

47 Clone Libraries of Ribosomal RNA Gene Sequences for Characterization of Bacterial and Fungal Communities

M. B. Leigh¹ · L. Taylor¹ · J. D. Neufeld²

¹Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK, USA

mb.leigh@uaf.edu

²Department of Biology, University of Waterloo, Waterloo, Canada

jneufeld@uwaterloo.ca

1	Introduction	3971
1.1	Important Caveats	3972
2	Approach	3973
2.1	Overview of Procedure	3973
2.2	Nucleic Acid Extraction	3974
2.3	PCR or RT-PCR Amplification	3975
2.4	PCR for Bacterial 16S rRNA Genes	3975
2.5	PCR for Archaeal 16S rRNA Genes	3976
2.6	PCR for ITS + Partial 18S rRNA Genes	3977
2.7	RT-PCR for Bacterial 16S rRNA	3979
2.8	RT-PCR for Archaeal 16S rRNA	3980
2.9	RT-PCR for Fungal ITS	3980
2.10	Preparation of Insert for Cloning (Optional)	3981
2.11	Gel Purification	3981
2.12	3' Adenylation of Purified PCR Products	3981
2.13	Cloning	3982
2.14	Steps in the Procedure	3982
2.15	Screening Clones for Insert	3983
2.16	Steps in the Procedure	3984
2.17	Screening for Unique Clones Using RFLP	3984
2.18	DNA Sequencing	3985
3	Notes on Sequencing Primer Selection	3985
3.1	Sequence Data Analyses	3985
4	Research Needs	3986
4.1	Equipment, Reagents and Materials	3988
4.1.1	Equipment	3988

4.1.2 Reagents and Kits 3989

4.1.3 Materials 3989

4.2 Time Considerations 3989

4.3 Troubleshooting 3990

Abstract: Constructing clone libraries from phylogenetic marker genes is currently the most widely used means of assessing microbial community composition and diversity. Although more expensive and time-consuming than community fingerprinting techniques, sequence analysis of clone libraries provides an unparalleled level of phylogenetic resolution due to the long read lengths generated by Sanger sequencing technology. In this chapter, we provide the principles and methodologies of clone library construction and sequence analyses for the purpose of investigating bacterial, archaeal and fungal community composition. Protocols for generating libraries from either ribosomal RNA (rRNA) genes or rRNA transcripts are provided. Each of the multiple steps involved in cloning are discussed, including DNA extraction, PCR amplification of phylogenetic marker genes, cleanup and preparation of the insert, ligation of the insert into a plasmid vector, transformation of competent *Escherichia coli* cells and screening of clones to verify the presence of cloned inserts, followed by DNA sequence analyses. Primers useful for phylogenetic analyses of bacterial, archaeal and fungal communities are provided. Research needs and emerging high-throughput community sequencing methods are briefly highlighted.

1 Introduction

Environmental microbiology research often requires the assessment of microbial community composition and diversity. Cultivation-based approaches, although valuable in many other ways, are not well suited to this task due to the bias imposed by laboratory growth conditions (Torsvik et al., 1998). A variety of direct molecular biological methods targeting ribosomal RNA (rRNA) have emerged that circumvent culture bias to provide phylogeny-based assessments of community composition and diversity. Community fingerprinting tools such as terminal restriction fragment length polymorphism (T-RFLP; See ► Chapter 58, Vol. 5, Part 3) and denaturing gradient gel electrophoresis (DGGE; See ► Chapter 60, Vol. 5, Part 3) offer valuable, inexpensive and high-throughput measures of community composition, but are limited in their ability to provide taxonomic identities of particular phylotypes. Furthermore, fingerprint profiles are not reliable as estimators of community diversity per se (Blackwood et al., 2007; Neufeld and Mohn, 2005). Clone libraries constructed from phylogenetic marker genes are currently the most widely used means of assessing microbial community composition and diversity simultaneously. However, clone library construction and sequencing are labor-intensive and expensive, and thus time and research budgets often limit the number of sequenced clones that are obtained from a sample. In this protocol, we provide an overview of the principles of clone library construction and sequence analyses for the purpose of investigating bacterial, archaeal and fungal community composition. Although we focus on libraries of ribosomal genes useful for taxonomic identification, the same protocol may be applied to investigate other genes of interest, such as those involved in biodegradation.

The construction of clone libraries for community analysis first requires a selection of appropriate target genes. The small subunit (16S or 18S) rRNA gene has been used extensively as a phylogenetic marker, as it contains both variable and conserved regions, the latter of which permit universal primer design and facilitate subsequent sequence alignment. Since the pioneering work of Norman Pace (Pace et al., 1986), 16S rRNA genes or transcripts have been used as a target for determining the phylogenetic affiliations of organisms from within the domains Bacteria (Nocker et al., 2007) as well as Archaea. For fungi, 18S rRNA genes have been used frequently, but do not provide species-level discrimination; hence, the more variable

intergenic spacer (ITS) region of the nuclear ribosomal operon has been preferred in many studies. Inclusion of partial 5' 28S rRNA gene sequence in the same fungal amplicon permits species-level diagnosis as well as approximate phylogenetic placement of unknown fungi that lack species-level matches in public databases.

The selected target genes are amplified by the polymerase chain reaction (PCR) or by reverse transcription-PCR (RT-PCR) from extracts of DNA or RNA, respectively, which have been obtained from a mixed community. Using DNA is advantageous because it is relatively simple to extract and more stable than RNA. However, analyzing RNA is desirable because it is generally thought to represent metabolically active cells in the community due to the positive correlation observed between metabolic activity and the RNA:DNA ratio (Kerkhof and Ward, 1993; Muttray and Mohn, 1999). Several bacterial 16S rRNA clone libraries (Akob et al., 2007; Mills et al., 2005) and one fungal ITS study (Anderson and Parkin, 2007) generated from DNA and RNA extracts of the same community all produced marked differences in relative abundance of different taxa, and may be valuable for comparisons of abundant versus active community members. In other cases, RNA and DNA libraries produced similar results (Nogales et al., 2001), which suggests that in some cases active and predominant community members may be the same.

Before sequencing of PCR or RT-PCR products can be performed, two critical steps are needed: the mixture of amplicons with different sequences must be separated from each other, and additional copies of each must be made. Clone library construction is a means for achieving these goals. Cloning involves the ligation of the amplified genes into a plasmid vector, followed by the transformation of competent cells (typically *Escherichia coli*) with the recombinant vector. The most common approach to ligation of the PCR-amplified genes into a vector, so-called "T-A" cloning, relies on the presence of "sticky ends" generated by *Taq* DNA polymerase. The polymerase adds a single 3' adenosine overhang to PCR amplicons and these overhangs ligate to cloning vectors by virtue of a vector-specific 3' thymidine overhang, catalyzed by DNA ligase or topoisomerase I. Transformed cells are spread-plated onto a selective medium for easy identification of transformants. Selection generally involves antibiotic resistance genes expressed when cells have taken up the transgenic vector. The TOPO TA cloning system uses the pCR4-TOPO vector which encodes antibiotic resistance (both kanamycin and ampicillin), and also allows direct selection of recombinants via disruption of a gene (*ccdB*) lethal to *E. coli*, eliminating the need for blue/white screening because cells that contain non-recombinant vector are killed upon plating. Colonies derived from individual transformed cells are then picked and regrown to produce larger quantities of plasmid for extraction (or colony PCR) and subsequent sequence analyses. Once sequences are obtained and verified for base-call accuracy, sequences may be compared to databases such as Genbank (Benson et al., 2007), the Ribosomal Database Project (RDP-II) (Cole et al., 2005), GreenGenes (DeSantis et al., 2006) and EMBL (<http://www.ebi.ac.uk/embl/>) (Kulikova et al., 2007) to determine the taxonomic affiliations of the source organisms.

1.1 Important Caveats

As with all PCR-based systems for community analysis, clone libraries are subject to biases introduced during PCR amplification and/or cloning. If primers match imperfectly to different taxa or exhibit different temperature-dependent specificities, amplification may occur disproportionately and bias downstream assessments of community composition. Assessing

primer specificity is an ongoing process and several 16S rRNA gene primers have been reevaluated recently. For example, improved sequences were proposed for 27f (Frank et al., 2008), different annealing temperatures have been suggested for 63f (Sipos et al., 2007), and the effectiveness of domain-specific primers for bacteria and archaea have also been reevaluated (Baker et al., 2003). Widely used fungal primers performed well in tests of their phylogenetic breadth (Fierer et al., 2005), yet there are clearly lineages that will evade detection with these primers (Taylor and McCormick, 2008).

Cloning bias may also be a factor, and may occur through a number of different mechanisms. Because many commercial vectors contain promoters (e.g., *lac* promoter) upstream of the insertion site, transcribed rRNA inserts could potentially impede the growth of *E. coli*, such as by disrupting ribosomal function. Ligation efficiency is also impacted by the length of insert (Sambrook and Russell, 2006), resulting in shorter sequences being cloned at higher frequency. Although the lengths of 16S rRNA genes vary due to variable region indels, it is unknown whether such a size-specific bias measurably affects 16S or 18S rRNA gene cloning. One study compared the fungal community recovered using TOPO TA cloning with that obtained using the low-bias, blunt-cloning Lucigen PCR-SMART vector, and found little evidence for biases due to size, insert expression or TA ligation (Taylor et al., 2007).

The taxonomic identification of organisms based solely on the sequence of a single gene may also be problematic. Microbial taxonomists generally recommend that species-level identification of bacteria involve polyphasic analyses, including multiple genetic and phenotypic analyses (Stackebrandt et al., 2002). Furthermore, there is not complete agreement on a species concept for prokaryotic microorganisms (Cohan, 2002; Rosselló-Mora and Amann, 2001). The relationship between 16S rRNA gene sequence and DNA-DNA hybridization is not a linear one and so reliable bacterial species identifications cannot be based solely on 16S rRNA sequences (Rosselló-Mora and Amann, 2001). The 16S rRNA sequence similarity among multiple species of a bacterial genus (e.g., fluorescent *Pseudomonas*) ranges from 93.3 to 99.9% (Moore et al., 1996). A sequence similarity cutoff of 97% for bacterial OTUs is frequently used because most organisms with <97% 16S rRNA sequence similarity generate DNA-DNA hybridization values of less than 70% and are thus considered members of different species, although organisms with greater than 97% similarity may or may not be members the same species (Goris et al., 2007). The situation is simpler in fungi, because horizontal gene transfer is much less common in eukaryotes than prokaryotes, and high similarity in the ITS region (e.g., 99% or above) virtually guarantees a species-level match. On the other hand, many fungal sequences in public databases are misidentified (Vilgalys, 2003), and, as with prokaryotes, environmental surveys often yield clones with only distant matches to characterized strains (O'Brien et al., 2005; Taylor and McCormick, 2008).

Because of the expense and time-intensiveness of clone libraries, it is usually unfeasible to sequence sufficient clones to detect all phylotypes within a community. As a result, rarefaction analyses are used to assess the extent of community coverage, and various statistical estimators can be applied to predict total species richness in undersampled communities (Hughes et al., 2001).

2 Approach

2.1 Overview of Procedure

The first step in clone library construction is to obtain a DNA or RNA extract from a mixed microbial community of interest, such as soil, sediment, water or other source. Phylogenetic

markers are then amplified using PCR or RT-PCR. Amplicons are purified and inserted into a plasmid vector containing antibiotic resistance genes. Competent *E. coli* cells are transformed with the recombinant vectors and grown on media containing antibiotics, which selects for growth of *E. coli* containing the circularized vector. Traditionally, colonies are further screened based on blue or white color, resulting from the expression of the intact *lacZ* gene, which is expressed when insert is absent (blue colonies are caused by hydrolysis of X-gal to generate a blue precipitate). However, TOPO-TA cloning as described here is based on disruption of the lethal *ccdB* gene when insert is present, eliminating the need for blue/white screening. Insert-containing colonies may then be further screened by PCR using primers located on the vector to confirm the presence and size of inserts. Restriction digestion and agarose gel electrophoresis of extracted plasmid from individual clones is another approach to verifying the presence and size of inserts. An optional screening step with restriction fragment length polymorphism (RFLP) of extracted plasmids or insert PCR products may be used to identify replicate clones in a library, to reduce sequencing costs. Cloned inserts may be sequenced directly from PCR amplification products or from extracted plasmid as template. Insert sequences may be aligned, assembled in trees, used to estimate community diversity, compared across samples, and subjected to database searches for the determination of taxonomic affiliation.

2.2 Nucleic Acid Extraction

The purpose of this step is to obtain DNA or RNA from a mixed microbial community of interest. It is beyond the scope of this protocol to provide detailed instructions on nucleic acid extraction, but some recommendations are provided. A variety of kits and published protocols are available for DNA or RNA extraction. Empirical testing is often required to identify a protocol that is effective for a particular environmental sample. Widely used kits for rapid DNA extraction from soils or sediments are the Fast DNA Spin kit for soil (Bio101, La Jolla, CA), the Power Soil DNA extraction kit (MoBio Laboratories, Carlsbad, CA) and the SoilMaster DNA Extraction kit (Epicentre Biotechnologies, Madison, WI). For rapid coextraction of DNA and RNA, the published protocol of Griffiths et al. (2000) is recommended. The MoBio Soil RNA kit also provides reasonable yields of DNA and so can be used for co-extraction. The manual freeze-grinding DNA extraction protocol published by Zhou et al. (1996) and the DNA and RNA coextraction by Hurt et al. (2001) are more labor intensive but may prove effective, particularly when kits provide poor results or low yields. DNA and RNA coextraction methods can be followed by DNase or RNase treatment to obtain the desired nucleic acid for subsequent analysis.

Notes: Challenges commonly faced during nucleic acid extraction include the coextraction of humic acids and other contaminants that can inhibit PCR. The presence of humics in extracts is evident by a brown color (although extracts containing inhibitory levels may appear colorless) or by their absorbance at the same wavelengths used to quantify nucleic acids (250–300 nm). Dilution of DNA extracts prior to PCR may sufficiently reduce inhibitor concentrations and enable PCR amplification. Precipitation of nucleic acids with polyethylene glycol (PEG) can also significantly reduce the concentration of humic materials. PEG precipitation may be used in place of the final precipitation in most extraction protocols, or performed on the extract itself (Griffiths et al., 2000). Glycogen (20–60 µg per precipitation; Roche) is an effective carrier for DNA and RNA precipitation and may be combined with ethanolic or PEG precipitations to create a visible pellet and provide quantitative recoveries from low-biomass environments (Neufeld et al., 2007). Due to nucleic acid contamination of glycogen that is

purified from shellfish (data not shown), linear polyacrylamide may be preferable for routine applications. If DNA is being extracted from crude oil or highly oil-contaminated samples, the use of a kit designed for stool samples (QiaAmp Stool Kit, Qiagen Inc., Valencia, CA) can effectively remove many PCR inhibitors (Tanaka et al., 2002).

2.3 PCR or RT-PCR Amplification

PCR (for DNA samples) or RT-PCR (for RNA samples) is performed prior to cloning in order to obtain sufficient quantities of the specific target from the mixed community. Example protocols are provided below for PCR and RT-PCR amplification of bacterial and archaeal 16S rRNA and fungal ITS genes or transcripts. Different primer sets are used to target different domains and/or taxa. Consult [Table 1](#) for commonly used primers for amplification of rRNA genes or transcripts from bacteria, archaea and fungi. Different primers will require specific annealing temperatures and consulting with existing literature, or empirical testing of a range of annealing temperatures using a temperature gradient thermal cycler is recommended. Note that shorter amplicons offer increased cloning efficiency, however the resulting sequences provide less sequence information for taxonomic identification and sequence alignment.

2.4 PCR for Bacterial 16S rRNA Genes

1. Prepare PCR mixtures using the following recipe ([Table 2](#)). The volumes listed in [Table 2](#) are for single reactions and a first step is to combine PCR reagents into a “master mix” sufficient for distributing into multiple reaction tubes. In some cases, multiple replicate reactions per sample may be needed to ensure adequate quantities of PCR product for cloning. Always prepare a negative control with no added DNA, and a positive control with a DNA template known to produce successful PCR results. To reduce the risk of contamination, PCR amplifications should be prepared in a clean space or PCR workstation (equipped with air filters and UV lights). As an additional precaution, aerosol barrier (filter) tips are recommended.
2. Amplify using the following thermal cycler program (min:sec). Note that by beginning each PCR program with an initial 95°C pause (not shown), PCR tubes may be loaded directly into the heated block. This effectively serves as a simplified “hot-start” to reduce non-specific amplification that may be associated with loading the tubes at room temperature and heating more slowly.

Denaturation

95°C 5:00

25 cycles

95°C 1:00

55°C* 1:00

72°C 1:40

Final extension

72°C 10:00

*Note: Optimal annealing temperature depends on primer used. Lower annealing temperatures (as low as 50°C) may reduce amplification bias, particularly if using primer 63f (Sipos et al., 2007). It has also been suggested that additional diversity may be recovered if PCR is performed across a range of annealing temperatures (i.e., 48–58°C

■ **Table 1**

PCR and sequencing primers useful for bacterial, archaeal and fungal clone libraries and phylogenetic characterization

	Primer sequence	Reference
Bacterial 16S rRNA^a		
27f	5'-AGAGTTTGATCMTGGCTCAG-3'	Lane (1991)
27f-YM + 3	5'-AGAGTTTGATYMTGGCTCAG-3' plus 5'-AGAATTGATCTTGGTTCAG-3' plus 5'-AGAGTTTGATCCTGGCTTAG-3' plus 5'-AGGGTTCGATTCTGGCTCAG-3'	Frank et al. (2008)
63f	5'-CAGGCCTAACACATGCAAGTC-3'	Marchesi et al. (1998)
519r	5'-G(AT)ATTACGCGGC(GT)GCTG -3'	Lane (1991)
1392r	5'-ACGGGCGGTGTGTRC-3'	Lane (1991)
1492r	5'-TACCTTGTTACGACTT-3'	Lane (1991)
Archaeal 16S rRNA		
A571F	GCYTAAAGSRICCGTAGC	Baker et al. (2003)
UA1204R	TTMGGGGCATRCIKACCT	Baker et al. (2003)
A751	CCGACGGTGAGRGRYGA	Baker et al. (2003)
UA1406R	ACGGGCGGTGWGTRCAA	Baker et al. (2003)
Fungal ITS/LSU		
ITS1-F	5'-CTTGGTCATTAGAGGAAGTAA-3'	Gardes and Bruns (1993)
ITS4	5'-TCCTCCGCTATTGATATGC-3'	White et al. (1990)
TW13	5'-GGTCCGTGTTCAAGACG-3'	White, unpublished (http://plantbio.berkeley.edu/~bruns/tour/primers.html)
Vector primers^b		
M13f	5'-GTTTCCCAGTCACGACGTTG TA-3'	TOPO TA Cloning kit
M13r	5'-CAGGAAACAGCTATGACC-3'	TOPO TA Cloning kit

^aFor additional bacterial and archaeal 16S rRNA primers, see Baker et al. (2003)

^bM13f and M13R are provided in limited quantities in TOPO TA Cloning kit

with primers 27F and 1492R) and then PCR products pooled and subjected to gel purification (Brodie, 2005).

3. Verify success of PCR amplification using agarose gel electrophoresis (1% agarose, 100V, 30–45 min) based on the predicted amplicon size (will vary slightly among different organisms).

2.5 PCR for Archaeal 16S rRNA Genes

PCR amplification of archaeal 16S rRNA genes may be conducted using either of the two primer pairs (A571F/UA1204R or A751/UA1406R, ▶ [Table 1](#)), which were developed by Baker et al. (2003) and have demonstrated enhanced Archaeal domain specificity.

■ **Table 2****Components of a single 25- μ l PCR mix for amplification of bacterial 16S rRNA genes**

Reagent	Volume (μ l)	Final concentration
Water (sterile, nuclease free)	19.18	
10X PCR buffer	2.5	1X
MgCl ₂ (50 mM)	0.75	1–4 mM
dNTP mixture (25 mM each)	0.2	0.2 mM
Primer 27f (10 μ M)	0.5	0.2 μ M
Primer 1392r or 1492r (10 μ M)	0.5	0.2 μ M
<i>Taq</i> DNA polymerase (5U/ μ l)	0.125	0.03 U/ μ l
BSA (10 mg/ml) ^a	0.25	0.1 mg/ml
DNA sample (1–50 ng/ μ l)	1.0	0.04–2.0 ng/ μ l
Total volume	25	

^aBSA is a recommended additive to improve PCR product yields from soil or sediment samples, but may not be necessary for other samples, such as water

1. Prepare the following PCR mix (► [Table 3](#)) for archaeal 16S rRNA amplification.
2. Amplify using the following thermal cycler program (min:sec) (Baker et al., 2003)
 - Denaturation
 - 94°C 5:00
 - 25 cycles
 - 94°C 1:00
 - 55°C 1:00
 - 72°C 1:00
 - Final extension
 - 72°C 10:00
3. Verify success of PCR amplification using agarose gel electrophoresis (1% agarose, 100 V, 30–45 min).

2.6 PCR for ITS + Partial 18S rRNA Genes

1. Prepare PCR amplification mixtures using the recipe in ► [Table 4](#) Multiple replicate reactions per sample may be needed to ensure adequate quantities of PCR product for cloning.
2. Amplify using the following thermal cycler program (min:sec)
 - Denaturation
 - 95°C 2:00
 - 25 cycles
 - 94°C 0:30
 - 56°C 0:40
 - 72°C 3:00
 - Final extension
 - 72°C 10:00

■ **Table 3**

Components of a single 25- μ l PCR mix for amplification of archaeal 16S rRNA genes (Baker et al., 2003)

Reagent	Volume (μ l)	Final concentration
Water (sterile, nuclease free)	19.18	
10X PCR buffer	2.5	1X
MgCl ₂ (50 mM)	0.75	1.5 mM
dNTP mixture (25 mM each)	0.2	0.2 mM
Primer A571F or A751 (10 μ M)	1.25	0.5 μ M
Primer UA1204R or UA1406R (10 μ M)	1.25	0.5 μ M
<i>Taq</i> DNA polymerase (5U/ μ l)	0.2	0.04 U/ μ l
BSA (10 mg/ml) ^a	0.25	0.1 mg/ml
DNA sample (1–50 ng/ μ l)	1.0	0.04–2.0 ng/ μ l
Total volume	25	

^aBSA is a recommended additive to improve PCR product yields from soil or sediment samples, but may not be necessary for other samples, such as water

■ **Table 4**

PCR mixture for ITS + partial 18S rRNA genes

Reagent	Volume (μ l)	Final concentration
Water (sterile, nuclease free)	18.025	
10X PCR buffer	2.5	1X
MgCl ₂ (50 mM)	0.75	1.5 mM
dNTP mixture (25 mM each)	0.1	0.10 mM each
Primer ITS1-F (10 μ M)	1.25	0.5 μ M
Primer TW13 (10 μ M)	1.25	0.5 μ M
<i>Taq</i> DNA polymerase (5 U/ μ l)	0.125	0.03 U/ μ l
DNA sample (1–50 ng/ μ l)	1	0.04 – 2.0 ng/ μ l
Total volume	25	

- Verify success of PCR amplification using agarose gel electrophoresis (1% agarose, 100 V, 30–45 min) based on the predicted amplicon size.

Notes: The primer ITS4 (White et al., 1990) can be used in place of TW13 if a shorter amplicon is preferred or difficulty is encountered at the PCR stage. However, the partial 18S rRNA sequence region allows for deep phylogenetic placement of unknown fungi that may be impossible with the ITS alone, since the ITS can only be aligned across groups of closely related fungi, typically at the genus level. If product yield is unacceptably low, the number of PCR cycles can be increased to 30. However, we advise against this because the proportion of chimeric sequences increases with higher cycle numbers, and the relative proportions of

legitimate taxa in the final PCR pool become distorted due to the kinetic bias described previously (Suzuki and Giovannoni, 1996).

2.7 RT-PCR for Bacterial 16S rRNA

This method is a multi-step process using primer 519r for reverse transcription of 16S rRNA, followed by PCR of the cDNA. The universal bacterial 16S rRNA primer 519r is used in conjunction with 27f to generate an approximately 520-bp product. Longer RT-PCR amplicons can be generated using other primers but are more challenging, due to the tendency for reverse transcriptase to dissociate from the RNA template. This protocol is written specifically for use with Superscript RNA polymerase (Invitrogen Corp., Carlsbad, CA), however other RNA polymerases are available and the reader is referred to product information for appropriate protocols.

Extreme caution should be exercised when working with RNA to ensure that RNase is not present in reagents (purchase RNase free reagents) or plasticware (certified RNase free) and that contamination is avoided. Note that autoclaving is insufficient to inactivate RNase enzymes. Aerosol barrier (filter) tips should be used and gloves worn at all times to protect the sample from RNases on skin. RNA template should be DNAase-treated to remove contaminating DNA, which is commonly coextracted with RNA. Extracts and reagents should be kept cold at all times to minimize RNase activity.

1. Prepare the following mixture to anneal primer to template (► Table 5). Include a positive control RNA extract and an RNA-free negative control. For a no-RT control, prepare an additional control containing sample RNA template and follow through the entire process except without the addition of reverse transcriptase enzyme. Amplification of the no-RT control indicates the presence of DNA in RNA samples.
2. Incubate at 65°C for 5 min, then chill on ice.
3. Prepare the following mixture (► Table 6) and add 7 µl to each reaction tube on ice from the previous step.
4. Incubate at 42°C for 2 min.
5. Add 1 µl Superscript II enzyme (Invitrogen, Cat. No. 18064022) to each tube from previous step to complete reverse transcription (RT) mixture. If RNA quantity is below 1 ng then dilute enzyme 1:4 in RNase-free water and then add 1 µl of diluted enzyme to each reaction as directed in product information. Mix well by pipetting up and down.

■ Table 5

Reaction mixture for bacterial 16S rRNA RT-PCR primer annealing

Reagent	Volume (µl)	Final concentration
Primer 519r ^a (10 µM)	0.4	0.4 µM
dNTPs (25 mM)	1	2.3 mM
Water (sterile, nuclease-free)	4.6	
Template RNA (1 ng–5 µg of total RNA)	5	0.1 ng/ µl – 0.5 µg/ µl
Total volume	11	

^aOther reverse primers, such as 1392r or 1492r may be used. Amplification of shorter products may be more successful, due to the tendency of RNA polymerase to dissociate during the transcription of longer fragments

Table 6

Reverse transcription reaction mixture (excluding reverse transcriptase)

Reagent	Volume (μl)	Final concentration
5X first strand buffer	4	2.9X
0.1 M Dithiothreitol (DTT)	2	0.03 M
RNAseOUT (Invitrogen Corp.) ^a	1	
Total volume	7 μl	

^aRNAseOUT (Invitrogen, Cat. No. 10777–019) is required if using <50 ng starting RNA

- Incubate at 42°C for 50 min to produce cDNA then 70°C for 15 min to inactivate the enzyme.
- Amplify 1 μl of RT-product (cDNA) generated in Step 6 using same PCR method described in 16S rRNA PCR protocol above, except with primer 519r in place of 1392r (or 1492r). Include positive and negative (no DNA) control in PCR, in addition to controls generated during RT (described in Step 1).
- Verify the success of the RT-PCR reaction using agarose gel electrophoresis. Amplicons should be ~520 bp in size. If a product is produced from the no-RT control, this indicates that DNA contamination is present since the contaminating DNA will be amplified in the PCR step. If this is the case, then further DNase treatment of template RNA is required.

2.8 RT-PCR for Archaeal 16S rRNA

The same multi-step process described above for RT-PCR of bacterial 16S rRNA can be followed for archaeal 16S rRNA when modified to use one of the reverse primers (UA1204R or UA1406R) targeting archaeal 16S rRNA (🔗 Table 1) for reverse transcription in conjunction with subsequent PCR using the same reverse primer with the appropriate forward primer (A571F/UA1204R or U751/UA1406R). Refer to above RT-PCR protocol for bacterial 16S rRNA for important tips on RNA handling.

2.9 RT-PCR for Fungal ITS

The protocols described above for RT-PCR of bacterial or archaeal 16S rRNA can be followed for fungal ITS when modified to use the universal eukaryote ITS rRNA primer ITS4 for reverse transcription in conjunction with subsequent PCR using primers ITS4 and ITS1-F. Following reverse transcription with primer, amplify 1 μl of RT-product (cDNA) using the same PCR method described in fungal ITS rRNA PCR protocol above, but with ITS4 rather than TW13 to generate a 500–800 bp product. Refer to above RT-PCR protocol for bacterial 16S rRNA for important tips on RNA handling.

Notes: PCR or RT-PCR should be performed within a short time prior to cloning, as fresh PCR products yield best cloning results. Multiple low volume PCR amplifications (12.5 or 25 μl) produce higher yields than a single larger volume reaction, therefore it is recommended that to obtain sufficient quantities of PCR product multiple low volume reactions be prepared instead of scaling up the single-reaction volume.

2.10 Preparation of Insert for Cloning (Optional)

This step is optional, but improved cloning results may be achieved when PCR products are purified prior to cloning by agarose gel electrophoresis to remove nonspecific products and primer dimers, followed by 3' adenylation of the purified products. The 3'A overhang on the insert is essential for successful cloning using TA cloning method, which relies on these sticky ends to anneal to the vector. *Taq* DNA polymerase naturally adds 3'A to amplicons but the single stranded 3'A overhang is vulnerable to loss during electrophoresis. If this is a suspected cause of low cloning efficiency, 3'A overhangs can be quickly and easily added using the method below immediately before proceeding to cloning.

2.11 Gel Purification

1. If the volume of combined replicate PCR amplifications exceeds 50 μ l, the DNA should be concentrated to enable loading into a single well in an agarose gel for gel purification. DNA can be easily concentrated using the QIAquick PCR purification kit and eluting DNA in 30 μ l water or buffer.
2. Prepare a 1% agarose gel with a large-toothed comb to accommodate the PCR product in as few lanes as possible (ideally one). Leave empty wells between samples to prevent cross contamination during gel cutting. Load and run the gel at 100V for \sim 45 min. Stain using ethidium bromide or other fluorophores such as SYBR Green or SYBR Safe (Invitrogen Corp., Carlsbad, CA).
3. Briefly visualize the gel with a UV transilluminator, working quickly to avoid damaging DNA by UV exposure. Wearing personal protective equipment to shield all skin and eyes from UV light and ethidium bromide, excise bands using sterile razor blade or scalpel. Change to a new sterile blade between each sample. Place the excised gel slices into pre-weighed 1.5-ml microfuge tubes.
4. Gel purify using QIAquick Gel Purification kit (Qiagen), and perform final elution step using 30 μ l of water, or use Minelute PCR Purification Kit (Qiagen) to elute in a lower volume (e.g., 10 μ l) to obtain higher DNA concentrations.
5. Determine the DNA concentration of purified products. DNA concentrations ranging from 10 to 100 ng/ μ l are desirable at this stage.

Notes: An alternative to gel purification is the use of size fractionation columns. For fungal libraries, we typically use ChromaSpin columns (ChromaSpin 400, Clontech, Mountain View, CA) prior to TOPO cloning. The frequency of short inserts in our libraries is less than 1% meaning that checking clones for insert sizes becomes unnecessary.

2.12 3' Adenylation of Purified PCR Products

This step is optional, but may improve cloning efficiency by replacing any single stranded 3'A lost from inserts during gel purification. The 3' adenylation step should be performed immediately prior to cloning.

1. Prepare the following master mix ([Table 7](#)).
2. Dispense 6.6 μ l of master mix into PCR tubes.

Table 7

Master mix for 3' adenylation of inserts

Reagent	Volume (μl)	Final concentration
10X PCR buffer	2.5	1X
MgCl ₂ (50 mM)	1.5	11.4 mM
dATP (10 mM)	2.5	3.8 mM
Taq DNA polymerase (5U/μl)	0.1	0.08 U/μl
Total volume	6.6	

3. Add gel-purified PCR product and water (up to 20 μl combined volume of water and sample) to the 6.6 μl of master mix.
4. Incubate on a heat block or in water bath at 72°C for 10 min.

2.13 Cloning

The protocol below makes use of the TOPO TA cloning kit for sequencing with TOP10 chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA) and follows the kit instructions with minor modifications. Note that SOC. medium and competent *E. coli* cells (chemically competent or electrocompetent) may be prepared in the laboratory but these protocols are not provided here. Please contact the authors if protocol recommendations are needed.

2.14 Steps in the Procedure

1. Mix reagents in Table 8 in the order in which they are listed. Half reactions may be performed if desired.
2. Mix gently and incubate for 5–30 min at room temperature, then place on ice. Increased incubation time may increase cloning efficiency, especially for large or dilute inserts. During the incubation, thaw one vial of OneShot chemically competent *E. coli* cells on ice for each transformation, warm a vial of SOC. medium (kit provided) to room temperature and pre-warm selective agar plates (LB containing ampicillin or kanamycin) to 37°C.
3. Add 4 μl of the cloning reaction into a vial of chemically competent *E. coli* (kit provided) and mix gently by stirring or flicking tube. Do not mix by pipetting up and down as this may damage the cells.
4. Incubate on ice for 30 min. Meanwhile, warm water bath up to 42°C.
5. Heat shock cells for 30 s at 42°C (no shaking), then immediately transfer to ice.
6. Add 250 μl of room-temperature SOC medium. Cap tube and shake horizontally (200 rpm) at 37°C for 1 h.
7. Spread 10–50 μl of each transformation mix on pre-warmed selective plate and incubate overnight at 37°C. Spreading a range of volumes from the same sample onto separate plates will help ensure that isolated colonies are obtained.

■ Table 8

Ligation reaction mixture

Reagent	Volume (μl)
Insert (PCR product to be cloned)	0.5 – 4.0 (10–100 ng total DNA ^a)
Salt solution (kit provided)	1
Sterile water	(for final volume of 5 μl with insert and salt)
TOPO vector	1
Total volume	6

^aThe ratio of insert to vector should be 3:1 to maximize diversity recovery. One microliter of TOPO vector contains 3.9 fmol (10 ng) DNA. For a ~1,500 bp PCR product, 10–12 fmol (10–12 ng) should be used (Brodie, 2005)

8. Pick colonies with sterile toothpicks or narrow-bore pipette tips and place in LB freezing medium containing ampicillin or kanamycin in deep 96-well growth block or 96-well plate. Avoid picking tiny satellite colonies if present. The use of kanamycin (bacteriocidal) rather than ampicillin (bacteriostatic) in selective plates reduces the risk of resistant satellite colonies.
9. Cover growth block with breathable sealing tape and grow overnight at 37°C.
10. When cells have grown (liquid in wells is turbid), the clone library can be frozen at either –80°C or –20°C. Airpore tape should be removed and replaced with aluminum foil tape to maintain moisture prior to freezing. It may be desirable to aliquot a portion of each well into separate 96-well plates prior to freezing and store in different freezers to prevent library loss in the event of freezer malfunction.
11. To re-grow clones as colonies, thaw plates and use flame-sterilized 96-pin colony replicator to stamp onto Omnitrays containing selective agar medium. This format is convenient for shipping libraries, such as to contract sequencing services.

Note: Transformation of *E. coli* by electroporation uses electrocompetent cells, also available with purchase of TOPO TA cloning kits, or readily produced by established protocols. The efficiency of electroporation is higher by several-fold and the procedure is rapid for small numbers of samples.

2.15 Screening Clones for Insert

It is important to screen clones to estimate the proportion of clones carrying inserts of the expected size. Although the positive selection strategy of the TOPO TA kit effectively screens for clones with inserts, short non-target inserts can also create antibiotic resistant clones. Screening can be performed with PCR and agarose gel electrophoresis and/or restriction digestion of plasmids. Screening tests may be done either directly from original colonies or from clones following growth in 96-well plates. It is recommended that at least ten clones be screened per library to obtain an estimated percentage of successful clones.

The following PCR screening method involves PCR amplification with vector primers M13f and M13r. For screening using restriction digestion, the reader is referred to the TOPO TA kit product information.

2.16 Steps in the Procedure

1. Prepare the PCR mix in [▶ Table 2](#), except using vector primers M13f and M13r (100 μ M, 0.5 μ l each). To add template, touch a colony with toothpick or narrow bore pipette tip and insert into the PCR tube containing master mix, or pipette a small quantity (0.2–0.5 μ l) of liquid culture from 96-well plates into the PCR mix. Note that the initial denaturation cycle (5 min or more) is important to lyse the *E. coli* cells and release the plasmids into the PCR mix.
2. Analyze results using agarose gel electrophoresis (1–1.5% agarose, 100 V for 1.5–2 h) with a DNA ladder that includes reference bands near the predicted amplicon size. When screening pCR4-TOPO vector using M13f and M13r, clones with correct inserts will produce bands approximately 165 bp longer than the original PCR products. Clones lacking an insert will produce bands of \sim 165 bp. Other bands will often be present and do not affect results. Refer to the TOPO PCR4 vector data provided in product information of the TOPO TA kit for details regarding location of priming sites on the vector.

Notes: Another option for screening is restriction digestion of plasmids. Briefly, plasmids are extracted from clones, digested using *Eco*RI restriction endonuclease and then analyzed using agarose gel electrophoresis. In conjunction with the plasmid restriction map the presence of insert can be ascertained.

2.17 Screening for Unique Clones Using RFLP

Sequencing costs can be reduced by identifying sets of clones containing nearly identical inserts and choosing only one or two representatives of each set of clones for sequence analysis. This approach is particularly attractive for fungal ITS clone libraries due to the fact that sequence variation in the non-coding ITS1 and ITS2 regions is sufficiently great that restriction sites usually vary among even related species. Separate digestion of PCR-amplified clone inserts with at least two different 4-cutter enzymes is recommended to reduce the frequency with which shared restriction patterns mask species differences. Details of fungal ITS RFLP analyses have been previously published (Gardes and Bruns, 1993). In brief, amplified inserts should be digested to completion with 4-cutter enzymes such as *Hae*III or *Mbo*I and run on high-resolution agarose gels (e.g., 2% Nusieve agarose plus 1% SeaKem agarose). A representative of each banding pattern can then be sequenced.

Notes: Using the TOPO PCR4 vector, clones with identical inserts can produce slightly different RFLP patterns, depending on the direction in which the insert was ligated into the vector. Hence, some redundant sequencing may occur with this approach. To circumvent this problem, using a restriction enzyme that cuts adjacent to all inserts (e.g., *Eco*RI) can ensure that insert orientation does not have a noticeable impact on RFLP patterns. Another difficulty is that different fungal species will occasionally produce similar or identical ITS RFLP patterns when using a single enzyme. This unintended lumping of taxa may be reduced by utilizing multiple enzymes (three are usually sufficient) in separate digests, and by sequencing several representatives of the most abundant restriction fragment types.

2.18 DNA Sequencing

Sequencing of clone libraries is most commonly performed using the dideoxynucleotide-chain-terminating DNA sequencing method of Sanger et al. (1977). There are many competitively-priced contract sequencing services (as low as 1–5 US dollars per reaction) or if equipment is available sequencing may be done in-house. A detailed protocol for sequencing is beyond the scope of this protocol. Readers are referred to *Short Protocols in Molecular Biology, 5th Edition* for Sanger sequencing methods (Ausubel et al., 2002).

3 Notes on Sequencing Primer Selection

Sequencing is generally performed as single extension from a vector primer such as (M13f, M13r, T3 or T7). However it should be kept in mind that DNA randomly inserts in the forward or reverse direction. Sequence reads typically range from 600 to 800 bp, so when longer inserts are being analyzed ~50% of clone sequences will be from the 5' end of the gene and ~50% from the 3' end, precluding alignment of all sequences. A recommended alternative is to sequence with the forward 16S rRNA primer used to generate inserts (e.g., 27f) or an internal primer, which guarantees that the same portion of the gene will be sequenced from each clone.

3.1 Sequence Data Analyses

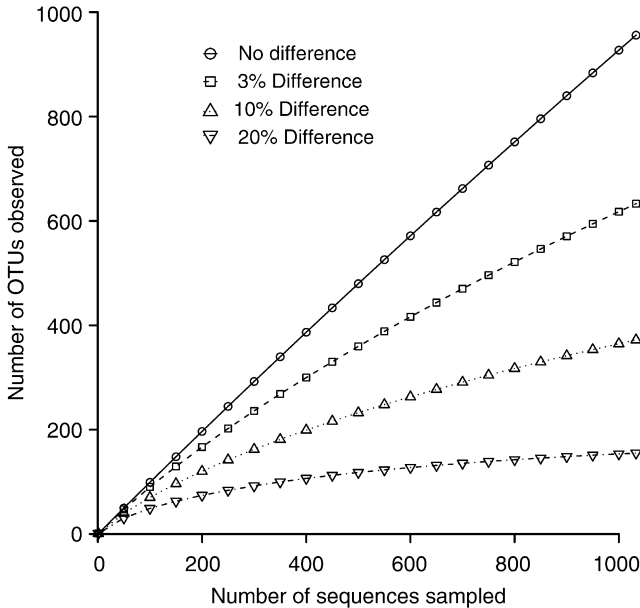
1. Raw sequence data should be manually and carefully checked for basecalling accuracy, and trimmed to exclude low quality sequence regions, typically at the beginning and end of the sequences. Software such as Sequencher (Gene Codes Corp.), Aligner (CodonCode Corp.) or Phred/Phrap (Ewing et al., 1998) (<http://www.phrap.org/index.html>) can automatically trim sequences based on quality. For bacterial and archaeal sequences, the RDP pipeline (<http://rdp.cme.msu.edu/pipeline/>) provides automated trimming using Phred and vector sequence removal using LUCY (Chou and Holmes, 2001). We also recommend removing the original PCR primer sequences as these do not necessarily represent exact copies of the original environmental sequences.
2. Sequences should be checked for chimeras. Chimeras are hybrid PCR products derived from two or more different organisms and are artifacts that can be formed during PCR. It has been reported that as many as 1 out of 20 sequences in GenBank are chimeras or contain sequencing errors or other anomalies (Ashelford et al., 2005). Several free programs for chimera checking are available, including MALLARD (Ashelford et al., 2006), Pintail (<http://www.bioinformatics-toolkit.org/>), Bellerophon (<http://foo.maths.uq.edu.au/%7Ehuber/bellerophon.pl>) and ChimeraCheck (<http://rdp8.cme.msu.edu/cgis/chimera.cgi?su=SSU>). However, none of these programs are effective for fungal ITS regions because they depend upon multiple sequence alignments, which are impossible when analyzing diverse fungal sequences. At present, a crude but moderately effective approach is to compare separate BLAST searches using the ITS1 versus ITS2 regions of each clone sequence.
3. To determine taxonomic affiliation of bacteria and archaea based on 16S rRNA the RDP-II (<http://rdp.cme.msu.edu/index.jsp>) (Cole et al., 2005; Wang et al., 2007) or GreenGenes (<http://greengenes.lbl.gov>) (DeSantis et al., 2006) databases are recommended. The RDP-II is frequently updated with 16S rRNA sequences from GenBank, which are aligned

and classified according to the higher-order taxonomy proposed in Bergey's Taxonomic Outline of the Prokaryotes, and thus provides more reliable taxonomic identifications than NCBI GenBank BLAST searches. Sequences can be compared to the RDP database using Sequence Match function to identify nearest matches in the database. The RDP Classifier function classifies the sequence into Bergey's hierarchical taxonomy with confidence levels for each taxonomic level provided.

4. For identification of fungi using the ITS region, BLAST searches against Genbank are most commonly used. Another option is to search all GenBank sequences via a database and web site designed for fungal matching (<http://www.borealfungi.uaf.edu>). This site provides taxonomic information on top hits as well as a curated version of GenBank data from which unidentified, environmental sequences have been excluded. Another useful database of well identified fungi is maintained by the Unite group: <http://unite.ut.ee/>.
5. Microbial communities in the environment are highly diverse, and it is unlikely that all community members present will be detected using clone libraries. For this reason, it is important to examine the extent of community coverage obtained before drawing conclusions regarding the species richness, diversity or relative abundance of community members detected. Rarefaction curves are a valuable means to determine coverage and are constructed by plotting the number of different operational taxonomic units (OTUs) on the y-axis versus the number of clones sequenced on the x-axis (Schloss and Handelsman, 2005). If the curve reaches an asymptote, then the community has been sampled completely and all OTUs have been recovered. The shape of the rarefaction curve is affected by the definition of OTU applied, e.g., 95, 97 or 100% sequence similarity. As exemplified in [▶ Figs. 1](#) and [▶ 2](#), complete community coverage is rarely obtained in environmental microbiology studies. [▶ Fig. 1](#) provides an illustration of how changing the percent sequence similarity cutoff for OTU designation impacts the shape of the curve, while [▶ Fig. 2](#) illustrates the major impact of the spatial sampling scheme on diversity estimates. Statistically derived indices such as Chao1 and other tools are valuable for estimating species richness and diversity of undersampled communities using tools described below. To perform these calculations and to obtain other estimates of diversity, a number of programs are available. Measurements of α diversity (diversity within one sample), including OTU (species) richness estimations (Chao1 or ACE) and diversity indices (e.g., Shannon, Simpson) can be calculated using programs such as EstimateS (<http://viceroy.eeb.uconn.edu/estimates>), DOTUR (Schloss and Handelsman, 2005) (<http://www.plantpath.wisc.edu/fac/joh/dotur.html>), or the newest version MOTHUR (http://schloss.micro.umass.edu/mothur/Main_Page). Comparison of diversity and/or composition between two and more communities (β diversity) can be examined using programs such as EstimateS, UniFrac (Lozupone et al., 2007, 2006), SONS (Schloss and Handelsman, 2006b) LIBSHUFF (Schloss et al., 2004; Singleton et al., 2001) and TreeClimber (Schloss and Handelsman, 2006a). Tools for comparing microbial structure from clone library data were recently reviewed and evaluated by Schloss (2008).

4 Research Needs

The goal of many ecological analyses is to estimate and compare microbial diversity and composition across multiple samples and with varying physicochemical parameters. However,

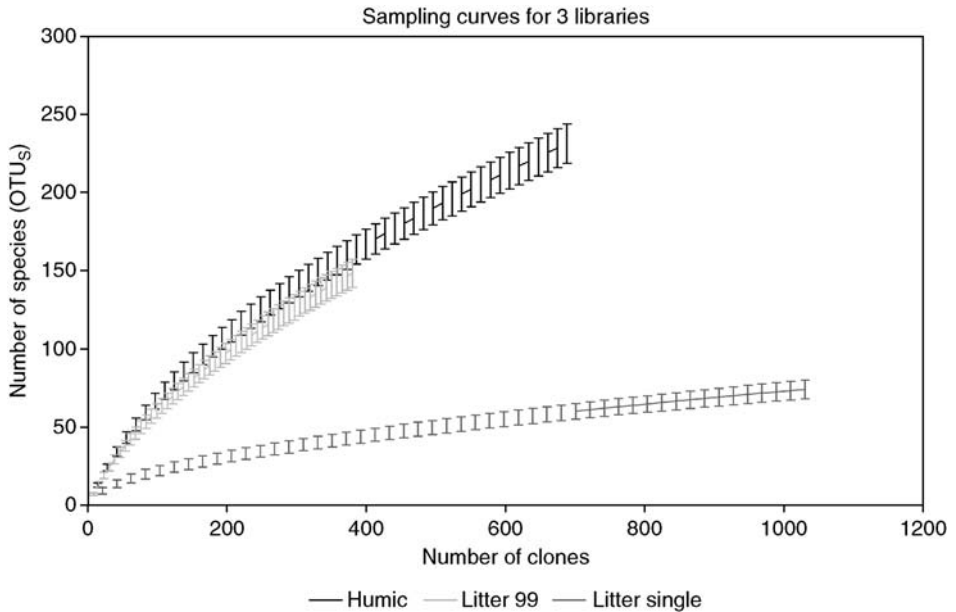


■ **Figure 1**

Rarefaction curves generated from a 16S rRNA gene library from Alaskan soil at different sequence identities. Error bars represent the 95% confidence intervals. This figure was provided courtesy of Patrick Schloss from a dataset described in Schloss and Handelsman (2006c).

it is important to be aware of bias that will skew the clone library profiles from the true microbial community composition. Biases are introduced at both the PCR stage of community analysis (Acinas et al., 2005; Becker et al., 2000; Polz and Cavanaugh 1998; Suzuki and Giovannoni, 1996; Suzuki et al., 1998) and also in the ligation, transformation and colony picking stages in cloning approaches, but these latter biases have not been well investigated (Taylor et al., 2007).

Strategies for characterization of microbial communities are evolving rapidly. In particular, high density microarrays such as the Phylochip (Brodie et al., 2006) are likely to see increasing use, complementing the use of clone libraries for community comparisons. Arrays offer the advantages of high sensitivity and competitive costs relative to sequencing of clone libraries (DeSantis et al., 2007), and have the advantage of circumventing limitations of cloning. High-throughput sequencing from methods such as serial analysis of ribosomal sequence tags (SARST) (Neufeld et al., 2004) and next-generation sequencing platforms such as 454 pyrosequencing, Illumina and the SOLiD platform from Applied Biosystems (Huber et al., 2007; Hudson, 2008) are clone-free (except traditional SARST) and offer much higher throughput than Sanger-sequencing, but currently at the cost of short sequence reads. Although short sections of the 16S rRNA gene are sufficiently variable to allow discrimination of closely related bacterial taxa, this is not the case with fungi. Only the ITS regions are sufficiently variable to allow discrimination of closely related species. However, ITS regions cannot be aligned well across genera or families, and thus do not enable the construction of broad phylogenetic trees. Thus, the short sequences currently available via pyrosequencing may not be appropriate for the analysis of fungal microbial communities.



■ **Figure 2**

Numbers of fungal OTUs detected in three clone libraries from the same site. Two soil horizons (humic and litter) were removed from 99 soil cores (top curves) or a single core (bottom curve, litter only). DNA was extracted separately from each core, then combined prior to PCR of the ITS region, cloning and sequencing. Sequences were grouped at 97% identity to enumerate OTUs. Rarefaction analyses of each library were carried out in EstimateS. The lower slope in the curve from the single soil core indicates lower total fungal diversity than seen in the pooled cores, but saturation has not been achieved after sequencing 1,000 clones from the single core (Taylor and Herriott, unpublished data).

Sanger sequencing has become increasingly affordable in recent years, with sequence reads of \$1 US being cited, when 96-well plates are provided for sequencing. Although alternative approaches for assessing and comparing multiple microbial communities will become increasingly important in the near future, Sanger sequencing of clone libraries will likely remain a competitive strategy for characterization of selected microbial communities by providing long (>750 bp) sequence reads at increasingly affordable rates.

4.1 Equipment, Reagents and Materials

4.1.1 Equipment

- PCR thermal cycler
- Agarose gel electrophoresis apparatus and power supply
- UV transilluminator
- Microcentrifuge (refrigerated if working with RNA)

- Water bath (42°C)
- 37°C shaking and non-shaking incubator
- Pipettors
- Spectrophotometer or NanoDrop (Thermo Scientific)
- Autoclave

4.1.2 Reagents and Kits

- TOPO TA Cloning Kit for Sequencing with TOP10 chemically competent *E. coli* (Invitrogen, Carlsbad, CA, Cat. No. K4575–01)
Note: competent cells and SOC medium can be made easily to reduce costs (Frank et al., 2008). Refer to TOPO TA cloning kit protocol for SOC medium composition.
- Basic PCR or RT-PCR reagents
- Reagents for agarose gel electrophoresis (agarose, TAE, ethidium bromide, loading buffer)

4.1.3 Materials

- Pipette tips (sterile)
- Aerosol barrier tips, sterile, nuclease free (for 16S rRNA gene amplification)
- Toothpicks (sterile)
- Petri plates (sterile)
- Sterile 96-well plates, standard (Fisher Scientific Cat. No. 21–377–205) or 2.0-ml deep well (Dot Scientific, Burton, MI, Cat. No. SB2434-NS7)
- Breathable sealing tape for 96-well plates (Dot Scientific, Burton, MI, Cat. No. T391)
- Aluminum foil sealing tape for 96-well plates (Dot Scientific, Burton, MI, Cat. No. 626-PCR)
- Omnitrays (single-well plates) (Nunc 242811)
- Syringe filters and syringes (for antibiotics)

4.2 Time Considerations

The total time required to construct a clone library, beginning with DNA extraction and ending with a library of clones requires approximately 6 days. DNA extraction protocols vary between 2 h and 2 days to complete, and preparation of media and supplies for cloning require 0.5–1 days. Preparation of insert, including PCR or RT-PCR amplification, agarose gel electrophoresis and gel purification require 1–2 days. Performing the cloning reactions requires 1 day. Picking colonies may require a few hours depending on how many clones are desired. Growth in liquid media after picking colonies requires an additional day. Screening clones for insert requires an additional 0.5–1 day. Sequencing of the library requires additional time which varies from a few days to 2 weeks, depending on the contract service or protocol used.

4.3 Troubleshooting

Low cloning efficiency

- Perform control cloning reaction as described in TOPO TA cloning kit to verify that reagents and methods are working properly.
- Increase concentration of PCR product used for cloning by generating more replicate PCR reactions or concentrating existing sample.
- Incubate insertion reaction for longer period of time, up to 30 min.
- Electroporate electrocompetent cells for increased transformant yields.
- Perform 3'adenylation reaction as indicated previously to incorporate lost overhanging adenines.

Desired Insert Not Present in Clones

- Gel purify or perform size fractionation of PCR products prior to cloning

Acknowledgments

MBL was supported by the National Science Foundation under Award 0626544. DLT was supported by the National Science Foundation under Grant no. 0317144 and JDN acknowledges a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

References

- Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* 71: 8966–8969.
- Akob DM, Mills HJ, Kostka JE (2007) Metabolically active microbial communities in uranium-contaminated subsurface sediments. *FEMS Microbiol Ecol* 59: 95–107.
- Anderson I, Parkin P (2007) Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *J Microbiol Methods* 68: 248–253.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* 71: 7724–7736.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* 72: 5734–5741.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (2002) *Short Protocols in Molecular Biology*, Vol 5. Wiley.
- Baker G, Smith J, Cowan D (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55: 541–555.
- Becker S, Boger P, Oehlmann R, Ernst A (2000) PCR bias in ecological analysis: A case study for quantitative *Taq* nuclease assays in analyses of microbial communities. *Appl Environ Microbiol* 66: 4945–4953.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2007) GenBank. *Nucleic Acids Res* 36: D25–D30.
- Blackwood CB, Hudleston D, Zak DR, Buyer JS (2007) Interpreting ecological diversity indices applied to T-RFLP data: insights from simulated microbial communities. *Appl Environ Microbiol* 73: 5276–5283.
- Brodie E (2005) PCR and TOPO TA Cloning for community analysis. Greengenes: 16S rRNA gene database and workbench compatible with ARB <http://greengenes.lbl.gov/Download/Protocols/>.

- Brodie EL, DeSantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL, Hazen TC, Richardson PM, Herman DJ, Tokunaga TK, Wan JMM, Firestone MK (2006) Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* 72: 6288–6298.
- Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* 17: 1093–1104.
- Cohan F (2002) What are bacterial species? *Annu Rev Microbiol* 56: 457–487.
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM (2005) The ribosomal database project (RDPII): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33(Database Issue): D294–D296.
- DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* 53: 371–383.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069–5072.
- Ewing B, Hillier L, Wendl M, Green P (1998) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175–185.
- Fierer N, Jackson JA, Vilgalys R, Jackson RB (2005) Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl Environ Microbiol* 71: 4117–4120.
- Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 74: 2461–2470.
- Gardes M, Bruns T (1993) ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Mol Ecol* 2: 113–118.
- Goris J, Konstantindis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57: 81–91.
- Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and RNA-based microbial community composition. *Appl Environ Microbiol* 66: 5488–5491.
- Huber JA, Mark Welch D, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. *Science* 318: 97–100.
- Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Resour* 8: 3–17.
- Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* 67: 4399–4406.
- Hurt RA, Qiu X, Wu L, Roh Y, Palumbo AV, Tiedje JM, Zhou JZ (2001) Simultaneous recovery of RNA and DNA from soils and sediments. *Appl Environ Microbiol* 67: 4495–4503.
- Kerkhof L, Ward BB (1993) Comparison of nucleic acid hybridization and fluorometry for measurement of the relationship between RNA/DNA ratio and growth rate in a marine bacterium. *Appl Environ Microbiol* 59: 1303–1309.
- Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57: 81–91.
- Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, Baldwin A, Bates K, Bhattacharyya S, Bower L, Browne P, Castro M, Cochran G, Duggan K, Eberhardt R, Faruque N, Hoad G, Kanz C, Lee C, Leinonen R, Lin Q, Lombard V, Lopez R, Lorenc D, McWilliam H, Mukherjee G, Nardone F, Garcia-Pastor MP, Plaister S, Sobhany S, Stoehr P, Vaughan R, Wu D, Zhu W, Apweiler R (2007) EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Res* 35: D16–D20.
- Lane DJ (1991) 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*. E Stackebrandt, M Goodfellow (eds.) Chichester, UK: John Wiley & Sons, Inc., pp. 115–175.
- Lozupone C, Hamady M, Kelley S, Knight R (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73: 1576–1585.
- Lozupone C, Hamady M, Knight R (2006) UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* 7: 371–384.
- Marchesi J, Sato T, Weightman A, Martin T, Fry J, Hiom S, Dymock D, Wade W (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microbiol* 64: 795–799.
- Mills HJ, Martinez RJ, Story S, Sobecky PA (2005) Characterization of microbial community structure in Gulf of Mexico gas hydrates: comparative analysis of DNA- and RNA-derived clone libraries. *Appl Environ Microbiol* 71: 3235–3247.
- Moore ERB, Mau M, Arnscheidt A, Böttger EC, Huston RA, Collins MD, van de Peer Y, de Wachter R, Timmis KN (1996) The determination and

- comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationships. *Syst Appl Microbiol* 19: 478–492.
- Muttray AF, Mohn WW (1999) Quantitation of the population size and metabolic activity of a resin acid degrading bacterium in activated sludge using slot-blot hybridization to measure the rRNA:rDNA ratio. *Microbial Ecol* 38: 348–357.
- Neufeld JD, Mohn WW (2005) Assessment of microbial phylogenetic diversity based on environmental nucleic acids. In *Molecular Identification, Systematics, and Population Structure of Prokaryotes*. Stackebrandt E. (ed.) Heidelberg: Springer-Verlag.
- Neufeld JD, Vohra J, Dumont MG, Lueders T, Manefield M, Friedrich MW, Murrell JC (2007) DNA stable-isotope probing. *Nat Protocols* 2: 860–866.
- Neufeld JD, Yu Z, Lam W, Mohn WW (2004) Serial analysis of ribosomal sequence tags (SARST): a high-throughput method for profiling complex microbial communities. *Environ Microbiol* 6: 131–144.
- Nocker A, Burr M, Camper AK (2007) Genotypic microbial community profiling: a critical technical review. *Microb Ecol* 54: 276–289.
- Nogales B, Moore ERB, Llobet-Brossa E, Rossello-Mora R, Amann R, Timmis KN (2001) Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl Environ Microbiol* 67: 1874–1884.
- O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Appl Environ Microbiol* 71: 5544–5550.
- Pace N, Stahl D, Lane D, Olsen G (1986) The analysis of natural microbial-populations by ribosomal-RNA sequences. *Adv Microb Ecol* 9: 1–55.
- Polz ME, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* 64: 3724–3730.
- Rosselló-Mora R, Amann R (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* 25: 39–67.
- Sambrook J, Russell DW (2006) *The condensed protocols from Molecular cloning: a laboratory manual*, Vol. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467.
- Schloss PD (2008) Evaluating different approaches that test whether microbial communities have the same structure. *Isme J* 2: 265–275.
- Schloss P, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* 71: 1501–1506.
- Schloss A, Handelsman J (2006a) Introducing TreeClimber, a test to compare microbial community structures. *Appl Environ Microbiol* 72: 2379–2384.
- Schloss P, Handelsman J (2006b) Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microbiol* 72: 6773–6779.
- Schloss PD, Handelsman J (2006c) Toward a census of bacteria in soil. *PLoS Comput Biol* 2:e92.
- Schloss PD, Larget BR, Handelsman J (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* 70: 5485–5492.
- Singleton DR, Furlong MA, Rathbun SL, Whitman WB (2001) Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl Environ Microbiol* 67: 4374–4376.
- Sipos R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K, Nikolausz M (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol Ecol* 60: 341–350.
- Stackebrandt E, Frederiksen W, Garrity GM, Grimont P, Kampfer P, Mea Maiden (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* 52: 1043–1047.
- Suzuki MT, Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* 62: 625–630.
- Suzuki M, Rappe MS, Giovannoni SJ (1998) Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Appl Environ Microbiol* 64: 4522–4529.
- Tanaka Y, Sogabe N, Okumura K, Kurane R (2002) A highly selective direct method of detecting sulphate-reducing bacteria. *Lett Appl Microbiol* 35: 242–246.
- Taylor D, Herriott I, Long J, O'Neill K (2007) TOPO-TA is OK: a test of phylogenetic bias in fungal environmental clone library construction. *Environ Microbiol* 9: 1329–1334.
- Taylor D, McCormick M (2008) Internal transcribed spacer primers and sequences for improved characterization of basidiomycetous orchid mycorrhizas. *New Phytol* 177: 1020–1033.
- Torsvik V, Daae FL, Sandaa RA, Øvreas L (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol* 64: 53–62.

- Vilgalys R (2003) Taxonomic misidentification in public DNA databases. *New Phytol* 160: 4–5.
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) A Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16): 5261–5267.
- White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR protocols A guide to methods and applications*. MA Innis, DH Gelfand, JJ Sninsky, White JW (eds) New York: Academic Press, pp. 315–322.
- Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62: 316–322.

