Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Enrichment for microbes living in association with plant tissues

J.-Y. Jiao[1,2], H.-X. Wang[1,2], Y. Zeng[1] and Y.-M. Shen[1]

1 Kunming Institute of Botany, Chinese Academy of Sciences, Yunnan, China
2 Graduate School of the Chinese Academy of Sciences, Beijing, China

Keywords
endophytes, *Maytenus hookeri*, metagenome, microbe enrichment, 16S rDNA, *Trewia nudiflora*, uncultured microbes.

Correspondence
Y. Zeng, Kunming Institute of Botany, Chinese Academy of Sciences, Heilongtan, Kunming 650204, Yunnan, China.
E-mail: biochem@mail.kib.ac.cn

2005/0572: received 23 May 2005, revised 12 August 2005 and accepted 15 August 2005

doi:10.1111/j.1365-2672.2006.02830.x

Abstract

Aims: To investigate a cultivation-independent method of enrichment for microbes living in association with plant tissues.
Methods and Results: A large quantity of leaves or seeds was enzymatically hydrolyzed, and the pellets were collected by differential centrifugation. Enzyme concentration, buffer and incubation time were optimized for release of plant-associated microbes. The relative abundance of plant nuclear DNA and bacterial DNA in the enriched sample was estimated by PCR amplification of genome-specific marker genes. The efficiency of microbe enrichment was estimated from the proportion of bacterium-derived clones and their restriction fragment length polymorphism (RFLP) types as detected by 16S rRNA gene-based techniques. With a higher ratio of bacterial to plant nuclear DNA, the enriched samples showed a considerably enhanced proportion of bacterium-derived clones and a wider sequence diversity of those clones.
Conclusions: The method described here proved to be remarkably effective in enriching for bacteria living in association with plant tissues.
Significance and Impact of the Study: The method can be applied to study plant-associated microbes in the field of environmental molecular ecology and environmental metagenomics.

Introduction

Plants are complex habitats for colonization by different kinds of microbes. Recent research in molecular ecology has shown that the vast majority of microbes in a natural community is refractory to cultivation and remains unexplored (Ward *et al.* 1990; Pace 1997). A desire to understand the physiology and ecological roles of these unknown microbes has led to the development of culture-independent metagenomic approaches (Handelsman *et al.* 1998; Rondon *et al.* 2000; Handelsman 2004). Encouraging reports come from the identification of polyketide synthase gene clusters from uncultured bacterial symbionts of *Paederus* beetles (Piel 2002) and from the marine sponge *Theonella swinhoei* (Piel *et al.* 2004). In both cases the uncultured microbes are highly enriched in association with their hosts, making it feasible to construct an ideal metagenomic library containing sufficient copies of interesting genomes from the uncultured microbes. In a few plants, microbes are also found inhabiting particular sites, such as the root nodules of legumes (Bauer 1981) and leaf galls (Miller *et al.* 1983). For most plants, however, little is known about the diversity and colonization sites of microbes, so metagenomic libraries contain a high proportion of plant genomic DNA. Therefore, the first and essential step is to enrich for the microbes living in association with plant tissues.

A Nycodenz density gradient was successfully used to separate bacterial cells from soil particles (Courtois *et al.* 2001, 2003). Enrichment for genomes of metabolically active cells has been extensively studied (Schloss and Handelsman 2003; Chowdhury *et al.* 2004; Handelsman

2004), but enrichment for uncultured microbes from plant tissues has not been reported before.

In this paper, we report an effective method to enrich for microbes (cultured and uncultured) in host plant tissues, and examine its efficiency by 16S rDNA-based analyses. The shrub *Maytenus hookeri* was selected for this initial study because the maytansinoids (19-membered macrocyclic lactams related to ansamycin antibiotics of microbial origin) were isolated from this plant (Yu *et al.* 2002). The available evidence suggests production of the core structure of the plant maytansinoids by an associated microbe. However, an intensive study of plant-associated microbial isolates has repeatedly failed to reveal a microbial producer of maytansinoids from maytansinoid-producing plants (Cassady *et al.* 2004). We first attempted to enrich for plant-associated microbes of cultured- and uncultured-origin, and then used seeds of another maytansinoid-producing plant (*Trewia nudiflora*) to confirm the feasibility of our method for microbe enrichment in plant tissues.

Materials and methods

Hydrolyzation of plant cell walls and differential centrifugation

Leaves of *M. hookeri* were collected in the Kunming Botanical Garden and washed with tap water. For each sample, 40 g of fresh leaves were ground to a fine powder in liquid nitrogen and suspended in 200 ml of 0.25 mol l^{-1} sucrose solution containing macerozyme R-10 and cellulase R-10 (Yakult Honsha Co., Ltd., Tokyo, Japan). Macerozyme at 1·5% plus 0·5% cellulase or 3·5% macerozyme plus 1·5% cellulase were applied in each treatment. As a control, the leaf powder (40 g) was suspended in 0·25 mol l^{-1} sucrose only. All samples were incubated at 40°C with gentle agitation for 3 h and then subjected to separation by differential centrifugation. The pellets from five centrifugation steps at 200 ***g***, 500 ***g***, 800 ***g***, 1000 ***g***, 3000 ***g*** (20 min for each step) were collected. A 0.5 g sample of each pellet was extracted for DNA using the procedure of Li *et al.* (2001) modified by adding lysozyme (1.5 mg ml^{-1}) and incubating the mixture at 37°C for one hour before starting lysis.

Estimation of DNA abundance in each pellet

The DNA isolated from the above pellet consists of host plant nuclear DNA, host plastidial DNA and, if any, plant-associated microbial DNA. In order to estimate the relative abundance of different genomic DNAs, marker genes including the entire ITS (nucleus-specific), the *trnL-trnF* spacer (plastid-specific) and the 16S–23S rRNA gene spacer (bacterium-specific) were each amplified by PCR using an Eppendorf Mastercycler gradient. Ten-fold serial dilutions (10-, 100-, 1 000- fold or higher increments) of all the original DNAs (0·2–0·5 $\mu g\ \mu l^{-1}$) were first prepared and amplified with the appropriate universal primers (White *et al.* 1990; Taberlet *et al.* 1991; Jensen *et al.* 1993). The highest dilution that worked for PCR amplification of the marker gene was obtained (Table 1). Considering the PCR inhibitors possibly present in each DNA sample, amplification with dilute DNA sample was performed if there were no PCR products from the starting template. Data were used only qualitatively, i.e. the absence or presence of the expected band(s) was recorded (~0·75 kb for the plant nuclear ITS; ~1·0 kb for *trnL-F* including the intron, the 3′ exon, and the intergenic spacer; 0·3–0·9 kb for most bacterial 16S-23S rRNA gene spacers) (White *et al.* 1990; Taberlet *et al.* 1991; Jensen *et al.* 1993). PCR was carried out in 25 μl volumes containing 2·5 μl of 10 × PCR buffer, 2·5 μl of each primer (5 $\mu mol\ l^{-1}$), 0·5 μl of 10 mmol l^{-1} dNTPs mix, and 0·6 U of *Taq* polymerase (TaKaRa). The PCR parameters were as follows for amplification of the nuclear ITS and the *trnL-trnF* spacer: one cycle of 3 min at 94°C, linked to 30 cycles of 1 min at 94°C, 1 min at 52°C, 2 min at 72°C, followed by 10 min at 72°C to complete primer

Table 1 PCR amplification patterns of the marker genes in each pellet from differential centrifugation

	200 ***g***			500 ***g***			800 ***g***			1000 ***g***			3000 ***g***		
Pellets marker genes	N	P	B	N	P	B	N	P	B	N	P	B	N	P	B
Control	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	++	/	/	/
1.5% macerozyme 0.5% cellulase	+++	+++	++	++	++	+++	–	–	++	–	++	+++	–	–	+
3.5% macerozyme 1.5% cellulase	++	++	+++	++	+	+++	–	++	+++	–	++	+++	–	+	+++

N, nuclear rDNA (ITS region); P, plastidial *trnL-trnF* spacer; B, bacterial 16S-23S rRNA gene spacer. The highest dilution that worked for PCR amplification of the marker gene was recorded as + (1- or 10-fold), ++ (100-fold), and +++ (1000-fold), respectively. Absence of band(s) was recorded as –. No pellet was obtained in the control from centrifugation at 3000 ***g***.

extension. PCR of the 16S-23S rRNA gene spacer was carried out using the following program: one cycle of 2 min at 98°C, linked to 30 cycles of 1·5 min at 94°C, 1·5 min at 49°C, 1·5 min at 72°C, followed by 10 min at 72°C to complete primer extension. PCR products were visualized by 1·0% agarose gel electrophoresis in 1 × TAE buffer after staining with ethidium bromide (10 μg ml^{-1}). PCR was repeatedly performed in all samples.

To test the specificity of the primers, bacterial DNAs of several species (*Actinosynnema pretiosum*, *Bacillus subtilis*, *Streptomyces* sp.CS) were amplified using the plant nuclear ITS sequence; no amplicon of the expected size (~0.75 kb) was detected. However, a bacterial 16S-23S rRNA gene spacer was easily amplified from the plant DNA because of the presence of complex nuclear and plastid DNA.

Preparation of the enriched samples

Based on the results obtained from the experiments described above, two enriched samples were prepared with different enzyme treatment. For each sample, fresh leaves (40 g) were ground to a fine powder in liquid nitrogen, and then suspended in the respective enzyme solution under aseptic conditions. For Method I, the leaf powder was suspended in 200 ml of 0·25 mol l^{-1} sucrose solution containing 3.5% macerozyme plus 1·5% cellulase and incubated at 40°C with gentle agitation for three hours. The mixture was then separated by two-step differential centrifugation. First, the mixture was centrifuged at 200 ***g*** for 5 min, the supernatant was collected and the pellet (i.e. 200 ***g***-pellet) was discarded. Second, the above supernatant was centrifuged at 3000 ***g*** for 20 min and the pellet (i.e. 3000 ***g***-pellet) was collected. Optionally, the 200 ***g***-pellet was suspended in 0·25 mol l^{-1} sucrose solution, the centrifugation step at 200 ***g*** was repeated three times and all the supernatants were pooled for centrifugation at 3000 ***g***. For Method II, 400 ml of enzyme solution (0·1% macerozyme, 1·0% cellulase, 0·7 mol l^{-1} mannitol, 5 mmol l^{-1} N-morpholinoethanesulfonic acid, 9 mmol l^{-1} $CaCl_2$, 65 μmol l^{-1} KH_2PO_4, pH 5·7) were mixed with leaf powder, and the suspension was aseptically incubated at 28°C with gentle agitation for 12 h. The mixture was then subjected to the two-step differential centrifugation used in Method I. Alternatively the centrifugation step at 200 ***g*** was repeated three times and the supernatants were pooled for centrifugation at 3000 ***g***. The collected 3000 ***g***-pellets were used as the enriched samples for further analyses. In addition, Method II was repeatedly performed with 40 g of *T. nudiflora* seeds collected in Xishuangbanna Botanical Garden (Yunnan, China), to verify that the method of microbe enrichment is effective with a different plant tissue. Seeds, in terms of components and cell structures, are quite different from leaves.

Assessing the efficiency of microbe enrichment by 16S rDNA-based techniques

DNA was extracted individually from 0·5 g of the enriched samples as described before (Li *et al.* 2001) with some modifications by adding lysozyme (3·0 mg ml^{-1}), L-lysine (0·15 mol l^{-1}), EGTA (6 mmol l^{-1}, pH 8·0) and incubating the mixture at 37°C for one hour before starting lysis. After confirming DNA quality by spectrophotometry and gel electrophoresis, relative abundance of different DNAs in the controls and the enriched samples was estimated by marker gene PCR as described above. The DNA was then used for 16S rDNA library construction. The control DNA was extracted directly from 2·0 g of *M. hookeri* leaves or 2·0 g of *T. nudiflora* seeds.

Nearly full-length 16S rDNA was PCR-amplified using a bacterium-specific forward primer (27f, 5′-AGA GTT TGA TCM TGG CTC AG-3′) at nucleotide positions 8-27 (*Escherichia coli* numbering) and a reverse primer (1492r, 5′-TAC GGY TAC CTT GTT ACG ACT T-3′) at positions 1513–1492 (Lane *et al.* 1985), for 20 cycles only to minimize bias associated with high cycle numbers (Suzuki and Giovannoni 1996). The PCR followed the program: one cycle of 3 min at 95°C, linked to 20 cycles of 40 s at 95°C, 40 s at 50°C, 1·5 min at 72°C, and a final 10 min of elongation at 72°C. DNA was ligated into the pGEM®-T vector (Promega) and introduced into $CaCl_2$-competent *E.coli* DH5α. From each library, more than 100 positive clones were randomly picked from the plates and used for restriction fragment length polymorphism (RFLP) analysis.

PCR reactions were performed directly on *E. coli* cells. Each positive clone was inoculated in a PCR buffer using a sterile toothpick for amplification of the insert. Vector-specific universal T7/SP6 promoter sequences were used as primer pairs and the resulting PCR products were checked for size and quality on 0.8% agarose. Because plastid-derived sequences were expected, T7/SP6-amplified PCR products were first digested with *Pvu*II (NEB), which cuts in plastid 16S rDNA (Sessitsch *et al.* 2002). Clones from which the *Pvu*II could digest inserts were recognized as a group of plastid-derived clones. The undigested 16S rDNAs were then digested separately with *Hha*I and *Hae*III (5 μl of the PCR product with five units of *Hha*I or *Hae*III in 20 μl total volume for 3 h at 37°C). Digests were electrophoresed in 2.5% agarose gels. Restriction patterns were compared and nearly identical patterns were grouped. Each phylotype was defined as a group of sequences with identical *Hha*I and *Hae*III patterns. A representative clone of each RFLP type was partially sequenced at Shanghai GeneCore BioTechnologies Co., Ltd., China. DNA sequences were preliminarily analyzed by the BLASTN (2.2.11 version) program (Altschul

Table 2 Estimation of microbial diversity based on 16S rDNA sequence analyses

	M. hookeri leaves			*T. nudiflora* seeds	
Clones	Control	Method I	Method II	Control	Method II
Number of clones detected	111	114	114	105	97
*Pvu*II-insensitive clones	4	10	104	14	94
RFLP types	4	7	25	7	30
Phylogenetic types*	1 chl, 1 mt, 2 bac	1 chl, 6 bac	1 chl, 24 bac	1 mt, 6 bac	30 bac
% bacterial clones†	1.8 (2/111)	7.9 (9/114)	90.3 (103/114)	12.4 (13/105)	97.0 (94/97)

*A representative clone of each RFLP type was partially sequenced and phylogenetically grouped by BLASTN at NCBI. Phylotypes were designated as chl (chloroplast-affiliated), mt (mitochondrion-affiliated) and bac (bacterium-affiliated), respectively.
†Percentage of 16S rDNA clones potentially affiliated with bacteria.

et al. 1997) of NCBI in the GenBank database to identify their putative close phylogenetic relatives. Plastidial and mitochondrial small subunit ribosomal RNA gene sequences of *M. hookeri* and *T. nudiflora* were deposited in GenBank and given accession numbers AY855838, AY855839, and DQ000201, respectively.

Results

Differences in PCR amplification of the marker genes

As summarized in Table 1, the three marker genes were amplified in roughly uniform abundance for all pellets from the control and for the enzyme-treated 200 ***g***-pellets. In the enzyme-treated pellet from centrifugation at 1000 ***g*** or 3000 ***g***, bacterial 16S–23S rRNA gene spacer can be amplified with 1000-fold dilute template. Conversely, the plant nuclear ITS gene was undetectable by PCR in the enzyme-treated pellets from centrifugation at 800 ***g***, 1000 ***g*** and 3000 ***g*** (Table 1). Differences in DNA amplification were observed for the 3000 ***g***-pellet derived from the treatment with 3·5% macerozyme plus 1·5% cellulose.

Efficiency of microbe enrichment in plant tissues

Before cloning and RFLP typing, different DNAs extracted from the enriched samples and the controls were estimated for relative abundance by PCR amplification of three marker genes (data not shown). Dilution of Method II DNA at the highest 10 000-fold resulted in amplification of the bacterial 16S–23S rRNA gene spacer, whereas neither the plant nuclear ITS region nor the *trnL-trnF* fragment were detected with the same template. On the contrary, only the plant nuclear ITS gene was amplified when diluting the control DNA at the highest 10 000-fold. Compared to the control, the enriched sample showed a higher ratio of bacterial to plant nuclear DNA. Similar results were obtained with *T. nudiflora* seeds (results not shown). To compare the enriched samples with their respective controls for microbial diversity, we employed 16S rRNA gene-based techniques including 16S rDNA cloning, RFLP analysis of 16S rDNA for rapid comparison of rDNAs (Moyer *et al.* 1994) and sequence analysis. The efficiency of microbe enrichment was deduced from the percentage of clones affiliated with bacteria and the diversity of their RFLP patterns.

Because plastid-derived sequences were expected, T7/SP6-amplified PCR products were first digested with *Pvu*II to identify a restriction site in chloroplast 16S rDNA sequences that is not found in most eubacterial 16S rDNA genes (Sessitsch *et al.* 2002). As revealed in this study, a *Pvu*II site is indeed present in plastid 16S rDNA sequences of *M. hookeri* (AY855838) and *T. nudiflora* (DQ000201). Clones from which the *Pvu*II could not digest inserts were recognized as a group of *Pvu*II-insensitive clones that might be potentially of bacterial origin. Table 2 showed that the majority (90–97%) of Method II clones were *Pvu*II-insensitive, potentially affiliated with bacteria. As expected, the control showed a high abundance of plant organelle ribosomal sequences. Among *Pvu*II-insensitive clones of *M.hookeri*, unique RFLP types were observed for 25 in Method II, with only seven RFLPs in Method I and four in the control. Likewise, *Pvu*II-insensitive clones of *T.nudiflora* showed 30 RFLPs for Method II and seven for the control (Table 2). Based on sequence analyses, Method II showed a substantially higher diversity of bacteria than Method I and the control. Phylogenetic grouping revealed that our 16S rDNA clone libraries were dominated by *γ-Proteobacteria*. A few clones of the libraries were affiliated with uncultured or unclassified bacteria (Table 3). However, as only partial 16S rDNA sequences (∼0·6 kb) were used, this phylogenetic placement is tentative.

Discussion

Plant-associated microbes include those on the surface of plant organs as well as endophytic microbes. Endophytic

Table 3 Sequence analysis of eubacterial 16S rDNA clones from *M. hookeri* leaf tissues*

Clone	Number	Closest match in NCBI database	Identity (%)	Putative phylum
Control				
M82	1	*Pseudomonas* sp. (AJ575816)	99	*γ-Proteobacteria*
M104	1	Uncultured Methylobacteriaceae (AY360529)	97	*α-Proteobacteria*
Method I				
37-3	1	*Hafnia alvei* (AY572428)	99	*γ-Proteobacteria*
40-3	1	Unclassified bacterium (AY561547)	94	*γ-Proteobacteria*
77-3	1	*Citrobacter freundii* (AB210978)	98	*γ-Proteobacteria*
78-3	4	Uncultured soil bacterium (AF423250)	98	*γ-Proteobacteria*
84-3	1	*Providencia* sp. (AY870456)	98	*γ-Proteobacteria*
148-3	1	Enterobacteriaceae (AY538694)	99	*γ-Proteobacteria*
Method II				
3N2	6	*Erwinia* sp. (AY660693)	98	*γ-Proteobacteria*
3N4	28	Uncultured *Pantoea* sp. (AY942953)	98	*γ-Proteobacteria*
3N10	4	Unclassified bacterium (AY822505)	99	*γ-Proteobacteria*
3N12	2	Unclassified bacterium (AY770431)	99	*γ-Proteobacteri*
3N13	1	Unclassified bacterium (AY822500)	98	*γ-Proteobacteria*
3N16	1	*Enterobacter agglomera* (AF130951)	91	*γ-Proteobacteria*
3N22	2	Uncultured bacterium (AY958878)	99	*γ-Proteobacteria*
3N24	2	*Erwinia rhapontici* (Z96087)	97	*γ-Proteobacteria*
3N28	1	*Pseudomonas* sp. (AJ575816)	99	*γ-Proteobacteria*
3N44	2	*Erwinia rhapontici* (AJ233417)	97	*γ-Proteobacteria*
3N45	1	*Sphingomonas yunnanensis* (AY894691)	97	*α-Proteobacteria*
3N48	2	Unclassified bacterium (AF479376)	99	*γ-Proteobacteria*
3N53	31	*Pantoea agglomerans* (AY849936)	99	*γ-Proteobacteria*
3N65	4	*Rahnella aquatilis* (AY253919)	97	*γ-Proteobacteria*
3N74	1	*Klebsiella pneumoniae* (AF076033)	97	*γ-Proteobacteria*
3N76	2	*Pseudomonas* sp. (AY131221)	99	*γ-Proteobacteria*
3N77	1	*Pseudomonas oleovorans* (D84018)	97	*γ-Proteobacteria*
3N82	2	*Pseudomonas* sp. (AF375845)	99	*γ-Proteobacteria*
3N83	2	*Pantoea agglomerans* (AY924376)	98	*γ-Proteobacteria*
3N97	1	Uncultured bacterium (AY958985)	98	*γ-Proteobacteria*
3N105	1	Unclassified bacterium (AY822535)	99	*γ-Proteobacteria*
3N107	1	Unclassified bacterium (AY770422)	98	*γ-Proteobacteria*
3N108	4	*Pantoea agglomerans* (AY924375)	97	*γ-Proteobacteria*
3N111	1	*Pseudomonas* sp. (AY269867)	99	*γ-Proteobacteria*

*Tentative phylogenetic placement and percent identity values were determined by BLASTN, and approximately 600 bp of the 16S rDNA sequence for each clone was applied for alignment.

bacteria reside in plant tissues mainly in intercellular spaces, rarely in intracellular spaces and inside vascular tissues (Thomas and Graham 1952). In order to release all microbes living in association with plant tissues, it is necessary to fully separate middle lamellas of plant cells, but simultaneously to maintain the bacterial cells intact during separation. A mild and more specific treatment with enzymes hydrolyzing plant cell walls was applied for this purpose. On the other hand, in order to obtain sufficient endophytic bacterial DNAs, relatively large amounts of starting material are required because an extremely large proportion of the total DNA is assumed to originate from the host plant genome. Hence, 40 g of fresh leaves or seeds were ground with liquid nitrogen and treated with enzymes at different concentrations. Though liquid nitrogen may affect bacterial cell integrity, endophytic bacteria were effectively enriched by Method II. This is probably due to a potential protection provided by the host cell structures that completely surround the endophytic bacteria.

Theoretically, the DNA isolated from plant tissues may derive from the plant nuclei, the plastids, the mitochondria and, if any, the plant-associated microbes. The mitochondrial DNA could be ignored in this study because of its very low occurrence in our samples. It was reported that 8000 ***g***-force for 20 min precipitated most plant mitochondria (Nagahashi and Hiraike 1982). Marker genes including the entire ITS (nucleus-specific), the *trnL-trnF* spacer (plastid-specific) and the 16S–23S rRNA gene spacer (bacterium-specific) were used because they

are genome-specific, highly conserved and most frequently present in their respective genomes. It is well known that the nuclear ribosomal RNA genes including the ITS region exist in hundreds or thousands of copies for a eukaryotic genome (Buchanan *et al.* 2000). Ribosomal RNA genes may be present in multiple copies in bacterial genomes (Cole and Girons 1994; Farrelly *et al.* 1995). Though the plastidial *trnL-trnF* spacer is a single copy sequence, hundreds of DNA molecules may be found in a plastid and dozens to hundreds of plastids may reside in a leaf cell. Sometimes plastid DNA may account for up to 10–20% of the leaf total DNA (Buchanan *et al.* 2000). Bacterial 16S–23S rRNA gene spacer was frequently amplified from the plant DNA because of the presence of complex nuclear and plastid DNA. On the one hand, the plant nuclear DNA may serve as template for the amplification of the 16S-23S rRNA gene spacer and *trnL-trnF* spacer, resulting in PCR bands of the expected size. On the other hand, the 16S–23S rRNA gene spacer is inevitably amplified from the plastid DNA. However, the plastidial 16S–23S rRNA gene spacer can be distinguished from its bacterial counterpart by the size of the amplicon. Because of introns (0·5–2·5 kb) present in two *trn* genes, the plastidial 16S-23S rRNA gene spacer is usually more than 1·0 kb in size (Buchanan *et al.* 2000). However, the bacterial 16S–23S rRNA gene spacer often ranges from 0·3 to 0·9 kb (occasionally more than 1·0 kb) (Jensen *et al.* 1993).

DNAs extracted from the pellets of five centrifugation steps were serially diluted and used as templates for marker gene PCR amplification. The highest dilution that worked for PCR amplification of the marker gene was obtained, so there should be more marker DNA copies present if amplified at the highest 1000-fold dilution than for the highest 10-fold dilution. The bacterial 16S–23S rRNA gene spacer was predominantly amplified in the enzyme-treated 1000 ***g***- or 3000 ***g***-pellet, whereas no nuclear ITS products were detectable (Table 1). The nuclear ITS region is specific and has a large number of copies for the plant nuclear DNA. In this respect, one can reasonably assume that in the DNA from enzyme-treated 1000 ***g***- or 3000 ***g***-pellet, amplification of the 16S–23S rRNA gene spacer contributed significantly to the existing bacterial DNA rather than to the plant nuclear DNA. On the other hand, the plant nuclear DNA may serve as a potential template for amplification of the 16S–23S rRNA gene spacer and *trnL-trnF*, resulting in the expected bands. Nearly equal amplification of both nuclear ITS and 16S–23S rRNA gene spacer, or both nuclear ITS and *trnL-trnF*, cannot be interpreted as a uniform abundance of their corresponding genomic DNAs. This is supported by the control pattern in which the three marker genes were almost equally amplified (Table 1). Conceivably, the plant nuclear DNA may constitute the majority of the DNA in the control and in the enzyme-treated 200 ***g***- or 500 ***g***-pellet (Table 1). Given all the observations, we suggest that the step of discarding the enzyme-treated pellet from centrifugation at 200 ***g*** or 500 ***g*** was crucial to enhance the ratio of bacterial to plant nuclear DNA in the 3000 ***g***-pellet. Therefore, enzyme treatment of plant tissues and subsequent differential centrifugation are potentially effective in reducing the plant genomic DNA content and in enriching for plant-associated bacteria.

Enzyme concentration, buffer and incubation time were further optimized for preparation of the enriched samples. Mannitol buffer (0·7 mol l^{-1}) was used in Method II as an osmotic stabilizer in order not to break up the protoplasts, because protoplasts should be more easily separated from bacteria by differential centrifugation. Pellets from centrifugation at 200 ***g*** were removed for the purpose of reducing the plant nuclear DNA content and enhancing the ratio of bacterial to plant nuclear DNA in the subsequent DNA extracts. A repeat suspension of the 200 ***g***-pellet was attempted to increase the bacterial yield. In addition, L-lysine and EGTA were used as DNase inhibitors in the DNA extraction procedure (Peterson *et al.* 1997) to avoid degradation of the bacterial DNA during lysozyme incubation. As shown in Table 3, there was no overlap in the phylogenetic affiliation of Method I clones and Method II clones. From each sample only a limited number of clones (~100 clones) were applied for RFLP typing, thereby not representing the whole biodiversity of the current 16S rDNA library. Moreover, as only partial 16S rDNA sequences (~0.6 kb) were used, this phylogenetic placement is tentative.

Cultivation-dependent approaches have been widely applied for the analysis of species diversity of endophytic bacteria. Most bacterial species isolated from internal plant tissues belonged to *Pseudomonas*, *Bacillus*, *Enterobacter* and *Agrobacterium* (Hallmann *et al.* 1997). Recently, a broad phylogenetic spectrum of endophytic bacteria was detected from potato and other plants by a 16S rRNA-based cultivation-independent method (Sessitsch *et al.* 2002; Dent *et al.* 2004). Despite an increasing number of reports on the diversity of bacterial endophytes in plants, few studies have focused on metagenomic analysis of these microbes. Metagenomic analysis involves DNA isolation from an environmental sample, library construction and functional analysis (Handelsman 2004). Efficiently extracting bacterial DNA from host plant tissues is central to constructing a metagenomic library that harbours sufficient copies of interesting genomes from plant-associated bacteria. As revealed in this study, Method II for microbe enrichment in plant tissues was remarkably effective in enhancing the ratio of bacterial to plant nuclear DNA and increasing the species diversity of

plant-associated bacteria. The strategy described here can be applied to studying plant-associated bacteria in the field of environmental molecular ecology and environmental metagenomics.

Acknowledgements

We thank Professor David A. Hopwood, John Innes Centre, for critical reading of the manuscript, and Tao Yang for his helpful discussion. Anonymous reviewers are gratefully acknowledged for their comments. This work was supported by the National Natural Science Foundation of China (30430020) and the Chinese Academy of Sciences (KSCX2-SW-313).

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Bauer, W. (1981) Infection of legumes by Rhizobia. *Annu Rev Plant Physiol Plant Mol Biol* **32**, 407–449.

Buchanan, B.B., Gruissem, W. and Jones, R.L. (2000) *Biochemistry & Molecular Biology of Plants*. Rockville, MD: The American Society of Plant Biologists.

Cassady, J.M., Chan, K.K., Floss, H.G. and Leistner, E. (2004) Recent developments in the maytansinoid antitumor agents. *Chem Pharm Bull* **52**, 1–26.

Chowdhury, S.P., Khanna, S., Verma, S.C. and Tripathi, A.K. (2004) Molecular diversity of tannic acid degrading bacteria isolated from tannery soil. *J Appl Microbiol* **97**, 1210–1219.

Cole, S.T. and Girons, I.S. (1994) Bacterial genomics. *FEMS Microbiol Rev* **14**, 139–160.

Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P. and Simonet, P. (2001) Quantification of bacterial subgroups in soils: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ Microbiol* **3**, 431–439.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J. *et al.* (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* **69**, 49–55.

Dent, K.C., Stephen, J.R. and Finch-Savage, W.E. (2004) Molecular profiling of microbial communities associated with seeds of *Bera vulgaris* subsp. *Vulgaris* (sugar beet). *J Microbiol Methods* **56**, 17–26.

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.

Hallmann, J., Quadt-Hallmann, A., Mahaffee, W.F. and Kloepper, J.W. (1997) Bacterial endophytes in agricultural crops. *Can J Microbiol* **43**, 895–914.

Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol* **68**, 669–685.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**, R245–R249.

Jensen, M.A., Webster, J.A. and Straus, N. (1993) Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl Environ Microbiol* **59**, 945–952.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.I. and Pace, N.R. (1985) Rapid determination of 16S rRNA sequences for phylogenetic analysis. *Proc Natl Acad Sci USA* **82**, 6955–6959.

Li, H.W., Luo, J.H., Hemphill, J.K., Wang, J.J. and Gould, J.H. (2001) A rapid and high yielding DNA miniprep for cotton (*Gossypium* sp.). *Plant Mol Biol Rep* **19**, 183a–183e.

Miller, M.I., Scott, A. and Gardner, I.C. (1983) Leaf nodule development in *Psychotria kirkii* Hiern (Rubiaceae). *Ann Bot* **52**, 791–802.

Moyer, C.F., Dobbs, F.C. and Karl, D.M. (1994) Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi seamount, Hawaii. *Appl Environ Microbiol* **60**, 871–879.

Nagahashi, J. and Hiraike, K. (1982) Effects of centrifugal force and centrifugation time on the sedimentation of plant organelles. *Plant Physiol* **69**, 546–548.

Pace, N.R. (1997) A molecular view of microbial diversity and biosphere. *Science* **276**, 734–740.

Peterson, D.G., Boehm, K.S. and Stack, S.M. (1997) Isolation of milligram quantities of nuclear DNA from tomato (*Lycopersicon esculentum*), a plant containing high levels of polyphenolic compounds. *Plant Mol Biol Rep* **15**, 148–153.

Piel, J. (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc Natl Acad Sci USA* **99**, 14002–14007.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N. and Matsunaga, S. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proc Natl Acad Sci USA* **101**, 16222–16227.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A. *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**, 2541–2547.

Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14**, 303–310.

Sessitsch, A., Reiter, B., Pfeifer, U. and Wilhelm, E. (2002) Cultivation-independent population analysis of bacterial endophytes in three potato varieties based on eubacterial and *Actinomycetes*-specific PCR of 16S rRNA genes. *FEMS Microbiology Ecology* **39**, 23–32.

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

Taberlet, P., Gielly, L., Pautou, G. and Bouvet, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol* **17**, 1105–1109.

Thomas, W.D. and Graham, R.W. (1952) Bacteria in apparently healthy pinto beans. *Phytopathology* **42**, 214.

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.

White, T.J., Bruns, T.D., Lee, S. and Taylor, J.W. (1990) Amplification and direct sequencing of fungal ribosomal DNA for phylogenetics. In *PCR Protocols-A Guide to Methods and Applications* eds Innis, M.A., Gelfund, D.H., Sninsky, J.J. and Whites, T.J. pp. 315–322. San Diego, CA: Academic Press.

Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.Q., Xu, J., Moss, S.J. *et al.* (2002) The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. *Proc Natl Acad Sci USA* **99**, 7968–7973.