Environmental Microbiology (2008) **10**(10), 2684–2691 doi:10.1111/j.1462-2920.2008.01689.x

Enriching plant microbiota for a metagenomic library construction

Hao-Xin Wang,[1,2] Zhao-Liang Geng,[1,2] Ying Zeng[1]* and Yue-Mao Shen[1*†]

[1]*State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, the Chinese Academy of Sciences, 132 Lanhei Rd, Kunming 650204, Yunnan, China.*

[2]*Graduate School of the Chinese Academy of Sciences, Beijing 100039, China.*

Summary

Plant microbiota (the microorganisms that live in any associations with plant tissues) represents a rather unexplored area of metagenomic research compared with soils and oceans. Constructing a metagenomic library for plant microbiota is technically challenging. Using all the biomass without pre-enrichment could lead to vast proportions of the host plant DNA in the metagenomic library, doubtless obliterating the microbial contribution. Therefore, the first and essential step is to enrich for the constituent microorganisms from plant tissues. Here, a strong enrichment for plant microbiota was achieved by coupling SDS (sodium dodecyl sulfate) with NaCl, creating a predominantly microbial metagenomic library that contains 88% bacterial inserts. 16S rDNA sequence analysis revealed that the metagenomic DNA of enrichments originates from very diverse microorganisms. At least 74 distinct ribotypes (at a 97% threshold) from seven different bacterial phyla were identified and mainly distributed among *Actinobacteria* and *Proteobacteria*. Additionally, a simplified version of Amplified Ribosomal DNA Restriction Analysis (ARDRA) was developed for a quick and efficient assessment of the enriching procedures. This work opens further insight into the great biotechnical potential of plant microbiota, holding more potential for drug discovery through a metagenomic strategy, and paving the way for recovery and biochemical characterization of functional gene repertoire from plant microbiota.

Received 29 January, 2008; accepted 17 May, 2008. *For correspondence. E-mail biochem@mail.kib.ac.cn; Tel. (+86) 871 5223210; Fax (+86) 871 5150227. †Present address: School of Life Sciences, Xiamen University, No. 422 South Siming Road, Xiamen, Fujian 361005, China.

Introduction

Recent research in molecular ecology has shown that the vast majority of microorganisms in a natural community is refractory to cultivation and remains unexplored (Ward *et al*., 1990; Pace, 1997). The as yet uncultured microorganisms hold an unlimited and intriguing potential for development of novel genes and biocatalysts for use in biotechnology. Metagenomics is shown to be one of the principal strategies applied to access and investigate this potential (Handelsman, 2004; Streit *et al*., 2004). To date, metagenomics studies have been conducted for a variety of common environments (e.g. soil, water, etc.) and extreme environments (e.g. glaciers, geysers, etc.) as well. Other interesting microbial habitats like marine sponges, insects, and the rumen of ruminants or even the human body have also explored in metagenomics (Piel *et al*., 2004; Ferrer *et al*., 2005; Schirmer *et al*., 2005; Tringe *et al*., 2005; Turnbaugh *et al*., 2007).

Representing a unique kind of environmental niches, plants serve as complex habitats for colonization by different kinds of microbes. In a few plants, microbes are found inhabiting particular sites such as nodules and leaf galls. For most plants, however, little is known about the diversity and colonization sites of microbes. The microbial biomass may account for only a minute part of all biomass in the host plant. The microorganisms that live in any associations with plant tissues (known as plant microbiota) have not been investigated for metagenomics as yet. This is largely due to a bottleneck generated by the methodology used for microbe enrichment from a host plant. Constructing a metagenomic library for plant microbiota is technically challenging. Without prior enrichment procedures, the metagenomic library could contain an extremely high proportion of plant-derived DNA, doubtless obliterating the microbial contribution and leading to a 'waste' library. Therefore, the first and essential step is to enrich for the plant microbiota from plant tissues.

The tropic tree *Mallotus nudiflorus* (also known as *Trewia nudiflora*) was selected for this 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

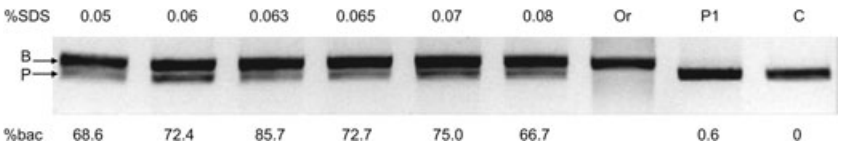

Fig. 1. Effects of 0.9% NaCl plus SDS on microbe enrichment as detected by ARDRA. Numbers on the top line refer to the concentration (%) of SDS used in the preparation. In the agarose gel the upper band represents the undigested 16S rDNA (~1.5 kb) potentially affiliated with bacteria (B-arrow), while the lower band harbouring sequences of about 1.3 kb corresponds to the digested 16S rDNA of plastid origin (P-arrow). '%bac' is the percentage of 16S rDNA clones potentially affiliated with bacteria. 'C' is the control and 'Or' is the original 16S rDNA prior to PvuII digestion.

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 from *M. nudiflorus* stem barks, and then construct a metagenomic library from the microbial enrichment to screen genes potentially involved in the biosynthesis of maytansinoids or other polyketides. In our previous study, the constituent microorganisms were enriched by the enzymatic hydrolysis and subsequent differential centrifugation (Jiao *et al*., 2006). Whereas this is a fairly crude approach, it increased the representation of prokaryotes in the enrichment from plant leaves or seeds. Our further efforts were focused on how to specifically enrich for plant microbiota: that is, to remain as close as possible to its native community structure while eliminating the plant DNA. Here, the whole new approach created a predominantly microbial metagenomic library that contains 88% bacterial inserts.

Results

A strong enrichment for plant microbiota by coupling SDS (sodium dodecyl sulfate) with NaCl

After release of constituent microorganisms by homogenization, centrifugation at 200 *g* removed the majority of cell debris and nuclei derived from *M. nudiflorus* stem barks. The deep-green pellet P1 collected by the subsequent centrifugation at 5000 *g* was highly abundant in plant DNA especially the plastid DNA, as evidenced by a complete PvuII digestion of the amplified 16S ribosomal DNA (rDNA) (Fig. 1). Extensive research on disruption of plastids by detergents revealed that 0.1% SDS (sodium dodecyl sulfate) gave rise to a pale grey pellet P2, indicating a dominant lysis of plastids. Lower concentrations of SDS were further investigated with respect to microbial cell integrity. While SDS at concentration of 0.01% hardly broke plastids, 0.05% and 0.08% SDS produced a light green and a pale grey P2 respectively (Fig. S1). Analyses of the PCR-amplified 16S rDNA library revealed only 1.4% bacterium-affiliated clones in the P2 derived from the preparation of 0.08% SDS, indicating a relatively high proportion of plastid ribosomal sequences in the P2. Considering that plastid DNA is usually attached to the plastid membrane (Sato *et al*., 1993), it is conceivable that DNA-containing plastid debris was collected in the P2 along with microbial cells by centrifugation at 5000 *g*. Thus, the next task was to separate plastids and their debris from microbial cells prior to the final centrifugation.

Salts are widely used to precipitate plant polysaccharides and cell debris. We found that NaCl or $CaCl_2$ at concentrations ranging from 0.5% to 3.0% caused a distinct precipitation in P1 suspension (Fig. S2). The precipitate was discarded and the resulting P2 was applied for 16S rDNA-based analyses. About 12% bacterium-affiliated clones were detected in the P2 derived from 0.9% NaCl or 0.5% $CaCl_2$ preparations. Nevertheless, less than 5% bacterial clones were observed for other concentrations (data not shown). Attempts to aggregating plant cell debris by salts did increase the proportion of bacterial clones in the 16S rDNA library of the P2 when compared with P1 (0.6%, Fig. 1), although only marginally. To some extent, the results suggested a favourable effect of NaCl or $CaCl_2$ on enriching for the plant-associated microbes.

Based on the results obtained from the experiments described above, enrichments were performed for 0.9% NaCl plus SDS varying from 0.05% to 0.08% according to the final protocol presented in *Experimental procedures*. 16S rRNA genes were amplified by PCR with DNA extracted from each pellet P, and then digested with PvuII or applied for a 16S rDNA library construction. As shown in Fig. 1, coupling SDS with NaCl conferred a much higher ratio of bacterial to plastidial DNA in the enriched samples. This is supported by the relative intensity between the undigested (B-arrow) and digested (P-arrow) bands of the corresponding 16S rDNA by PvuII, as the *M. nudiflorus* plastid rDNA (DQ000201) is sensitive to PvuII digestion (Jiao *et al*., 2006). In addition, the majority (66–85%) of 16S rDNA clones were PvuII-insensitive, potentially affiliated with bacteria. On the contrary, the amplified 16S rDNA of both the control and the direct homogenization (P1) was almost completely digested by PvuII because of its plastidial origin, and very few clones were detected to be bacterial origin (Fig. 1). Enrichments were repeatedly performed for 0.9% NaCl plus 0.063% SDS (Fig. S3) and defined as our final protocol.

Table 1. Microbe enrichment assessed by 16S rDNA sequence analysis.

Sources	% bac[a]	ARDRA types	Phyla or phylum
The final enrichment	96.8	113	*Actinobacteria*, *Proteobacteria*, *Firmicutes*, *Gemmatimonadetes*, *Bacteroidetes*, *Planctomycetes*, *Deinococcus-Thermus*
Enzyme-based	70.1	26	*Proteobacteria*
P1[b]	0.6	–	–
Control	0	–	–

a. The percentage of 16S rDNA clones potentially affiliated with bacteria.
b. P1 refers to the pellet derived from direct homogenization as described in *Experimental procedures*.
–, not applicable.

Totally 0.4 g of the enriched pellet was obtained from 1.6 kg of fresh stem barks using the final protocol described in *Experimental procedures*. Scanning electron microscopic and fluorescent microscopic analyses, including Acridine Orange staining, found the organisms in the enrichment to be primarily prokaryotic (Fig. S4). To verify if our method is effective for microbe enrichment from other plants, we randomly collected stem barks of four different trees grown in Kunming Botanical Garden. Following the final protocol, NaCl plus SDS resulted in a moderate enrichment in two of those plants as revealed by PvuII-digested patterns of the respective 16S rDNA (Fig. S5).

A considerably enhanced microbial diversity in the enrichment

The microbial diversity in the final enrichments for a metagenomic study was analysed through PCR-amplified 16S rDNA libraries generated by using primers specific for *Bacteria*. Phylogenetic diversity of the plant microbiota was surveyed from more than 180 eubacterial 16S rDNA clones. As shown in Table 1, more than 95% of the 16S rDNA clones were potentially of bacterial origin and 113 Amplified Ribosomal DNA Restriction Analysis (ARDRA) types were detected in the enrichment-derived 16S rDNA library. The metagenomic DNA of final enrichments was demonstrated to originate from very diverse microorganisms. At least 74 distinct ribotypes (at a 97% threshold) from seven different bacterial phyla were identified, and mainly distributed among *Actinobacteria* and *Proteobacteria* (Fig. 2). A few clones were affiliated with uncultured or unclassified bacteria. In addition, half of sequences were singlets, and the most common ribotype accounts for 4.7% of the clones. Based on the sequence analysis (Table S1), 61.3% of the 16S rDNA has a similarity below 97% with the best hits in database, indicating that most of the microbes in *M. nudiflorus* microbiota might be previously unknown phylotypes.

To address how representative is the metagenomic library of the actual microbial community on the stem barks, rarefaction curves were built by DOTUR (Schloss and Handelsman, 2005). As shown in Fig. 3, rarefaction curves failed to reach saturation when using sequence identity of 97% (species level) or 95% (genus level) as cut-off. However, the curve of 80% cut-off (phylum level) was more flat. Results demonstrated that the sampling was sufficient in phylum level and insufficient in species and genus level. Coverage estimators such as Chao1 predicted the total number of bacterial OTUs (operational taxonomic units) in this sample to be more than 169 when a 97% identity cut-off is used and more than 195 when a 98% identity cut-off is used. Besides the Chao1 might underestimate true richness at low sample sizes (Hughes *et al*., 2001), the 16S rDNA amplification might also decrease the OTUs observed because not all of rRNA genes can be amplified with the same 'universal' primers. Thus the 'true' richness of plant microbiota is believed far more than the OTUs we sampled and predicted by Chao1 estimator. The entire list of 16S rRNA gene analysis by BLASTN can be found in Tables S1 and S2. On the other hand, 26 unique ARDRA types mapping primarily to only one phylum (*Proteobacteria*) were observed for our previous enzyme-based Method II (Table 1). As compared with the control and the direct homogenization (P1), the final enrichment showed a considerably enhanced proportion of bacterium-derived clones and a much wider species and phyla diversity of those clones.

The metagenomic library dominated by prokaryotic inserts

A fosmid library for the plant microbiota harbouring entirely 1.37×10^6 clones was generated from the final enrichments. Based on the restriction digest analysis, the average insert size of the library was examined to be 34.5 kb (Fig. S6; Table S3). The overall insert size of the library was therefore roughly analogous to 10 000 copies of *Escherichia coli* genomes or 5000 copies of *Streptomyces* genomes. In this respect, the metagenomic library contained an estimated coverage of more than 1000 genome types when a uniform phylogenetic distribution is hypothesized in the community.

The end sequences (~700 bp) of 187 of random fosmids were available to assess the range of prokaryotic DNA inserts in the metagenomic library. As expected, 166 (88.8%) of the clones were determined to be prokaryotic at *E*-value of $\leq 10^{-4}$. A few of clones were considered not assignable (9.6%) or *Eukaryota* (1.6%). The results of BLASTX were provided in Table S4. The putative proteins encoded by end sequences were subjected to assign into

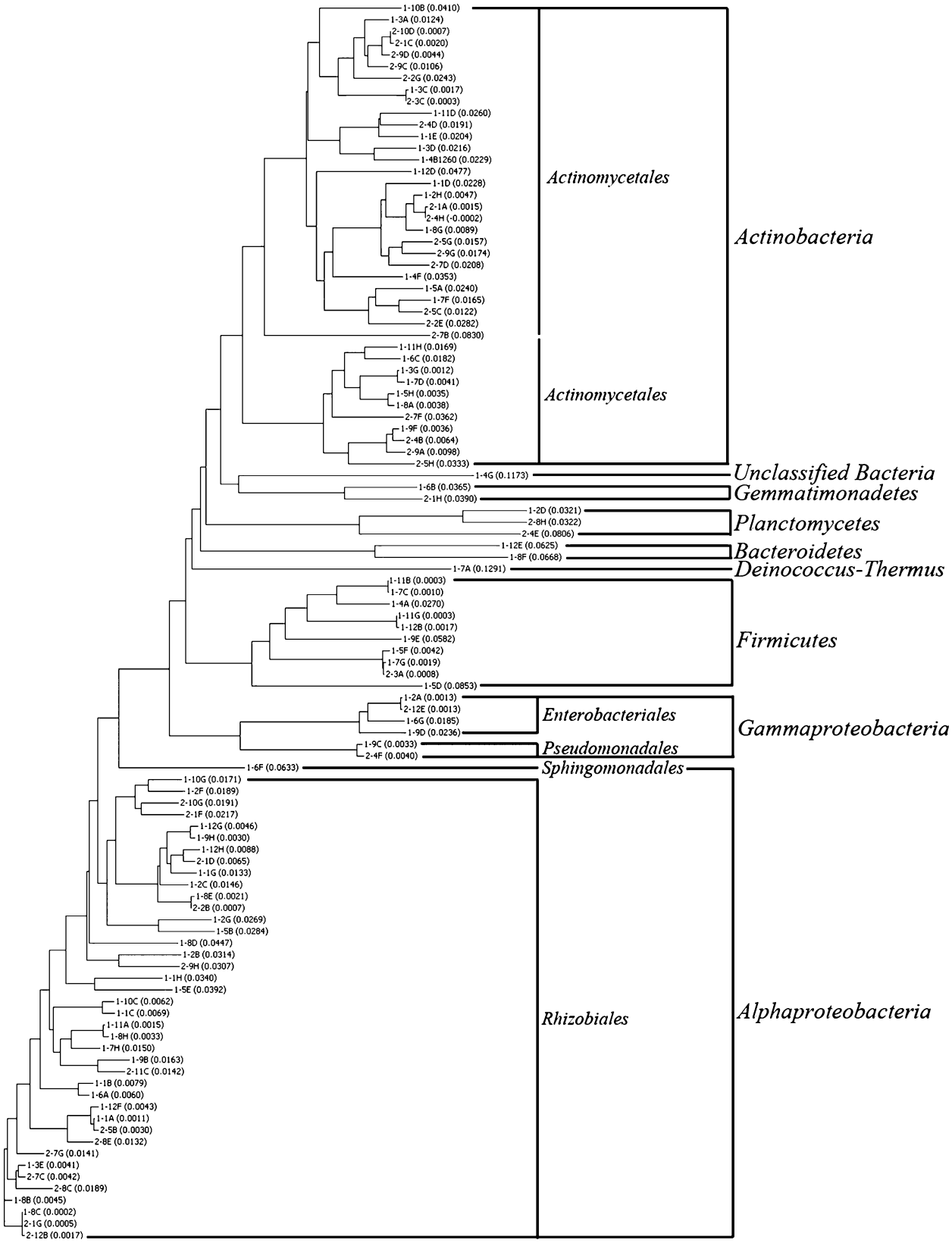

Fig. 2. A neighbour-joining phylogenetic tree rendered with AlignX in Vector NTI Suite. 16S rDNA sequence analysis revealed that the metagenomic DNA originates from very diverse microorganisms.

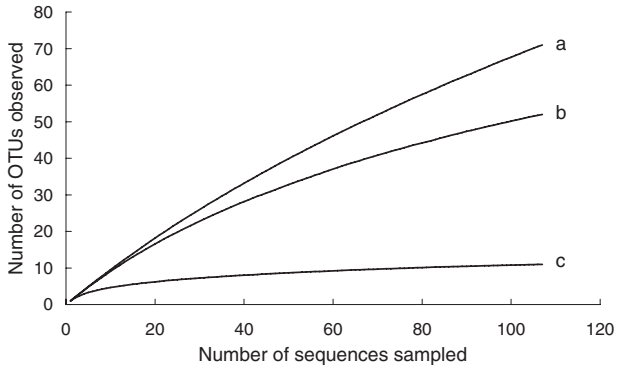

Fig. 3. Rarefaction curves generated by DOTUR at different levels of sequence identity. Cut-off of 97% identity (a) is typically used to the same species, 95% cut-off (b) to the same genus and 80% cut-off (c) to the same phylum.

specific functional roles using COGnitor (Tatusov *et al*., 2001). More than half (68%) of the predicted proteins were assigned into certain COGs (clusters of orthologous groups of proteins) covering all the 18 functional categories in COG database (Fig. S7). Estimated from end sequences, the metagenomic library was very rich in genomes of high GC content. Most of inserts (72%) were observed for GC content to reach beyond the 70% level (up to 82.3%). The result is consistent with a dominant distribution in the library by *Actinobacteria* and α-*Proteobacteria* as revealed by 16S rRNA gene analyses. Additionally, both fosmid end and 16S rDNA sequence analyses seemed to achieve similar patterns of phylogenetic diversity contained in the metagenomic library and in its source DNA (Table 2). The most abundant groups are *Proteobacteria*, *Actinobacteria*, *Firmicutes* (low-GC Gram-positive), accounting for more than 90% OTUs (Fig. S8). It is very encouraging that the metagenomic library contained a large proportion of prokaryotic inserts and broad microbial diversity, which reinforces the feasibility of our enriching procedures for plant microbiota.

The metagenomic clones have been screened and characterized to allow evaluation of its suitability for ectopic expression. In an extraordinary effort we finally found a candidate fosmid harbouring the target genes out of approximately 75 000 clones (unpublished data). The fosmid has completely been sequenced and used for further screening the whole gene cluster from the library.

Discussion

Plant microbiota represents a rather unexplored area of metagenomic research, in contrast to soils and oceans. Applying metagenomics to access and probe for its genetic and biotechnical potential requires a metagenomic library construction first. To achieve this there is the need to enrich for the constituent microorganisms from large quantities of plan tissues, given an extremely small biomass of the plant microbiota. The enzyme-based procedure described in our previous study might easily bias the original community structure owing to microbial propagation during the 12 h incubation at 28°C (Jiao *et al*., 2006). To address the problem we attempt to develop a more specific approach. In the present work, kilogram quantities of stem barks were thoroughly homogenized in water. The resulting mixture was centrifuged at low speed and the supernatant was collected. To the supernatant, the salt and detergent were added to further remove plant DNA especially the plastid DNA. The principle underlying this type of enrichment is probably that the plastids lysed by SDS co-precipitate with plant cell debris during the aggregation stimulated by NaCl or $CaCl_2$. Incubation at 4°C for 1 h appears to be most suitable for a clear precipitation; otherwise longer incubation time might reduce the microbial yield in the enrichment presumably by increasing the likelihood of the occurrence of microbe sedimentation and/or lysis by SDS.

In this article we reported a simplified version of ARDRA for a fast and efficient evaluation of microbe-enriching procedures. The ratio of bacterial to plastidial DNA in the sample was quickly estimated from the relative intensity between the undigested and digested bands of the corresponding 16S rDNA by PvuII. Because most eubacterial 16S rDNA are devoid of the recognition site for PvuII, after digestion they retain the size equal to that of 16S rDNA prior to digestion and form the undigested band on elec-

Table 2. Microbial diversity in the metagenomic library and its source DNA.

Phylum or sub-phylum	End sequence survey[a]	16S rDNA survey[b]
Proteobacteria		
α-*Proteobacteria*	26.5% (44)	38.7% (41)
β-*Proteobacteria*	3.0% (5)	ND
γ-*Proteobacteria*	6.0% (10)	5.7% (6)
δ-*Proteobacteria*	2.4% (4)	ND
Actinobacteria	56.0% (93)	37.7% (40)
Firmicutes	2.4% (4)	9.4% (10)
Deinococcus-Thermus	1.2% (2)	0.9% (1)
Chloroflexi	0.6% (1)	ND
Planctomycetes	0.6% (1)	2.8% (3)
Cyanobacteria	0.6% (1)	ND
Acidobacteria	0.6% (1)	ND
Gemmatimonadetes	ND	0.9% (1)
Bacteroidetes	ND	1.9% (2)
Unclassified *Bacteria*	ND	1.9% (2)

a. End sequence survey in the metagenomic library was carried out from 166 of fosmids harbouring prokaryotic DNA inserts based on BLASTX analyses.
b. 16S rDNA survey in the metagenomic DNA prior to library construction was performed from 106 of rDNA clones with available sequences. The figures in parentheses refer to the number of clones belonging to a unique group.
ND, not detectable.

trophoresis (Fig. 1, B-arrow). However, the *M. nudiflorus* plastid rDNA is sensitive to PvuII (Jiao *et al*., 2006) and produces a sequence slightly shorter than the original one after digestion. The digested band (Fig. 1, P-arrow) representing plastid rDNA can be expected in the gel to be very close to the undigested band. Given that the PCR-amplified rDNA yield is theoretically proportional to the copies of starting DNA in the PCR reaction (Keohavong *et al*., 1988), the relative intensity between the undigested and digested bands probably reflects the ratio of bacterial to plastidial DNA in the sample.

For each new plant species being study, the enrichment procedure for plant microbiota has to be adjusted and optimized. $CaCl_2$ proved to be an alternative when NaCl failed to precipitate the suspension from stem barks of *Eucommia ulmoides* and *Vernicia montana* (Fig. S5). Delicate calculations on SDS concentration and further refinement of the method need to be performed as well for microbe enrichment more efficiently from diverse plant resources. By the enriching method presented here, the ratio of bacterial to plastidial DNA was considerably enhanced and a much wider species and phyla diversity of 16S rDNA clones was observed in the final enrichment. Furthermore, the metagenomic library contained a large proportion of prokaryotic inserts and broad microbial diversity. The method can be applied to study biology and chemical biology relevant to the uncultivable microorganisms associated with plants, holding more potential for drug discovery through a metagenomic strategy. Thus, this work opens further insight into the great biotechnical potential of plant microbiota, paving the way for recovery and biochemical characterization of its functional gene repertoire.

Experimental procedures

Homogenization and separation

Stem barks of *M. nudiflorus* (L.) Kulju & Welzen (*Euphorbiaceae*), grown in a tropical rain forest, were collected in Xishuangbanna Botanical Garden (Yunnan, China). The fresh barks in kilogram quantities were washed with tap water, cut into small pieces and thoroughly homogenized in MilliQ water. The homogenate was filtered through two layers of bandage and centrifuged at 200 *g* for 5 min at 4°C. The supernatant was subsequently centrifuged at 5000 *g* for 10 min (4°C) to collect the pellet (termed P1). Each aliquot of P1 was suspended in MilliQ water for extensive studies to get rid of as many plant organelles as possible. Reagents including 0.1% SDS, 0.1% Tween80, 0.1% Triton X-100, NaCl (at a final concentration of 0.1%, 0.5%, 0.9%, 1.5%, 3.0%), $CaCl_2$ (at a final concentration of 0.5%, 3.0%) were independently investigated. The P1 suspension treated with a reagent, after standing at 4°C for 0.5 h or 1 h, was centrifuged at 200 *g* for 5 min (4°C). The pellet was discarded and the supernatant was further centrifuged at 5000 *g* for 10 min at 4°C to collect the pellet (termed P2).

Final protocol for microbe enrichment

Fresh stem barks of 40 g were chopped and homogenized (18 000 r.p.m. $\times$ 1 min $\times$ 4) in 300 ml of MilliQ water. The suspension was filtered through two layers of bandage and centrifuged at 200 *g* for 5 min at 4°C. To the supernatant of 260 ml add NaCl of 2.34 g (at a final concentration of 0.9%) and 1.64 ml of 10% SDS (at a final concentration of 0.063%). Mix gently and incubate at 4°C for 1 h to expect an automatic precipitation. The upper phase was carefully transferred to a clean bottle while not disturbing the precipitate. The bottle of the upper phase was centrifuged at 5000 *g* for 10 min (4°C) to collect the pellet. The supernatant was decanted and the pellet was re-suspended in 400 ml of MilliQ water. NaCl and SDS were added in the same way as described above. Incubate at 4°C for another 1 h and then the pellet was collected by centrifugation at 5000 *g* for 10 min at 4°C. This pellet (termed P) was highly enriched for the constituent microorganisms. Totally 1.6 kg of fresh stem barks generated 0.4 g of the final enrichments.

In comparison, the enzyme-based enrichment was performed from 40 g of *M. nudiflorus* stem barks exactly according to our previous Method II (Jiao *et al*., 2006). For the control, DNA was extracted directly from 2.0 g of the stem barks. Meanwhile, stem barks of four different trees grown in Kunming Botanical Garden were collected to verify if our method is effective for microbe enrichment from other plants. The trees include *Daphniphyllum glaucescens* Bl. (*Daphniphyllaceae*), *V. montana* Lour. (*Euphorbiaceae*), *Liquidamba formosana* Hance (*Hamamelidaceae*), *E. ulmoides* Oliv. (*Eucomiaceae*).

Assessing the enriching procedures by 16S rDNA-based techniques

DNA was extracted from different pellets as described in the section below. For ARDRA of 16S rRNA sequences, 27F and 1492R primer pair was used to target bacterial genomes (Lane *et al*., 1985). The efficiency of microbe enrichment was deduced from the percentage of clones affiliated with bacteria and from the diversity of their ARDRA patterns. Aside from ARDRA exactly done as described in our previous work (Jiao *et al*., 2006), a simplified version of ARDRA was developed in this study for a fast and efficient assessment. Briefly, the PCR-amplified 16S rDNA was first digested by PvuII and then separated in 1% agarose gels. Two largest bands close together can be expected in the gel. The undigested band harbouring sequences of about 1.5 kb is composed completely of bacterial 16S rDNA. Slightly shorter sequences (~1.3 kb) usually derived from plastid rDNA form the digested band. The ratio of bacterial to plastidial DNA in the sample was quickly estimated from the relative intensity between the undigested and digested bands. All the 16S rDNA sequences were checked for chimera by the program Chimera Check (Cole *et al*., 2003) at the Ribosomal Database Project II, and no obvious chimeras were identified. All 16S rDNA sequences were phylogenetically classified by BLASTN (Altschul *et al*., 1990) from the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) and RDP Classifier (http://rdp.cme.msu.edu/classifier). A neighbour-joining phylogenetic tree was rendered with AlignX in Vector NTI Suite.

Metagenomic library construction and evaluation

Each 30 mg pellet was dispersed in 1 ml of L buffer (Kieser *et al.*, 2000) and treated with lysozyme (2 mg ml^{-1}) plus Achromopeptidase (0.5 mg ml^{-1}) for 1 h at 37°C. The pellet was collected by centrifugation at 12 000 *g* for 10 min and then re-suspended in 1 ml of proteinase K (0.5 mg ml^{-1}) solution [500 mM Tris (pH 8.0), 10 mM NaCl, 20 mM EDTA, 1% SDS] for incubation of 6 h at 55°C. The lysate was extracted twice with phenol-chloroform, precipitated with isopropanol and treated with DNase-free RNase A (10 mg ml^{-1}). The DNA was further purified by 0.7 M NaCl/1% CTAB.

The final enrichments from stem barks of 1.6 kg yielded a total of 20 μg of metagenomic DNA. The microbial diversity in the metagenomic DNA was analysed through 16S rDNA-based techniques. From the same source DNA of 16 μg, a fosmid library with entirely 1.37×10^6 clones was generated using a copy control fosmid library production kit (vector pCC1FOS) of Epicentre (Madison, WI), according to the manufacturer's instruction. Average insert size was examined by restriction digest analysis. Fifteen random fosmids were digested by BamHI, PvuII or XhoI respectively. To assess the range of prokaryotic DNA inserts in the metagenomic library, the ends (~700 bp) of about 200 random fosmids were sequenced at Invitrogen Biotechnology (Shanghai, China). All sequences were searched against the NCBI non-redundant databases.

Nucleotide sequence accession numbers

16S rRNA sequences from *M. nudiflorus* microbiota have been submitted to the GenBank databases under Accession No. EU289415–EU289520.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (30430020).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

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.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., *et al.* (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* **31:** 442–443.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Martins Dos Santos, V.A.P., *et al.* (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* **7:** 1996–2010.

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

Hughes, J.B., Hellmann, J.J., Ricketts, T.H., and Bohannan, B.J. (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67:** 4399–4406.

Jiao, J.Y., Wang, H.X., Zeng, Y., and Shen, Y.M. (2006) Enrichment for microbes living in association with plant tissues. *J Appl Microbiol* **100:** 830–837.

Keohavong, P., Wang, C.C., Cha, R.S., and Thilly, W.G. (1988) Enzymatic amplification and characterization of large DNA fragments of genomic DNA. *Gene* **71:** 211–216.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. Norwich, UK: John Innes Foundation Press.

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.

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

Piel, J., Hui, D., Fusetani, N., and Matsunaga, S. (2004) Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ Microbiol* **6:** 921–927.

Sato, N., Albrieux, C., Joyard, J., Douce, R., and Kuroiwa, T. (1993) Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids. *EMBO J* **12:** 555–561.

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E.F., and Hutchinson, C.R. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissolute*. *Appl Environ Microbiol* **71:** 4840–4849.

Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71:** 1501–1506.

Streit, W.R., Daniel, R., and Jaeger, K.E. (2004) Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Curr Opin Biotechnol* **15:** 285–290.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22–28.

Tringe, S.G., Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., *et al.* (2005) Comparative metagenomics of microbial communities. *Science* **308:** 554–557.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007) The human microbiome project. *Nature* **449:** 804–810.

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.

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

Supplementary material

The following supplementary material is available for this article online:

Fig. S1. Pellets showing green to pale grey in colour by SDS lysis. P1 is the pellet derived from direct homogenization. Triton and Tween refer to Triton X-100 and Tween80 respectively.

Fig. S2. NaCl or $CaCl_2$ caused a distinct precipitation in P1 suspension incubated at 4°C for 1 h.

Fig. S3. Enrichments repeatedly performed for 0.9% NaCl plus 0.063% SDS. DNA extracted from each P was applied to a quick ARDRA assessment by PvuII digestion. Digests were resolved in 1% agarose gel and visualized by ethidium bromide staining. 'B-arrow', bacterial rDNA; 'P-arrow', plastid rDNA.

Fig. S4. Fluorescent (Acridine Orange staining) microscopic and scanning electron microscopic images of the enrichment. Bar = 10 μm.

Fig. S5. PvuII-digested patterns from enrichments of four different plant species. NaCl was found to fail in precipitating the suspension from stem barks of *Vernicia montana* and *Eucommia ulmoides*. Instead, precipitation can be triggered on $CaCl_2$ treatment. Further modifications are needed for microbe enrichment more efficiently from those plants. 'C' is the control and 'O' is the original 16S rDNA prior to PvuII digestion.

Fig. S6. The restriction patterns of fosmids. M, DNA marker DL15000+2000 (Takara).

Fig. S7. COGs of putative proteins encoded by fosmid ends rendered with COGnitor program. The Capitals represent functional categories, and the percentage following is the relative proportion of this category.

Fig. S8. Microbial diversity patterns in the metagenomic library and its source DNA.

Table S1. Sequence analysis of 16S rDNA clones derived from the metagenomic DNA of microbe enrichment from *M. nudiflorus* stem barks.

Table S2. Ribotypes grouping with a 97% identity threshold.

Table S3. Restriction digest analysis of fosmids. The average insert size of the metagenomic library was 34.5 kb due to the vector of 8 kb.

Table S4. BLASTX and BLASTN of end sequences from random fosmids in the metagenomic library for the plant microbiota.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

