MOLECULAR ANALYSIS OF THE BACTERIAL DIVERSITY IN A SPECIALIZED CONSORTIUM FOR DIESEL OIL DEGRADATION

Douglas Antonio Alvaredo Paixão(2), Maurício Rocha Dimitrov(3), Rodrigo Matheus Pereira(4), Fábio Raphael Accorsini(5), Maria Benincasa Vidotti(5) & Eliana Gertrudes de Macedo Lemos(6)

SUMMARY

Diesel oil is a compound derived from petroleum, consisting primarily of hydrocarbons. Poor conditions in transportation and storage of this product can contribute significantly to accidental spills causing serious ecological problems in soil and water and affecting the diversity of the microbial environment. The cloning and sequencing of the 16S rRNA gene is one of the molecular techniques that allows estimation and comparison of the microbial diversity in different environmental samples. The aim of this work was to estimate the diversity of microorganisms from the Bacteria domain in a consortium specialized in diesel oil degradation through partial sequencing of the 16S rRNA gene. After the extraction of DNA metagenomics, the material was amplified by PCR reaction using specific oligonucleotide primers for the 16S rRNA gene. The PCR products were cloned into a pGEM-T-Easy vector (Promega), and *Escherichia coli* was used as the host cell for recombinant DNAs. The partial clone sequencing was obtained using universal oligonucleotide primers from the vector. The genetic library obtained generated 431 clones. All the sequenced clones presented similarity to phylum *Proteobacteria*, with *Gammaproteobacteria* the most present group (49.8% of the clones), followed by *Alphaproteobacteria* (44.8%) and *Betaproteobacteria* (5.4%). The *Pseudomonas* genus was the most abundant in the metagenomic library, followed by the *Parvibaculum* and the *Sphingobium* genus, respectively. After partial sequencing of the 16S rRNA, the diversity of the bacterial consortium was estimated using DOTUR software. When comparing these sequences to the
database from the National Center for Biotechnology Information (NCBI), a strong correlation was found between the data generated by the software used and the data deposited in NCBI.

Index terms: PCR, cloning, 16S rRNA, metagenomics, soil, biosurfactants.

RESUMO: ANÁLISE MOLECULAR DA DIVERSIDADE BACTERIANA DE UM CONSÓRCIO DEGRADADOR DE ÓLEO DIESEL

O óleo diesel é um composto derivado do petróleo, constituído basicamente por hidrocarbonetos. Condições precárias no processo de transporte e armazenagem desse produto contribuem significativamente para derrames acidentais, ocasionando sérios problemas ecológicos no solo e água, alterando assim toda a diversidade microbiológica do ambiente. A estratégia de clonagem e sequenciamento do gene 16S rRNA é uma das técnicas moleculares que permitem estimar e comparar a diversidade microbiana de diferentes amostras ambientais, seja elas impactadas ou não. O objetivo deste trabalho foi estimar a diversidade de microrganismos pertencentes ao domínio Bacteria em um consórcio degradador de óleo diesel por meio de sequenciamento parcial do gene 16S rRNA. Após extração do DNA metagenômico, o material foi amplificado por reação de PCR com oligonucleotídeos iniciadores específicos para o gene 16S rRNA. Os produtos da reação de PCR foram clonados em vetor pGEM T Easy (Promega) e transformados em células competentes de Escherichia coli. O sequenciamento parcial dos clones foi feito com oligonucleotídeos universais do vetor. A biblioteca obtida gerou 431 clones. Todos os clones mostraram similaridade com o filo Proteobacteria, onde as Gamma proteobacteria compreenderam o grupo de maior representatividade, com 49,8 % dos clones, seguida das Alphaproteobacteria, com 44,8 %, e das Betaproteobacteria, com 5,4 %. O gênero Pseudomonas destacou-se como representante com maior frequência de clones na biblioteca, seguido pelos gêneros Parvibaculum e Sphingobium. Após o sequenciamento parcial do gene 16S rRNA, a diversidade bacteriana do consórcio foi estimada utilizando-se o software DOTUR. Essas sequências, quando comparadas com as do banco do National Center for Biotechnology Information (NCBI), mostraram grande correlação entre os dados gerados pelo software utilizado e aqueles depositados no NCBI.

Termos de indexação: PCR, clonagem, 16S rRNA, metagenoma, solo, biosurfactantes.

INTRODUCTION

Brazil is one of the largest producers of petroleum oil in the world. According to the Brazilian National Oil Production agency (ANP), in 2008 Brazil produced approximately 96,371,606 million m³ of this oil. As a direct consequence of these high production, concerns regarding the risk of accidents such as oil spillage and environmental contamination became prominent, especially concerns regarding possible contamination of soil and subterranean water sources. According to research done by a Brazilian state agency responsible for water distribution in 2007 (Cetesb), fuel sale stations are those with the highest potential to generate soil and water contamination, being responsible for 77% of the notifications concerning contamination of water sources (Cetesb, 2007).

Diesel oil is composed of a mixture of alkenes and aromatic compounds which are frequently found as contaminants in soil following accidental oil spillage (Gallego et al., 2001). To restore the original quality of the contaminated soil area, many techniques are used in the process of removing the hydrocarbons. Some of these techniques use physical and chemical processes, such as the incineration and washing of the affected soil. The use of microorganisms capable of degrading toxic compounds by the process denominated bioremediation can contribute to the recovery of contaminated soil (Bento et al., 2005b). This technique uses endogenous microorganisms previously isolated from contaminated areas to promote the removal of pollutants from natural environments and the conversion of these pollutants into inert substances (Bamforth & Singleton, 2005). It is a method that can be very useful and practical for the removal of certain pollutants, although successful results depend on the combining of the microorganisms with adequate environmental factors at the site, such as nutrient, oxygen and pH levels, as well as local temperatures.

A varied group of microorganisms stand out regarding their capacity to process and decompose
Beijerinckia, genera species from genera hydrocarbons derived from oil. These include bacteria Preto County, São Paulo State (Brazil), at 21 ° 06 ' S, collected from a contaminated site located in Ribeirão. However, the use of microbial consortium has been found as being more effective in the degradation of these compounds (Jacques et al., 2007).

Many genetic tools are used to typify microbial communities from many environments without the need of cultivation (Kovalchuk et al., 2004). Of these, the sequencing of gene 16S rRNA is of much use. This specific gene is formed by a primary structure composed of areas which alternate between constant and variable regions, and allows a broad investigation for the determination of phylogenetic relationships (Ludwig & Schleifer, 1994). The precision of the phylogenetic inference obtained depends not only on the number of bases that are compared, but also on the areas that are compared (Hurst et al., 2002). Thus, the partial sequencing of the 16S rRNA gene is an important tool for analysis of the diversity of microorganisms existing in bacterial communities from environmental samples. The use of this technique has provided very good results in estimating and comparing bacterial communities from different samples and different environments (Kuske et al., 1997; Pereira et al., 2006; Silveira et al., 2006; Val-Moraes et al., 2009).

Obtaining better knowledge and understanding the microorganism diversity pertaining to bacterial communities from affected environments is important and imperative to attain economical and strategic benefits such as those that would be obtained with the discovery of microorganisms that present potential for being used in biotechnological processes for bioremediation of contaminated areas. The present work used partial sequencing of the 16S rRNA gene to estimate the diversity of microorganisms of the Bacteria domain in a consortium specialized in diesel oil degradation.

MATERIALS AND METHODS

Obtaining the microbe consortium

The microbe consortium assessed was developed and provided by researcher Maria Benincasa Vidotti, PhD, from the Department of Biology applied to the agricultural and animal husbandry development of the FCAV/UNESP, who kindly allowed its use in the present study.

The soil used for obtaining the consortium was collected from a contaminated site located in Ribeirão Preto County, São Paulo State (Brazil), at 21 ° 06 ' S, 47 ° 49 ' O, which held, in the past, a factory that manufactured automotive lubricants. The subsamples of the soil were collected randomly, at 0–20 cm of depth, following a zigzagging pattern. After the collected samples were homogenized, and from the total amount of soil collected (approximately 50 g), 1 g was added to 50 mL of sterile mineral Bushnell-Haas Broth medium (B.H.B) composed of (g L-1): MgSO4, 0.2; KH2PO4, 1.0; CaCl2, 0.02; (NH4)2HPO4, 1; KNO3, 1; FeCl3, 0.05 (Bushnell & Haas, 1941), and maintained for 12 h at 30 °C under constant agitation at 150 rpm. Afterwards, the suspensions were left to rest for one hour for decantation of the solid phase. For enrichment, a 500 μL aliquot of the supernatant was inoculated into 50 mL of the sterile B.H.B medium containing 0.1 % (v/v) of diesel oil, sterilized by filtration through 0.22 μm membranes (Millipore), which was used as the only source of C. The consortium began to be harvested after seven days of the culture’s development.

Extraction of metagenomic DNA from the consortium and amplification of the 16S rRNA gene

The extraction of the metagenomic DNA from the consortium was done using a FastDNA® SPIN Kit for Soil (BIO 101-QUANTUM BIOTECHNOLOGIES), following the manufacturers instructions. Metagenomic DNA was used for the amplification of the 16S rRNA gene by PCR. In the reaction, the fD1 (position 8–27) and rD1 (position 1524–1540) initiating oligonucleotides from the 16S rRNA gene in Escherichia coli, strain K12, were used as indicated by Weisburg et al. (1991). Conditions added were modified as follows: PCR Buffer 1X [20 mmol L-1 Tris-HCl (pH 8.4), 50 mmol L-1 KCl], 200 mmol L-1 of each dNTP, 2 mmol L-1 MgCl2, 1.25 U Taq DNA Polimerase, 5 pmol of each initiator, 35 ng of metagenomic DNA and pure water completing 50 mL (final volume) of the reaction. For the PCR reaction an MJ Research Inc. thermocycler, model PTC - 200 was used with the following program: 94 °C for 2 min, 35 cycles at 94 °C for 30 sec, 55 °C for 50 sec, 72 °C for 2 min and 72 °C for 5 min. The amplicons generated were confirmed by electrophoresis in 1 % Agar gel containing 0.5 μg mL-1 ethidium bromide, buffer TBE (Tris 89 mmol L-1; 89 mmol L-1 boric acid and 2.5 mmol L-1 EDTA, pH 8.3) and visualized by a photo documenter Gel Doc 1000 (Bio Rad, USA) with UV light.

Cloning and partial sequencing of the 16S rRNA gene

The amplified product of the 16S rRNA gene was cloned in a pGEM-T Easy vector (Promega, Madison, WI, USA) following the manufacturer’s instructions. Competent DH5α Escherichia coli cells were used for the transformation (Hanahan, 1983). The transforming clones were selected in Petri dishes containing a Luria Bertani medium (LB) supplemented with ampicilin (100 μg mL-1).
The selected clones were harvested in 1 mL of a Circle Grow (CG) medium containing ampicillin (100 µg mL⁻¹) and were then cultivated for 22 h at 37 °C under constant agitation at 250 rpm. The plasmid DNA selected from the clones was extracted according to the method suggested by Sambrook & Russell (2001) and purified using MultiScreen filters (Millipore). The plasmids containing the 16S rDNA fragments were sequenced using 0.5 µL of DYEnamic ET Terminator Cycle Sequencing Kit (GE Healthcare); 10 pmol of the T7 sequence promoter initiating oligonucleotide (20 mer); 150 ng of plasmid DNA in a final volume of 10 µL. The sequencing reaction was done using micro plates which held 96 wells. The conditions in the thermocycler were as following: 2 min at 95 °C, 40 cycles of 95 °C for 30 sec, 50 °C for 15 sec, and 60 °C for 2 min. The samples were submitted to precipitation using 75 % isopropyl alcohol and then rinsed with 70 % ethanol. The amplicons were precipitated using MultiScreen filters (Millipore). The plasmids containing the 16S rDNA were analyzed by a Phred/Phrap program (Ewing et al., 1998a; Ewing & Green, 1998b) used for the detection of the level of confidence for each produced sequence. The sequencing process was repeated and analyzed using this program, obtaining ≥ 20 sequences of acceptable quality. The DNA sequences were then compared to the bank of sequences based on the “Gene IDs”. Each distinct sequence obtained was defined as an operational taxonomic unit (OTU) (Figures 1 and 2).

**Results and Discussion**

The metagenomic DNA isolated in the study was of satisfactory quality, found to be free of contamination by protein and RNAs, and produced molecular sizes of close to 20 Kb. The amplification of this material using the fD1/rD1 initiators generated 1,500 pb fragments, which were used for the construction of the 16S rRNA gene metagenomic library. The library generated 480 clones, of which 431 presented enough quality for the analysis and identification of phylogenetic groups.

Following the proposal of the second edition of *Bergey's trust* Manual, which classifies the taxonomic levels in Domain, Phylum, Class, Order, Family, Genera and Species, the classification employed by the *Classifier* program grouped the library sequences in the following manner: 1 phylum, 3 classes, 5 orders, 8 families and 11 genus. All the 16S rDNA sequences identified belong to the *Proteobacteria* phylum, with 419 (97.2 %) of these distributed between three distinct *Alpha*, *Beta* and *Gammaproteobacteria* classes. The other 12 sequences (2.8 %) were not able to be classified at the taxonomic class level.

Of the three identified classes, *Gammaproteobacteria* was the one with the highest number of components (49.8 % of the clones), followed by *Alphaproteobacteria* with 44.8 % and *Betaproteobacteria* with 5.4 %.

The comparison of the 431 16S rDNA sequences from the consortium library to the international gene bank (NCBI) showed that 55 sequences were distinct, representing 14 % of all the sequenced clones. This distinction was identified based on comparison to the sequence codes known as “Gene IDs”. Each distinct sequence obtained was defined as an operational taxonomic unit (OTU) (Figures 1 and 2).

The fifty five distinct sequences were compared to the ribosomal 16S rDNA sequence data bank of Database Project II with 95 % confidence, and showed similarity to genus: *Variovax, Flavimonas Stenotrophomonas, Cupriavidus, Sphingomonas, Bosea, Bordetella*, *Thermomonas, Achromobacter*, *Pseudomonos, Parivbaculum*, and *Sphingobium*. These last three genus were, respectively, the taxons presenting the highest frequencies regarding the number of clones from the library (Figure 3). Partial sequencing of the 16S rRNA gene is a reliable technique, and has been used frequently for the identification of microorganisms, these being from cultures (Toledo et al., 2009), or available from environmental samples (Nercessian et al., 2008).
On the other hand, the curve with grouping criterion that presented 90% of similarity attained its plateau when coming close to 20 taxonomic units. These results reveal that the number of 16S rDNA clones sequenced was close to the total number of individuals which composed the bacterial community sampled. These results agree with those observed when comparing the clone sequences with the gene bank of the National Center for Biotechnology Information (NCBI), in which the number of identified OTUs was 55. The correlation of the obtained data compared to that from the international gene bank, added to the data generated by the DOTUR program, confirms that the estimates regarding the diversity of microorganisms in the consortium was attained.

The assessment of the bacterial diversity through the sequencing of 16S rRNA shows that organisms from phylum *Proteobacteria* are a common presence in different types of soil (Kuske et al. 1997; Pereira, et al. 2006; Silveira et al, 2006; Val-Morais et al. 2009). Besides this find, studies of bacterial diversity reported in published literature.

All genus identified in this study had already been found to be involved in the degradation of hydrocarbons. The *Pseudomonas* group was the genus that most consistently appeared in the library, pertaining to 32.2% of the total clones, followed by genus *Parvibaculum* with 29% and *Sphingobium* with 12%.

The DOTUR software allowed the determination of the number of operational taxonomic units (OTUs). The results were also used for the construction of the rarefaction curve. The design of the rarefaction curve can be used to determine if the size of the samples taken was sufficient to reveal the total diversity of a community (Roesch et al., 2007). In the present study, two rarefaction curves were tested, represented by 97% and 90% of similarity, as criterion for grouping (figure 4). The analysis of the figure showed an initial ascendancy of both curves, which tended to attain a plateau when at approximately 60 taxonomic units in the curve that represented 97% of similarity.

On the other hand, the curve with grouping criterion that presented 90% of similarity attained its plateau when coming close to 20 taxonomic units. These results reveal that the number of 16S rDNA clones sequenced was close to the total number of individuals which composed the bacterial community sampled. These results agree with those observed when comparing the clone sequences with the gene bank of the National Center for Biotechnology Information (NCBI), in which the number of identified OTUs was 55. The correlation of the obtained data compared to that from the international gene bank, added to the data generated by the DOTUR program, confirms that the estimates regarding the diversity of microorganisms in the consortium was attained.

The assessment of the bacterial diversity through the sequencing of 16S rRNA shows that organisms from phylum *Proteobacteria* are a common presence in different types of soil (Kuske et al. 1997; Pereira, et al. 2006; Silveira et al, 2006; Val-Morais et al. 2009). Besides this find, studies of bacterial diversity

Figure 1. Phylogenetic tree based on the partial sequences of 16S rDNA from the consortium, generated by the MEGA program, version 4. The distance matrix was calculated using the “Neighbor-Joining” algorithm and the nucleotide substitution method “Kimura 2-P”, with a “bootstrap” of 1000 repetitions.
in soils sustaining intensive culture systems (Pereira et al. 2006; Val-Morais et al. 2009), as well as in areas of environmental preservation in which human interference does not occur, show that this phylum is present in the most predominant bacterial communities (Silveira et al. 2006). Similar results
can be observed when assessing soils that present a record of contamination by hydrocarbons (Wunsche et al. 1995; Ozaki et al. 2006).

The *Proteobacteria* phylum is considered to be one of the largest and most diverse of the Bacteria domain, occurring in the most varied environments, with approximately 528 genus already described (Garrity et al., 2004). Furthermore, the components of this phylum are considered as being efficient in hydrocarbon degradation, including aromatic, aliphatic and other components (Atlas, 1981). The diversity of this phylum in many different types of environments is associated to the large morphologic and metabolic variety of the organisms that compose this group (Pereira et al 2006), such as chemolithotrophs, chemoorganotrophs and phototrophs which use diverse strategies for obtaining energy (Canhos et al. 1997). Some studies also have associated the presence of these groups of bacteria to certain nutritional conditions in the soil, since some bacteria from class *Alphaproteobacteria* develop with more ease in fertile soil (Mccaig et al. 1999). Microbiologic evaluations such as the present add to the knowledge regarding ecology and bacterial diversity attained in this field of study, since it allows the assessment and monitoring of bacterial groups in the most varied environments, especially regarding different types of soil.

The identification of microorganisms that use hydrocarbons as a C source and that present tolerance to contaminants has been reported in areas with a history of contamination by petroleum hydrocarbons and in many different types of soil (Richard & Vogel, 1999; Van Beilen et al., 2002; Widada et al. 2002). The predominance of the *Pseudomonas* genus was expected, based on the number of published studies that describe the frequency of these in contaminated soils (Wunsch et al., 1995; Pirillo, 2008; Higashioka et al. 2009), as well as their efficiency in the degradation of hydrocarbons (Whyte et al., 1997, Obuekwe et al., 2008, Bishnoi et al., 2009). The tolerance and the predominance of this genus of bacteria in environments contaminated by hydrocarbon compounds could be associated to specific catabolic pathways which involve different genes and enzymes which transform complex molecules contained in petroleum into common intermediate products from their catabolic routes. Furthermore, a large number of *Pseudomonas* species which are capable of using petroleum hydrocarbons as a C source have been studied and their characteristics described (Whyte et al., 1997; Yuste et al., 2000; Dinkla et al., 2001).

Genus *Parvibaculum* is known as being composed by many species which are good consumers of surfactants, using these as sources of C and energy (Schlcheck et al., 2003). The production of this compound by other members of the consortium probably was beneficial for the permanence of this distinct group of microorganisms. Studies show that when associated in consortiums, microorganisms increase their potential for producing biosurfactants (Bento et al., 2005a). Furthermore, diesel and crude oil are considered to be excellent C sources for the production of biosurfactants by microorganisms (Ilori et al., 2005), with *Pseudomonas* species being excellent producers of biosurfactants (Rahman & Gakpe, 2008). According to Schleheck et al. (2000), with the LAS surfactant (Alquilbenzeno Linear Sulfonado) available as the only source of C and energy, it was easily degraded and assimilated by the isolated *Parvibaculum lavamentivorans* DS-1.

The present study allowed the identification of organisms which, isolated or associated in consortium, could potentially be explored in biotechnological processes for bioremediation which involve the degradation of hydrocarbons. Although no new group of microorganisms was identified in the processes for decomposing hydrocarbon residue, it was possible to quantify, in sampling terms, the taxons which were represented in more significant numbers.

**CONCLUSIONS**

1. The consortium assessed demonstrated a predominance of organisms from phylum *Proteobacteria*, with the Alpha, Beta and Gamma classes as the most prominent.

2. The predominant clones in the library showed similarities to genus *Pseudomonas*, *Parvibaculum* and *Sphingobium* respectively.

3. Partial sequencing of gene 16S *rRNA*, together with the analysis done by the DOTUR software, was capable of estimating diversity of the bacterial communities in the assessed consortium.

**LITERATURE CITED**


