Molecular community analysis of microbial diversity
Ingela Dahllöf

New technologies that avoid the need for either gene amplification (e.g. microarrays) or nucleic acid extraction (e.g. in situ PCR) have recently been implemented in microbial ecology. Together with new approaches for culturing microorganisms and an increased understanding of the biases of molecular methods, these techniques form the most exciting advances in this field during the past year.

Addresses
Department of Marine Ecology NERI, Fredriksborgvej 399 4000, Roskilde, Denmark;
e-mail: ind@dmu.dk

0958-1669/02/$ — see front matter
© 2002 Elsevier Science Ltd. All rights reserved.
DOI 10.1016/S0958-1669(02)00314-2

Abbreviations
DGGE double-gradient gel electrophoresis
FISH fluorescence in situ hybridisation
IS-PCR in situ polymerase chain reaction
ITS internal transcribed spacer
PCR polymerase chain reaction
TGGE temperature-gradient gel electrophoresis
T-RFLP terminal restriction fragment length polymorphism

Introduction
The study of microbial diversity and community analysis has leapt forward since the advent of DNA sequencing, which in turn has revolutionised our understanding of microbial phylogeny [1]. The concept of species in microbiology will only be touched upon here, but there are many good articles on this subject [2,3••,4]. The development of molecular techniques has made it common to investigate community diversity using the rRNA gene (rDNA) or the rRNA itself. The rapidly growing rDNA sequence data bank, accessible via the Internet (www.ncbi.nlm.nih.gov/entrez/), now makes it possible to compare sequences from across the world. Being able to determine microbial diversity at a high-resolution level (groups, species and strains) without the need for cultivation will further our understanding of several issues; for example, it will help us to determine structure–function relationships and to analyse the interactions formed between microbes and the abiotic environment and other organisms. The basic approach for molecular diversity analysis is shown in Figure 1 and is discussed in detail below. Briefly, isolates from a community can be subject to direct analysis, using in situ methods, or nucleic acid can be extracted for analysis using microarrays or dot-blot hybridisation. A gene might also be amplified using the polymerase chain reaction (PCR) and evaluated using a range of techniques including pattern analysis, cloning and sequencing, probe hybridisation and microarrays.

As the amount of available information grows, it becomes increasingly important to ask what the aim of mapping diversity is, and to identify the traps that can be fallen into when interpreting and extrapolating our findings using the current techniques? The aim of this review is to provide the reader with some insight into the various approaches published during the past year, so that an appropriate choice of methods for each application can be made.

Choice of gene
The use of 16S rDNA or rRNA is currently the most common approach for community analysis and therefore deserves close scrutiny. It is obvious that the phylogenetic properties of 16S, as well as the large amount of sequences available, offer a considerable advantage, but there are also disadvantages. For example, the heterogeneity of 16S between multiple copies within one species [5,6] hampers pattern analysis [7], and can confuse the interpretation of diversity from clone libraries and sequences retrieved from banding patterns. The extent of 16S heterogeneity does vary between different regions [8,9••], but so does resolution [10]. It has also been shown that 16S lacks resolution at the species level, most recently in Bacillus [11]. By contrast, the region between the different rRNA subunits, the internal transcribed spacer (ITS) [12] used in ribosomal intergenic spacer analysis (RISA) [13], was shown to distinguish closely related strains where 16S lacks resolution [14,15]. However, there is also evidence of heterogeneity in the size and sequences within the ITS regions of one species [6].

Caution and thoroughness, as reported in Zhou et al. [16••], are extremely important if 16S or any other heterogeneous gene is used to draw ecological conclusions concerning diversity and abundance. The use of other genes, such as that for the σ factor rpoB [11,17], which appears to be present in only one copy and has shown higher discrimination between species for some groups [11], can therefore be recommended both for pattern analysis and clone libraries. The lower number of sequences currently available in the gene libraries is, at the present time, a limitation for this type of gene when it comes to species identification, but could be overcome rapidly as sequencing is no longer such a daunting task.

Functional genes — such as the gene for nitrate assimila-

nase A [18], nodD for rhizobia communities [19••], and

pmoA and msxA found in methanotrophic bacteria [20,21]
— are other alternatives used in diversity studies, especially when investigating structure–function relationships. Both 16S and pmoA in methanotrophs retrieve the same phylo-
genetic information [22], whereas amoA has been shown to give similar, but not identical, evolutionary relationships for ammonium oxidisers when compared with 16S [23••].
Figure 1

Common approaches to the analysis of microbial diversity. Boxes with bold frames indicate preparatory steps necessary before detection and analysis can be performed; boxes with thin frames represent detection techniques. FISH [25] and IS-PCR [26*] can be used directly, whereas for the other techniques the DNA is first extracted. Microarrays [31*,32*,40,52] or dot-blot hybridisation [27,48,53] can use whole DNA or the DNA can be taken through the polymerase chain reaction (PCR) to amplify a gene fragment. The community structure can then be evaluated through pattern analysis (fingerprinting) of the PCR product, through cloning and sequencing, through hybridisation with probes or to microarrays or using a combination of these techniques. ARDRA, amplified rDNA restriction analysis; DGGE, double-gradient gel electrophoresis; FISH, fluorescent in situ hybridisation; ITS, internal transcribed spacer; SSCP, single-stranded conformation polymorphism; T-RFLP, terminal restriction fragment length polymorphism.

This type of gene needs to be further investigated with respect to heterogeneity and resolution.

Sampling and extraction

One aspect of sampling that is rarely mentioned concerns the effect of sample size and replication. In their study, Smit and colleagues used ten samples (each of 10 g) from each soil type investigated [24], but in contrast most studies use only a single extraction of less than 1 g to represent a sample type. If no previous knowledge of variation within a site is available, replication is necessary to enable comparison between sites. Sample size, replication and possible pooling of extracts are all dependent on the sought resolution of diversity. For example, is it possible to compare different soil types using 1 g samples without first proving that 1 g samples are representative for each soil?

The sample itself can be analysed directly using fluorescence in situ hybridisation (FISH), which was recently covered in a Current Opinion review [25], or using the in situ polymerase chain reaction (IS-PCR) [26**]. The drawback with FISH is the need for genes or RNA with a high-copy number. IS-PCR presents a solution to this problem, as single-copy genes can also be detected using this method. The advantage with both techniques is that heterogeneity within gene copies is not an issue and quantification can be made directly [27]. These techniques are most suitable when targeting a specific group or species, rather than aiming to map total diversity, as there are practical limits to how many probes and/or primer sets can be used.

Numerous studies have shown species bias for different DNA and/or RNA extraction methods [28*,29]. The extraction efficiency can be checked [29,30], although this is hardly feasible when working with a large number of samples. It is therefore important to choose the right extraction method for the community in question.

Extracted DNA can be used directly for slot-blot hybridisation or for analysis using microarrays. Direct DNA probing has been shown to be less sensitive than probing with PCR products [29,30], although sensitivity was increased for non-PCR amplified rRNA using fragmentation and a chaperone-detector probe strategy [31**]. The usefulness of a rapid combined extraction and fragmentation approach for both RNA and DNA, together with a portable microarray system, has also been demonstrated [32*]. Similarly, a gene array for functional genes has proved useful [33]. Microarrays have the potential to map total diversity, but they do not reveal unexpected species. For this reason they are most applicable for the determination of changes in an already characterised sample.

PCR

The PCR amplification step is known to introduce other biases, irrespective of the gene targeted. Primer specificity is a major stumbling block, especially when attempting to quantify a mixture of homologous target sequences [34]. All techniques that are based on PCR (cloning, pattern analysis and sequencing) will be affected by the biases introduced by PCR. It was recently shown that the original sample template is amplified during the initial 5–6 cycles of the PCR reaction [35], and that in the following cycles amplification occurs only on the PCR fragments produced earlier. This implies that sequences with a good primer match and high copy number will be selected for. Constructing truly general primers has proven difficult, and even single mismatches in the middle of the primer can cause a preferential selection [36]. Lowering the annealing temperature allows for mismatches and increases diversity in the PCR product [37*], but also increases the risk of unwanted by-products. Degenerate primers could be used [21], but these are not suitable for techniques like double-gradient gel electrophoresis (DGGE) as they produce multiple bands from one template, which gives the same problems as when using heterogeneous genes.

Pattern analysis

Pattern analysis, or fingerprinting, is often carried out by evaluating banding patterns of PCR products on gels, and
all of the techniques discussed below can be used to simply identify differences between communities. However, when using more refined analyses, such as diversity indices, it is necessary to ensure that each species is represented by one band only in order to determine changes in community structure correctly. Techniques that use enzyme digests — such as amplified rDNA restriction analysis (ARDRA), restriction fragment length polymorphism (RFLP), and single-stranded conformation polymorphism (SSCP) — produce multiple bands for single species [29], making community patterns difficult to evaluate further. These techniques are more suitable for screening clone libraries or isolates before sequencing.

Terminal restriction fragment length polymorphism (T-RFLP) gives only one band per species as only the fragment containing the fluorescently labelled primer site will be detected. The samples are normally run on long sequencing gels that give high resolution and sensitive detection. One disadvantage of this method, however, is how to select restriction enzymes when sequences are unknown. Using the known sequences in the gene library it was shown that many species share the same length of fragment even when an optimal enzyme was selected [38]. Separation through melting properties (DGGE or temperature-gradient gel electrophoresis [TGGE]) lacks the resolution of T-RFLP, thereby potentially hiding diversity [22,39]. Resolution can be improved, for example, by a double-gradient method where a gradient in polyacrylamide is run parallel to the denaturing gradient [10]. T-RFLP and DGGE/TGGE can both be recommended for pattern analysis without further sequencing if a non-heterogeneous gene is used, and with additional sequencing using heterogeneous genes.

Diversity analysis through hybridisation of probes to the PCR product, or of PCR products to microarrays with probes attached to them, are techniques where identification at the group or species level can be made directly without further sequencing. The sensitivity of microarrays is reported to be higher than for gels [40]. An alternative to microarrays is to use beads with attached probes combined with flow cytometry, through a probe-unique fluorescent labelling of the beads [41]. The disadvantage is, as in the earlier case of hybridisation and microarrays, that you essentially only find what you are looking for.

Samples with high diversity can cause resolution problems when using fingerprinting techniques [21]; however, by targeting a specific part of the population, either by using functional genes or group-specific parts of general genes [20,22], patterns may be resolved.

Bands from gels can be identified to species level through sequencing and comparison with the sequence library. Bands from longer PCR products give better identification possibilities, although this must be weighed against separation efficiency. The alternative is to create a clone library, which contains larger gene sequences and thereby provides a more positive identification. However, a large number of clones have to be sequenced to get an appreciation of the diversity [42*], and clone libraries are victims of the same biases as other PCR methods. Again the choice of gene and thoroughness of the study is important for a conclusive estimation. Retrieved sequences are often only compared with the library, but it is equally important to compare them to each other to rule out dissimilarities that can result from sequencing errors or heterogeneity of the gene.

**Culturing approaches**

The advent of molecular techniques opened new perspectives to microbial diversity and it was realised that previous culturing of environmental strains had been highly selective. Again, cloning of PCR products revealed a higher diversity than culturing [24,43], but there were also differences in the species identified by the techniques [24,44**]. Now we are also beginning to understand the selectivity of molecular techniques, based on all the steps: sampling, extraction, PCR, identification techniques, choice of genes and so on. The need for culturing is still high, not only to provide better background material for further development of molecular techniques, but also to enhance the understanding of function and structure. New isolation approaches, which mimic to a higher extent the environmental conditions from which the samples are drawn, are clearly the way forward. A study examining the influence of media on the isolation of *Pseudomonas* spp. showed that low-nutrient media was better-suited to low-nutrient soils [20], whereas dilution culturing gave a higher estimate of diversity [45]. High similarity to the original sample matrix, and the selection of single cells for cultivation, was successfully used to isolate novel halophiles from Red Sea brine [46**], as well as ammonia oxidisers [47].

**Conclusions**

There is no single technique available today that can catch the entire diversity of a microbial community. Biases are introduced at each treatment step, and only through an iterative process where culturing, *in situ* techniques and PCR-based methods are all used can we further our understanding. Meanwhile, it is important to be very clear about the question being addressed, and care should be taken with ecological conclusions drawn.

The ecological relevance of the community structure for the function of systems is a compelling reason to study microbial diversity. So far, most studies have been limited to descriptions of diversity and the reporting of new sequences. These studies are still important, but it is only through a combination of approaches that the connection of community structure to function can be made. Such studies are becoming more common [48], especially as the molecular techniques move into neighbouring fields such as ecology [49,50] and ecotoxicology [51]. Expression analysis of functional genes, for example, mRNA detection in conjunction with microarrays [52], together with
functional measurements such as substrate utilisation (e.g., using BIO-LOG® [44**]) and metabolic activity [48], provide new avenues towards understanding the relationship between structure and function.

Acknowledgements

M Taylor and R Case at the University of New South Wales, and the editors are gratefully acknowledged for their critical comments and proof reading.

References and recommended reading

Papers of particular interest, published within the annual period of review, that have significant new results or perspectives have been highlighted as:
• of special interest
•• of outstanding interest


This study demonstrates how the choice of universal 16S primers can affect diversity analysis.


The use of a specific gene to exclude all but the group of interest is successfully shown.


An important comparison of genes for phylogenetic studies.


An exciting alternative to FISH that can be used for all genes, active or not. This method opens the way to detection of functional gene expression.


A thorough study demonstrating the importance of the extraction step.


Introduces a sensitive microarray analysis avoiding the PCR step.


