Biotechnol Lett
DOI 10.1007/s10529-011-0693-6

ORIGINAL RESEARCH PAPER

Construction and validation of two metagenomic DNA libraries from Cerrado soil with high clay content

Alinne Pereira de Castro · Betania Ferraz Quirino · Heather Allen · Lynn L. Williamson · Jo Handelsman · Ricardo Henrique Krüger

Received: 16 May 2011 / Accepted: 1 July 2011

Abstract A challenge of metagenomic studies is in the extraction and purification of DNA from environmental samples. The soils of the Cerrado region of Brazil present several technical difficulties to DNA extraction: high clay content (>55% w/w), low pH (4.7) and high iron levels (146 ppm). Here we describe for the first time the efficient recovery and purification of microbial DNA associated with these unusual soil characteristics and the construction and validation of two metagenomic libraries: a 150,000 clones library with insert size of approximately 8 kb and a 65,000 clones library with insert size of approximately 35 kb. The construction of these metagenomic libraries will allow the biotechnological exploitation of the microbial community present in the soil from this endangered biome.

Keywords Clay · DNA extraction · Iron · Library · Metagenomics · Soil

Introduction

Cerrado is a savanna-like region that covers a large area of Brazil (Supplementary Fig. 1). Despite its biological importance, the Cerrado has been the focus of few studies about its microbial biological diversity.

Electronic supplementary material The online version of this article (doi:10.1007/s10529-011-0693-6) contains supplementary material, which is available to authorized users.

A. P. de Castro · R. H. Krüger (✉)
Laboratório de Enzimologia, Department de Biologia Celular, Instituto Central de Ciências Sul, Universidade de Brasília (UnB), Brasília, DF CEP 700910-900, Brazil
e-mail: kruger@unb.br

A. P. de Castro
e-mail: p.alinne@gmail.com

B. F. Quirino
Genomic Sciences and Biotechnology, Universidade Católica de Brasília, Brasília, DF, Brazil
e-mail: betaniaf@pos.ucb.br

B. F. Quirino
Embrapa-Agrienergy, Parque Estação Biológica, Brasília, DF 70770-901, Brazil

H. Allen · L. L. Williamson · J. Handelsman
Department of Plant Pathology, University of Wisconsin, 1630 Linden Dr. Madison, Madison, WI 53706, USA
e-mail: heather.Allen@ARS.USDA.GOV

L. L. Williamson
e-mail: lynnlwilliamson@msn.com

J. Handelsman
e-mail: jo.handelsman@yale.edu

Present Address:
J. Handelsman
Department of Molecular, Cellular & Developmental Biology, Yale University, New Haven, CT 06511, USA

Published online: 12 July 2011

To use a metagenomic approach to study Cerrado soil microbial communities, the first step is to obtain DNA in quantity and quality compatible with molecular biology protocols.

A number of protocols for the extraction of environmental DNA from soil have been published; however, each type of soil presents specific challenges. The extraction of microbial DNA from high clay content soils, such as those found in the Brazilian savanna, is particularly problematic because clay particles absorb organic matter, including DNA (Purdy 2005). Clay is a particle of small size and thus has a large surface area per unit weight and generally is coated with metal oxides (Jenne 1988) and organic matter (Davis 1984). These particles exhibit surface charges that attract negatively and positively charged ions (Brady and Weil 2002) and these characteristics can negatively affect the extraction and purification of microbial DNA associated with clayey soil (Krsek and Wellington 1999; Griffiths et al. 2000).

In metagenomic studies, environmental DNA is usually used for PCR amplification of the 16S ribosomal gene for characterization of the microbial community or for the construction of metagenome expression libraries with biotechnological purpose. Of these two applications for environmental DNA, the construction of libraries is more technically difficult and requires DNA of higher molecular weight and quality. In this work, we present the methods for the construction and the validation of two metagenomic libraries with DNA extracted from microorganisms associated with a high clay content (>55% w/w), high iron and low pH soil using a low copy number plasmid and a fosmid as vectors. The strategy described here should be applicable to obtain metagenomic DNA from similar soil types.

Materials and methods

Soil samples and physical–chemical analysis

Soil samples were collected in an area of Cerrado *sensu stricto* in the Brazilian Institute of Geography and Statistics; Brazil (15° 57′02.4″ S, 47° 52′32.1″ W). Samples were collected from the soil surface (0–10 cm) and were sieved through a 2 mm mesh to remove large particles and plant material. Soil samples were kept on ice until arrival in the laboratory where they were stored at −20°C. Soil physical–chemical characteristics were determined by standard methods by Soloquimica—soil analysis Ltda., Brazil (Table 1).

Table 1 Physical–chemical parameter's of Cerrado soil

Parameter	Cerrado *stricto sensu*
pH (H_2O)	4.7
Organic matter (g/kg)	57
Zn (ppm)	0.02
Fe (ppm)	146
Silt (g/kg)	150
Clay (g/kg)	550
Sand (g/kg)	300
K (ppm)	0.1

Microbial DNA extraction from Cerrado soil

For the construction of the small DNA inserts library in a low copy number plasmid, the extraction of soil microbial DNA was performed using the Power Soil DNA Isolation Kit (MOBIO Laboratories, Inc) following the manufacturer's protocol.

Direct extraction of microbial DNA from soil to construct the large DNA inserts fosmid library, was performed as described by Williamson et al. (2005) with modifications. Briefly, 50 g soil was mixed with 150 ml extraction buffer (100 mM Tris/HCl, pH 8.0, 100 mM sodium phosphate solution, 100 mM EDTA pH 8.0, 1.5 M NaCl, 1% w/v CTAB) in centrifuge tubes and subjected to two freeze–thaw cycles. In each cycle, tubes were immersed for 40 min in a slush of crushed dry ice and 2-propanol, followed by holding the sample at 65°C for 40 min. Then, 18 ml 20% (w/v) SDS and 9 ml 5 M guanidine isothiocyanate were added to the solution which was then held at 65°C for 2 h with occasional gentle mixing. Supernatants, collected after centrifuging the samples at 15,000×*g* for 20 min at 10°C, were combined with 25 ml chloroform/isoamyl alcohol (24:1 v/v) and mixed gently for 10 min. The aqueous phase was recovered by centrifugation at 15,000×*g* for 20 min at 10°C and precipitated with a 0.7 volume 2-propanol at room temperature for 20 min. The DNA was pelleted by centrifugation at 15,000×*g* for 40 min at 10°C, resuspended in 2 ml 10 mM Tris/HCl (pH 8.0) plus 10 mM EDTA (pH 8.0). The DNA was then extracted by adding an equal volume of Tris buffered

phenol/chloroform (pH 8.0) and pelleted by centrifugation at 10,000×*g* for 10 min at room temperature. The aqueous layer containing the DNA was removed and extracted again with chloroform-isoamyl alcohol (24:1 v/v).

The DNA was purified through electrophoresis in 1% low melting-temperature agarose gel (LMP Preparative Grade for Large Fragments; Promega) at 22 V with TAE buffer at (4° C) for 14 h. Sections containing DNA were cut from the unstained gel and this piece was loaded on to a 1% low melting-point agarose pulsed-field gel (Bio-Rad CHEFMapper) at 9 V/cm with a 0.5–0.5 s switch time, a linear ramping factor, and a 120° included angle for 5 h in TAE buffer (Maniatis et al. 1982). DNA larger than 30 kb was excised and extracted from the gel with GELase (Epicentre) according to the manufacturer's instructions.

Cloning of DNA fragments

Low copy number plasmid metagenomic library construction

Metagenomic DNA from the soil microbial community was partially digested with *Pst*I in buffer H (Promega) and the reaction was applied to a 1% low melting point agarose gel (Invitrogen) with TAE buffer. DNA fragments between 2 and 8 kb were excised and extracted from the gel with GELase (Epicentre) according to the manufacturer's directions.

The low copy number plasmid (pCF430) (Newman and Fuqua 1999) vector was isolated and purified with QIAprep Miniprep (QIagen, Valencia, CA) and digested with 10 U *Pst*I. The plasmid was dephosphorylated with shrimp alkaline phosphatase (Promega) according to the manufacturer's instructions. The dephosphorylated product was precipitated, self-ligated, and gel purified. With a clean scalpel, the portion of the gel that contained the non-stained, linear vector was excised and then extracted from the gel with QIAEX II Gel Extraction Kit (QIagen) according to the manufacturer's directions.

DNA ligation reactions between the soil microbial DNA inserts and the low copy number plasmid vector contained an insert/vector ratio of 3:1 and 3 U of T4 DNA ligase and buffer (Promega) was added to the reaction and incubated overnight at 16°C. Ligation products were transformed into electrocompetent *E. coli* TransforMax EPI300 and plated on LB medium containing tetracycline at 20 μg/ml. Plates were incubated overnight at 37°C. The diversity of inserts present in the library was evaluated by restriction digest analysis of randomly selected clones using the restriction enzyme *Pst*I, which occurs only once in the pCF430 plasmid.

Fosmid metagenomic library construction

To construct the library, 2.5 μg purified DNA was enzymatically treated to prepare 5′-phosphorylated blunt ends and ligated to pCCFos1 using a CopyControl Fosmid Library Production Kit (Epicentre) with minor modifications. Instead of following the protocol for titration of the phage particles, 10 μl packaged phage to 100 μl *E. coli* cells in sterile microfuge tubes were used. The EPI300-T1 (Epicentre) was used for library construction. After packaging into lambda phages, infected cells were plated as previously described (Gillespie et al. 2002) and plated on LB medium containing cloramphenicol at 10 μg ml^{1}. Fosmid DNA was evaluated by restriction digest analysis of randomly selected clones using *Not*I. After individual clone analysis, all transfected colonies were pooled together by scraping them from plates into LB containing cloramphenicol and mixing them with an equal volume of 15% (v/v) glycerol followed by storage at −80°C.

Validation of metagenomic libraries

Sequencing of low copy number plasmid and fosmid ends

Plasmid DNA and fosmid DNA from 20 clones each were isolated by using QIAprep Miniprep according to the manufacturer's instructions. The ends of plasmid clones were sequenced with *forward* pCF430 primer (5′-CTGTTTCTCCATACCCGTT-3′); and *reverse* pCF430 primer (5′-TGCAAGGCGATTAAGTTGG-3′) and the ends of fosmid clones were sequenced with PCC1FOS *forward* primer (5′-GGATGTGCTGCAAGGCGATTAAGTTGG-3′) and pCC1FOS *reverse* primer (5′-CTCGTATGTTGTGTGGAATTGTGAGC-3′). Plasmid end sequences had an average length of 400 bp, while fosmid end sequences has an average length of 500 bp. End sequences were analyzed using BLASTx.

16S rRNA gene sequence analysis

DNA was extracted from the low copy number plasmid- and fosmid-metagenomic libraries and used as templates to amplify 16S rRNA genes. For removal of contaminating *E. coli* chromosomal DNA in large-scale clone preparations of both low copy number plasmid and fosmid vectors, plasmid-safe ATP-dependent DNase (Epicentre) was used according to the manufacturer's directions.

16S rRNA gene fragments of approx. 1.4 kb were amplified with primers specific to the domain *Bacteria* 27F/1492R using as template the plasmid or fosmid DNA from metagenomic libraries and cloned into pGEM-T (Promega). Ligation products were transformed into electrocompetent *E. coli* TransforMax EPI300. Approximately 96 clones from each of the libraries were sequenced using ABI PRISM 377 (Applied Biosystems).

All sequences were processed into high quality sequences using Phred 20 for 400 bp. After removal of chimeric artifacts using the Bellerophon program (Huber et al. 2004), sequences were compared to the non-redundant database of sequences deposited at the National Center for Biotechnology using BLASTx (Altschul et al. 1997). Sequences were aligned using Muscle software and phylogenetic analyses were performed with Mega (Tamura et al. 2007), using the neighbor-joining model.

Accession numbers

All 16S rRNA gene sequences described in this study were deposited in GenBank and were assigned numbers HM437021-HM437166.

Results and discussion

Procedures for DNA extraction from soil are not universally applicable. For each soil type new protocols need to be developed or old ones tested and optimized. A number of soil composition characteristics such as amount of clay present can interfere with DNA extraction. Problems in extracting DNA from acidic soils have also been reported (He 2005). The Cerrado soil used in this study was classified as a clayey soil and contained at least 55% of clay as determined by physical–chemical tests (Table 1). The percentage of clay in the soil of the Brazilian Amazon ranges from 21 to 35% (O'Neill et al. 2009). The clay content reported in other studies is always lower than those of the Cerrado soil. For example, the soil of 14 sites located in the Czech Republic and France had clay percentages ranging between 2 and 37% (Sagova-Mareckova et al. 2008). In the United States, Alaskan soil was reported to have 10% clay (Williamson et al. 2005), while Wisconsin soil had 17% clay (Borneman et al. 1996). Studies with soils from Egypt and the Netherlands reported a clay content of approx. 10% (Messiha et al. 2009). Therefore, the clay content found in Cerrado soil, specifically Cerrado *sensu stricto* soil, is high compared to other sites.

As observed in Table 1, the Cerrado soil used in this study also shows a low pH of 4.7 and an elevated iron concentration for native soils of 146 ppm which can be a problem for DNA extraction. For humid tropical forest soils iron concentrations reported in the literature are between 18 and 50 ppm (Dubinsky et al. 2010). Iron concentrations close to 200 ppm are only found in soils of iron mining areas (Aremu et al. 2010). Iron reactivity with peroxides can lead to hydroxyl radical production which in turn can break down DNA molecules by Fenton reaction (Harrison and Arosio 1996).

In this study, we have extracted and purified DNA from the microbial community associated with a high clay, high iron and low pH soil. To test if the DNA extracted was compatible with common molecular biology applications, we constructed two metagenomic libraries using two different methodologies. For the construction of the low copy number plasmid metagenomic library, DNA was efficiently extracted from Cerrado soil using the Power Soil DNA Isolation kit. Many studies reported in the literature have made use of commercial kits to extract DNA from soil. However, the use of commercial kits for DNA extraction has limitations. Studies have shown that commercial kits can be biased depending on the methods and efficiency of cell lyses and DNA purification which may affect PCR performance. Importantly, kits are not suitable for large-scale DNA extraction or for the extraction of high molecular weight DNA compatible with fosmid library construction.

The low copy number plasmid metagenomic library from Cerrado soil constructed consisted of approximately 150,000 clones. To determine the size

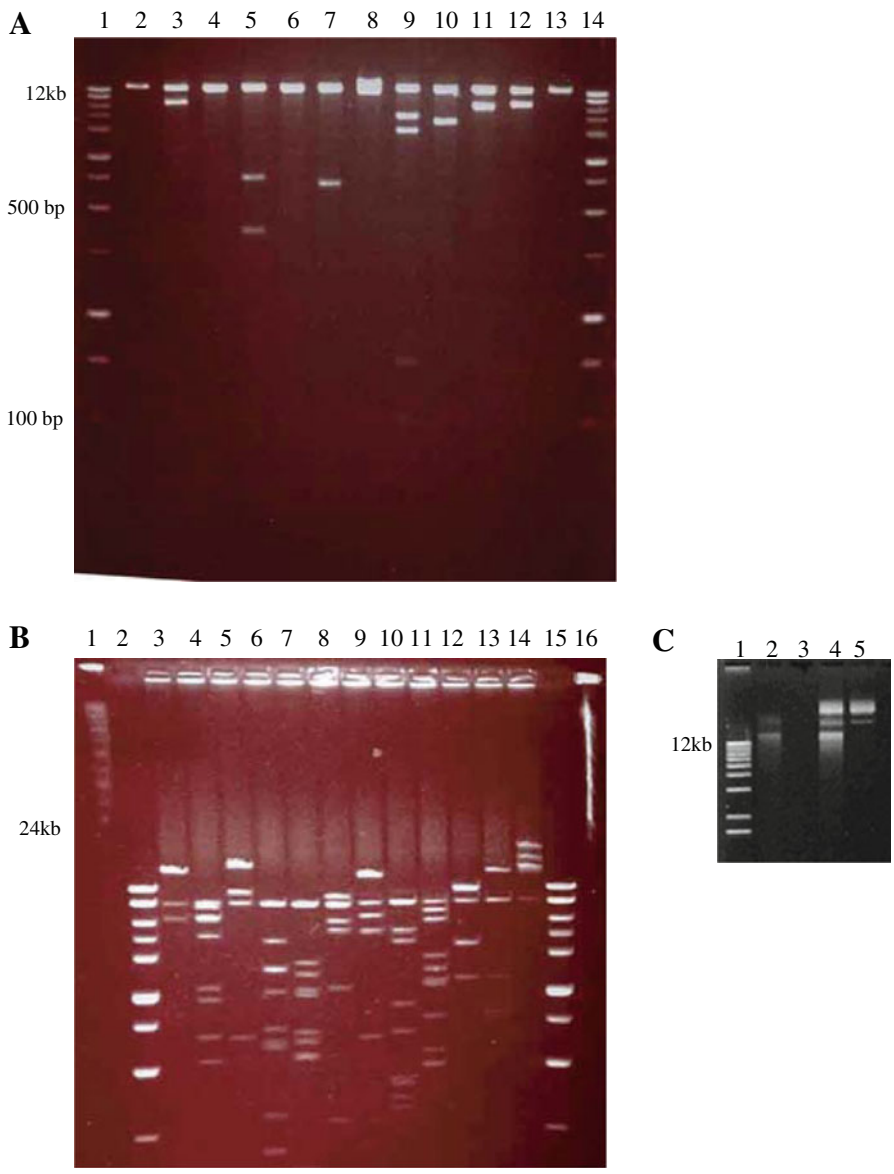

Fig. 1 **a** Agarose gel of *Pst*I digested low copy number plasmid metagenomic library clones from Cerrado soil. *Lanes 1 and 14*: 1 kb plus ladder; *Lanes 2 and 13*: Low copy number plasmid vector without insert; *Lanes 3–12*: Low copy number plasmid vector with insert. **b** PFGE gel of *Not*I-digested fosmid metagenomic library clones from Cerrado soil. *Lanes 1and 16*: Middle range size markers; *Lanes 2 and 15*: 1 kb plus ladder; *Lanes 3–14* Fosmid vector with insert. **c** Agarose gel of DNA libraries treated with Plasmid-Safe ATP-Dependent DNase. *Lane 1*: 1 kb plus ladder; *Lane 2*: *E.coli* DNA without Plasmid-Safe ATP-Dependent DNase; *Lane 3*: *E.coli* DNA with Plasmid-Safe ATP-Dependent DNase; *Lane 4:* DNA Pool of Fosmid library without Plasmid-Safe™ ATP-Dependent DNase; *Lane 5*: DNA Pool of fosmid library with Plasmid-Safe ATP-Dependent DNase

of inserts in this library, a total of 10 clones were randomly sampled and digested with *Pst*I and the average size of inserts was 8 kb (Fig. 1a), giving a total size of this library of 1.2 Gbp DNA. In contrast, the first operon size metagenomic DNA library from native Cerrado soil using a fosmid vector consisted of approximately 65,000 clones. Aiming to determinate the average size of cloned inserts in the library, a total 12 clones were randomly selected and digested with *Not*I. The average size of inserts was 35 kb (Fig. 1b), giving a total size of this library of 2.3 Gbp of chromosomal metagenomic DNA. To remove the contaminating *E. coli* host chromosomal DNA the preparations were treated with Plamid Safe ATP-dependent DNAse (Fig. 1c), see "Materials and methods" for details.

To validate the libraries constructed, two approaches were taken: end-sequencing of randomly selected clones and amplification and sequencing of 16S rRNA genes using the libraries' DNA as template. From the low copy number plasmid metagenomic library, the ends of a total of 20 clones were randomly chosen for end-sequencing. High quality sequences were obtained for only 14 clones. Of the 14 clones

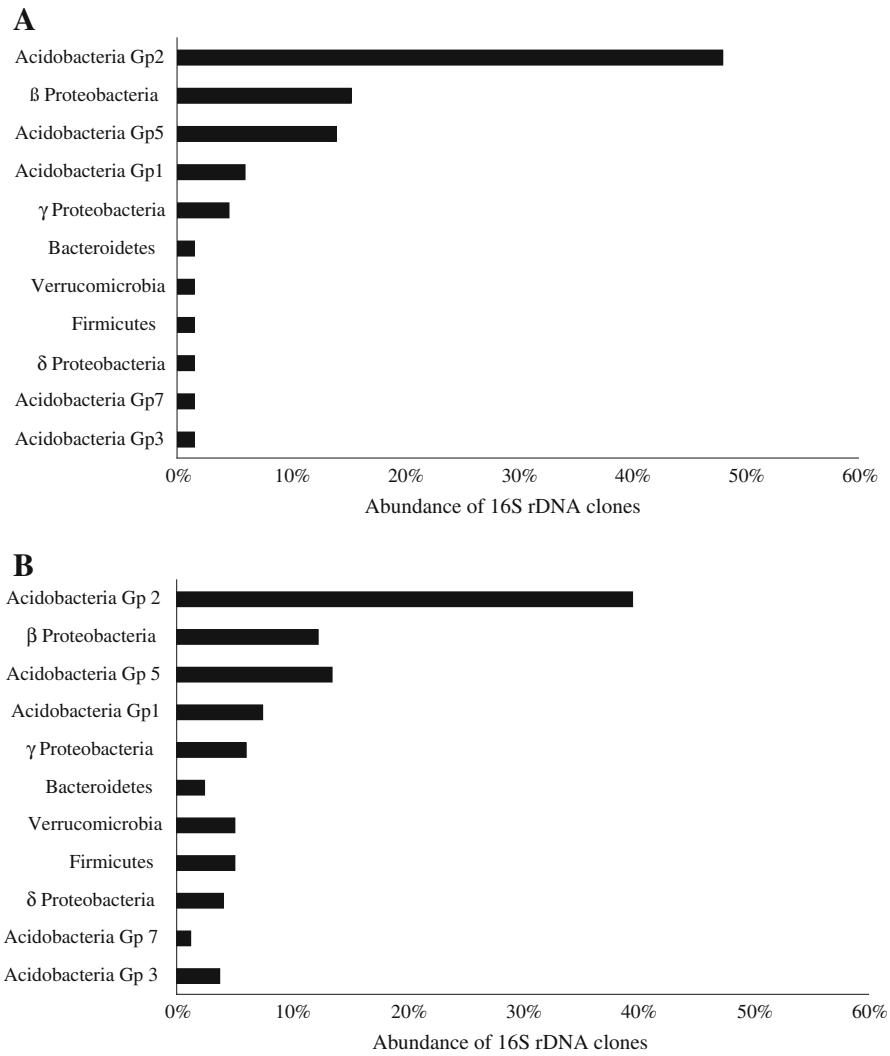

Fig. 2 Phylogenetic distribution of 16S rRNA gene clones amplified from Cerrado soil metagenomic DNA. **a** Low copy number plasmid. **b** Fosmid metagenomic library. 16S rRNA gene fragments of approximately 1.4 kb were amplified with primers specific for the domain Bacteria 27F/1492R using as template the DNA pool from plasmid and fosmid DNA from metagenomic libraries. The 16S rRNA gene amplicons were subsequently cloned into the TA cloning vector pGEM-T. The chimeric artifacts were removed using the Bellerophon software. The tool Classifier from RDP identified five bacterial phyla in both metagenomic libraries the Cerrado with a confidence threshold of 95%

sampled, seven sequences belonged to the Proteobacteria phylum (Supplementary Table 1). The remaining sequences of bacteria belonged to the phyla Actinobacteria, Firmicutes, Planctomycetes, Deinococcus-Thermus, and Acidobacteria.

For the fosmid metagenomic library, a total of 20 clones were randomly chosen to be analyzed by end-sequencing. Only two clones yielded low quality sequences and these were removed from analysis. High quality sequences from these reactions were further processed. Out of 18 clones sampled, 9 were ascribed to the phylum Proteobacteria (Supplementary Table 1). The remaining bacterial sequences belonged to phyla that included Acidobacteria, Cyanobacteria, Firmicutes, and Actinobacteria. Based on the analysis of these sequences, representatives of Archaea were not found in the both metagenomic libraries. This absence suggests scarcity of Archaea in the Cerrado soil samples studied, however, this needs to be further investigated.

The vast majority of DNA end-sequencing from both the low copy number plasmid metagenomic library and fosmid metagenomic library apparently belong to Proteobacteria. The data from our brief survey are in agreement with Rapp and Giovannoni (2003), who showed that this phylum was the most dominant phylum of microbial life in Cerrado soil. Moreover, this phylum is also prevalent and ubiquitous in other types of soils (Janssen 2006).

The second more refined, strategy for library validation and briefly surveying the diversity represented in the Cerrado soil libraries was to construct

and sequence clones from a 16S rRNA gene library generated using the DNA from constructed libraries as templates. A total of 192 16S rRNA gene clones were randomly chosen to be analyzed by sequencing. After removal of chimeric sequences and those of low quality, a total of 146 (81 sequences from fosmid library and 65 sequences from low copy number plasmid library) were analyzed. The tool Classifier from RDP identified five bacterial phyla in both metagenomic libraries the Cerrado with a confidence threshold of 95% (Fig. 2). There was an evident enrichment for Acidobacteria group 2. The great abundance of Acidobacteria in this soil sample can be explained by the acidity of this soil with pH near the 4.5 (Table 1). Sequences of the 16S rRNA gene of this group have been found in various locations around the world such as Asia, Australia, Americas and Europe suggesting that this phylum is an important constituent of many ecosystems, particularly soil.

Conclusion

Our results demonstrate the feasibility of extraction and cloning of DNA from soil samples that contain a high level of clay (55%) and iron (146 ppm) and at low pH (4.7). To our knowledge, this is the first study describing the extraction and purification of high quality DNA from soil with these characteristics. This is will allow the utilization of culture independent techniques to describe and access the metabolic potential of the microbial communities present in Cerrado soil, as well as those found in other soils throughout the world with similar characteristics.

Acknowledgment This work was supported by grants from FAPDF and CNPq.

References

Aremu MO, Atolaiye BO, Labaran L (2010) Environmental implication of metal concentrations in soil, plant foods and pond in area around the derelict udege mines of Nasarawa states, Nigeria. Bull Chem Soc Ethiop 24:351–360

Borneman J, Skroch PW, O'Sullivan KM, Palus JA, Rumjanek NG, Jansen JL, Nienhuis J, Triplett EW (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. Appl Environ Microbiol 62:1935–1943

Brady N, Weil RR (2002) The nature and properties of soils, 13th edn. Pearson education, Inc, Upper Saddler River, New Jersey

Davis JA (1984) Complexation of trace metals by adsorbed natural organic matter. Geochim Cosmochim Acta 48:679–691

Dubinsky EA, Silver WL, Firestone MK (2010) Tropical forest soil microbial communities couple iron and carbon biogeochemistry. Ecology 91:2604–2612

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002) Isolation of antibiotics turbomycin and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68:4301–4306

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Appl Environ Microbiol 66:5488–5491

Harrison PM, Arosio P (1996) The ferritins: molecular properties, iron storage function and cellular regulation. Biochim Biophys Acta 1275:161–203

He J, Xua Z, Hughes J (2005) Pre-lysis washing improves DNA extraction from a forest soil. Soil Biol Biochem 37:2337–2341

Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20:2317–2319

Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA geneand 16S rRNA genegenes. Appl Environ Microbiol 72:1719–1728

Jenne EA (1988) In: Baker, RA (ed) Trace inorganics in water. Advance Chemical Series, vol 73. Johnson Books, Boulder

Krsek M, Wellington EM (1999) Comparison of different methods for the isolation and purification of total community DNA from soil. J Microbiol Methods 39:1–16

Newman JR, Fuqua C (1999) Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli araBAD* promoter and the *araC* regulator. Gene 227:197–203

O'Neill B, Grossman J, Tsai MT, Gomes JE, Lehmann J, Peterson J, Neves E, Thies JE (2009) Bacterial community composition in Brazilian anthrosols and adjacent soils characterized using culturing and molecular identification. Microb Ecol 58:23–35

Purdy KJ (2005) Nucleic acid recovery from complex environmental samples. Methods Enzymol 397:271–292

Rapp MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394

Sagova-Mareckova ML, Cermak L, Novotna J, Plhackova K, Forstova J, Kopecky J (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Appl Environ Microbiol 74:2902–2907

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Williamson LL, Borlee BR, Schloss PD, Guan C, Allen HK, Handelsman J (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl Environ Microbiol 71:6335–6344