Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis

Friedrich v. Wintzingerode [a], Ulf B. Göbel [a,*], Erko Stackebrandt [b]

[a] *Institut für Mikrobiologie und Hygiene, Universitätsklinikum Charité, Dorotheenstr. 96, 10117 Berlin, Germany*
[b] *Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, 38124 Braunschweig, Germany*

Abstract

After nearly 10 years of PCR-based analysis of prokaryotic small-subunit ribosomal RNAs for ecological studies it seems necessary to summarize reported pitfalls of this approach which will most likely lead to an erroneous description on the microbial diversity of a given habitat. The following article will cover specific aspects of sample collection, cell lysis, nucleic acid extraction, PCR amplification, separation of amplified DNA, application of nucleic probes and data analysis.

Keywords: Microbial diversity; Nucleic acid; 16S rRNA gene; Polymerase chain reaction; Phylogenetic analysis

Contents

* Corresponding author. Tel.: +49 (30) 20934715; Fax: +49 (30) 2292741; E-mail: goebel@rz.charite.hu-berlin.de

1. Molecular rRNA-based strategies to determine microbial diversity

For many decades it has been obvious to microbiologists that the 4000 validly described prokaryotic species do not reflect the actual diversity of prokaryotic species in the environment. The prokaryotic species evolved more than 3.8 billion years ago and have continued to evolve to occupy almost any niche on the planet Earth. Any estimations of the number of prokaryotic species are mere guesses no matter whether the numbers range between a ten-fold or a million-fold increase. What we have learned from recent analyses of bacterial and archaeal symbionts of eukaryotic hosts and from environmental samples can be summarized as follows. (i) The cultured microorganisms represent only a small fraction of natural microbial communities and hence the microbial diversity in terms of species richness and species abundance is grossly underestimated. (ii) Our understanding of microbial diversity is not represented by the cultured fraction of the diversity. This recent increase in awareness of our inability to cope with microbial diversity is due to a quantum leap in methodologies (e.g. molecular cloning, polymerase chain reaction (PCR), DNA probing etc.) and in the development of concepts that allowed biologists to come to a unified view of the genealogy of all living material, i.e. the use of semantic molecules for phylogenetic studies. This development can best be demonstrated by the revolutionary progress made in the use of ribosomal RNAs and their genes as reliable phylogenetic markers for the assessment of the natural relatedness between isolated and uncultured prokaryotes combined with improved PCR and sequencing technology. The new discipline of molecular ecology offers the potential of determining the whole range of prokaryotic taxa without running into the problems of selective laboratory enrichment and growth media. Today, molecular genetic analysis of rDNA obtained from DNA extracted from natural habitats is routinely used in many laboratories worldwide and a broad range of otherwise similar strategies are applied without prior cultivation and isolation of the organisms.

Originally, emphasis was placed on targeting ribosomal RNAs from environmental specimens that were analyzed by filter hybridization of extracted rRNA [1] or by direct rRNA sequencing [2,3]. Alternatively, total DNA was shotgun cloned and the recombinant *Escherichia coli* clone library searched for ribosomal DNA (rDNA) inserts [4]. Later, rRNA was transcribed by reverse transcriptase and the rRNA genes were amplified in vitro. Amplified nucleic acids were separated either by cloning in *E. coli* [5,6] or by temperature or denaturing gradient gel electrophoresis (TGGE or DGGE) [7,8]. Several methods were developed for the assessment of genetic diversity which include sequence analysis of randomly picked clone inserts, hybridization with taxon-specific probes [9], restriction fragment length polymorphism (RFLP), and amplified ribosomal DNA restriction analysis (ARDRA) of clones [10,11] or separation of amplified rDNA by gel electrophoresis (DGGE or TGGE).

Microbial communities differ in both qualitative and quantitative composition. The relative proportion of their community members is subject to physico-chemical changes of the environment as well as changes caused by the physiological and metabolic changes caused by the organisms. Organisms that are abundant and culturable under certain conditions may develop into dormant and possibly uncultured forms. Due to the power of the PCR to amplify small amounts of DNA, organisms occurring in small numbers in an environment are now detectable. Also, the sample volume required for analysis is significantly reduced and micro-habitats are now open for investigation, e.g. termite guts [12], dental plaque [13], nitrogen-fixing root nodules [14], or lesions of cattle with dermatitis digitalis [15].

However, each physical, chemical and biological step involved in the molecular analysis of an environment is a source of bias which will lead to a distorted view of the 'real world'. After 10 years of molecular ecological studies it seems necessary to summarize reported pitfalls of the molecular ecological approach which will most likely lead to an erroneous description of the diversity of a given ecological niche.

2. Initial considerations

In contrast to eukaryotic life-forms prokaryotic organisms occur ubiquitously in the environment. Prokaryotic life has been observed under extreme conditions such as in permafrost soil, the antarctic region, deserts, alkaline niches and at temperatures up to 120°C in deep sea hydrothermal vents. Prokaryotic life does not require organic substrates. The complexity of microbial communities may differ significantly, ranging from a single species, e.g. the sulfur-oxidizing bacterial ectosymbiont associated with the marine nematode *Laxus* sp. [16], few species communities, e.g., iron-leaching species of the genus *Thiobacillus* [17], to highly diverse communities like those of agricultural soils [18], municipal waste-treatment plants [19] or peat [20].

The analytical strategy, e.g. choice of PCR primers and separation method for amplified DNA, required to obtain a representative description of the ecosystem under study largely depends on the expected number of its population members. The range of microbial diversity can be initially estimated from: (i) the number of different morphotypes determined by fluorescent microscopy following staining with specific dyes that bind DNA of living cells only, (ii) physical and chemical parameters of the sampling site like pH, temperature or supplied substrates, and (iii) prescreening of diversity by denaturing gel electrophoresis of amplified 16S rDNA.

3. Sample collection

This step, crucial for all subsequent analyses, is often ignored as a source of problems and pitfalls. Sampling may be less difficult for terrestrial ecosystems such as soil or habitats like waste water-treatment plants, attached-living microorganisms such as biofilms and microbial mats where sample volumes can be kept small and material can be stored on ice, frozen or processed immediately [9,19,21–24]. Extreme habitats, such as hot vents and deep marine sediments, as well as sampling sites which require extensive sampling effort, e.g. the marine environment, may require additional effort to collect, store or process samples at the sampling site. Special care must be taken during their transport to the laboratory to avoid loss of nucleic acids due to lysis of specimens.

Comparing different sample handling procedures, Rochelle and co-workers [25] reported significant variations in 16S rRNA gene types and diversity from anaerobic deep marine sediments, up to 503 m below the sea floor. Samples stored aerobically for up to 24 h before freezing contained mainly sequences representing the beta and gamma subgroups of Proteobacteria. Samples taken in parallel but stored anaerobically at 16°C contained sequences mainly representing alpha *Proteobacteria*. Sediment samples taken anaerobically and frozen within a period of 2 h contained the widest spectrum of sequence diversity. As to the cause of this variation, the authors suggested that enrichment of specific bacterial groups had occurred during storage before freezing. Ward et al. [5] and Gordon and Giovannoni [26] investigated bacterioplankton of the Sargasso Sea. Due to the low bacterial count in sea-water sample volumes of several hundreds of liters were concentrated by filtration on 0.1 μm or 0.2 μm membranes. The extended time required for filtration may have influenced the composition of the microflora. One way of circumventing the problem may be the release and stabilization of nucleic acids immediately after sample collection. Muralidharan and Wemmer [27] used lysis buffer containing Nonidet P-40, Tween 20 and sodium azide to store field-collected blood and tissue samples over several weeks at room temperature without significant loss in quality and quantity of released DNA. This method should be adopted for microbial cells as many organisms, at least the majority of Gram-negative bacteria, are sensitive to the detergents used.

4. Cell lysis and extraction of DNA

Lysis of microbial cells from environmental habitats marks a critical step in a PCR-mediated approach. Insufficient or preferential disruption of cells will most likely bias the view of the composition of microbial diversity as DNA or RNA, which is not released from the cells, will not contribute to the final analysis of diversity. On the other hand rigorous conditions required for cell lysis of Gram-positive bacteria should be avoided as this treatment may

lead to highly fragmented nucleic acids from Gram-negative cells. Fragmented nucleic acids are sources of artefacts in reverse transcription or PCR amplification experiments (see Section 6) and may contribute to the formation of chimeric PCR products [28]. In addition, various biotic and abiotic components of environmental ecosystems, such as inorganic particles or organic matter, affect lysis efficiency and may interfere with subsequent DNA purification and enzymatic steps.

Leff et al. [29] compared three different, widely used DNA extraction techniques for soil specimens. In the method of Ogram et al. [30] a bead beater was used to disrupt cells following incubation in sodium dodecyl sulfate (SDS) at 70°C. Purification of DNA was performed by phenol-chloroform extraction and CsCl-ethidium bromide density gradient ultracentrifugation. In the method of Tsai and Olsen [31] sediments were treated with lysozyme, and cells were lysed by rapid freezing and thawing. Following phenol-chloroform extraction, DNA was precipitated with isopropanol, and impurities were removed by gel filtration with Sephadex G-100 (Pharmacia). The method of Jacobsen and Rasmussen [33] is an indirect lysis approach. Cells were removed from sediments by a cation exchange resin (Chelex 100, Bio-Rad) and lysed by lysozyme and pronase treatment.

According to Leff et al. [29] the Ogram method resulted in the release of a significant amount of DNA but the DNA was badly sheared because of the bead beating procedure and, as determined by hybridization with an eubacterial-specific 16S rRNA probe [34], contained a smaller proportion of eubacterial DNA. The Tsai method revealed DNA of lower purity but the highest fraction of eubacterial DNA. The authors recommended the Jacobsen method for a PCR-mediated approach because of the low concentration of contaminants. However, recovery of cells may be differential, depending on the strength and nature of the attachment to the sediment.

Because recovery rates remain difficult to judge many investigators concentrate on combinations of direct lysis methods with subsequent PCR amplification. Several modifications of extraction protocols were developed to obtain high lysis efficiency for different ecosystems, which include mortar-and-pestle grinding in liquid nitrogen [21,35] and high-salt extraction buffer (1.5 M NaCl), extended heating (2–3 h) in the presence of SDS, hexadecyltrimethylammonium bromide (CTAB) and proteinase K [21]. The method described by Picard et al. [36] used both sonication, microwave heating and thermal shock for direct lysis of microorganisms in soil. This procedure should be applied with care as DNA is sheared into fragments ranging in size from 100 to 500 bp. Such highly fragmented DNA increases the formation of chimeric molecules during PCR as discussed below. Other authors stressed that certain components which are co-extracted from soil, mainly humic acids and other humic substances, strongly inhibit *Taq* polymerases [20,37–39] and other DNA modifying enzymes such as restriction endonucleases and DNase I [40]. In addition, these substances may interfere with DNA hybridization specificity [41]. Consequently, the contaminants must be removed from nucleic acid preparations, e.g. by agarose gel electrophoresis of nucleic acid extracts and subsequent excision of high molecular mass DNA [9,42] and ion-exchange chromatography with Qiagen-Tip 500 columns [39]. Alternatively, skimmed milk [35], CTAB [43] or polyvinylpolypyrrolidone (PVPP) [44,45] have been added to the soil prior to extraction.

In contrast to terrestrial habitats aquatic environments contain significantly lower levels of inorganic or organic particles and lysis protocols developed for pure cultures have been successfully applied [26,46].

5. Cell lysis and extraction of RNA

Assessment of the metabolically active fraction of the community should be done by analysis of RNA rather than DNA. However, extraction of RNA from environmental ecosystems requires special attention as RNAs are highly susceptible to degradation by RNases during the extraction procedures.

Moran and co-workers [47] developed a method for direct recovery of microgram quantities of rRNA from sediment, soil and water. Cells were initially lysed with lysozyme and the released RNA was selectively extracted with low pH buffered hot phenol, which also inhibits degradation by RNases. Humic substances were removed by subsequent gel fil-

tration of extracts with Sephadex G-75 spin columns. Another method for the recovery of RNA from low-biomass sediments is based on direct lysis by two cycles of boiling and one freeze-thaw cycle in alkaline phosphate buffer followed by extraction with guanidinium isothiocyanate-phenol-sarcosyl solution to inactivate nucleases [48]. Nucleic acids were ethanol-precipitated and DNA DNase I digested. Felske et al. [49] described a unique approach for extraction of rRNA from soils which is based on mechanical disruption of cells in the presence of PVPP, bovine serum albumin (BSA) and magnesium chloride and subsequent isolation of intact ribosomes by centrifugation. This method yields high quality rRNA.

6. PCR amplification

PCR amplification has become the method of choice for obtaining rRNA sequence data from microbial communities or pure cultures [50,51]. Full length 16S rDNA can be amplified either directly or after reverse transcription of rRNA with a set of primers binding to conserved regions of the 16S rRNA/rDNA [52,53]. Although it is a routine method for pure cultures, several problems arise when the methods are applied to environmental communities: (i) inhibition of PCR amplification by co-extracted contaminants, (ii) differential amplification, (iii) formation of artefactual PCR products. One has also to consider that (iv) contaminating DNA, and (v) 16S rRNA sequence variations due to *rrn* operon heterogeneity would unavoidably lead to a biased reflection of the microbial diversity.

6.1. Inhibition of PCR amplification

Humic acids or humic substances co-extracted with nucleic acids strongly inhibit DNA modifying enzymes. Using a commercial preparation of humic acids, Tebbe and Vahjen [39] found minimum inhibitory concentrations of 0.64, 0.16 and 0.08 $\mu g \ ml^{-1}$ for three *Taq* DNA polymerases from different suppliers (Boehringer Mannheim, Promega, and Perkin Elmer, respectively). These results were confirmed for Perkin Elmer Ampli*Taq* DNA polymerase [54] whereas ICI's Thermalase (KEBO, Albertslund, Denmark) was active in the presence of humic acid

concentrations of 8.33 $\mu g \ ml^{-1}$. Differences in the activity of DNA polymerases should be considered when samples with a expected high content of humic substances have to be analyzed.

Loss of nucleic acids during purification becomes a problem when only small sample volumes are available for processing. To circumvent extensive purification procedures, additives such as BSA and T4 gene 32 protein (gp32) were used, which reduce the inhibition effects of contaminants. BSA and gp32 were added to PCR reaction mixtures containing both known and poorly defined inhibitors which have been reported to contaminate environmental DNA preparations [55]. The presence of either 400 ng μl^{-1} of BSA or 150 ng μl^{-1} gp32 in the PCR assay led to a 10–1000-fold higher tolerance towards $FeCl_3$, fulvic acids, tannic acids, or extracts from feces, fresh water, or marine water. Use of BSA and gp32 together offered no more relief of inhibition than either alone at its optimal level. The author strongly recommended the use of non-acetylated BSA as the acetylated, nuclease-free preparation inhibits PCR itself [56].

Few data are available on the specific inhibition of reverse transcriptases or enzymes with both reverse transcriptase and DNA polymerase activity, such as r*Tth* DNA polymerase (Perkin Elmer). Felske et al. [49] failed to apply RT-PCR with r*Tth* DNA polymerase on RNA directly extracted from soil. Stinear et al. [57] reported lower sensitivity in RT-PCR amplification of *Cryptosporidium parvum* heat shock protein 70 mRNA in water samples with higher total organic carbon.

Dilution of the DNA template to minimize inhibitors in the amplification reaction is not recommended as very low DNA concentrations may influence the PCR efficiency (see Section 6.2).

6.2. Differential PCR amplification

In PCR amplification of 16S rDNA/rcDNA from complex microbiota a mixture of homologous molecules serve as template. Amplified DNA can only reflect quantitative abundance of species if the amplification efficiencies are the same for all molecules. This requires several assumptions [58]: (i) all molecules are equally accessible to primer hybridization, (ii) primer-template hybrids form with equal efficien-

cies, (iii) extension efficiency of DNA polymerase is the same for all templates, and (iv) limitations by substrate exhaustion equivalently affect the extensions of all templates. These assumptions seem difficult to hold as universal primers employed for the amplification of rDNA/rcDNA often contain degeneracies which may influence the formation of primer-template hybrids.

An effect of genome size and *rrn* gene copy number on PCR was found by Farrelly et al. [59], when 16S rDNA of mixtures of four different bacterial species with known genome sizes and *rrn* operon numbers (*Pseudomonas aeruginosa*, *E. coli*, *Bacillus subtilis* and *Thermus thermophilus*) were amplified with a primer pair which resulted in the formation of a DNA fragment of 500 bp. In contrast to mixtures of *P. aeruginosa* and *E. coli* with *T. thermophilus*, the ratio of amplified products for mixtures of *B. subtilis* and *T. thermophilus* showed greater deviations from the predicted ratio calculated from the number of *rrn* genes per equimolar amounts of DNA. Since information on genome size and *rrn* gene copy number is completely lacking for all of the uncultured microbial diversity the authors conclude that with the methods currently available, quantification of microbial communities from analysis of 16S rDNA clone libraries is not possible. In a similar approach Suzuki and Giovannoni [58] observed a bias in amplification of a mixture of environmental 16S rDNA clones which was strongly dependent on the choice of primers and number of cycles of replication. While mixtures of two templates amplified with the 519F-1406R (900 bp amplicon) primer pair yielded amplicons of the same ratio as the templates, the second primer pair 27F-338R (300 bp amplicon) resulted in a strong bias towards 1:1 mixtures of gene products, regardless of the initial proportions of templates. This bias was reduced by decreasing the number of replication cycles from 35 to 10. A possible explanation for the discrepancy in the ratio of end-products is that reannealing of gene products progressively inhibits the formation of template-primer hybrids when primers with a high amplification efficiency are used. The authors suspect this PCR-produced bias to be small if the environmental DNA contains highly diverse templates. In these cases it would be unlikely that the amplification of any gene will produce amplicons in an inhibiting concentration. However, the template diversity would be significantly reduced if non-universal primer pairs were applied, e.g. primers specific for certain bacterial genera.

Hybridization efficiency and specificity of primers influence the PCR amplification of mixed 16S rDNA templates. Suboptimal binding of the primer will result in less efficient amplification of the respective DNA. Especially domain-specific or universal primers must have uniform hybridization efficiency to guarantee the amplification of all target 16S rDNAs. Brunk et al. [60] examined the small subunit (SSU) rRNA sequences in the Ribosomal Database Project (RDP) [61] using a computer algorithm which simulates hybridization between DNA sequences to evaluate the efficiency and specificity of a number of published and widely used domain-specific or universal probes and to select alternative oligonucleotides. As derived from calculated hybridization potentials three published universal probes (primer) hybridized well with all sets of SSU rRNA sequences except with DNA of the Crenarchaeota, two of three eukaryotic probes also showed high specificity whereas only three of seven bacterial probes had exclusively high affinities to bacterial sequences. All of the investigated published probes for *Archaea*, *Euryarchaeota* and *Crenarchaeota* had low specificities. Similar observations were reported by Zheng et al. [62] when six universal 16S rRNA probes were evaluated for stability of probe-target duplexes with representatives of all three domains. All probes showed domain-specific variations in dissociation temperatures. In accordance with this, Paster et al. [63] reported a 4-base mismatch in the binding site of the eubacterial probe Eub338 [34] for a member of the genus *Cristispira*.

By using dot-blot hybridization with taxon-specific probes Rainey et al. [64] found that composition of PCR-amplified environmental 16S rDNA clone libraries significantly changed when the same batch of isolated DNA and the same cloning vector, but two different pairs of primers were used.

In conclusion, the choice of primer for universal or taxon-specific PCR amplification of 16S rDNA from complex microbiota may influence the recovery of target sequences. In order to reduce problems arising from primer efficiency and specificity we strongly recommend the use of computer algorithms such as

CHECK_PROBE, which is implemented in the RDP database [65], PROBE_MATCH, implemented in the ARB software package [66], or that described by Brunk et al. [60]. These software programs are designed for the analysis of 16S rRNA targeted probes and are based on almost all available 16S rRNA sequences. These tests enable investigators to check published primers with the currently available 16S rRNA sequence dataset and allow the design of new probes/primers. Additionally, specific software for the analysis of probes, supplied by the commercial EMBL® and GenBank® databases, can be used.

Beside the described effects of genome size, *rrn* gene copy numbers and choice of primers for the PCR amplification of mixed 16S rDNA templates, the varying mol percent G+C composition of 16S rRNA genes is suspected to cause differential amplification. Genes with a higher G+C content dissociate with a lower efficiency leading to preferential strand separation of genes with a lower G+C content during the denaturation step and may therefore result in a preferential amplification of templates with a lower G+C content. Reysenbach et al. [67] found preferential amplification of yeast rRNA genes when mixed with DNA from two hyperthermophilic archaea strains. This selectivity could be reduced by adding 5% acetamide as a denaturant to the reactions, which also minimized non-specific primer annealing. Baskaran et al. [68] developed a protocol for the uniform PCR amplification of a mixture of DNA with varying G+C content, ranging from 44% to 80%, which includes the addition of dimethylsulfoxide (DMSO) in combination with betaine and the use of a mixture of Klentaq1 and *Pfu* DNA polymerase. Application of this procedure on PCR amplification of 16S rDNA from complex microbial communities appears to be promising since the G+C contents of rDNAs of known prokaryotes vary between 50 and 66%.

An effect of DNA concentrations on the amplification efficiency of 16S rDNAs from environmental DNA was recently reported. Chandler and co-workers [69] observed significant changes in the composition of 16S rDNA clone libraries when diluted or undiluted environmental DNA was used as PCR template. They propose that very low DNA concentrations in the range of a few to tens of picograms generate random fluctuations in PCR effi-

ciency, which led to the observed contrast in clone libraries.

DNA-associated molecules which resist standard deproteinizing procedures during DNA purification could be a source of diminished amplification efficiency of 16S rDNA of Gram-positive bacteria as these molecules could cause loops in template strands which inhibit elongation by DNA polymerase during PCR. Interference of those molecules with DNA modifying enzymes has been reported by Waterhouse and Glover [70] who observed different hybridization patterns when DNA of the Gram-positive bacterium *B. subtilis* was prepared with three different isolation procedures and hybridized with a rDNA probe. They suggested that a non-protein molecule, presumed to be peptidoglycan, remained bound to DNA after either phenol-chloroform or potassium acetate was used as deproteinizing agent protecting certain restriction enzyme sites from digestion.

6.3. Formation of PCR artefacts

The appearance of PCR artefacts is a potential risk in the PCR-mediated analysis of complex microbiota as it suggests the existence of organisms that do not actually exist in the sample investigated. Several types of PCR artefacts have been reported: (i) chimeras between two different homologous molecules, (ii) deletion mutants due to stable secondary structures, and (iii) point mutants due to misincorporation by DNA polymerases.

6.3.1. Formation of chimeric molecules

In vitro recombination of homologous DNA leading to chimeric molecules composed of parts of two different sequences has been widely observed and is not restricted to 16S rDNA amplification from complex microbiota. Chimeras between two different DNA molecules with high sequence similarity (i.e. homologous genes) can be generated during the PCR process as DNA strands compete with specific primers during the annealing step. Shuldiner et al. [71] discovered formation of chimeric DNA during PCR analysis of two different non-allelic preproinsulin genes of *Xenopus laevis* which are very similar to each other (i.e. 94% in the coding region). The authors suspected partial extension by *Taq* DNA

polymerase due to regions of stable secondary structures, subsequent annealing of two DNA fragments and complete extension to cause chimeric molecules that consist of preproinsulin I at its 5′-end, and preproinsulin II at its 3′-end. Choi et al. [13] found seven chimeric molecules out of 81 analyzed 16S rDNA sequences (8.6%), when the microbial community of a subgingival plaque sample was investigated by amplification and cloning of partial 16S rRNA genes. Meyerhans et al. [72] reported that 5.4% of all amplified molecules would be chimeric if PCR was used to co-amplify two distinct HIV-1 *tat* genes. The frequency of such recombinants could be decreased 2.7-fold by a 6-fold increase in *Taq* DNA polymerase elongation time. These findings were supported by Wang and Wang [73] who observed a decrease in chimera formation with increasing elongation time (2–5 min), when mixtures of two different 16S rRNA genes were amplified. The authors also found a positive correlation between frequency of chimeras and both number of PCR cycles and sequence similarity between mixed templates. PCR of mixtures of templates with 99.3% sequence similarity resulted in 30% chimeras after 30 cycles, 20.9% after 20 cycles, and 4.8% after 10 cycles. The frequency of chimera formation was reduced by 50% after 30 cycles if templates with 82% sequence similarity were used. Similar frequencies were reported by Ford et al. [74] who revealed that 30% of the amplicons were chimeric molecules when PCR co-amplifications with 35 cycles were used for analysis of genes encoding murine immunoglobulin (Ig) λ light chain variable (V) regions.

In addition to incomplete strand synthesis during the PCR process DNA damage has been suggested to promote the formation of chimeric molecules in PCR co-amplification of templates with high sequence similarities. Pääbo et al. [75] investigated the influence of template breaks caused by restriction enzyme digestion, UV irradiation, sonication, and depurination, on PCR co-amplification of cow lysozyme type 2b and 3 genes. All types of DNA damage were shown to support production of recombinant PCR products. Since rigorous cell lysis conditions for DNA preparation from environmental samples are likely to cause damages similar to those described by Pääbo et al. [75] these findings could have significant impact on 16S rDNA amplification from com-

plex microbiota. This is in accordance with the finding of Liesack et al. [28] who reported the occurrence of chimeric 16S rDNA in PCR when low molecular mass DNA (4–6 kb) was used as template in the analysis of a mixed culture of two strict barophilic bacteria. Brakenhoff et al. [76] detected reverse transcription-dependent chimeric cDNA clones in RT-PCR analysis of the human γ-crystallin gene family. It was suspected that prematurely terminated cDNA generated during reverse transcription of mRNA hybridized to intact mRNA and served as primer for further reverse transcription by *Taq* DNA polymerase during PCR amplification. This PCR artefact caused by the reverse transcriptase activity of *Taq* DNA polymerase is also a potential risk for RT-PCR-based 16S rRNA analysis of environmental microbial ecosystems. Incomplete in vitro rcDNA synthesis has been detected in members of the genus *Streptomyces* and several thermophilic organisms which exhibit specific post-transcriptional modifications of the 16S rRNA that cause reverse transcriptase to terminate rcDNA synthesis.

6.3.2. Formation of deletion mutants

It is well known that PCR templates containing stable secondary structures often yield very low amplification efficiency or deletion mutagenesis in PCR products [77–79]. As ribosomal RNAs usually exhibit intensive secondary structures [80], RT-PCR could lead to deletion mutants, which would be excluded from subsequent analysis as amplified 16S rRNA genes are often size selected to avoid cloning or gel electrophoresis of non-specific amplicons.

To circumvent problems arising from template secondary structures during *Taq* DNA polymerase PCR Chou [81] recommended the use of *E. coli* single-strand DNA binding protein in PCR reactions or the application of DNA polymerases, which have been described to have a higher processivity than *Taq* DNA polymerase.

6.3.3. Formation of point mutants

Since its first application in PCR reactions *Taq* DNA polymerase is known to have an intrinsic misincorporation rate during strand synthesis, which can lead to base substitutions [82]. As reviewed by Eckert and Kunkel [83], the observed error frequencies for *Taq* DNA polymerase-based PCR can range

from approximately one error per 290 nucleotides (3×10^{-3}) to one error per 5411 nucleotides (2×10^{-4}), depending on the reaction conditions used. Ford et al. [74] reported an even lower error rate of about 2.6×10^{-5}/bp per cycle. Similar values were measured for AMV reverse transcriptase [84]. Stewart et al. [85] found a misincorporation rate of one nucleotide per 700 bases for r*Tth* DNA polymerase (Perkin Elmer), in PCR amplification of the 8-kb DNA genome of the human papilloma virus HPV16. Several thermostable DNA polymerases contain a $3' \rightarrow 5'$ exonuclease (proofreading) activity, which results in a significant lower misincorporation rate during strand synthesis. Compared to *Taq* DNA polymerase, which lacks the proofreading activity, PCR amplification with the proofreading DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus* (Pfu) leads to a 10-fold improvement in the misincorporation rate [86].

Such a minute error rate seems to have little impact on the phylogenetic evaluation of PCR-amplified 16S rRNA genes as the maximum misincorporation rate would lead to five wrong nucleotides for the entirely gene (about 1500 bp), corresponding to 0.3% sequence divergence. However, if RT-PCR and/or several PCR cycles are performed, i.e. nested PCR with group-specific primers after domain-specific amplification, or recombinant clones were analyzed by PCR amplification of the insert and subsequent cycle sequencing with *Taq* DNA polymerase, misincorporation can accumulate leading to a higher error rate. The presence of misincorporated (or misinterpreted) nucleotides is highly problematic when they are located at sites which have been selected as a probe target or when small differences in sequence are used in strain discrimination. Evaluation of the stability of secondary structure may help to identify such erroneous nucleotides.

6.4. Contaminating DNA

Contaminating DNA, containing the specific target sequence of the PCR reaction involved, can lead to both amplification in negative controls without external DNA being added and co-amplification in experimental reactions. Direct analysis of those artificially mixed amplicons by sequencing or hybridization would lead to ambiguous results, whereas clon-

ing or gel electrophoresis and subsequent analysis would simulate sequence diversity which actually does not exist.

Non-specific DNA can be introduced in PCR reactions as tube-to-tube contaminants, i.e. amplification products of previous reactions are unintentionally transferred to fresh reactions, or by contaminated reagents. Several reports described the latter contaminations as bacterial DNA [87–89]. Amplification of ribosomal RNA genes appears to be extremely sensitive to contaminating bacterial DNA as universally conserved regions of bacterial genes serve as target sequences. Maiwald et al. [90] characterized DNA contaminating *Taq* DNA polymerase which was amplified during PCR with a primer set for the *Legionella* 5S rRNA gene. Their results indicate that not the bacterium used for production of the recombinant enzyme but contaminating soil bacteria were the origin of the foreign DNA.

Several strategies have been developed to avoid or eliminate DNA contaminations, which include both laboratory organization and decontamination systems. Niederhauser et al. [91] tested the reliability of several decontamination procedures and found UV treatment and pre-PCR uracil DNA glycosylase digestion the most effective.

6.5. 16S rRNA sequence variations due to rrn operon heterogeneity

The number of rRNA gene regions (*rrn* operons) located on prokaryotic chromosomes differs widely. Although there is no correlation between *rrn* operon copy number and genome size, an observed trend is that slow-growing organisms have fewer copies than more rapidly growing bacteria [92].

Only one rRNA gene was reported for *Bradyrhizobium japonicum* strains [93], and three *rrn* loci were located within the genome of *Dichelobacter nodosus* [94], whereas Stewart et al. [95] found 10 copies in *Bacillus subtilis* and Johansen et al. [96] estimated 12 and nine copies in two strains of *B. cereus*. For the type strain of *Paenibacillus polymyxa* a number of at least 12 copies was postulated by Nübel et al. [8]. The authors also detected an extensive sequence heterogeneity of 10 variant nucleotide positions in the 16S rRNA genes, when a 347-bp fragment of the 16S

rDNA, containing variable regions V6 and V8, was PCR-amplified and analyzed by TGGE. RT-PCR of rRNA, as well as whole-cell hybridizations, revealed a predominant representation of particular sequences in ribosomes of exponentially growing cultures. Rainey et al. [97] reported the presence of 15 different sequences in cloned 16S rRNA genes of *Clostridium paradoxum*, which is the highest known number of *rrn* operon copies in prokaryotes so far. The majority of the cloned genes contained intervening sequences (IVSs), which are located in the variable region I of the 16S rDNA and varied in length from 120 to 131 nucleotides. These IVSs were absent from mature 16S rRNA as shown by Northern hybridization and analysis of RT-PCR products. Similar significant heterogeneity was found in the two genes encoding 16S rRNA from the halophilic archaeon *Haloarcula marismortui*, which differ by 74 nucleotide substitutions, thus exhibiting 5% overall sequence divergence [98], and in the two 16S rRNA genes of *Thermobispora bispora*, which differ in 98 nucleotide positions (6.4% sequence divergence) together with six regions of deletion-insertions [99]. The previously discussed 16S rDNA sequence heterogeneities are in contrast to other findings, which indicate identical or nearly identical *rrn* operons for *B. subtilis* [100], *Rhodobacter sphaeroides* [101], and *Haemophilus influenzae* Rd [102]. However, a recent computer analysis of sequences deposited in GenBank® revealed a level of intraspecific and intrastrain sequence variations that cannot be explained by experimental errors indicating higher sequence divergences in *rrn* operons of one organism than previously expected [103].

In addition to 16S rDNA sequence heterogeneities within one organism several studies have demonstrated the presence of a single type of IVSs in 16S and 23S rRNA genes [104–107]. Ralph and McClelland [108] found evidence that insertions in the 23S rRNA genes of *Leptospira* species are mobile elements which can be horizontally transferred, when phylogenetic trees of the 16S rDNA were compared with those derived from the IVSs.

In conclusion, 16S rRNA genes of some Bacteria and Archaea reflect the occurrence of inter- and intraspecific *rrn* operon heterogeneities. These differences can interfere with the analysis of 16S rDNA clone libraries or gel electrophoresis patterns derived from environmental ecosystems as it is not clear whether one 16S rDNA sequence represents a distinct organism or is just one representative gene of the entire 16S rRNA operon of an organism. Because it is likely that IVSs are introduced into 16S rRNA genes by lateral transfer their inclusion in phylogenetic analyses can lead to erroneous results. Such sequence idiosyncrasies should therefore be excluded prior to phylogenetic analysis. Furthermore, the reported IVSs could lead to the exclusion of the respective 16S rDNAs as PCR is often followed by size selection of the amplicons to avoid subsequent analyses of non-specific products. The observed differences between 16S rDNA and mature 16S rRNA hinder the design of 16S rRNA targeted probes from 16S rDNA data as sequence variations present in the rDNA might be absent in the rRNA.

7. Separation of amplified 16S rRNA genes

For almost all analyses of microbial ecosystems amplified 16S rRNA genes have to be separated prior to subsequent sequencing and/or hybridization as they constitute a heterogeneous mixture of sequences. Exceptions are 'one-species' microbial communities like the highly specific symbiosis between a sulfur-oxidizing bacterium and a marine nematode [16] and co-cultures of two microorganisms of different kingdoms (*Bacteria* and *Archaea*), which allowed specific PCR amplification [109].

Cloning in *E. coli* is the most widely used method to separate PCR-amplified DNA identical in length but different in sequence. However, the influence of cloning systems on the composition of 16S rRNA gene libraries from environmental ecosystems is poorly investigated. Scharf et al. [110] were the first to describe a method for directly cloning PCR products, which requires amplification primers with restriction enzyme sites added to their 5′ ends, and subsequent cleavage of the amplicons to generate 'sticky ends' for ligation into the vector. Several investigators used this approach for the analysis of PCR-amplified 16S rRNA genes from complex microbial ecosystems. Liesack and Stackebrandt [9] and Choi et al. [13] used PCR primer with attached *Bam*HI and *Sal*I sites to clone amplified 16S rRNA genes of an Australian soil sample and a subgingival

plaque sample in the vector involved. Both restriction enzymes cut rarely only in 16S rRNA genes [53]. This is especially important as internal cleavage sites would lead to shorter amplicons, which are eliminated by size fractionation electrophoresis techniques. In addition, these short sequences may lack sites used as targets for subsequent identification by hybridization with specific oligonucleotides. Other cloning methods avoid restriction enzyme digestion such as blunt-end cloning, TA cloning, or ligation-independent cloning (LIC), based on specific T4 DNA polymerase digestion of 3′ ends [111].

Rainey et al. [64] reported a changing distribution of taxon-specific clones in 16S rDNA clone libraries which were derived from the same batch of DNA but generated with different cloning systems. As it is unlikely that complex mixtures of amplified 16S rRNA genes are cloned with uniform efficiency, it has to be assumed that cloning systems generally influence the abundance of single sequences in 16S rRNA gene libraries. Furthermore, the use of *E. coli* strains with an active DNA repair system for transformation of recombinant vector molecules can lead to the formation of artificial 16S rRNA genes. During the PCR denaturation-annealing step heteroduplexes between two complete strands of different 16S rRNA genes could be formed due to the high sequence homology. Cloning of heteroduplexes in bacteria capable of DNA mismatch repair would lead to independent repair to one strand or the other, at each point of mismatch resulting in mosaic sequences of two genes [112].

Beside cloning in *E. coli*, amplified 16S rRNA genes can be separated electrophoretically in polyacrylamide gels containing denaturing gradients. In those gels amplicons with identical length migrate in dependence on their respective primary sequences and base compositions. Muyzer et al. [7] used DGGE to separate amplified 16S rDNA from microbial mats and biofilms. Whereas in DGGE a denaturing gradient is formed by urea and formamide, TGGE is based on a temperature gradient. In contrast to the laborious cloning procedures these approaches facilitate a rapid estimation of microbial diversity as complexity of resulting separation patterns reflects the heterogeneity of amplified 16S rRNA genes, which makes them an ideal tool to monitor changes in community compositions [113].

During PCR a GC-rich sequence (30–40-mer) has to be attached to guarantee sequence-specific strand separation. This GC-rich sequence can cause incomplete strand synthesis during PCR leading to multiple bands for one template [8]. The occurrence of additional bands interferes with the analysis of highly complex microbiota as separation patterns will not appear as distinct bands. This requires additional cloning steps for subsequent sequence analysis.

A novel source of potential problems has recently been found by Ward-Rainey, Rainey and Stackebrandt (unpublished). Amplification of planctomycete 16S rDNA with conserved eubacterial PCR primers led to the co-amplification of additional genes of the same size as the 16S rDNA fragment. Following cloning these genes were identified as genes encoding proteins. Failure to purify the PCR fragments by cloning would result in ambiguous, often indeterminable sequencing patterns.

8. Analysis of 16S rRNA sequence data

The ultimate goal of a PCR-mediated analysis of 16S rRNA molecules from complex microbiota is the retrieval of sequence information, which allows determination of microbial diversity, i.e. cultured and uncultured microorganisms, by comparative 16S rRNA sequence analysis. Sequencing of amplified and separated 16S rRNA genes can be performed by radioactive or non-radioactive standard techniques with both universal sequencing primers and internal 16S rDNA primers [114,53].

The quality of results obtained by comparative 16S rRNA sequence analyses strongly depends on the available dataset. Although about 5000 full and partial 16S rRNA and 16S rDNA sequences of cultivated microorganisms and environmental clones have been released this number reflects only a minor part of the expected microbial diversity. 16S rRNA genes retrieved from environmental samples often exhibit a low sequence similarity to known sequences making their phylogenetic affiliation difficult. This leads to the question whether environmental sequences represent uncultured, novel microorganisms or whether they cannot be assigned to known taxa due to the fact that for many cultivated microorgan-

isms 16S rRNA sequences are not available or of low quality (i.e. partial sequences and/or many ambiguous bases in released sequences).

Liesack and Stackebrandt [9] recovered 16S rDNA clone sequences from an Australian terrestrial soil and assigned them to a new phylum within the domain Bacteria sharing common ancestry with members of the phyla *Chlamydia* and *Planctomycetaceae*. Later phylogenetic analysis of the full 16S rDNA sequence of *Verrucomicrobium spinosum* [115] revealed a specific affiliation of these clone sequences to this species. In another example, novel environmental clone sequences from the Atacama desert, Chile (Rainey, Friedman and Stackebrandt, unpublished), together with clone sequences from Australian soil [116] and peat [20], could be affiliated to the *Rubrobacter radiotolerans* lineage within the class *Actinobacteria*.

As discussed above, amplification of a mixture of 16S rRNA genes may lead to the formation of chimeric molecules. These chimeras have to be recognized and excluded from further phylogenetic analyses as they do not reflect true microbial diversity. Several methods have been developed for the recognition of chimeric 16S rRNA genes, which rely on checking the complementarity of helical regions or performing comparative sequence analyses of different sections of the 16S rDNA. Liesack et al. [28] were the first who used this approach to detect chimeric 16S rDNA assembled from the 16S rDNAs of two moderately related (90% sequence similarity) strictly barophilic bacterial strains A and B. The hybrid rDNA exhibited 99.5% similarity to strain A in the first 1095 nucleotides while 100% similarity with strain B was found in the last 505 bases, which resulted in two additional mismatches and one U-G base pair in the 16S rRNA helix 984–990/1215–1221 (*E. coli* nomenclature). However, Kopczynski et al. [117] detected chimeric 16S rDNA sequences in a clone library, derived from a cyanobacterial mat, which exhibited little or no secondary structural abnormality. These chimeras seemed to be generated from previously recovered environmental 16S rRNA genes of uncultivated microorganisms as judged by different sequence homologies of the 5′ and 3′ halves. The authors pointed out that chimeras formed from uncultivated species whose 16S rRNA sequence is unknown could be recognized only by

establishing different phylogenetic affiliations for separate sequence domains.

Robison-Cox et al. [118] developed a method for the detection of chimeric SSU rRNA sequences, which is based on determining nearest neighbors to an aligned query sequence over two defined sequence domains (aligned similarity method, ASS). Results were compared with those obtained by the CHECK_CHIMERA method [65], which is also based on pairwise similarity analysis but uses unaligned sequences. Both methods were able to detect both known and artificial SSU rRNA chimeras, i.e. chimeras which were formed from two different, authentic sequences, when the parental sequences are quite different. If the similarity between parental sequences is increased from 82 to 96% the confidence of detection of chimeras by both methods decreases from 95 to 50%. As generation of chimeric SSU rRNA sequences is more likely to occur between genes with a high similarity this would mean that the most difficult chimeras to detect are those which are most readily formed. The detection of chimeric sequences is also complicated by the occurrence of authentic SSU rRNA sequences that behave like chimeras. Those sequences have been previously discussed by Sneath [119] as natural chimeras. In conclusion, the authors recommended the combined use of all available methods for the detection of SSU rRNA chimeras, as a single method, especially those which are based on the nearest-neighbor approach, is not sufficient.

9. Cross-checking the results

The PCR-mediated analysis of 16S rRNA is a powerful tool for the determination of microbial diversity of environmental ecosystems. Although there is no general guideline for a 'good PCR-mediated analysis of 16S rRNA from environmental samples' we recommend comparison of results of different nucleic acid extractions, PCR amplification and cloning experiments. Alternatively a mixture of nucleic acids obtained with different extraction methods could be used for multiple PCR amplifications and cloning procedures as these steps are most likely to introduce an experimental error.

Specific oligonucleotides derived from environ-

mental 16S rRNA sequences should be used as dye-labeled probes for 16S rRNA targeted in situ hybridization of fixed sample material [120]. This can answer the question whether a 16S rRNA sequence recovered by the PCR-mediated approach represents an active organism or belongs to a dormant or dead cell as 16S rRNA in situ hybridization preferentially detects actively growing microorganisms with a sufficient content of target molecules. Furthermore, in contrast to the analysis of 16S rRNA gene libraries or DGGE/TGGE pattern, in situ hybridization provides a powerful tool for quantitative analyses as single cells can be specifically detected and counted under the microscope.

The optimal validation of the identification of certain phylogenetic groups in an environmental sample is the enrichment and cultivation of the organisms involved. However, as today only a minor part of the expected microbial diversity can be cultured ex situ, only a few investigators have reported the isolation of microorganisms whose 16S rRNA sequences have been previously detected [17,121–123]. With the growing 16S rRNA sequence dataset, which enables more and more accurate phylogenetic affiliations and possibly information of metabolic pathways from the phylogenetically nearest neighbor, improvements in ex situ culture techniques should become easier.

Finally, certain phylogenetic groups like the methanogens or the methanotrophs exhibit a restricted metabolic potential, which is determined by characteristic functional genes. The detection of those specific genes, e.g. methanol dehydrogenase structural genes of methanotrophs, has been used to verify results obtained by 16S rRNA sequence analyses [124].

References

[1] Stahl, D.A., Flesher, B., Mansfield, H.R. and Montgomery, L. (1988) Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. Appl. Environ. Microbiol. 54, 1079–1084.

[2] Pace, N.R., Stahl, D.A., Lane, D.J. and Olsen, G.J. (1986) The analysis of natural microbial populations by ribosomal RNA sequences. Adv. Microb. Ecol. 9, 1–55.

[3] Seewaldt, E. and Stackebrandt, E. (1982) Partial sequence of 16S ribosomal RNA and the phylogeny of *Prochloron*. Nature 295, 618–620.

[4] Schmidt, T.M., De Long, E.F. and Pace, N.R. (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. 173, 4371–4378.

[5] Ward, D.M., Weller, R. and Bateson, M.M. (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345, 63–65.

[6] Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. and Field, K.G. (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345, 60–63.

[7] Muyzer, G., De Waal, E.C. and Uitterlinden, A.G. (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.

[8] Nübel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R.I., Ludwig, W. and Backhaus, H. (1996) Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. J. Bacteriol. 178, 5636–5643.

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

[10] Köhler, G., Ludwig, W. and Schleifer, K.H. (1991) Differentiation of lactococci by rRNA gene restriction analysis. FEMS Microbiol. Lett. 84, 307–312.

[11] Weidner, S., Arnold, W. and Pühler, A. (1996) Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. Appl. Environ. Microbiol. 62, 766–771.

[12] Ohkuma, M. and Kudo, T. (1996) Phylogenetic diversity of the intestinal bacterial community in the Termite *Reticulitermes speratus*. Appl. Environ. Microbiol. 62, 461–468.

[13] Choi, B.K., Paster, B.J., Dewhirst, F.E. and Göbel, U.B. (1994) Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. Infect. Immun. 62, 1889–1895.

[14] Mirza, M.S., Hameed, S. and Akkermans, A.D.L. (1994) Genetic diversity of *Datisca cannabina*-compatible *Frankia* strains as determined by sequence analysis of the PCR-amplified 16S rRNA gene. Appl. Environ. Microbiol. 60, 2371–2376.

[15] Choi, B.K., Nattermann, H., Grund, S., Haider, W. and Göbel, U.B. (1997) Spirochetes from digital dermatitis lesions in cattle are closely related to treponemes associated with human periodontitis. Int. J. Syst. Bacteriol. 47, 175–181.

[16] Polz, M.F., Distel, D.L., Zarda, B., Amann, R.I., Felbeck, H., Ott, J.A. and Cavanaugh, C.M. (1994) Phylogenetic analysis of a highly specific association between ectosymbiotic, sulfur-oxidizing bacteria and a marine nematode. Appl. Environ. Microbiol. 60, 4461–4467.

[17] Goebel, B.M. and Stackebrandt, E. (1996) Molecular analysis of the microbial diversity in a natural acidic environment. In: Biohydrometallurgical Processing (Jerez, C.A., Vargas, T., Toledo, H. and Wiertz, J.V., Eds.), pp. 43–52. University Press, University of Chile, Santiago.

[18] Borneman, J., Skroch, P.W., O'Sullivan, K.M., Palus, J.A.,

Rumjanek, N.G., Jansen, J.L., Nienhuis, J. and Triplett, E.W. (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. Appl. Environ. Microbiol. 62, 1935–1943.

[19] Schuppler, M., Mertens, F., Schön, G. and Göbel, U.B. (1995) Molecular characterization of nocardioform actinomycetes in activated sludge by 16S rRNA analysis. Microbiology 141, 513–521.

[20] Rheims, H., Rainey, F.A. and Stackebrandt, E. (1996) A molecular approach to search for diversity among bacteria in the environment. J. Ind. Microbiol. 17, 159–169.

[21] Zhou, J., Bruns, M.A. and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62, 316–322.

[22] Bond, P.L., Hugenholtz, P., Keller, J. and Blackall, L.L. (1995) Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. Appl. Environ. Microbiol. 61, 1910–1916.

[23] Pedersen, K., Arlinger, J., Ekendahl, S. and Hallbeck, L. (1996) 16S rRNA gene diversity of attached and unattached bacteria in boreholes along the access tunnel to the Aspo hard rock laboratory, Sweden. FEMS Microbiol. Ecol. 19, 249–262.

[24] Ferris, M.J., Muyzer, G. and Ward, D.M. (1996) Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl. Environ. Microbiol. 62, 340–346.

[25] Rochelle, P.A., Cragg, B.A., Fry, J.C., Parkes, R.J. and Weightman, A.J. (1994) Effect of sample handling on estimation of bacterial diversity in marine sediments by 16S rRNA gene sequence analysis. FEMS Microbiol. Ecol. 15, 215–225.

[26] Gordon, D.A. and Giovannoni, S.J. (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. Appl. Environ. Microbiol. 62, 1171–1177.

[27] Muralidharan, K. and Wemmer, C. (1994) Transporting and storing field-collected specimens for DNA without refrigeration for subsequent DNA extraction and analysis. BioTechniques 17, 420–422.

[28] Liesack, W., Weyland, H. and Stackebrandt, E. (1991) Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed culture of strict barophilic bacteria. Microb. Ecol. 21, 191–198.

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

[30] Ogram, A., Sayler, G.S. and Barkay, T. (1987) The extraction and purification of microbial DNA from sediments. J. Microbiol. Methods 7, 57–66.

[31] Tsai, Y.L. and Olsen, B.H. (1991) Rapid method for direct extraction of DNA from soil and sediments. Appl. Environ. Microbiol. 57, 1070–1074.

[33] Jacobsen, C.S. and Rasmussen, O.F. (1992) 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.

[34] Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R. and Stahl, D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56, 1919–1925.

[35] Volossiouk, T., Robb, E.J. and Nazar, R.N. (1995) Direct DNA extraction for PCR mediated assays of soil organisms. Appl. Environ. Microbiol. 61, 3972–3976.

[36] Picard, C., Ponsonnet, C., Paget, E., Nesme, X. and Simonet, P. (1992) Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. 58, 2717–2722.

[37] Tsai, Y.-L. and Olson, B.H. (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for PCR. Appl. Environ. Microbiol. 58, 2292–2295.

[38] Smalla, K., Cresswell, N., Mendoca-Hagler, L.C., Wolters, A. and van Elsas, J.D. (1993) Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. J. Appl. Bacteriol. 74, 78–85.

[39] Tebbe, C.C. and Vahjen, W. (1993) 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.

[40] Porteous, L.A. and Armstrong, J.L. (1991) Recovery of bulk DNA from soil by a rapid, small-scale extraction method. Curr. Microbiol. 22, 345–348.

[41] Steffan, R.J. and Atlas, R.M. (1988) DNA amplification to enhance detection of genetically engineered bacteria in environmental samples. Appl. Environ. Microbiol. 54, 2185–2191.

[42] Moré, M.I., Herrick, J.B., Silva, M.C., Ghiorse, W.C. and Madsen, E.L.. (1994) Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl. Environ. Microbiol. 60, 1572–1580.

[43] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1990) Current Protocols in Molecular Biology, Vol. 2. Greene/Wiley-Interscience, New York.

[44] Holben, W.E., Jansson, J.K., Chelm, B.K. and Tiedje, J.M. (1988) DNA probe method for the detection of specific microorganisms in the soil bacterial community. Appl. Environ. Microbiol. 54, 703–711.

[45] Barns, S.M., Fundyga, R.E., Jeffries, M.W. and Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. USA 91, 1609–1613.

[46] Somerville, C.C., Knight, I.T., Straube, W.L. and Colwell, R.R. (1989) Simple, rapid method for direct isolation of nucleic acids from aquatic environments. Appl. Environ. Microbiol. 55, 548–554.

[47] Moran, M.A., Torsvik, V.L., Torsvik, T. and Hodson, R.E. (1993) Direct extraction and purification of rRNA for ecological studies. Appl. Environ. Microbiol. 59, 915–918.

[48] Ogram, A., Sun, W., Brockman, F.J. and Frederickson, J.K. (1995) Isolation and characterization of RNA from low-biomass deep-subsurface sediments. Appl. Environ. Microbiol. 61, 763–768.

[49] Felske, A., Engelen, B., Nübel, U. and Backhaus, H. (1996) Direct ribosome isolation from soil to extract bacterial rRNA

for community analysis. Appl. Environ. Microbiol. 62, 4162–4167.

[50] Stackebrandt, E. (1992) Molecular microbial ecology. In: Encyclopedia of Microbiology (Lederberg, I., Ed.), Vol. 3, pp. 171–179, Academic Press, New York.

[51] Stackebrandt, E., Liesack, W. and Witt, D. (1992) Ribosomal RNA and rDNA sequence analyses. Gene 115, 255–260.

[52] Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L. and Pace, N.R. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA 82, 6955–6959.

[53] Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173, 697–703.

[54] Jacobsen, C.S. (1995) Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and pcr amplification assay. Appl. Environ. Microbiol. 61, 3347–3352.

[55] Kreader, C.A. (1996) Relief of amplification in PCR with bovine serum albumin or T4 gene 32 protein. Appl. Environ. Microbiol. 62, 1102–1106.

[56] McKeown, B.J. (1994) An acetylated (nuclease-free) bovine serum albumin in a PCR buffer inhibits amplification. BioTechniques 2, 247–248.

[57] Stinear, T., Matusan, A., Hines, K. and Sandery, M. (1996) Detection of a single viable *Cryptosporidium parvum* oocyst in environmental water concentrates by reverse transcription-PCR. Appl. Environ. Microbiol. 62, 3385–3390.

[58] Suzuki, M.T. and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630.

[59] Farrelly, V., Rainey, F.A. and Stackebrandt, E. (1995) Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. 61, 2798–2801.

[60] Brunk, C.F., Avaniss-Aghajani, E. and Brunk, C.A. (1996) A computer analysis of primer and probe hybridization potential with bacterial small-subunit rRNA sequences. Appl. Environ. Microbiol. 62, 872–879.

[61] Larsen, N., Olsen, G.J., Maidak, B.L., McCaughey, M.J., Overbeek, R., Macke, T.J., Marsh, T.L. and Woese, C.R. (1993) The ribosomal database project. Nucleic Acids Res. 21, 3021–3023.

[62] Zheng, D., Alm, E.W., Stahl, D.A. and Raskin, L. (1996) Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. Appl. Environ. Microbiol. 62, 4504–4513.

[63] Paster, B.J., Pelletier, D.A., Dewhirst, F.E., Weisburg, W.G., Fussing, V., Poulsen, L.K., Dannenberg, S. and Schroeder, I. (1996) Phylogenetic position of the spirochetal genus *Cristispira*. Appl. Environ. Microbiol. 62, 942–946.

[64] Rainey, F.A., Ward, N., Sly, L.I. and Stackebrandt, E. (1994) Dependence on the taxon composition of clone libraries for PCR amplified, naturally occurring 16S rDNA, on the primer pair and the cloning system used. Experientia 50, 796–797.

[65] Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J. and Woese, C.R. (1996) The Ribosomal Database Project (RDP). Nucleic Acids Res. 24, 82–85.

[66] Strunk, O. and Ludwig, W. (n.d.) ARB Software Package. Technical University, Munich.

[67] Reysenbach, A.-L., Giver, L.J., Wickham, G.S. and Pace, N.R. (1992) Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. 58, 3417–3418.

[68] Baskaran, N., Kandpal, R.P., Bhargava, A.K., Glynn, M.W., Bale, A. and Weissman, S.M. (1996) Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome Res. 6, 633–638.

[69] Chandler, D.P., Fredrickson, J.K. and Brockman, F.J. (1997) Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. Mol. Ecol. 6, 475–482.

[70] Waterhouse, R.N. and Glover, L.A. (1993) Differences in the hybridization pattern of *Bacillus subtilis* genes coding for rRNA depend on the method of DNA preparation. Appl. Environ. Microbiol. 59, 919–921.

[71] Shuldiner, A.R., Nirula, A. and Roth, J. (1989) Hybrid DNA artefact from PCR of closely related target sequences. Nucleic Acids Res. 17, 4409.

[72] Meyerhans, A., Vartanian, J.-P. and Wain-Hobson, S. (1990) DNA recombination during PCR. Nucleic Acids Res. 18, 1687–1691.

[73] Wang, G.C.Y. and Wang, Y. (1996) The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology 142, 1107–1114.

[74] Ford, J.E., McHeyzer-Williams, M.G. and Lieber, M.R. (1994) Chimeric molecules created by gene amplification interfere with the analysis of somatic hypermutation of murine immunoglobulin genes. Gene 142, 279–283.

[75] Pääbo, S., Irwin, D.M. and Wilson, A.C. (1990) DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. 265, 4718–4721.

[76] Brakenhoff, R.H., Schoenmakers, J.G.G. and Lubsen, N.H. (1991) Chimeric cDNA clones: a novel PCR artifact. Nucleic Acids Res. 19, 1949.

[77] Cariello, N.F., Thilly, W.G., Swenberg, J.A. and Skopek, T.R. (1991) Deletion mutagenesis during polymerase chain reaction: dependence on DNA polymerase. Gene 99, 105–108.

[78] Pallansch, L., Beswick, H., Talian, J. and Zelenka, P. (1990) Use of an RNA folding algorithm to choose regions for amplification by the polymerase chain reaction. Anal. Biochem. 185, 57–62.

[79] Holstege, F.C.P., van der Vliet, P.C. and Timmers, H.T.M. (1994) Instability of human TATA-binding protein CAG triplet repeats during amplification by PCR. Biochim. Biophys. Acta 1219, 157–159.

[80] Gutell, R.R., Larsen, N. and Woese, C.R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. 58, 10–26.

[81] Chou, Q. (1992) Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by *E. coli* SSB. Nucleic Acids Res. 20, 4371.

[82] Gelfand, D.H. (1992) *Taq* DNA polymerase, PCR technology.

In: Principles and Applications for DNA Amplification (Erlich, H.A., Ed.), pp. 17–22. Freeman and Company, New York.

[83] Eckert, K.A. and Kunkel, T.A. (1991) DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl. 1, 17–24.

[84] Roberts, J.D., Bebenek, K. and Kunkel, T.A. (1988) The accuracy of reverse transcriptase from HIV-1. Science 242, 1171–1173.

[85] Stewart, A.C.M., Gravitt, P.E., Cheng, S. and Wheeler, C.M. (1995) Generation of entire human papillomavirus genomes by long PCR: frequency of errors produced during amplification. PCR Methods Appl. 5, 79–88.

[86] Lundberg, K.S., Shoemaker, D.D., Adams, M.W.W., Short, J.M., Sorge, J.A. and Mathur, E.J. (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. Gene 108, 1–6.

[87] Böttger, E.C. (1990) Frequent contamination of Taq polymerase with DNA. Clin. Chem. 36, 1258–1259.

[88] Rand, K.H. and Houck, H. (1990) *Taq* polymerase contains bacterial DNA of unknown origin. Mol. Cell. Probes 4, 445–450.

[89] Schmidt, T.M., Pace, B. and Pace, N.R. (1991) Detection of DNA contamination in *Taq* polymerase. BioTechniques 11, 176–177.

[90] Maiwald, M., Ditton, H.-J., Sonntag, H.-G. and von Knebel Doeberitz, M. (1994) Characterization of contaminating DNA in *Taq* polymerase which occurs during amplification with a primer set for *Legionella* 5S ribosomal RNA. Mol. Cell. Probes 8, 11–14.

[91] Niederhauser, C., Höfelein, C., Wegmüller, B. and Candrian, U. (1994) Reliability of PCR decontamination systems. PCR Methods Appl. 4, 117–123.

[92] Krawiec, S. and Riley, M. (1990) Organization of the bacterial chromosome. Microbiol. Rev. 54, 502–539.

[93] Kündig, C., Beck, C., Hennecke, H. and Göttfert, M. (1995) A single rRNA gene region in *Bradyrhizobium japonicum*. J. Bacteriol. 177, 5151–5154.

[94] La Fontaine, S. and Rood, J.I. (1996) Organization of ribosomal RNA genes from the footrot pathogen *Dichelobacter nodosus*. Microbiology 142, 889–899.

[95] Stewart, G.C., Wilson, F.E. and Bott, K.F. (1982) Detailed physical mapping of the ribosomal RNA genes of *Bacillus subtilis*. Gene 19, 153–162.

[96] Johansen, T. and Kolsto, C.R. (1996) Variable numbers of rRNA gene operons in *Bacillus cereus* strains. FEMS Microbiol. Lett. 136, 325–328.

[97] Rainey, F.A., Ward-Rainey, N.L., Janssen, P.H., Hippe, H. and Stackebrandt, E. (1996) *Clostridium paradoxum* DSM 7308[T] contains multiple 16S rRNA genes with heterogenous intervening sequences. Microbiology 142, 2087–2095.

[98] Mylvaganam, S. and Dennis, P.P. (1992) Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic Archaebacterium *Haloarcula marismortui*. Genetics 130, 399–410.

[99] Wang, Y., Zhang, Z. and Narendrakumar, R. (1997) The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. J. Bacteriol. 179, 3270–3276.

[100] Ogasawara, N., Nakai, S. and Yoshikawa, H. (1994) Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. DNA Res. 1, 1–14.

[101] Dryden, S.C. and Kaplan, S. (1990) Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. Nucleic Acids Res. 18, 7267–7277.

[102] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496–512.

[103] Clayton, R.A., Sutton, G., Hinkle J.R., Bult, C. and Fields, C. (1995) Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. Int. J. Syst. Bacteriol. 45, 595–599.

[104] Linton, D., Dewhirst, F.E., Clewley, J.P., Owen, R.J., Burnens, A.P. and Stanley, J. (1994) Two types of 16S rRNA genes are found in *Campylobacter helveticus*: analysis, applications and characterization of the intervening sequence found in some strains. Microbiology 140, 847–855.

[105] Linton, D., Clewley, J.P., Burnens, A., Owen, R.J. and Stanley, J. (1994) An intervening sequence (IVS) in the 16S rRNA gene of the eubacterium *Helicobacter canis*. Nucleic Acids Res. 22, 1954–1958.

[106] Ralph, D. and McClelland, M. (1993) Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. Proc. Natl. Acad. Sci. USA 90, 6864–6868.

[107] Rainey, F.A., Ward, N.L., Morgan, H.W., Toalster, R. and Stackebrandt, E. (1993) Phylogenetic analysis of anaerobic thermophilic bacteria: aid for their reclassification. J. Bacteriol. 175, 4772–4779.

[108] Ralph, D. and McClelland, M. (1994) Phylogenetic evidence for horizontal transfer of an intervening sequence between species in a spirochete genus. J. Bacteriol. 176, 5982–5987.

[109] Mägli, A., Rainey, F.A. and Leisinger, T. (1995) Acetogenesis from dichloromethane by a two-component mixed culture comprising a novel bacterium. Appl. Environ. Microbiol. 61, 2943–2949.

[110] Scharf, S.J., Horn, G.T. and Erlich, H.A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 233, 1076–1078.

[111] Aslanidis, C. and de Jong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 18, 6069–6074.

[112] Borriello, F. and Krauter, K.S. (1990) Reactive site polymorphism in the murine protease inhibitor gene family is de-

lineated using a modification of the PCR reaction (PCR+1). Nucleic Acids Res. 18, 5481–5487.

[113] Santegoeds, C.M., Nold, S.C. and Ward, D.M. (1996) Denaturing gradient gel electrophoresis used to monitor the enrichment culture of aerobic chemoorganotrophic bacteria from a hot spring cyanobacterial mat. Appl. Environ. Microbiol. 62, 3922–3928.

[114] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[115] Ward-Rainey, N., Rainey, F.A. and Schlesner, H. (1995) Assignment of hitherto unidentified 16S rDNA species to a main line of descent within the domain bacteria. Microbiology 141, 3247–3250.

[116] Stackebrandt, E., Liesack, W. and Goebel, B.M. (1993) Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J. 7, 232–236.

[117] Kopczynski, E.D., Bateson, M.M. and Ward, D.M. (1994) Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. 60, 746–748.

[118] Robison-Cox, J.F., Bateson, M.M. and Ward, D.M. (1995) Evaluation of nearest neighbor methods for detection of chimeric small-subunit rRNA sequences. Appl. Environ. Microbiol. 61, 1240–1245.

[119] Sneath, P.H.A. (1993) Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. Int. J. Syst. Bacteriol. 43, 626–629.

[120] Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59, 143–169.

[121] Huber, R., Burggraf, S., Mayer, T., Barns, S.M., Rossnagel, P. and Stetter, K.O. (1995) Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis. Nature 376, 57–58.

[122] Wyss, C., Choi, B.K., Schüpbach, P., Guggenheim, B. and Göbel, U.B. (1996) *Treponema maltophilum* sp. nov., a small oral spirochete isolated from human periodontal lesions. Int. J. Syst. Bacteriol. 46, 745–752.

[123] Teske, A., Sigalevich, P., Cohen, Y. and Muyzer, G. (1996) Molecular identification of bacteria from a coculture by denaturing gradient gel electrophoresis of 16S ribosomal DNA fragments as a tool for isolation in pure cultures. Appl. Environ. Microbiol. 62, 4210–4215.

[124] Holmes, A.J., Owens, N.J.P. and Murrell, J.C. (1995) Detection of novel marine methanotrophs using phylogenetic and functional gene probes after methane enrichment. Microbiology 141, 1947–1955.