire operons is required (Handelsman et al., 1998); and isolation of small fragments may increase the frequency of formation of PCR artifacts (Zhou et al., 1996).

Numerous methods for isolation of bacterial DNA from soils have been published in a variety of journals over the past 20 years, with significant advances being made over Torsvik’s original procedure (Torsvik, 1980). Her procedure involved separation of bacterial cells from particulates by differential centrifugation, followed by lysis and subsequent separation of bacterial DNA or RNA from soil organic carbon by a series of chromatographic separations. As with most first generation procedures, this was time consuming and inefficient; the original procedure required sample sizes of 60–90 g and at least 3 days for completion (Torsvik, 1980; Holben et al., 1988).

Second generation methods are streamlined and require smaller sample sizes, thereby allowing simul-
taneous processing of multiple samples (Tsai and Olson, 1991; Kuske et al., 1998). Most current procedures are based on the direct lysis of cells in the presence of the soil (rather than first separating cells from sand, silt and clay particles), followed by extraction of nucleic acids from soils and their subsequent purification (Ogram et al., 1987). Small sample sizes (from 250 mg to 1 g) are used such that much of the procedure may be conducted in micro centrifuge tubes, and recovery efficiencies of over 80% have been reported (Zhou et al., 1996; Kuske et al., 1998). However, although direct lysis of bacterial cells and subsequent purification of the released nucleic acid typically yields higher amounts of DNA than the cell fractionation method, it also extracts much more humic compounds than the original cell fractionation method (Torsvik, 1980). A variety of purification methods to remove humates have been described, including selective precipitations and extractions (Guo et al., 1997), various chromatographic matrices (including ion exchange, size exclusion, and hydrophobic interaction matrices) (Ogram, 1998), capture by magnetic beads (Jacobsen, 1995), and agarose gels (Young et al., 1993). Times required for complete processing typically range from 4 to 12 h. Isolation of RNA from soils is somewhat more difficult due to its labile nature (Ogram et al., 1995) and alternative methods have been developed to circumvent these problems (Felske et al., 1996). A few companies (e.g. MoBio, www.mobio.com; Bio101, www.Bio101.com) now market commercial kits for purification of soil DNA which work quite well for many applications, and provide very rapid purifications (less than 30 min).

Even though significant advances have been made in the efficiency and speed of DNA purification from soils, this is still the limiting step in many cases. It is likely that the primary reason so many procedures for soil DNA purification have been published is that most procedures are optimized for a specific soil. This implies that any given procedure may not be universally applicable to all soils (Zhou et al., 1996). Soils are highly variable with regard to physical and chemical properties, particularly with regard to the structure of organic carbon. Co-purification of organic carbon, particularly humic and fulvic acids, is the primary obstacle to efficient recovery of DNA suitable for amplification by PCR. A procedure that efficiently separates humic compounds from one soil may not do so from another soil. Operator variability is also likely to be a factor in the variability of DNA recovery between laboratories, prompting the development of new procedures by individual laboratories. Even within laboratories, considerable variability between quality and amount of isolated DNA is sometimes observed.

A single method for purification of DNA from all environmental matrices would be of great benefit to the advancement and general use of molecular genetic tools in soil microbial ecology. This would be a great step toward automating the procedures, and for standardizing results observed between laboratories. A procedure that is equally efficient in all matrices, efficiently lyses all target organisms, requires little time to complete, and is amenable to processing multiple samples simultaneously would be highly desirable.

Development of a generally accepted strategy for evaluating lysis and purification efficiency would assist in developing a universal purification procedure. A variety of approaches have been used, including those based on recovery of exogenous labeled nucleic acids (Ogram et al., 1995; Lee et al., 1996) or cells (Steffan et al., 1988) added to a sample, and those based on estimations of purification from crude DNA extracts from a given sample (Zhou et al., 1996; Sandaa et al., 1998). None of these approaches is without drawbacks, and the discipline would be well served if a benchmark method for assessing purification efficiency was developed.

Development of a universally accepted method for cell lysis and purification of nucleic acids extracted from environmental samples would be a valuable prelude to the development of a universal DNA/RNA purification procedure. A significant step toward an automated lysis/purification system is currently being developed by environmental scientists at the US Department of Energy's Pacific Northwest Laboratories (Chandler et al., 1999), with the potential to add on-line rapid PCR. Similar methods for rapid isolation of nucleic acids and analysis are being developed currently as a part of the US Department of Defense's research initiatives for detecting biological warfare agents (http://www.darpa.mil/), and it is likely that this research need will help drive technologies for environmental DNA isolation at a rapid pace.

2. Assessment of bacterial community composition

In her 1980 paper, Torsvik alluded to the potential use of DNA:DNA reassocation kinetics to assess the diversity of uncultured bacteria in soils. This pointed toward her next landmark work, published 10 years later (Torsvik et al., 1990a, 1990b). In this series of papers (see also Torsvik et al., 1994), DNA:DNA reassociation kinetics were used to estimate that over 10,000 species of bacteria (most of which had never been cultivated in the laboratory) were present in one gram of forest soil. DNA:DNA reassociation kinetics was originally developed to assess the complexity of mammalian genomes (Britten and Kohne, 1968), and was commonly used to determine the molecular weight of bacterial genomes (Gillis et al., 1970). This approach is based on the assumption that more com-
plex denatured DNA reassociates at a slower rate than less complex denatured DNA, and that the kinetics of reassociation are proportional to the genomic complexity. Reassociation kinetics has a number of limitations when applied to analysis of soil bacterial diversity, and consequently has rarely been used to assess soil bacterial diversity (Ritz et al., 1997). It is technically difficult, requires large amounts of DNA, is time consuming, and usually requires a UV spectrophotometer dedicated to this analysis for prolonged periods. This approach has not been used as a routine method for assessment of bacterial diversity, and it is unlikely that it is sufficiently sensitive to assess the changes in diversity that would be of interest to most soil ecologists. Nonetheless, the advantage of this approach is that it may be the only means developed to date by which the total number of bacterial species within a soil sample may be estimated (assuming the efficiency of lysis is not limiting).

Other methods derived from analysis of complex single organism genomes have been applied to comparisons of soil communities, with limited success. These include DNA:DNA cross hybridization (Ritz and Griffiths, 1994; Xia et al., 1995) and random amplified polymorphic DNA (RAPD) fingerprinting (Xia et al., 1995). Cross hybridization provides a reasonably good means of comparing two communities (just as it is used to compare similarities between two bacterial strains; Stackebrandt and Goebel, 1994), although it suffers from a lack of sensitivity that makes quantitative comparison of communities possessing similar structures difficult. RAPD analysis has been used for quickly scanning differences between communities (as with scanning complex genomes; Williams et al., 1990), but is semi-quantitative at best and may lack reproducibility.

Perhaps the greatest advance in microbial ecology methods during the second half of the twentieth century came when Norman Pace (for a review, see Hugenholtz et al., 1998) translated Carl Woese’s concepts of rRNA phylogeny (Woese, 1987) to the analysis of natural microbial communities. This greatly expanded our knowledge of the diversity of soil bacteria and provided the basis for most current methods of diversity analysis. Most of these methods are based on the diversity of 16S rDNA sequences present in a sample, and include cloning and sequence analysis (Liesack and Stackebrandt, 1992), amplified rDNA restriction analysis (Moyer et al., 1994), denaturing gradient gel electrophoresis (Muyzer et al., 1993), and terminal-restriction fragment length polymorphism analysis (Liu et al., 1997). These approaches are useful for analyzing the richness (i.e. numbers of species) of small segments of a soil community through the use of PCR primers specific for a given phylogenetic subset of the total community, such as the α-Proteobacteria. They are not appropriate, however, for assessing the total bacterial species richness or species evenness (i.e. relative proportions of individual members in different species) due to the very complex nature of a soil microbial community and limitations of PCR. It is not possible using current rDNA-based technology to account for potentially greater than 10,000 distinct species in a single sample.

Among current limitations of rDNA-based methods for analysis of soil bacterial diversity are the potential formation of PCR artifacts (Wang and Wang, 1997), the potential discrepancy between the quantitative composition of rRNA genes within the template (sample) DNA and the final amplification product (Farrell et al., 1995; Suzuki and Giovannoni, 1996; Polz and Cavanaugh, 1998; Suzuki et al., 1998), the time and cost required to screen several phylogenetic groups simultaneously, and the tedious, time and cost involved in sequencing and sequence analysis (if identification is desired). This information may or may not be related to the current concept of the definition of a bacterial species, dependent on the individual species (16S rDNA sequences are not used to define species; strains belonging to a given species as currently defined may have different 16S rDNA sequences) (Stackebrandt and Goebel, 1994). Methods are required that will allow the rapid and simultaneous analysis of bacterial composition or diversity (including species richness and evenness) of multiple samples and multiple phylogenetic groups, and database analysis packages required to reduce and analyze these data must be developed. The relationship between the information obtained and numbers of individual species in the sample must be clearly defined.

These methods are moving toward more quantitative approaches, but not without difficulty. Absolute numbers of individuals belonging to a given phylogenetic group, and relative numbers of individuals within various groups, are difficult to generate using PCR-based approaches. Quantitative PCR methods have been applied for this purpose (Lee et al., 1996; Chandler, 1998; Felske et al., 1998; Johnsen et al., 1999), but they are tedious and are not appropriate for enumerating large numbers of target species. Quantitative PCR methods based on the analysis of rRNA genes are subject to all of the uncertainties associated with PCR, and with quantitative uncertainty of the numbers of rRNA genes (rrn operons) per genome. Copy number ranges from 1 to 14 in different bacteria, making quantitative relationships between rrn copies and cell number difficult to define without having first characterized the target organism. Estimates relating rrn concentrations in a soil sample and cell number may be made based on relationships between genome sizes and rrn copy number (Fogel et al., 1998).

Solutions to these problems will likely be found in
the near future, and it may be that the relatively new approaches of genomics and proteomics will provide the vehicles. Genomics will probably provide new approaches for analysis of soil communities, and the associated approaches of high throughput sequencing and bioinformatics will undoubtedly play important roles in microbial ecology over the next 10 to 20 years. As technologies for sequence determination advance, our current reliance on PCR may be replaced with the complete sequencing and analysis of soil community genomes. Torsvik's use of reassociation kinetics to define the complexity of a community genome is an early application of genomic approaches to study community complexity. Genomic approaches are already opening up previously untapped genetic resources from soils for industrial and pharmaceutical applications (Handelsman et al., 1998), and DNA microarrays are currently being developed for rapid analysis of microbial communities (Guschin et al., 1997; Kelly et al., 1999; Murray et al., 1999).

Microarray assays are not based in PCR, and may eventually provide easily accessible quantitative data regarding the relative representation of individual phylogenetic groups in soils or for the relative activities of different phylogenetic groups within a sample or between samples. In general, microarrays are constructed by fixing tens of thousands of oligonucleotides to a glass or nylon membrane, which is then hybridized to a sample. For community structure analysis, oligonucleotides directed toward various phylogenetic groups of interest are fixed to the support. For analysis of individual strains, oligonucleotides directed toward open reading frames detected from the genome of a given strain are fixed. This approach has great potential for elucidating a range of aspects of soil microbial ecology, including identification of differences in community structure between different samples, identification of which phylogenetic groups may be active or inactive during a particular treatment (Guschin et al., 1997), and identification of differences between strains isolated from different environments, which may lead to identification of possible differences in niche occupation (Murray et al., 1999).

rDNA analysis provides a powerful approach for identification of bacterial groups within soil communities, but is limited with regard to elucidating the physiological states of individual strains or their activities, and ecological roles within the community. It is likely that molecular genetic methods will contribute to answering these questions, and RNA based methods have been shown to provide powerful tools for assessing in situ activities (Fleming et al., 1993; Devereux et al., 1996). Determination of activities of specific genes other than rRNA, as assessed by measurement of specific mRNA concentrations (Fleming et al., 1993), is currently limited by the relatively small database available on the diversity of bacterial genes present contributing to a given function in a community. New approaches, such as those currently used in bioprospecting (searching for novel catabolic, biosynthetic, or antibiotic functions in environmental samples, Handelsman et al., 1998) will be necessary for identifying undiscovered, and potentially unculturable, genotypes. Efficient cloning strategies for large fragments of soil DNA and high throughput sequencing may yield information of the metabolic diversity of non-culturable strains, which could provide important clues to the ecological roles of non-cultivated strains.

3. Conclusions

Much research over the past 20 years has been devoted to the development of basic methods that will allow routine application of molecular genetic tools to analysis of soil microbial communities. It is likely that significant effort over the next 20 years will also be devoted to overcoming methodological limitations of analyzing diverse mixed genomes. Accurate routine quantitative analysis of aspects of community structure is currently the greatest obstacle to be overcome, and will require advances in DNA isolation methodology and community structure analysis. Many of the methods currently in use were originally developed for genetic analysis of higher organisms rather than of soil communities, and it is likely that advances in methodologies for mammalian and plant genetics, such as those in the newly emerging fields of genomics and proteomics, will continue to provide new methods for analysis of complex soil communities. It is also likely that new methods developed specifically for analysis of microbial communities will ultimately be required for molecular genetics to become completely quantitative for routine use in soil microbial ecology laboratories. It is apparent that molecular genetics does not represent a panacea for all difficulties in soil community analysis (and has never been represented as such), and a polyphasic approach incorporating various disciplines, including conventional isolation techniques and screening functional capabilities, in addition to molecular genetics may be required to completely describe individual aspects of soil microbial ecology.

References

Chandler, D.P., Schuck, B.L., Brockman, F.J., Bruckner-Lea, C.J.,


Suzuki, M., Giovannoni, S., 1996. Bias caused by template annealing
in the amplification of mixtures of 16S rRNA genes by PCR. Applied and Environmental Microbiology 62, 625–630.


