Isolation of High Molecular Weight DNA from Forest Topsoil for Metagenomic Analysis

Mei-Fong Pang, Noorlidah Abdullah, Choon-Weng Lee and Ching-Ching Ng*

Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Received 21 Jan 2008 / Accepted 2 May 2008

Abstract. In this study, we employed a modified method to extract DNA from forest topsoil that was suitable for construction of large insert soil metagenomic library. The DNA extraction method used produced considerable DNA yield with DNA fragments ranging from 48 kb up to 290 kb. The recovery of soil DNA suitable for PCR and metagenomic library construction is difficult because soil DNA is often co-purified with polyphenolics and contaminants that interfere with the downstream applications. PCR amplification of 16S and fungal 18S SSU rRNA genes from the extracted soil DNA suggesting that the DNA isolated using this modified method contained low concentration of PCR inhibitory substances and had sufficient purity for PCR without the need of further purification. Sequence analysis of PCR amplicons revealed this extraction method can efficiently capture a wide range of microorganisms including the hard-to-lyse Gram-positive bacteria and fungi. We have also successfully constructed a metagenomic fosmid library with insert size of between 23.1 kb – 40 kb. This metagenomic library will serve as basis for screening of novel biocatalysts from the soil metagenome.

Keywords. Metagenomic, High molecular weight (HMW) DNA, Metagenomic library construction

INTRODUCTION

The microbial diversity of soil is extremely complex. It has been estimated that there are approximately 2.6 X 10²⁹ prokaryotes residing in soil (Whitman et al., 1998). Traditional methods to identify and characterize indigenous soil bacteria are based on direct cultivation and isolation of microorganisms that involve morphological examination through the microscope (Torsvik et al., 1990). However, most of the soil microorganisms need special handling in order to grow them in laboratory conditions and up to 99.9% remain uncultured (Hugenholtz and Pace, 1996; Lorenz and Schleper, 2002).

In order to circumvent limitations of the culture method, DNA-based approaches, have been adopted to explore the entire microbial community (Nordgard et al., 2005). In theory, metagenomic DNA recovered from a sample represents the total DNA of all indigenous soil microorganisms (Handelsman et al., 1998; Rondon et al., 1999). Metagenomic coupled with PCR techniques and bioinformatics, is a powerful tool to study the phylogeny and taxonomy of particular genes in a particular community (Torsvik and Ostreus, 2002), including numerous unculturable microbes (Theron and Cloete, 2000). Cloning of metagenomic DNA in an appropriate vector and then transforming the resulting recombinant molecules into a surrogate host enables us to access to the gene pool of unculturable microorganisms (Rondon et al., 2000).

Recovery of metagenomic DNA that is suitable for PCR and metagenomic library construction remains a challenge (Daniel, 2005) because coextracted polyphenolics substances in isolated soil DNA will interfere with downstream applications (Tsai and Olson, 1992). Isolation of high molecular weight (HMW) metagenomic DNA is crucial for metagenomic library construction in order to improve the possibilies of retaining the gene clusters involved in biosynthetic pathways in clones (Bertrand et al., 2005). Furthermore, extreme efficiencies of DNA extraction are needed to assure representative of the entire metagenome (Handelsman et al., 1998; Martin-Laurent et al., 2001).

In this study, we employed a modified DNA extraction method from Yeates and colleagues (1998) which was suitable for construction of large insert soil metagenomic libraries.

*Author for Correspondence.
Mailing address: Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia. Tel: +60-(0)3-79675872; Fax: +60-(0)3-79675908; Email: ccng@um.edu.my
The quality of the DNA extract was evaluated based on the DNA fragments size and the effects of this DNA extraction method on PCR. Also, assessment of soil microbial DNA and construction of metagenomic library were performed.

**MATERIALS AND METHODS**

**Soil samples.** Forest topsoil was collected from the upper 5 cm layer in Rimba Ilmu, University of Malaya. Rimba Ilmu is a tropical botanical garden modelled after a rainforest canopy concept. Soil was sieved through a 2 mm mesh to remove roots and large particles.

**DNA extraction from forest topsoil.** The soil metagenomic DNA extraction was carried out using a modified chemical-enzymatic lysis method from Yeates et al. (1998). Soil (20 g) was suspended in 50 ml of DNA extraction buffer [100 mM Tris-HCl (pH 8.0), 100 mM sodium EDTA (pH 8.0), 1.5 M NaCl]. Lysozyme (1 mg, 10 mg/ml) was added and the sample was incubated at 37 °C for 1 h. Sample was further incubated at 65 °C with SDS (2 ml, 20%, w/v) and proteinase K (15 µl, 20 mg/ml) for 2 h. The mixture was then centrifuged at 6,000 rpm for 10 min to remove soil residue. Supernatant was transferred into a clean tube, and then precipitated by using half-volume of PEG (30%, w/v) / NaCl (1.6 M) and incubated at room temperature for another 2 h.

DNA was pelleted and resuspended in 20 ml of TE (10 mM Tris-HCl, 1 mM sodium EDTA, pH 8.0). The crude metagenomic DNA extract was further purified by using equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). Subsequently the metagenomic DNA was precipitated with 0.1 volume NaCl (5 M) and 2.5 volume of absolute ethanol. The DNA pellet was recovered by centrifugation at 13,000 rpm for 30 min. DNA pellet was washed by 70% (v/v) ethanol. The DNA pellet was recovered by centrifugation at 13,000 rpm for 30 min. DNA pellet was washed by 70% (v/v) ethanol. DNA was pelleted and resuspended in 20 ml of TE (10 mM Tris-HCl, 1 mM sodium EDTA, pH 8.0). The crude metagenomic DNA extract was further purified by using equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). Subsequently the metagenomic DNA was precipitated with 0.1 volume NaCl (5 M) and 2.5 volume of absolute ethanol. The DNA pellet was recovered by centrifugation at 13,000 rpm for 30 min. DNA pellet was washed by 70% (v/v) ethanol. DNA was embedded into an agarose plug by mixing with equal volume of molten 2% (v/v) Low Melting Point (LMP) agarose (BRL Life Technologies, Inc, USA), and was stored in 0.5X Tris-Acetate-EDTA (TAE) at 4 °C for pulsed-field gel electrophoresis (PFGE) later.

**Metagenome DNA size determination by PFGE.** The size of the metagenomic DNA was determined by PFGE. The agarose plug was embedded into 1% (v/v) agarose gel immersed in 0.5X TAE buffer in a Bio-Rad CHEF-DR II apparatus at 9 °C. The PFGE conditions were set according to the manufacturer's instructions. The 1% pulsed field gel was run at 4.5 Volts/cm using ramped pulsed times from 5 to 120 s for 48 h. Lambda PFGE Marker and Low Range PFGE Marker (NEB Biolabs) were used as reference points.

**PCR efficacy analysis.** The quality of the soil metagenomic DNA extract was assessed through PCR. The target genes for PCR amplification included the prokaryotic 16S rRNA gene (27F : 5'- AGA GTT TGA TCA TCM (M) GGG CTC AG-3' and 1525R : 5'- AAG GAG GTG W(T)C CA(R) CC -3') (Lane, 1991); fungal small-subunit rRNA gene (nu-SSU-0817 : 5'- TTG ACA TGG AAT AAT RRA ATA GGA -3' and nu-SSU-1536 : 5'- ATT GCA ATG CYC TAT CCC CA -3') (Borneman and Hartin, 2000) and actinomycetes type II polyketide synthase (PKS) gene targeting the ketosynthase alpha (KS) domain (540F : 5'- GG(I) TGC AC(S) TG(I) GG(I) (M)T(S) GAC -3' and 1100R : 5'- CCG AT(S) GC(I) CC(S) AG(I) GAG TG -3') (Wawrik et al., 2005). The soil metagenomic DNA (0.798 µg/µl) was subjected to serial dilution of 1:10, 1:20, 1:30, 1:40, 1:50 and 1:100 prior to PCR to determine the PCR efficacy. The PCR reaction (total volume 50 µl) consisted of 1 µl of DNA template, 5 µl of 10X PCR buffer (Finnzymes Oy, Finland), 1 µl of 10 mM dNTPs mix (Finnzymes Oy, Finland), 1 U of Tag DNA Polymerase (Finnzymes Oy, Finland) and 5% (v/v) of DMSO (Sigma-Aldrich Inc, USA). PCR conditions were as follows: (i) prokaryotic 16S rRNA gene: 5 min at 95 °C followed by 32 cycles of 1 min at 95 °C, 1 min at 60 °C and 1.5 min at 72 °C followed by a 15 min extension step at 72 °C; (ii) fungal small-subunit rRNA gene: 5 min at 95 °C followed by 40 cycles of 30 s at 95 °C, 30 s at 50 °C and 30 s at 72 °C followed by a 15 min extension step at 72 °C; (iii) actinomycetes type II PKS KSα domain: 5 mins at 95 °C followed by 40 cycles of 1 min at 95 °C, 1 min at 60 °C and 1 min at 72 °C followed by a 15 min extension step at 72 °C. PCR was performed using GeneAmp PCR System 2400 (Perkin-Elmer Corporation, USA).

**Assessment of soil metagenomic DNA.** The resulting PCR products were gel purified by using GeneClean kit (Qbiogene, USA) and cloned into pGEM-T Easy Vector (Promega, USA). Recombinant clones were identified by blue-white selection on X-Gal/IPTG ampicillin (100 µg/ml) Luria-Bertani plate. Positive clones with inserts were amplified using plasmid-targeting primers M13F (5'- GGT TTT AAA GAC TTG GAA ACA CAA ATT TAA ACA GAC GCT TAC CAC AGT TCA TTC AAC GAT GTT GAT CAC CAT-3') and M13R (5'- AGC GGA TAA GTC ACG CTA CTC ATT CAC GAC ACC CGT TTA ACG AAC CCA GAT CCT TCC AC-3') to determine the inserts. The colony PCR conditions were as follows: initial denaturation for 10 min at 94 °C followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 45 s, extension at 72 °C for 1.5 min and final extension at 72 °C for 5 min. PCR products were restricted with HaeIII (NEB, UK) and Alul (NEB, UK) to determine RFLP profiles of the clones to avoid selection of redundant clones. Clones with unique RFLP patterns were sent for DNA sequencing (Solgene, Korea) by using M13 forward and M13 reverse primers. The generated prokaryote 16S rRNA and fungal 18S rRNA gene sequences were compared with sequences deposited in NCBI nucleotide database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) by nucleotide blast program to obtain suitable closest match. The generated actinomycetes type II PKS KSα conserved domain amino acid sequences were also compared with the
AsPac J. Mol. Biotechnol., Vol. 16 (2), 2008

37

Isolation of HMW DNA from soil available sequences in the NCBI protein database by using translated nucleotide query.

**Nucleotide accession numbers.** DNA sequences obtained in this study were submitted to GenBank under the accession numbers EU445192 to EU445233.

**Metagenomic library construction.** A forest soil metagenomic library was constructed using CopyControl pCC1FOS (Epicenter, Madison) with some modifications from the manufacturer's instructions. Before metagenomic library construction, the metagenomic DNA was size-selected using PFGE. First, metagenomic DNA-containing agarose plug was embedded into 1% (w/v) of LMP agarose. The metagenomic DNA was separated using the PFGE. DNA fragment sizes between 30-40 kb were excised from the gel and digested with Gelase (Epicenter, Madison). The size-selected metagenomic DNA was end-repaired using End Repair Enzyme Mix (Epicenter, Madison) before ligation. A 1:10 molar ratio of insert/vector was used for the ligation. Both vector and DNA insert were heated to 55 °C for 10 min and then placed on ice for 2 mins before the addition of 2 U DNA ligase (Epicenter, Madison). Ligation was carried out at 16 °C overnight. Packaging of the forest soil metagenomic fosmid clones was carried out using the MaxPlax Lambda Packaging Extracts (Epicenter, Madison). Analysis of the insert size of the fosmid metagenomic library clones was done by restriction digestion by *Not* I prior to PFGE.

**RESULTS**

**Extraction and purification of soil DNA.** The size of the isolated metagenomic DNA was determined by PFGE.

DNA fragments sizes ranged from 48 kb up to 290 kb were isolated (Figure 1).

**PCR efficacy analysis.** Bacterial 16S rRNA genes were successfully amplified after 20 to 50-fold dilution of the metagenomic soil DNA (Figure 2). Fungal 18S rRNA genes can be PCR amplified with 20 to 40-fold dilution of metagenomic DNA. Type II PKS KSα genes can be PCR amplified with 10 to 30-fold dilution of metagenomic DNA (Figure 2 and Figure 3).

**Assessment of soil metagenomic DNA.** To evaluate effects of this DNA extraction method on the soil metagenomic DNA, we opted for different sets of PCR primers targeting prokaryotic 16S rRNA gene and eukaryotic microorganisms 18S rRNA gene. PCR primers specifically targeting actinomycetes type II PKS KSα domain were also used. PCR amplification was successful with all the primers tested (Figure 2 and Figure 3). Analysis of the 16S rRNA gene sequences indicated that microbial diversity in Rimba Ilmu forest topsoil were mainly consisted of uncultured members of Acidobacteria, Proteobacteria, Candidate Division TM6, Firmicutes, Actinobacteria and Gemmatimonadetes. While analysis of 18S rRNA gene sequences demonstrated that fungi from Basidiomycota and Ascomycota were commonly found. These results revealed that the microbial community in the Rimba Ilmu soil metagenome composed of a wide range of uncultured prokaryotic microorganisms (Table 1). Amino acid sequence analysis of the actinomycetes
Isolation of HMW DNA from soil, 1996; Miller

DNA shearing (Liesack and Stackebrandt, 1992; Leff produce considerable yield of DNA but often cause severe disruption methods such as bead-beating and sonication lysis method were used instead of bead-beating. Physical purification steps. We modified the DNA extraction method to study the physiology of unculturable microorganisms metagenomic library is currently used as a genomic approach to study the large pool of genomic information. DNA is a prerequisite in order to perform molecular analysis. will interfere with downstream applications, therefore purity recovery of HMW and high quality soil metagenomic library was estimated to contain about 127,000 Mbp (0.127 Gbp) of DNA.

DISCUSSION

Recovery of HMW and high quality soil metagenomic DNA is a prerequisite in order to perform molecular analyses to access the large pool of genomic information of unculturable soil microbes (Ward et al., 1990; Zhou et al., 1996). Isolation of HMW DNA is important to reduce the risk of chimera formation during PCR amplification (Liesack 1996). Isolation of HMW DNA is important to reduce the risk of contaminants will interfere with downstream applications, therefore purity of metagenomic DNA is important for molecular analysis. This DNA extraction method produced considerable DNA yield (79.8 µg/g of soil, data not shown) compared to previous studies which produced similar DNA yield but with smaller DNA fragment size (< 23 kb) (Zhou et al., 1996; Miller et al., 1999).

Although previous studies showed nycodenz extraction technique followed by lysis of biomass immobilised in agarose plugs could recover DNA fragments in excess of 1 Mb (Berry et al., 2003), this DNA extraction method was laborious. In addition, nycodenz extraction technique recovers DNA biased towards certain cellular morphologies (Courtois et al., 2001). Since the presence of contaminants will interfere with downstream applications, therefore purity of metagenomic DNA is important for molecular analysis. This DNA extraction method produced relatively pure DNA with A260/A280 ratio of 1.35) or PVPP (A260/A280 ratio of 1.23) (Zhou et al., 1996). Although numerous DNA extraction methods have been reported for the isolation of DNA from soil (Bruce et al., 1992; Kuske et al., 1998; Yeates et al., 1998; Miller et al., 1995; Yeates et al., 1998) which is not suitable for large-insert metagenomic library construction. This modified enzymatic and chemical lysis DNA extraction method was gentler and able to recover HMW DNA (up to 290 kb) in comparison with Yeates et al. (1998) (up to 20 kb). Besides, this DNA extraction method produced considerable DNA yield (79.8 µg/g of soil, data not shown) compared to previous studies which produced similar DNA yield but with smaller DNA fragment size (< 23 kb) (Zhou et al., 1996; Miller et al., 1999).

Table 1. Comparison of 16S and fungal 18S rRNA genes PCR product to known nucleotide sequences in GenBank database.

<table>
<thead>
<tr>
<th>Phylogenetic Affiliations</th>
<th>Clone</th>
<th>Closest match in Genbank database</th>
<th>Accession no.</th>
<th>Nucleotide similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidobacteria</td>
<td>R98</td>
<td>Uncultured Acidobacteria AKYG1861, 16S rRNA gene</td>
<td>AY921997</td>
<td>97%</td>
</tr>
<tr>
<td>R38</td>
<td>Uncultured Acidobacteria AKHY1002, 16S rRNA gene</td>
<td>AY922388</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>R40</td>
<td>Uncultured δ-proteobacterium GMA-N01, 16S rRNA gene</td>
<td>AY582894</td>
<td>92%</td>
</tr>
<tr>
<td>R60</td>
<td>Uncultured β-proteobacterium AKYH1130, 16S rRNA gene</td>
<td>AY921994</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>R36</td>
<td>Uncultured α-proteobacterium Blmi11, 16S rRNA gene, clone</td>
<td>AY318160</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>R24</td>
<td>Uncultured actinomycetin GASP-WDOS2_G10 16S rRNA gene</td>
<td>EF075495</td>
<td>94%</td>
</tr>
<tr>
<td>Candidate division TM6</td>
<td>R34</td>
<td>Uncultured division TM6 bacterium MW3.210WL 16S rRNA gene</td>
<td>AY040398</td>
<td>90%</td>
</tr>
<tr>
<td>Gemmatimonadetes</td>
<td>R55</td>
<td>Uncultured bacterium BS128 16S rRNA gene</td>
<td>AB240268</td>
<td>91%</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>R58</td>
<td>Uncultured Firmicutes bacterium GASP-WA1W2_E07, 16S rRNA gene</td>
<td>EF072410</td>
<td>92%</td>
</tr>
<tr>
<td>Bacteria</td>
<td>R10</td>
<td>Uncultured forest soil bacterium DUNssu098, 16S rRNA gene</td>
<td>AY913321</td>
<td>93%</td>
</tr>
<tr>
<td>Fungi; Basidimycota</td>
<td>1889</td>
<td><em>Megacollybia platypilla</em> isolate AFTOL-ID 560, 18S rRNA gene</td>
<td>AY786053</td>
<td>98%</td>
</tr>
<tr>
<td>Fungi; Ascomycota</td>
<td>1885</td>
<td><em>Sordariomycete</em> sp. pgg-lsf, 18S rRNA gene</td>
<td>EF023708</td>
<td>98%</td>
</tr>
<tr>
<td>1887</td>
<td><em>Penicillium glabrum</em> strain ALI 218, 18S rRNA gene</td>
<td>AF290254</td>
<td>99%</td>
<td></td>
</tr>
</tbody>
</table>

Molecular characterization of metagenomic library. A forest soil metagenomic library, designated as RI1, was constructed. RI1 consisted of 3624 fosmid clones with insert sizes ranged from 23.1 to 40 kb (1% of clones examined). Representative of clones examined are shown on Figure 4. This forest soil metagenomic library was estimated to contain about 127,000 Mbp (0.127 Gbp) of DNA.

Figure 4. Fosmid clones were picked randomly and subjected to Nol digestion to determine the insert size. Most of the insert size was between 23.1 kb – 40 kb. The vector size was 8 kb. Lane 1: Low Range PFG marker (NEB), lane 2 to lane 17 fosmid clones digested by Nol.
1999; Bertrand 	extit{et al}, 2005; Desai 	extit{et al}, 2007), most of the soil DNA extraction methods require extensive purification steps to remove the contaminants before molecular analyses. Purification steps such as gel-plus-minicolumn and gel-plus-concentrator methods (Zhou et al, 1996), Sephadex G-200 spin column purification (Miller et al, 1999), cesium chloride (CsCl) density gradient (Bertrand et al, 2005) are laborious, time-consuming and result in significant DNA loss. Previous experiments showed 1000 – 10,000-fold dilution of soil DNA was required for successful amplification of the 16S rRNA gene from soil DNA although PCR amplification was possible with 10-fold dilution after Sephadex G-200 column purification (Miller et al, 1999). In this work, the purity of the soil metagenomic DNA was evaluated through PCR efficacy analysis. This is because Taq polymerase is sensitive to contaminants such as humic acid (Zhou et al, 1996). Soil metagenomic DNA was subjected to serial dilution prior to PCR to minimise the PCR inhibitory effects (Tsai and Olson, 1992). The PCR efficacy analysis clearly showed that our metagenomic DNA isolated contained relatively low concentration of PCR inhibitory substances and has sufficient purity for PCR without the need of further purification as compared to other DNA extraction methods (Borneman et al, 1996; Zhou et al, 1996; Miller et al, 1999; Bertrand et al, 2005). Sequence analysis of PCR amplicons revealed that this modified DNA extraction method with addition of lysozyme and Proteinase K in the lysis step efficiently captured a wide range of microbial DNA including the hard-to-lyse Gram-positive bacteria and fungi.

Molecular characterization of RI1 metagenomic library showed most of the insert size was between 23.1 kb – 40 kb (Figure 4) with an average DNA insert size of 32 kb. We performed size-selected DNA insert (ranged between 30 kb to 40 kb) prior to metagenomic library construction. This was satisfactory insert size since the average DNA insert size of metagenomic fosmid library clones from other studies were about 35 kb (Ginolhac et al, 2004; Lee et al, 2006; Lim et al, 2005). The maximum insert size for fosmid clone was approximately 40 kb. This indicated that our DNA extraction method could produce DNA of high quality with sufficient purity not only for PCR but also for cloning purposes. In addition, our preliminary end-sequencing results of some metagenomic clones indicated that inserts of RI1 metagenomic library were affiliated with genes related to co-factor and enzyme biosynthesis, energy production and conversion, transcription regulators and membrane fusion protein. Of particular interest is the presence of inserts which shared similarity with genes related to self-defense mechanisms and antibiotic resistance (data not shown).

The main purpose of metagenomic DNA extraction was to obtain high quality and maximal DNA recovery to represent the entire soil metagenome. We had demonstrated that the DNA extraction method was simple, rapid and efficient in recovering high quality DNA suitable for metagenomic analysis and metagenomic library construction.

ACKNOWLEDGEMENT

We would like to thank Dr. Chong-Lek Koh for his advice and help and Ms Siew-See Lee for her technical assistance. This work is funded by IRPA 09-02-008-BTK/ER-36 (Metagenomik) from the Ministry of Science, Technology and Innovation (MOSTI), Malaysia. We thank MOSTI National Science Fellowship (NSF) for giving Mei-Fong, Pang a scholarship.

REFERENCES


