Journal of Microbiological Methods 75 (2008) 457–463

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Probe-based negative selection for underrepresented phylotypes in large environmental clone libraries

Jenni Hultman [a,b,*], Miia Pitkäranta [a], Martin Romantschuk [b], Petri Auvinen [a], Lars Paulin [a]

[a] Institute of Biotechnology, Viikinkaari 4, 00014 University of Helsinki, Finland
[b] Department of Ecological and Environmental Sciences, University of Helsinki, Niemenkatu 73, 15140 Lahti, Finland

ARTICLE INFO

Article history:
Received 13 March 2008
Received in revised form 23 June 2008
Accepted 21 July 2008
Available online 25 July 2008

Keywords:
Clone library
Hybridization
PCR
Probes
Sequencing
Selection

ABSTRACT

Studies based on cloning and sequencing to investigate microbial diversity in a vast range of samples has become widespread in recent years. Results have revealed immense microbial diversity in many different environments, but also dominance of a few sequence types in the constructed clone libraries. Here we describe a method to enrich the clone libraries by avoiding sequencing of known, abundant sequence types, instead focusing on novel, rare ones. The protocol is based on gridding the PCR products from clone libraries on membranes and hybridisation of species-specific probes. Clones that do not give positive hybridisation results are sequenced. This method was used for fungal clone libraries from compost samples. Altogether 1536 clones were gridded and six probes used. From these clones, 59% hybridised with a probe, and therefore, only 41% of the clones were sequenced. In addition, 384 samples were sequenced to verify the hybridisation results. The numbers of false-negative (5.2%) and false-positive (3.9%) hybridisations were low. This method provides a mean of lowering the costs of sequencing projects and speeding up the process of characterising microbial diversity in environmental samples. The method is especially suitable for samples with a few dominating sequence types.

1. Introduction

In recent years, many large sequencing projects describing microbial communities in environmental samples have been carried out. Studies of environments, including the Sargasso Sea (Venter et al., 2004), human intestines (Eckburg et al., 2005), agricultural soil and the deep sea (Tringe et al., 2005; Sogin et al., 2006), have examined the total microbial genomic DNA or genes coding for ribosomal RNA (rRNA). In our recent project, fungal community rRNA was sequenced from samples taken from two municipal waste composting facilities before and after wood ash amendment (Hultman et al., submitted for publication). Results from this project showed that some fungal phylotypes found by PCR and sequencing were abundant, representing over 78% of all the clones sequenced in certain library and therefore, finding new, minor phylotypes became relatively expensive with this method. Similar results showing high dominance of certain phylotypes in environmental libraries have been reported by others (Hunt et al., 2004; Neubert et al., 2006). In many environmental sequencing projects, screening methods such as restriction fragment length polymorphism (RFLP), amplified ribosomal DNA restriction analysis (ARDRA) or denaturing gradient gel electrophoresis (DGGE) have been used prior to sequencing to lower the number of sequenced clones. While these methods do lower the number of sequenced clones, they also skew the results when only the most abundant sequence types are sequenced (DGGE) and some groups may be overlooked. With community fingerprinting methods the numerically rare phylotypes are not generally detected (Bent and Forney, 2008). RFLP tends to underestimate the true diversity as microbes from different species or genera can share an identical restriction pattern (Dunbar et al., 2001). Likewise, a single DGGE band can contain several different ribotypes (Costa et al., 2006) and one ribotype can also be represented by a larger group of bands (Kisand and Wikner, 2003). Most phylotypes in communities are present in low numbers and difficult to find by random sampling as dominating phylotypes make up minority of the diversity (Curtis and Sloan, 2005). With parallel sequencing, it has become relatively easy to gain information on microbial communities (Huber et al., 2007), but the method is still rather expensive and data analysis requires extensive computing resources.

Oligonucleotide fingerprinting has been used to detect microbes in soil environments (Valinsky et al., 2002a,b) and diagnostic microarrays are widely used in environments ranging from marine sediments to air filtrates (reviewed in Loy and Bodrossy, 2006). Likewise, quantitative PCR targeting various fungal species has served to detect and quantify known organisms in various environments (Haugland et al., 2004; Fierer et al., 2005). The main problem with these methods is that they do not give new information, detecting only the species for which the probes were designed.

In the present study, the platform has been converted for screening purposes, and a macroarray containing spotted clone library DNA is

* Corresponding author. Institute of Biotechnology, Viikinkaari 4, 00014 University of Helsinki, Finland. Tel.: +358 9 191 59734; fax: +358 9 191 58952.
E-mail address: jenni.hultman@helsinki.fi (J. Hultman).

0167-7012/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2008.07.016

used. Probes designed for taxa known to be abundant in the studied environment are hybridised on the library array. Phylotypes for which no specific probe exists give a negative hybridisation result and are subsequently characterised by sequencing. To our knowledge this is the first study to use this converted protocol to find the under-represented phylotypes in large environmental clone libraries. The procedure of probe design, hybridisation and sequencing can be repeated until the desired library coverage is achieved (Fig. 1).

2. Materials and methods

The plasmid libraries used were constructed from the ash-amended composting plants. Briefly, DNA was extracted from five samples taken from two compost sites, full-scale and pilot-scale, from a 500 mg sample using a bead-beating procedure with a FastPrep instrument using a FastDNA SPIN Kit for soil (MP Biomedicals, Illkirch, France) according to the manufacturer's instructions. The DNA yield was estimated by agarose gel electrophoresis and visualised with EtBr. PCR was performed in three replicates with FUN18f (5'-TTG CTC TTC AAC GAG GAA T-3', Hultman et al., submitted for publication) and ITS4 (White et al., 1990) primers flanking the fungal internal transcribed spacer (ITS) area and located at the 5' end of the 18S ribosomal RNA gene and at the 3' end of the 28S rRNA gene, respectively. The PCR products were purified with a MicroSpin S-400 HR column (Amersham Biosciences, Buckinghamshire, UK), approximately 80 ng of purified PCR product was cloned into pDrive-vector and an *Escherichia coli* plasmid library was generated with a commercial T/A cloning kit according to the manufacturer's instructions (QIAGEN PCR Cloning^{plus}, QIAGEN, Hilden, Germany). Depending on the amount of transformants containing inserts, 96–768 colonies were picked and transferred to 1 ml of Luria Bertani broth supplemented with ampicillin (150 μg/ml) with a QPix robot (Genetix, Hampshire, UK). After overnight growth, the plasmids were extracted with Multiscreen96 Plasmid Plates (Millipore, MA, USA) according to the manufacturer's instructions. Inserts from the purified plasmids were amplified with universal primers flanking the cloning site. The PCR products were purified with MultiScreen PCR$_{384}$ purification plates (Millipore, MA, USA). Purified PCR fragments, from 96 to 768 clones per library, in total 4×384 (1536) PCR products, were gridded on nylon membranes (9×13 cm, Performa II, Genetix) in duplicate (Fig. 2) with an automated colony picker (QPix, Genetix) using a 384-pin gridding head. Bromphenol blue (~0.03% w/vol) was added to the PCR products to visualise the gridding result (Fig. 2A). The pins touched the PCR product and transferred it to the membrane six times, resulting in approximately 1 μl of transferred DNA solution. The membranes were UV cross-linked using energy of 5500×100 μJ (UV Stratalinker 2400 oven, Stratagene, CA, USA).

Based on the clone frequencies from clone library sequencing in our previous study with a similar sample material (Hultman et al., submitted for publication), probes were manually designed for the most abundant phylotypes using alignments obtained with ClustalW (Thompson et al., 1994) and Gap4 (Staden package, Staden et al., 2000). Altogether six probes were designed for the following common species: *Thermomyces lanuginosus*, *Rhizomucor pusillus*, *Candida ethanolica*, *Geotrichum candidum*, *Issatchenkia orientalis* and *Geosmithia emersonii* (Table 1). Publicly available DNA database reference data as well as clone library sequence data from target taxa and close relatives were used for probe design. The oligonucleotide probes (Oligomer, Helsinki, Finland) were labelled with digoxigenin (DIG Oligonucleotide Tailing Kit, Roche, Mannheim, Germany) according to the manufacturer's instructions. Probe specificity was confirmed with BLAST (Altschul et al., 1990) alignments. The specificities of these probes for these species were also confirmed via membrane hybridisations using membranes made of samples extensively characterised by rRNA gene library sequencing. The hybridisation conditions were adapted from the tailing kit (DIG Oligonucleotide Tailing Kit) and the hybridisation temperature was calculated by the ionic strength, G/C content and length of the oligonucleotide and was approximately 10 °C lower than the calculated Tm. The probe concentration was set to 6.75 pmol based on the probe specificity testing (data not shown).

The membranes were hybridised as follows: After a 2-h prehybridisation in hybridisation buffer (DIG easy hyb, Roche) at 42 °C, 6.75 pmol of each DIG-labelled probe was hybridised to the membrane in 5 ml of hybridisation buffer overnight at 42 °C. Following hybridisation, the membranes were washed twice for 5 min in 2×SSC, 0.1% SDS and twice for 15 min in 0.5×SSC, 0.1% SDS. Immunological detection was performed with a commercial kit according to the manufacturer's instructions (DIG nucleic acid detection kit, Roche). The membranes were then scanned and the images processed.

Hybridisation was performed with a mix of all six probes as well as with each probe separately. A grid of 384 squares was fitted on the membrane, and PCR products giving a positive hybridisation signal from both replicate spots were identified. Because the level of background was very low, manual detection of spots was straightforward and could

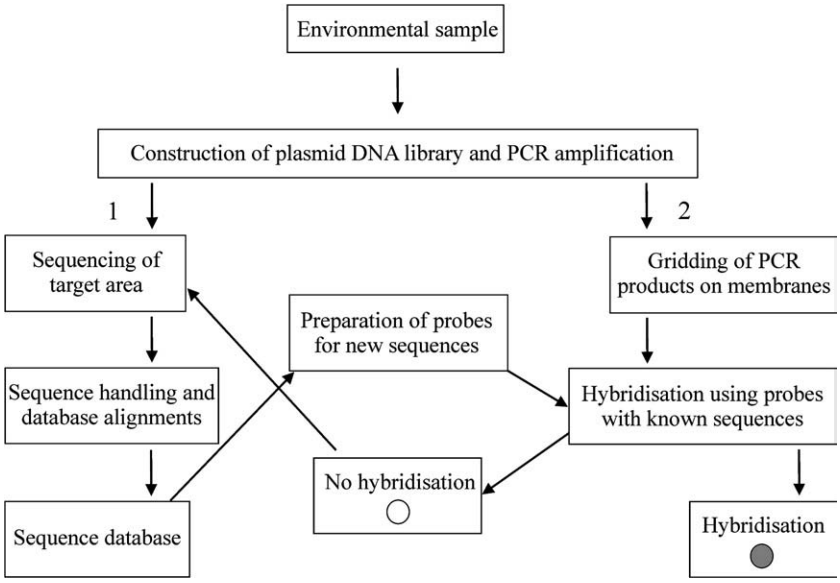

Fig. 1. Flow diagram of novel procedure. The left-hand side (1) describes the protocol used in sequencing projects and the right-hand side (2) the protocol with hybridisation. The expensive part of sequencing all of the clones is avoided, with only the clones not recognised by probes designed for the abundant phylotypes being sequenced.