



A simple boiling-based DNA extraction for RAPD profiling of landfarm soil to provide representative metagenomic content

J.H. Amorim¹, R.O. Vidal², G.V. Lacerda-Junior³, J.C.T. Dias³,
M. Brendel³, R.P. Rezende³ and J.C.M. Cascardo³

¹Departamento de Microbiologia, Universidade de São Paulo, São Paulo, SP, Brasil

²Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Ilhéus, BA, Brasil

³Laboratório Nacional de Luz Síncrotron, Campinas, SP, Brasil

Corresponding author: R.P. Rezende
E-mail: rezende.rachel@gmail.com

Genet. Mol. Res. 11 (1): 182-189 (2012)

Received September 27, 2011

Accepted December 8, 2011

Published January 27, 2012

DOI <http://dx.doi.org/10.4238/2012.January.27.5>

ABSTRACT. Landfarm soils are employed in industrial and petrochemical residue bioremediation. This process induces selective pressure directed towards microorganisms capable of degrading toxic compounds. Detailed description of taxa in these environments is difficult due to a lack of knowledge of culture conditions required for unknown microorganisms. A metagenomic approach permits identification of organisms without the need for culture. However, a DNA extraction step is first required, which can bias taxonomic representativeness and interfere with cloning steps by extracting interference substances. We developed a simplified DNA extraction procedure coupled with metagenomic DNA amplification in an effort to overcome these limitations. The amplified sequences were used to generate a metagenomic data set and the taxonomic and functional representativeness were evaluated in comparison with a data set

built with DNA extracted by conventional methods. The simplified and optimized method of RAPD to access metagenomic information provides better representativeness of the taxonomical and metabolic aspects of the environmental samples.

Key words: Landfarm area; Metagenomic; RAPD; Microbiota representativeness

INTRODUCTION

Landfarm soils are employed in industrial and petrochemical residue treatment by bioremediation (Maciel et al., 2007). A range of microorganisms and genes with potential biotechnological application may be present in environments such as these, where a selective pressure on metabolic pathways involved in the degradation of toxic compounds is intense (Schloss and Handelsman, 2003; Maciel et al., 2007). Detailed descriptions of taxa, metabolic pathways, protein/peptide interactions, and molecular relationships of all microorganisms from environmental microbiota present in landfarm soils are not possible using culture-dependent techniques, due to the lack of knowledge of culturing conditions for unknown microorganisms (Yun et al., 2004; Riesenfeld et al., 2004). Thus, a metagenomic approach has been indicated as a way to further access data contained in these ecosystems (Johnson and Slatkin, 2006; McHardy and Rigoutsos, 2007).

The term metagenomic derives from the concepts of meta-analysis and genomics (Schloss and Handelsman, 2003). This technology allows access to data at the phylogenetic level and the isolation and cloning of novel genes, with several biological functions, such as enzymatic, antimicrobial and antitumor (Streit and Schmitz, 2004; Roh et al., 2006; McHardy and Rigoutsos, 2007). Accessing metagenomic data is based on conventional DNA extraction methods that usually bias the representativeness of microbial taxa (Robe et al., 2003; Daniel, 2004; Roh et al., 2006). Besides, conventional cloning steps based on the cloning of polymerase chain reaction (PCR) products or large metagenomic DNA fragments are unfeasible in some aspects (Schloss and Handelsman, 2003). The interference of co-extracted substances in the PCR procedure or small number of vector copies and difficulties in connecting large DNA fragments to vectors are points to be considered (Robe et al., 2003; Roh et al., 2006; Amorim et al., 2008).

Another aspect that limits the study of metagenomic content concerns the bioinformatics methods used to analyze the genomic data generated (Rondon et al., 2000; Roh et al., 2006; Huson et al., 2007). Notably, most information is computed by methods that depend on sequences of the same gene to determine taxonomic and metabolic profiles of the environment (Rondon et al., 2000; Roh et al., 2006). Thus, they do not allow the determination of taxonomic and metabolic content at the same time (Rondon et al., 2000; Roh et al., 2006).

Here, we present a low-cost and practical alternative to accessing metagenomic content by boiling the soil sample and directly applying the random amplified polymorphic DNA (RAPD) technique to cover the metagenome of a landfarm area as much as possible, with a subsequent cloning step of the RAPD products into pTZ57R/T vectors, to finally construct a metagenomic library. A robust bioinformatic analysis was carried out focusing on open reading frames (ORFs) to compare the taxonomic and metabolic representativeness of metagenomic libraries constructed using traditional or boiling/RAPD-based approaches. The results presented in this study suggest new conditions to properly access metagenomic content.

MATERIAL AND METHODS

Soil sampling

A soil sample was collected at Landulpho Alves Refinery, in São Francisco do Conde, Bahia, Brazil. Sampling was performed as described elsewhere (Maciel et al., 2007).

DNA extraction and cloning of RAPD products

In a 1.5-mL sterile tube, 1 g landfarm soil was suspended in 1 mL sterile ultrapure water, followed by a boiling step of 20 min. Afterwards, the tube was centrifuged for 10 min at 13,400 rpm. The supernatant was kept and aliquots were diluted in sterile ultrapure water to 1:10, 1:50, 1:100, and 1:200, to optimize the best condition to obtain the highest yield of metagenomic RAPD product. The best dilution was directly used to perform RAPD reactions with the primers OPC01, OPK08, OPI07, OPN18, OPN19, OPK20, OPB13, OPS07, OPS10, D13, T17, R08, Q08, OPJ04 E T20 (Operon kit, Gibco), with the following reagents and conditions: 0.975 U *Taq* DNA polymerase; 3.5 mM MgCl₂; 0.48 mM dNTPs; 2.5 ng RAPD primer; 0.5 mM BSA and DNA (4 μL of best dilution) to a final volume of 12.5 μL. A conventional metagenomic DNA extraction procedure was performed as previously described (Amorim et al., 2008). The amount of DNA extracted by this method used as template in RAPD reactions was determined by screening dilutions, as done with DNA extracted by simple boiling. RAPD products were ligated into the T/A cloning site of the vector pTZ57R/T, according to manufacturer instructions (InsTAclone PCR cloning kit, Fermentas). Plasmids ligated to RAPD products were then transformed into chemically competent *Escherichia coli* TOP10 and recombinant bacterial colonies were selected by blue/white colony screening (Sambrook et al., 1989).

Sequencing and construction of metagenomic libraries

From 1600 selected colonies, 800 were used to clone DNA extracted by the conventional method and 800 were used to clone DNA extracted by simple boiling. The clones were individually cultured in Luria-Bertani medium containing ampicillin (100 μg/mL) in 96-well culture plates and then submitted to plasmid DNA extraction (Sambrook et al., 1989). Extracted plasmids were then used as templates to perform sequencing reactions, which were read in a MegaBace 1000 (GE Healthcare). Empty plasmids or plasmids containing inserts up to 100 bp were discarded. Plasmids containing inserts greater than 100 bp were used to compose metagenomic libraries.

Bioinformatics and statistics

Sequences were treated and filtered by Phred, Crossmatch (Ewing et al., 1998), Lucy and RepeatMasker (Smit et al., 2004; <http://www.repeatmasker.org>) and after they were mined by the Getorf program version 5.0.0 (from EMBOSS program family). ORFs with a minimum of 50 amino acids were further submitted to similarity analysis using BLASTP against the NR database. Sequences with a hit with a minimal e-value of 1e-7 were accepted. Selected sequences were then blasted against NR (NCBI non-redundant database), GO (Ashburner

et al., 2000; Gene Ontology <http://www.geneontology.org>) and KEGG (Kanehisa and Goto, 2000; Kyoto Encyclopedia of Genes and Genome <http://www.genome.jp/kegg>) to determine the taxonomic, functional and metabolic profiles, respectively. The results of these annotations were inserted into the MEGAN software (Huson et al., 2007). Statistical analyses were performed using two-way ANOVA, with statistical significance set at $P < 0.05$. MEGAN analysis corresponding to taxonomic, functional or metabolic profiles of clones from DNA extracted by the two different methods was performed.

RESULTS AND DISCUSSION

The aim of this study was to establish a simple and low-cost process to access the metagenomic content of impacted soils. The best conditions to allow cloning of RAPD products were determined. Hence, cloning RAPD products allowed not only the representation of different taxa, but also different biological functions and metabolic pathways.

The best conditions to obtain the highest metagenomic amplified DNA yield by RAPD were established using template DNA extracted by simple boiling at a 1:200 dilution (Figure 1A). The same conditions were shown to be more suitable for amplifying DNA extracted by the conventional method (Figure 1B). Using this dilution, co-extracted interference substances (Amorim et al., 2008) were also diluted, probably allowing better annealing of primers and efficient action of DNA polymerase. After cloning, sequencing and sequence treatment, metagenomic sequence sets were determined; 500 reads regarding DNA extracted by the conventional method, labeled as Land 1, and 570 reads regarding DNA extracted by simple boiling, labeled as Land 2.

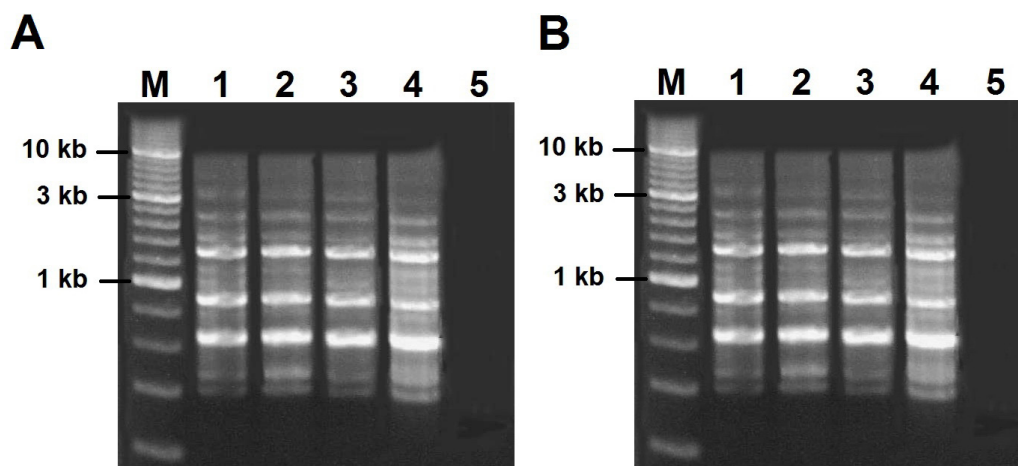


Figure 1. Optimization of RAPD conditions using **A.** DNA extracted by simple boiling method or **B.** DNA extracted by conventional method, as template. Reactions were performed with the following DNA dilutions: *lane 1* = 1:10; *lane 2* = 1:50; *lane 3* = 1:100; *lane 4* = 1:200. *Lane M* is the molecular weight marker, with molecular weights shown on the left side of the figure. *Lane 5* corresponds to the negative control of RAPD reactions.

Both, taxonomic and functional representativeness were affected by the DNA extraction method. As shown in Figure 2, the Land 1 dataset failed to represent several taxa pres-

ent in the landfarm soil studied. On the other hand, Land 1 was shown to represent more efficiently biological activities present in the environment studied (Figure 3). However, the representativeness regarding metabolic pathways was shown to be the same between the Land 1 and Land 2 datasets (Figure 4). Taken together, these results indicate that the extraction method affects metagenomic representativeness in the taxonomic, functional and metabolic fields. Moreover, extracting DNA by simple boiling seems to improve the determination of taxonomic representativeness of the environment studied.

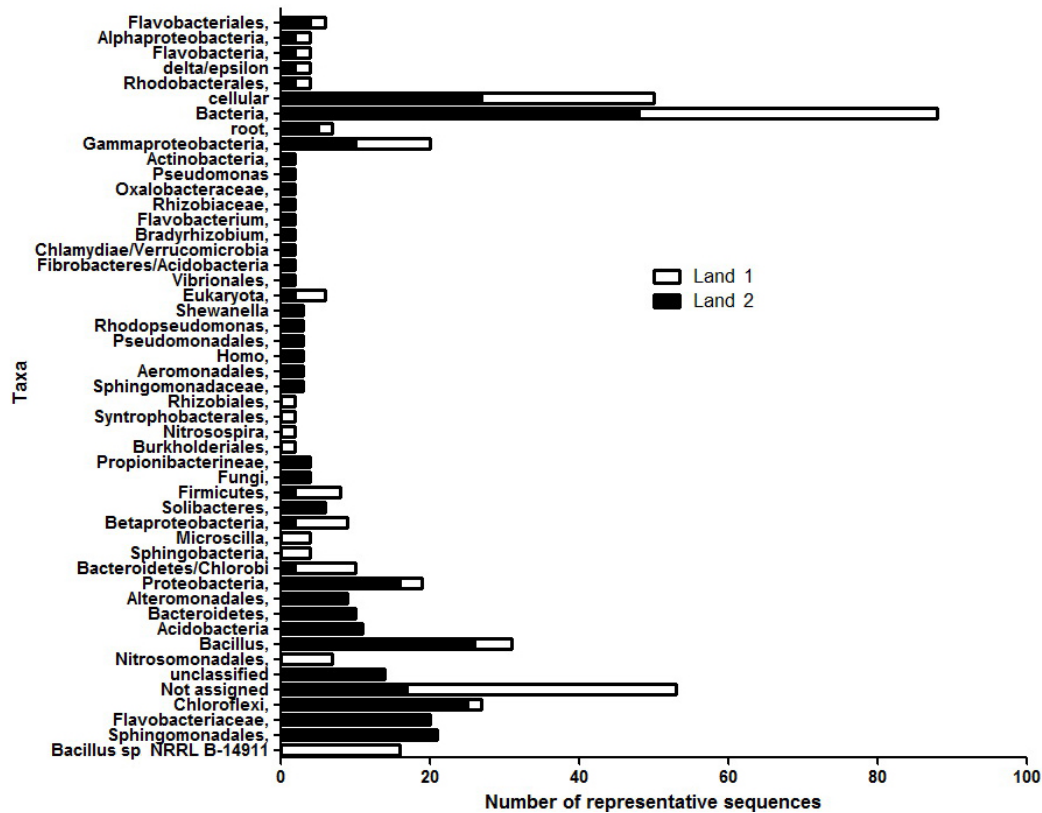


Figure 2. Representativeness of taxon content by Land 1 and Land 2 datasets. The number of sequences representative of each taxon is recorded. Land 2 dataset is shown to be more representative ($P < 0.05$).

The simple boiling of soil, applied in the construction of Land 2, allows lysis *in situ*, which can permit a more efficient access to DNA (Robe et al., 2003; Roh et al., 2006). As we have shown in Figure 2, it was demonstrated that Land 1 represents extra biological activities, which can be due to the generation of RAPD products using DNA from a lower number of genomes of representative taxa as template. This may allow primers to anneal in a larger number of gene families, subsequently amplifying them. The better representativeness of metabolic content observed in the Land 2 dataset confirms that disruption of bacterial cells *in situ* promotes more availability of DNA than using the conventional DNA extraction method. A pos-

sible cause of this may be the fact that in *ex situ* extraction methods a considerable number of bacterial cell DNA donors may be discarded in the process of separating cells and soil particles (Robe et al., 2003; Roh et al., 2006).



Figure 3. Representativeness of biological activity content by Land 1 and Land 2 datasets. The lighter the area is, the more sequences related to the corresponding biological activity are scored. Dark areas mean lack of representativeness.

Land 2 showed a higher diversity of microorganisms in comparison to Land 1 and was shown to be more representative (Figure 2). However, it is important to note that neither

represents an entire sample, as there are missing taxa in both samples. Reflecting the saturation problem, both samples show some exclusive biological activities and metabolic pathways, although Land 2 appears to be more representative (Figures 3 and 4). Regarding the biological activities, we noted the presence of some related to biological transport, such as ATP-binding cassette transporter, gluconate transporter and two-component sensor. The biosynthetic activity was also assigned in a significant number together with unknown biological processes. Together, these results indicate the potential of this environment as a possible reservoir of unclassified transport mechanisms and source of new molecules with biotechnological applications. For example, new substances involved in biosynthetic or catabolic pathways could contribute to improving some important industrial processes (Rondon et al., 2000; Roh et al., 2006; McHardy and Rigoutsos, 2007).

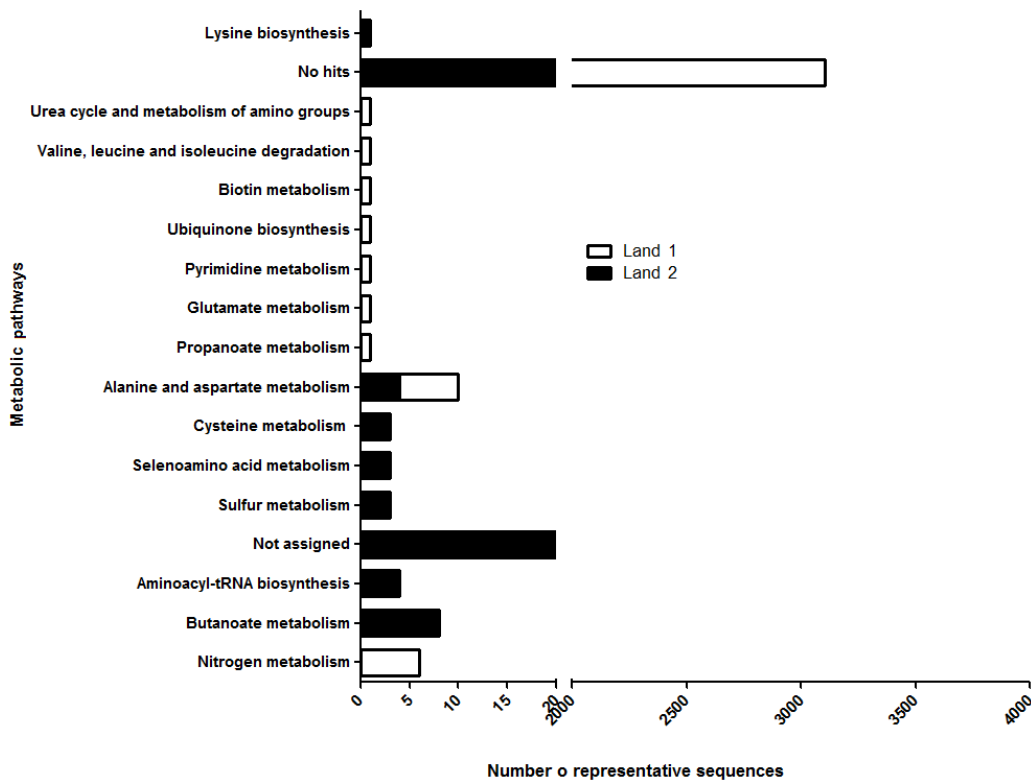


Figure 4. Representativeness of metabolic pathways contained in the environment studied. The number of sequences representative of each taxon is recorded. Land 2 dataset is shown to be more representative ($P < 0.05$).

Regarding the pathways in both samples, we noted the presence of some related to nitrogen metabolism, a substance regularly added as a nutrient to landfarms to favor microorganism metabolism and growth (Maciel et al., 2007). Although the presence of other pathways such as butanoate, propanoate and sulfur pathways, related to the degradation of toxic compounds, deserves attention, a predominant amount of no hits in this analysis corroborates the possibility

of this environment being a reservoir of new metabolic pathways and compounds, with the most diverse biotechnological applications, a result also found in the analysis of biological activities.

CONCLUSIONS

The simplified and optimized method presented here to access metagenomic data was shown to better determine the representativeness of the environment studied, in taxonomic and metabolic fields, allowing a more efficient profiling of the landfarm area. Data presented in this study suggest new conditions to properly access metagenomic content.

ACKNOWLEDGMENTS

We should always thank Professor Julio Cesar de Mattos Cascardo, a great scientist and best friend, who devoted his life to science and teaching. **Julio passed away in 2010. Research supported by Conselho Nacional de Pesquisa e Desenvolvimento (CNPq, #558272/2009-6) and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, scholarship granted to J.H. Amorin).**

REFERENCES

- Amorim JH, Macena TN, Lacerda GV Jr, Rezende RP, et al. (2008). An improved extraction protocol for metagenomic DNA from a soil of the Brazilian Atlantic Rainforest. *Genet. Mol. Res.* 7: 1226-1232.
- Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25-29.
- Daniel R (2004). The soil metagenome - a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15: 199-204.
- Ewing B, Hillier L, Wendl MC and Green P (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175-185.
- Huson DH, Auch AF, Qi J and Schuster SC (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17: 377-386.
- Johnson PL and Slatkin M (2006). Inference of population genetic parameters in metagenomics: a clean look at messy data. *Genome Res.* 16: 1320-1327.
- Kanehisa M and Goto S (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28: 27-30.
- Maciel BM, Dias JC, Dos Santos AC, Argolo Filho RC, et al. (2007). Microbial surfactant activities from a petrochemical landfarm in a humid tropical region of Brazil. *Can. J. Microbiol.* 53: 937-943.
- McHardy AC and Rigoutsos I (2007). What's in the mix: phylogenetic classification of metagenome sequence samples. *Curr. Opin. Microbiol.* 10: 499-503.
- Riesenfeld CS, Schloss PD and Handelsman J (2004). Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38: 525-552.
- Robe P, Nalin R, Capellano C, Vogel TM, et al. (2003). Extraction of DNA from soil. *Eur. J. Soil. Biol.* 39: 183-190.
- Roh C, Villatte F, Kim BG and Schmid RD (2006). Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Appl. Biochem. Biotech.* 134: 97-112.
- Rondon MR, August PR, Bettermann AD, Brady SF, 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.
- Sambrook J, Fritsch EF and Maniatis T (1989). *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory Press, New York.
- Schloss PD and Handelsman J (2003). Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* 14: 303-310.
- Smit AFA, Hubley R and Green P (2004). RepeatMasker Open-3.0. 1996-2007.
- Streit WR and Schmitz RA (2004). Metagenomics - the key to the uncultured microbes. *Curr. Opin. Microbiol.* 7: 492-498.
- Yun J, Kang S, Park S, Yoon H, et al. (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl. Environ. Microbiol.* 70: 7229-7235.