Analysis of 16S rDNA clone libraries: Part of the big picture

E. Stackebrandt, R. Pukall, G. Ulrichs, H. Rheims

DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, 38124 Braunschweig-Germany.

ABSTRACT

Analyses of 16S rDNA from cultured organisms and clone libraries have set the pace in molecular microbial ecological studies. Sequences led to the development of oligonucleotide probes and PCR primers used in the determination of community complexity and changes in structure. Today, clone libraries are part of a multi-structured strategy in which genes coding for proteins, mRNA and rRNA play equally important roles aimed at the understanding of complex ecological processes. This contribution summarizes general problems involved in the interpretation of clone libraries, and focuses on two particular problems of such analyses.

Introduction

The modern era of microbial diversity started with the analysis of nucleic acids recovered either directly from the environment or from isolated prokaryotic cells. Originating from the pioneering work of Pace and collaborators [6], the taxon composition of a microbial community was elucidated through analysis of ribosomal RNA molecules, firstly of 5S rRNA patterns and secondly of sequences of 16S rDNA fragments generated by shotgun cloning. Technical refinement included the cloning of PCR amplified 16S rDNA molecules and the use of taxon-specific PCR primers which narrow the range of retrieved sequences. A wide range of environments has been assessed with respect to sequence diversity. Knowledge about the 16S rDNA sequence diversity of cultured and as yet non-cultured organisms has stimulated the development of techniques which explore the *in situ* spatial and temporal structure of microbial diversity (e.g. by application of fluorescently labeled oligonucleotide probes and by recognition of DGGE and TGGE patterns [4]), and assessment of the metabolically active part of a community by probing ribosomal RNA with taxon-specific oligonucleotides [2].

Nevertheless, although the contribution of rDNA and rRNA to the elucidation of environmental connections must be considered significant, one molecule alone cannot nearly cover all facets of microbial ecology. Not only must the function of an ecosystem be deduced from analyses of genes expressed through mRNA and proteins, but the network of broad physiological interactions should also be verified by a quasi *in vitro* reconstitution by isolates. All these strategies must be accompanied by thorough physical and chemical analysis of the natural sample. In order to understand ecological interactions, strain richness and strain abundance must be assessed, and the physiological diversity of strains recognized. At present these goals are out of reach considering problems encountered with the analysis of clone libraries (see below) and the lack of knowledge of the relevance of the physiology

Microbial Biosystems: New Frontiers
Proceedings of the 8th International Symposium on Microbial Ecology
Bell CR, Brylinsky M, Johnson-Green P (ed)
Atlantic Canada Society for Microbial Ecology, Halifax, Canada, 1999.

of strains tested under laboratory conditions to their activity in the environment. Ecologists may hope to assess the biochemical reactions and to test the function of a microbial community - provided strain diversity is low, such as those defined by extreme conditions (temperature, pH value, low nutrients) [3]. However, even a thermophilic and acidophilic pond may contain representatives of nearly all major prokaryotic lines of descent [6].

Factors influencing the reliability of 16S rDNA clone libraries

Factors identified to change the relative proportion of naturally occurring taxa during the generation of rDNA clone libraries include (i) methodological shortcomings, (ii) problems intrinsic to the molecule of choice and (iii) problems due to the interpretation of data. As most of these pitfalls have been compiled and discussed in a recent review [16] the listing is kept as a brief compilation and individual points are not cited.

Methodological problems

Methodological problems include obtaining and handling of samples. Moving a natural sample to a mesocosm within less than two hours will change the relative ratio of taxa (Muyzer and Schäfer, pers. communication) and it can be expected that this effect will also occur during the sample transport to a laboratory. A severe bias is introduced by cell lysis and extraction of DNA as the release of nucleic acids depends markedly on the structure of the peptidoglycan and outer membranes of bacteria. Looking at the composition of clone libraries for which DNA was extracted from soil by gentle enzymatic lysis procedures, the occurrence of Gram-positive taxa is grossly underestimated in comparison with results obtained from isolation studies. On the other hand, rigorous lysis procedures may be detrimental to the intactness of genes, leading to the increased formation of chimeric structures or non-amplification of DNA targets.

Bias introduced by PCR amplification includes differences in the specificity of polymerases, inhibition of the reaction by interfering substances, differential PCR amplification and PCR artifacts (e.g. chimeric structures, formation of deletion mutants and point mutants). Differences in the clonability may be due to differences in commercially available cloning kits and in the relative ratio of PCR amplicons, in that products representing minor fractions of populations may be suppressed. Yet another problem has been described recently [15] which points out sequence diversity through possibly contaminated reagents.

Intrinsic problems

Intrinsic problems include the differences in genome size, and number of *rrn* operons. PCR amplification of *rrn* operons will be discriminated in DNA originating from strains with large genomes and a small number of *rrn* operons against DNA from strains with small genome sizes and a large number of *rrn* operons. As a microbial population is composed of strains ranging between these two extremes a significant shift of taxon representation is to be expected in a clone library. In addition, *rrn* operon microheterogeneity has been detected in many taxa, ranging from extensive differences in highly variable regions to point mutations scattered along the highly moderately variable regions [9].

Interpretation problems

Interpretation problems are connected with the analysis of sequences. Considering the high influence of biasing steps in the composition of clone libraries it is unlikely that the analyzed sequences represent the true diversity of environmental taxa. For obvious reasons most clone libraries are not substantially analyzed and most publications cover no more than about 100 clones (probably less than 0.1% of total). Sequences have to be compared to those published previously for cultured strains and clones from other libraries. Another problem arises from the preference of scientists to analyze different stretches and different lengths of the cloned 16S rDNA molecule. Although the selection may be justified from the taxon-specific location of variable regions [13], non-comparable stretches make it necessary to construct composite trees in which the branching order is highly tentative.

Diversity of culturable and cultured strains

A sequence defining a novel lineage is generally interpreted in terms of the presence of an as yet uncultured organism, but the availability of already cultured and easy to culture novel taxa is grossly underestimated. The number of only 4000 validly described species reflects not only the lack of rapid identification methods which would detect molecular differences in phenotypically similar organisms, but also our inability to keep pace with the description of novel isolates. Analysis of marine bacteria by 16S rDNA techniques has indicated a wide spectrum of novel organisms. Comparative investigation by the use of molecular methods and isolation work resulted in the recovery of novel strains which, following phylogenetic analysis, were found to represent previously undescribed taxa [1,14]. Another example of the relatively easy accessibility of new strains was shown when only slight modifications in the composition of media employed resulted in the isolation of novel types of planctomycetes [12]. It can be assumed that the vast majority of isolates are discarded during broad isolation studies due to the failure to quickly assess their genomic novelty.

Furthermore, non-public culture collections of the pharmaceutical industry most likely contain a high portion of novel species, the analysis of which would substantially increase the extent of prokaryotic novelty. No estimates of global prokaryotic species numbers can be given at present but it should be assumed that a major fraction of them is associated with eukaryotic hosts. Whether or not some of the novel lineages emerging from molecular environmental studies originate from host-associated systems must remain unanswered.

Sequence diversity may be due to *rrn* operon microheterogeneity

Microheterogeneity at the level of *rrn* operons has been reported for a number of described species, including *Clostridium paradoxum* [9], *Paenibacillus polymyxa* [5], as well as many *Desulfotomaculum* species (unpublished). The degree of diversity among the *rrn* operons is in the range of a few percent only, and, as shown for *C. paradoxum*, heterogeneity does not extend to the level of rRNA. If, however, the number of cells of such organisms is high in the environment investigated or sequences of the respective cells are artificially increased through biasing mechanisms described above, analysis of 16S rDNA clone libraries may be highly misleading.

Probing of 700 strains, isolated from mesocosms containing eutrophied Mediterranean seawater, indicated the presence of 30% gamma Proteobacteria. Sequence analysis of about 20 strains affiliated these organisms to different lineages, three of which grouped with

Alteromonas macleodii. Clone libraries were generated from different time points of one of these mesocosms and more than 100 clones were partially sequenced. Surprisingly, the vast majority of sequences (>70%) were closely related to that of *A. macleodii* [8]. Fig. 1 shows a selection of lineages. In order to determine whether the sequences originated from cultured organisms, probes were generated on the basis of sequences of certain phylogenetic clusters. While some probes hybridized selectively with amplified 16S rDNA targets of pure cultures, other probes showed cross-hybridization with the same isolate. Sequence analysis of amplified 16S rDNA of such organism revealed the presence of *rrn* operon microheterogeneity which represents yet another source of bias in the analysis of 16S rDNA clone libraries.

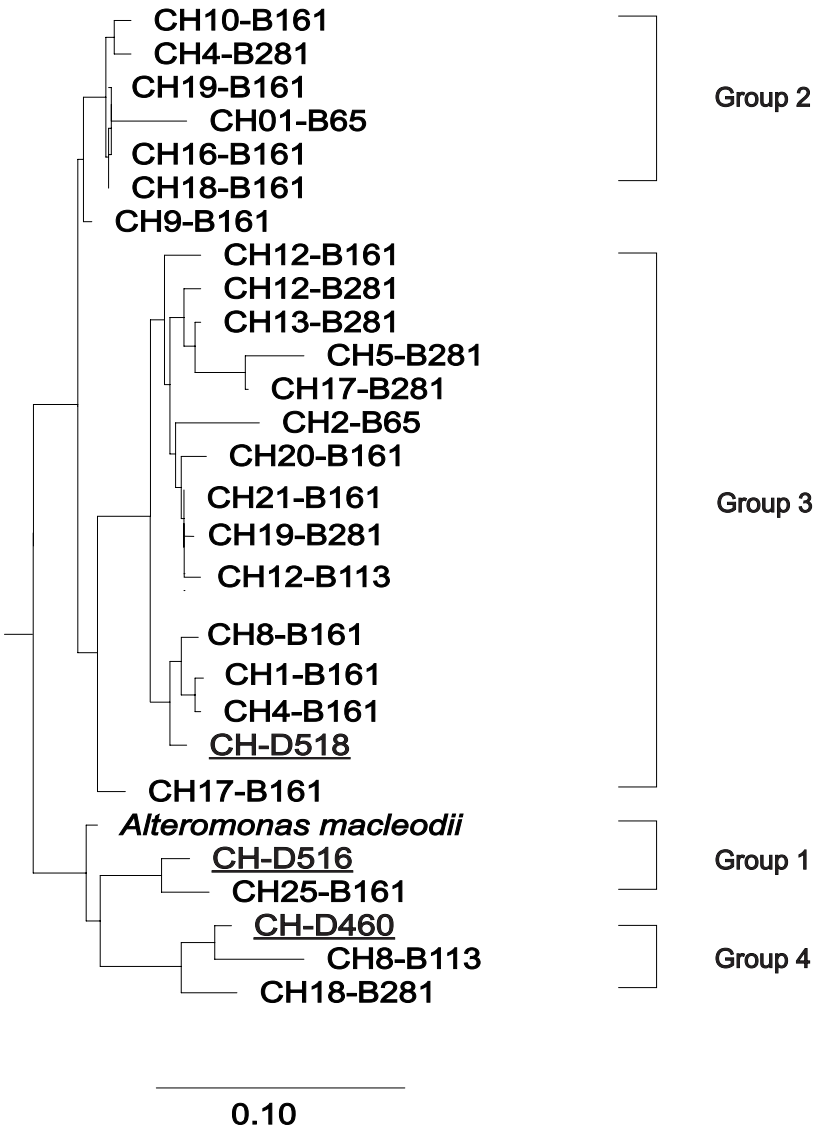

Fig. 1. Dendrogram of 16S rDNA relatedness showing the phylogenetic position of some partial sequences (300 nucleotides) of different clone libraries generated from DNA of a single mesocosm experiment to the nearest cultured organism, *Alteromonas macleodii*. Sequences of groups 2 and 3 were most likely obtained from a single taxon, represented by isolate CH-D518, due to *rrn* operon microheterogeneity [8]. The bar represents 10 nucleotide substitutions per 100 nucleotides.

Semi-quantitative assessment of taxon diversity

The combination of biases influencing the composition of 16S rDNA clone libraries leads to the question of the overall importance of these libraries. The majority of 16S rDNA sequences of organisms originally isolated from a given type of environment is not represented in clone libraries when this environment is subjected to molecular analysis. Assessment of prokaryotic diversity from the viewpoint of sequence composition of clone libraries would have resulted in an almost complete lack of knowledge of many of the cultured organisms, especially the Gram-positive bacteria. Nevertheless, analysis of clone libraries points towards the presence of organisms to be cultured in the future. Many of the novel sequences may indicate the presence of strains with new metabolic and biotechnological potential.

Despite the problems mentioned above similar novel clone sequences are represented in libraries generated from DNA of different environmental locations. This finding may be due to bias but may also indicate a semi-quantitative reflection of the naturally occurring diversity. An example is given by the marine Archaea which fall into one of two novel lineages, no matter which ocean and coastal environments were investigated [7]. Other examples are the worldwide occurrence of members of the *Acidobacterium*/*Holophaga* group and of the several groups branching deeply within the class *Actinobacteria* [11]. The fraction of these groups ranges between 1 and 23%, irrespective of differences in DNA isolation procedures, cloning systems, and DNA polymerases used [10]. A member of one

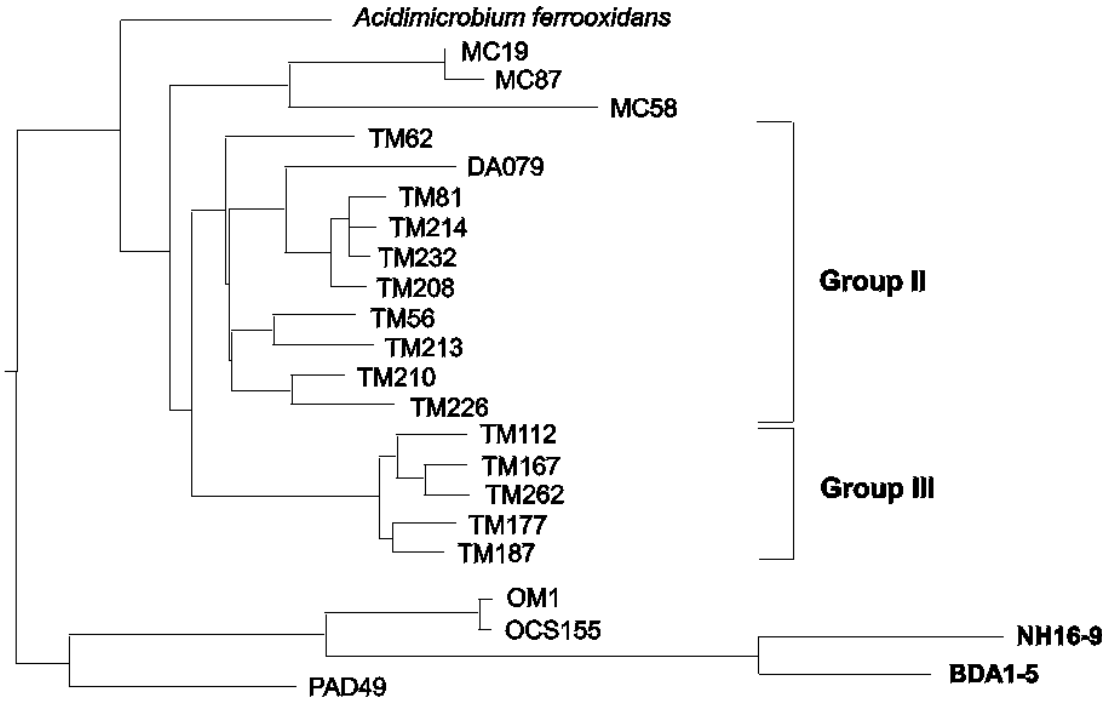

Fig. 2. Composite phylogenetic dendrogram indicating the relationship of actinobacterial clone sequences obtained from different environments and their most closely related cultivated organism. Clones are designated TM (for clones obtained from peat bog), DA (grassland soil), MC (forest soil), PAD (paddy field), OM, OCS, NH, BDA (marine samples). On the right hand side, group definitions are indicated as described in Rheims et al. [11].

of the actinobacterial clone groups (sequence DA079, group II, Fig. 2) was shown to represent 5.5% of the total ribosomal RNA isolated from that soil. This finding gives strong evidence for the importance these organisms have both in quantity and function.

Conclusion

16S rDNA clone libraries are embedded in a wide spectrum of strategies which include both analysis of cultured organisms and attempts to better understand the role of the as yet uncultured taxa. Emphasis is presently shifted towards the elucidation of dynamic changes and towards analysis of ribosomal RNA and messenger RNA, hence directed at the evaluation of function. The initial euphoric acceptance of clone libraries has been replaced by a more objective evaluation of their power to assess taxon diversity. 16S rDNA clone libraries, like libraries of genes coding for proteins are pieces of a jigsaw puzzle in which all pieces must fit to obtain a clear picture of ecological processes. However, as ecologists still do not know the complete picture, hardly anything is known about which pieces are missing. Progress made in any of the fields will be the stimulus for novel strategies and developments of even more sophisticated methods needed to understand the forces of evolution and survival.

References

1. Bowman, JP, McCammon, SA, Brown, MV, Nichols, DS, McMeekin, TA (1997) Diversity and association of psychrophilic bacteria in Antarctic sea ice. Appl Environ Microbiol 63:3068-3078
2. DeLong, EF, Wickham, GS, Pace, NR (1989) Phylogenetic stains: ribosomal RNA-based probes for the identification of single microbial cells. Science 243:1360-1363
3. Goebel, BM, Stackebrandt, E (1996) Molecular analysis of the microbial diversity in a natural acidic environment. In: Jerez, CA, Vargas, T, Toledo, H and Wiertz, JV, (Eds.) Biohydrometallurgical Processing, University Press, University of Chile, Santiago, pp. 43-52
4. Muyzer, G, De Waal, EC, Uitterlinden, AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695-700
5. Nübel, U, Engelen, B, Felske, A, Snaidr, J, Wieshuber, A, Amann, R, Ludwig, W, Backhaus, H (1996) Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. J Bacteriol 178:5636-5643
6. Pace, NR, Stahl, DA, Lane, DJ, Olsen, GJ (1986) The analysis of natural microbial populations by ribosomal RNA sequences. Adv Microb Ecol 9:1-55
7. Pukall, R, Stackebrandt, E (1998) Exploring the hidden microbial diversity of the marine environment. Recent research developments in microbiology, Research Signpost, India 2:535-551
8. Pukall, R, Päuker, O, Buntefuß, D, Ulrichs, G, Lebaron, P, Bernard, L, Guindulain, T, Vives-Rego, J, Stackebrandt, E (1998) High sequence diversity of *Alteromonas macleodii*-related cloned and cellular 16S rDNAs from a Mediterranean seawater mesocosm experiment. FEMS Microbiol Ecol 28:335-344

9. Rainey, FA, Ward-Rainey, N-L, Janssen, PH, Hippe, H, Stackebrandt, E (1996) *Clostridium paradoxum* DSM 7308^T contains multiple 16S rRNA genes with heterogeneous intervening sequences. Microbiology 142:2087-2095
10. Rheims, H, Felske, A, Seufert, S, Stackebrandt, E (1998) Molecular monitoring of an uncultured group of the class *Actinobacteria* in two terrestrial environments. J Microbiol Meth 36:65-75
11. Rheims, H, Spröer, C, Rainey, FA, Stackebrandt, E (1996) Molecular biological evidence for the occurrence of uncultured members of the actinomycete line of descent in different environments and geographic locations. Microbiology 142:2863-2870
12. Schlesner, H (1994) The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp. and other *Planctomycetales* from various aquatic habitats using dilute media. Syst Appl Microbiol 17:135-145
13. Stackebrandt, E, Rainey, FA(1995) Partial and complete 16S rDNA sequences, their use in generation of 16S rDNA phylogenetic trees and their implications in molecular ecological studies. In*:* Akkermans A.D.L,. van Elsas J.D, de Bruijn, F.J. Kluwer (eds.) Molecular Microbial Ecology Manual, Academic Publishers, The Netherlands, 3.1.1., 1-17
14. Suzuki, MT, Rappé, MS, Haimberger, ZW, Winfield, H, Adair, N, Ströbel, J, Giovannoni, SJ (1997) Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl Environ Microbiol 63:983-989
15. Tanner, MA, Goebel, BM, Dojka, MA, Pace, NR (1998) Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. Appl Environ Microbiol 64:3110-3113
16. Von Wintzingerode, F, Göbel, UB, Stackebrandt, E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev 21:213-229