JOURNAL
OF
MICROBIOLOGY
AND
BIOTECHNOLOGY

© The Korean Society for Microbiology and Biotechnology

Investigation of Bacterial Diversity in Membrane Bioreactor and Conventional Activated Sludge Processes from Petroleum Refineries Using Phylogenetic and Statistical Approaches

Silva, Cynthia C.[1], Ederson C. Jesus[2], Ana P. R. Torres[3], Maira P. Sousa[3], Vânia M. J. Santiago[3], and Valéria M. Oliveira[1]*

[1]*Division of Microbial Resources, Research Center for Chemistry, Biology and Agriculture (CPQBA), Campinas University - UNICAMP, CP 6171, CEP 13081-970, Campinas, SP, Brazil*
[2]*Center for Microbial Ecology, Michigan State University, East Lansing, MI, 488242, U.S.A.*
[3]*PETROBRAS R&D Center, Cidade Universitária, Quadra 7, Ilha do Fundão, Rio de Janeiro, RJ, CEP 21949-900, Brazil*

Bacterial diversity of two distinct wastewater treatment systems, conventional activated sludge (CAS) and membrane bioreactor (MBR), of petroleum refineries were investigated through 16S rRNA gene libraries. Sequencing and phylogenetic analysis showed that the bacterial community composition of sludge samples was distinct between the two wastewater treatment systems. MBR clones belonged predominantly to Class *Betaproteobacteria*, represented mainly by genera *Thiobacillus* and *Thauera*, whereas CAS clones were mostly related to Class *Alphaproteobacteria*, represented by uncultured bacteria related to Order *Parvularculales*. Richness estimators ACE and Chao revealed that the diversity observed in both libraries at the species level is an underestimate of the total bacterial diversity present in the environment and further sampling would yield an increased observed diversity. Shannon and Simpson diversity indices were different between the libraries and revealed greater bacterial diversity for the MBR library, considering an evolutionary distance of 0.03. LIBSHUFF analyses revealed that MBR and CAS communities were significantly different at the 95% confidence level ($P \leq 0.05$) for distances $0 \leq D \leq 0.20$. This work described, qualitatively and quantitatively, the structure of bacterial communities in industrial-scale MBR and CAS processes of the wastewater treatment system from petroleum refineries and demonstrated clearly differentiated communities responsible for the stable performance of wastewater treatment plants.

Keywords: Wastewater treatment, sludge, bacterial community, 16S rRNA genes library

*Corresponding author
Phone: +55 (19) 2139-2874; Fax: +55 (19) 2139-2852;
E-mail: vmaia@cpqba.unicamp.br

Petroleum refineries are great water consumers, since the refining process consists of a series of steps of petroleum amelioration classified as separation or conversion processes that use great volumes of water or aqueous solutions [43]. This situation contributes to worsening the picture of environmental pollution, with potential hazards to the water quality. As a consequence, strict regulations were created in Brazil to rule the use of water resources, leading the productive sectors to rationalize water consumption [40].

In that sense, petroleum refineries have made a considerable effort to treat their effluents, which may pose a difficult task because of the complexity of their chemical composition [37]. Compounds such as phenols and mercaptans are the main constituents of refineries of petrochemical industries, and may be harmful to human health and to the environment, even in low concentrations, as a consequence of the toxicity derived from the synergic effects of their interaction [6, 41]. In this context, treated effluents may still present recalcitrant compounds and some chronic toxicity, which make water reuse limited.

Biological treatment, or bioremediation, of wastewaters uses microorganisms to mineralize complex pollutant organic and inorganic compounds, transforming them into inert or nontoxic compounds. This technology has been broadly employed in petroleum refineries as an international practice, and has shown promising results for the efficient wastewater treatment and the establishment of the water reuse in such industries [37]. Additionally, the inefficiency, high costs, and operational complexity of the existing physicochemical treatments seem to stimulate the intense application of bioremediation. Conventional activated sludge (CAS) and membrane bioreactor (MBR) processes have

been widely used in the biological treatment of petrochemical wastewater [49, 57]. The CAS process usually presents three stages: sedimentation of gross solids in the influent followed by aerobic degradation of the organic matter, and then a secondary clarifier to remove the biomass. The MBR is one of the next-generation wastewater treatment processes and consists of a suspended growth-activated sludge system that utilizes microporous membranes for solid/liquid separation instead of secondary clarifiers. This system presents a high sludge age that gives the bacteria time to adapt to pollutant compounds and a low food/microorganism ratio that is expected to force bacteria to also mineralize poorly degradable organic compounds [27, 59], resulting in an effluent water quality that is significantly higher than the one generated by conventional treatment.

However, despite the great relevance of biological treatments to industries, most of the related studies have focused mainly on treatment performance and associated parameters, without giving proper attention to the microbial communities involved. Communities of bacteria present in activated sludge or biofilm reactors are responsible for most of the carbon and nutrient removal from the effluent and represent the major component of every biological wastewater treatment plant. Thus, a better understanding of the microbial ecology in these systems is required to reveal factors influencing the efficiency and stability of biological treatments and to develop strategies for improved process performance.

Microbial communities in environmental samples have been traditionally characterized by means of cultivation-based techniques or optical microscopy. These approaches allowed the selection of model organisms for the majority of microbiological processes in wastewater treatments [22]. Although very useful for taxonomical, physiological, and genetic studies, these techniques are inadequate for a proper description of the composition and dynamics of the microbial communities, since it is now well known that only a small fraction (0.1 to 10%) of the microbial diversity in nature can be recovered in the laboratory [1, 24].

Molecular tools developed in the last decades to overcome the limitations imposed by traditional cultivation techniques have been used to study bacterial diversity in wastewater biological treatments [33]. These studies demonstrated that most of the suggested model organisms are functionally not relevant in the bioreactors, and that other microorganisms, some of them yet uncultured, may be related to key ecological processes that take place *in situ*.

Based on previous knowledge that MBR and CAS processes produce effluent of different quality, this work aimed at a cultivation-independent characterization and comparison of the bacterial communities present in sludge samples originated from wastewater treatment plants of petroleum refineries operating in conventional activated sludge and membrane bioreactor systems.

MATERIALS AND METHODS

Conventional Activated Sludge and Membrane Bioreactor Processes

We investigated two wastewater treatment plants of petroleum refinery with operation in continuous flow. The influent of both plants had similar composition (data not shown). The first one has been in continuous operation for more than 10 years and is part of the industrial Wastewater Treatment Plant (WWTP) of Petrobras Refinery REVAP (São Paulo State, Brazil). It comprises a conventional activated sludge process as a secondary treatment, with an anoxic/oxic zone tank and a secondary clarifier. The treatment capacity of the WWTP is about 16,800 m³/d (700 m³/h) and the HRT (hydraulic retention time) is around 6.5 h. The second one is a submerged membrane bioreactor pilot that has been in continuous operation for 18 months and is part of the industrial WWTP of Petrobras Refinery REGAP (Minas Gerais State, Brazil). The treatment capacity of the MBR is about 24 m³/d (1 m³/h) and the HRT is around 8 h. This bioreactor operates with a microfiltration flat-sheet membrane module (Kubota, Osaka, Japan) with nominal porosity of 0.4 μm. This bioreactor effective volume is 8 m³ and aeration is continuously carried out. Filtration was intermittently carried out in the constant flow rate operation mode.

We considered the common analytical factors measured in the routine of the plants. The physicochemical analyses were performed during the month of sampling. COD (chemical oxygen demand), BOD (biochemical oxygen demand), VSS (volatile suspended solid), ammonium (NH_4^+-N) concentration, and alkalinity were measured according to the procedures described in Standard Methods [12]. Measures of pH were recorded by using a pH meter (Thermo Orion, Model 868; U.S.A.).

Sampling and DNA Extraction

Sludge samples were collected after 18 months of continuous operation at REGAP and after 10 years of continuous operation at REVAP, with logistic support of CENPES/PETROBRAS. Samples were collected in triplicate using 1-l sterilized bottles, which were kept on ice during transportation to the laboratory and stored at -20°C for subsequent community DNA extraction.

DNA extraction from sludge samples was carried out using a protocol based on Großkopf *et al.* [21] and Neria-Gonzáles *et al.* [42], with modifications, as follows: microbial cells were retrieved from sludge samples by centrifuging 4-ml aliquots at 13,400 ×g for 5 min. The pellets were suspended in 600 μl of PBS buffer and homogenized by vortex. The solution of lysozyme was added at a final concentration of 17 mg/ml and incubated at 37°C for 30 min, with manual shaking every 10 min. Proteinase K and SDS were subsequently added (final concentration of 0.7 mg/ml and 2%, respectively) and the solution was incubated at 60°C for 30 min, with manual shaking every 10 min. The tubes were subjected to 3 freeze–thaw cycles, being 2 min in liquid nitrogen followed by 2 min at 65°C. An equal volume of saturated phenol (pH 8.0) was added to the solution, followed by homogenization and centrifugation at 13,400 ×g for 5 min. The supernatant was extracted with an equal volume of chloroform:isoamylic alcohol (24:1) and centrifuged at 13,400 ×g for 5 min. For DNA precipitation, 5 M NaCl (10% of total volume) and 2 volumes of cold ethanol were added to the solution. The pellet was washed once with 70% ethanol, dried, and suspended in 25 μl of Milli-Q H_2O.

The yield and integrity of DNA obtained from sludge samples were confirmed through electrophoresis in 1% agarose gel.

16S rRNA Gene Libraries

For 16S rRNA gene library construction, the amplification was performed using the bacterial primer set 27f and 1100r [31]. Fifty μl reaction mixtures contained 50–100 ng of total DNA, 2 U of *Taq* DNA polymerase (Invitrogen), 1× *Taq* buffer, 1.5 mM MgCl$_2$, 0.2 mM of dNTP mix (GE Healthcare), and 0.4 mM each primer. The PCR amplifications were done using an initial denaturation step of 2 min at 95°C, followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, and 3 min at 72°C, followed by a final extension step at 72°C for 3 min, in an Eppendorf thermal cycler.

Cloning and Sequencing of 16S rRNA PCR Products

Two 16S rRNA gene libraries were constructed, one for each wastewater treatment plant, REGAP and REVAP. Three PCR replicates of each sample were pooled, purified using GFX PCR DNA and Gel Band Purification kit (GE Healthcare), and concentrated in a speed vacuum concentrator 5301 Eppendorf, A-2-VC rotor. The purified PCR product (200 ng) was ligated into a pGEM-T Easy Vector (Promega), according to the manufacturer's instructions, and transformed into *E. coli* JM109 competent cells. Approximately 200 positive clones were selected from each library for subsequent sequencing. The 16S rRNA inserts were amplified from plasmid DNA of selected clones using the universal M13 forward (5'-CGC CAG GGT TTT CCC AGT CAC GAC-3') and M13 reverse (5'-TTT CAC ACA GGA AAC AGC TAT GAC-3') primers. PCR was performed in a 50-μl reaction volume, containing 1–2 μl of an overnight clone culture, 0.4 μM each primer, 0.2 mM dNTP mix, 2 U *Taq* DNA polymerase (Invitrogen), 1× *Taq* buffer, and 1.5 mM MgCl$_2$. The amplification program consisted of an initial denaturation step at 94°C for 3 min, followed by 30 cycles of 94°C/20 s, 60°C/20 s, and 72°C/90 s. PCR products were purified as previously described for automated sequencing in the MegaBace DNA Analysis System 500 (GE Healthcare). The sequencing was carried out using 10f and 1100r primers [30] and the DYEnamic ET Dye Terminator Cycle Sequencing Kit (GE Healthcare), according to the manufacturer's recommendations.

Sequence Analysis

Partial 16S rRNA sequences obtained from clones were assembled in a contig using the phred/Phrap/CONSED program [15, 19]. Identification was achieved by comparing the contiguous 16S rRNA sequences obtained with 16S rRNA sequence data from reference and type strains, as well as environmental clones available in the public databases of GenBank (http://www.ncbi.nlm.nih.gov) and RDP (Ribosomal Database Project - Release 9; http://rdp.cme.msu.edu/) using the BLASTn and Classifier routines, respectively. The sequences were aligned using the CLUSTAL X program [56] and analyzed with MEGA software v.4 [53]. Evolutionary distances were derived from sequence-pair dissimilarities calculated as implemented in MEGA, using Kimura's DNA substitution model [28]. The phylogenetic reconstruction was done using the neighbor-joining (NJ) algorithm, with bootstrap values calculated from 1,000 replicate runs, using the routines included in the MEGA software.

Statistical Analyses

Physicochemical data were submitted to statistical analysis by using the program SISVAR [17]. The normal distribution of data was tested with the Shapiro–Wilk Test. Analysis of variance was carried out on the data and the difference between means was assessed with the Student's t-test with Bonferroni correction. The VSS variable was transformed to its neperian logarithm before being subjected to the analysis of variance.

Diversity index calculations (α-diversity measures) were performed individually for both 16S rRNA gene libraries with the programs MOTHUR version 1.1.0 and DOTUR v. 1.3 [46]. DOTUR was used to assign sequences to OTUs and to calculate collector's curves using 2 distance levels among sequences (3 and 20%), and MOTHUR was used to calculate, for each sludge sample, the Shannon (H') diversity index, Simpson index [34], and the nonparametric richness estimators ACE (abundance-based coverage estimator) [8] and Chao1 [7]. Chao1 richness estimates were based on singletons and doubletons as described by Chao [7], and ACE was based on the distribution of abundant (>10) and rare (≤10) species.

Statistical comparisons of coverage of bacterial 16S rDNA libraries from the two sludge samples were performed using the LIBSHUFF computer program [51], which used the formula of Good [18] to generate homologous and heterologous coverage curves from the clone libraries. Sequences were randomly shuffled 999 times between samples prior to the distance between the curves being calculated using the Cramér–von Mises statistical test [44]. The DNADIST program of PHYLIP [16], using Kimura's model [28] for nucleotide substitution, was employed to generate the distance matrix analyzed by LIBSHUFF.

Nucleotide Sequence Accession Numbers

The 16S rDNA partial sequences determined in this study for the environmental clones were deposited at the GenBank database under the accession numbers FJ933260 to FJ933490.

RESULTS AND DISCUSSION

Physicochemical Characteristics of Effluent Water in CAS and MBR Processes

Physicochemical parameters of processes were analyzed during 30 days in the month of sampling (September 2005 at REGAP and March 2006 at REVAP) and the average values are shown in Table 1. Evolution of chemical oxygen demand, biochemical oxygen demand, alkalinity, ammonia nitrogen (NH$_3$-N), pH, and volatile suspended solid were used to monitor the performance of the CAS and MBR processes.

Statistical tests were applied over the average values of parameters to assess if the CAS and MBR processes were significantly different. Results demonstrated stability and efficiency in the performance during the month of sampling for both processes. Only VSS and alkalinity values were significantly different (Table 1). The VSS parameter in the MBR was approximately 2.5 times more than in the CAS. This parameter reflects the quantity of biomass; therefore, this difference could be explained by the complete biomass retention that occurs in the MBR process. Ammonia, BOD, and COD were efficiently removed in both processes (Table 1). In spite of the significant difference between alkalinity

Table 1. Average values of operational parameter characteristics of CAS and MBR.

Parameters	CAS			MBR		
	Influent	Effluent	Removal %	Influent	Effluent	Removal %
SSV (mg/l)[a]	3,500±500			8,506±504		
pH	7.81±0.60	7.60±0.20		8.15±0.57	6.60±0.30	
NH4+-N (mg/l)	41.8±14.3	1,1±1,1	97.4	46.9±17.9	2,4±1.1	94.9
COD (mg/l)	412±155	63±29	84.7	458±53	68±27	85.2
BOD (mg/l)	165±44	5±5	97.0	148±51	3±1	98.0
Alkalinity (mgCaCO₃/l)[a]	247±64	10±5	96.0	395±79	28±14	92.9

Values are given as mean concentration±standard deviation.
[a]Statistical significance at the 5% level.

values, the high values found for both processes allowed pH to be maintained around 7–8, necessary for efficient ammonia removal by nitrification.

Composition of the Sludge Bacterial Communities

The bacterial community composition of sludges from the two different petrochemical wastewater systems was determined by analysis of 16S rRNA gene clone libraries. Conventional activated sludge and membrane bioreactor systems were studied operating in industrial scale. A total of 231 partial sequences were obtained from both libraries (110 sequences from MBR and 121 from CAS), each consisting of 850 nucleotides on average, which presented good quality for subsequent analyses. These 16S rRNA sequences were compared with sequences from reference and type strains, as well as environmental clones, available at the GenBank and RDP II databases.

In the MBR sample, 11 major bacterial taxa were identified. Seventy-eight sequences (71%) were related to the phylum *Proteobacteria*, being 38 (48.7%) affiliated to Class *Betaproteobacteria*, 28 to Class *Alphaproteobacteria*, 10 to class *Deltaproteobacteria*, and 1 to class *Gammaproteobacteria* (Fig. 1A and 1B). The remaining sequences were significantly less numerous, being related to the phyla *Acidobacteria* (11), *Actinobacteria* (3), *Bacteroidetes* (2), *Chloroflexi* (3), *Firmicutes* (1), *Nitrospirae* (8), *Planctomycetes* (1), *Chlorobi* (1), Candidate Division TM7 (1), and *Verrucomicrobia* (2) (Fig. 1A). In the CAS sample, 7 major bacterial taxa were identified. The phylum *Proteobacteria* was the most abundant one, with 77 sequences (64%) distributed among the classes *Alphaproteobacteria* (44 sequences), *Betaproteobacteria* (15 sequences), *Gammaproteobacteria* (7 sequences), and unaffiliated *Proteobacteria* (11 sequences) (Fig. 1A and 1B). The remaining sequences were related to the phyla *Acidobacteria* (2), *Actinobacteria* (8), *Chlorobi* (3), *Choroflexi* (15), *Firmicutes* (3), and *Nitrospirae* (10), and three clones were considered unaffiliated (Fig. 1A). All taxonomic groups found in both libraries have already been described in the literature as commonly occurring organisms in wastewater treatment systems [27, 29, 49].

The composition of phylotypes at the phylum level was similar between the libraries from distinct wastewater treatments. All phyla found in the CAS library were present in the MBR library (Fig. 1A). The phylum *Proteobacteria* was predominant in both libraries, corroborating literature data that report the predominance of this group in sludge samples from sewage and industrial wastewater treatment systems [27, 29]. However, the abundance of clones in *Alpha-* and *Betaproteobacteria* classes was discrepant between the libraries (Fig. 1B).

Differences were more pronounced when comparing the orders present in each library. In the MBR library, the most
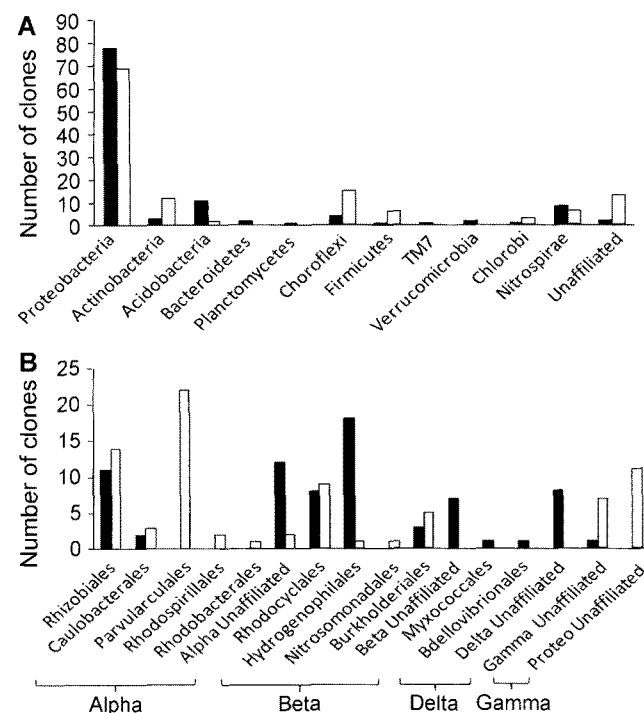

Fig. 1. Proportion of 16S rRNA gene clones from MBR (filled bar) and CAS (opened bar) processes among bacterial phyla (**A**) and orders within phylum *Proteobacteria* (**B**).
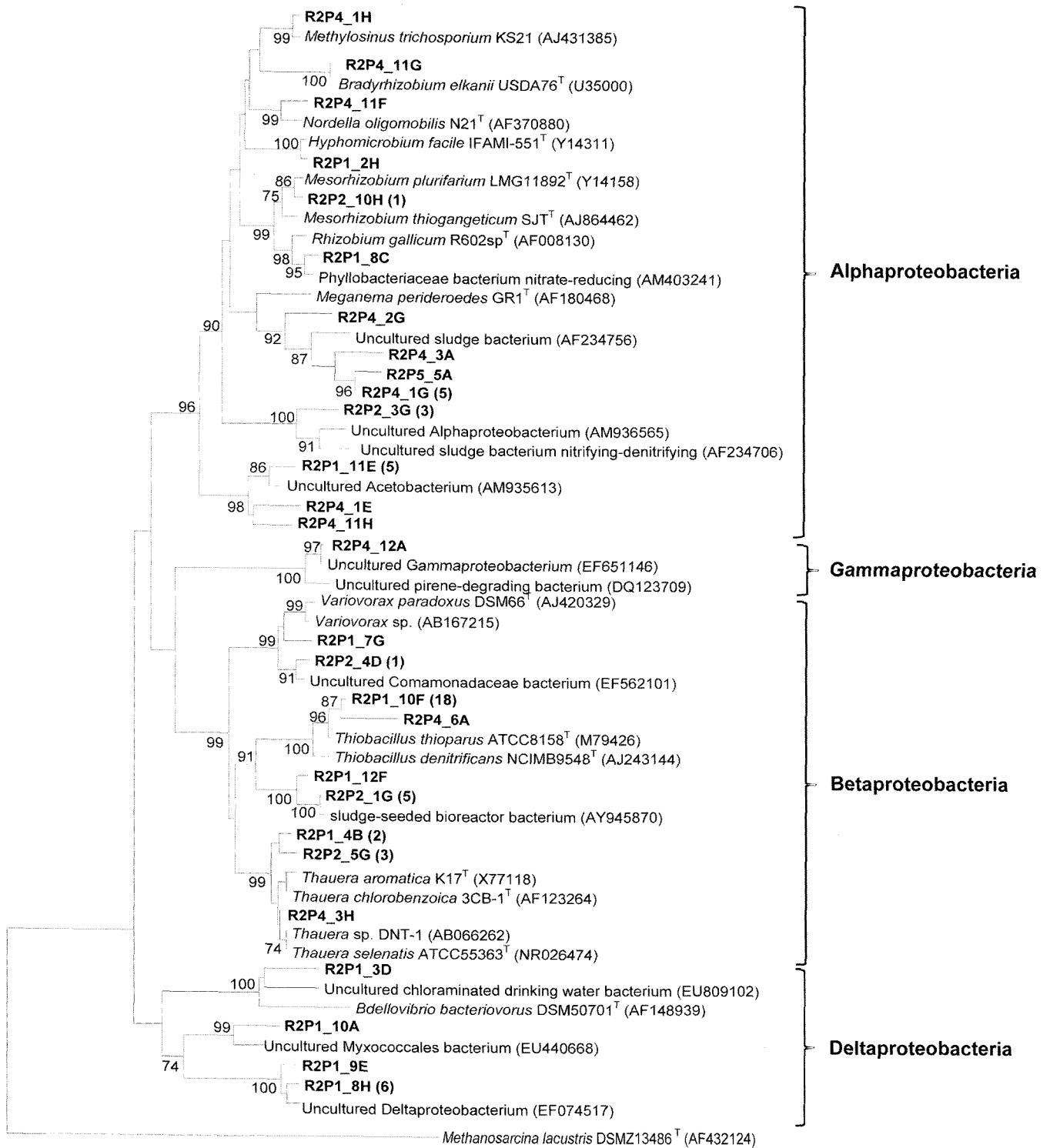Sequence classification was based on the RDP classifier and BLASTn results.

Fig. 2. Phylogenetic analysis based on partial 16S rRNA sequences from the MBR clone library (R2) representing members of phylum *Proteobacteria* and related species.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanosarcina lacustris* was used as outgroup.

abundant orders in the class *Betaproteobacteria* were *Hydrogenophilales* and *Rhodocyclales* (Fig. 1B), mainly represented by the genera *Thiobacillus* and *Thauera*, respectively (Fig. 2). These genera have been described in other studies of sludge diversity of wastewater treatment [5, 35, 48, 50, 52]. *Thauera* spp. have been reported as representing denitrifying bacteria in a wastewater treatment plant [55]. In addition, a recent work suggested that these bacteria may be explored as an important biological indicator for population optimization and performance monitoring in wastewater treatment processes [36]. Several researchers have demonstrated that *Thauera* species are related with aerobic and anaerobic degradation of aromatic compounds, such as phenol and toluene, by means of distinct catabolic pathways [35, 48]. The genus *Thiobacillus* is able to oxidize reduced inorganic or organic sulfur compounds commonly found in petroleum refining industry wastewaters (such as hydrogen sulfide, thiosulfate, methanethiol, ethanethiol, *etc.*) to sulfates [10, 52]. *Thiobacillus* strains have also been related to the degradation of polycyclic aromatic hydrocarbons (PAHs) [5]. Clones found in the MBR library were more closely related to *T. thioparus* (Fig. 2). These bacteria are used for the removal of hydrogen sulfide (H_2S) and other reduced sulfur compounds from petroleum refinery and sewage waste streams, in order to treat odorous emissions [2, 14]. Tanji *et al.* [54] described that immobilized cells of *Thiobacillus thioparus* were able to remove great quantities of dimethyl sulfide, methyl mercaptan, and hydrogen sulfide from sulfur-containing malodorous gas in laboratory-scale tower-type reactors. Clones affiliated to the order *Rhizobiales* in MBR library were as abundant as the ones affiliated to *Rhodocyclales* or *Hydrogenophyllales*, although more diverse, being related to bacteria capable of oxidizing methane (AJ431385), fixing nitrogen (AF041442), and reducing sulfate (AJ864462) (Fig. 2).

The results of the CAS library revealed that the dominant sequence types belonged to *Alphaproteobacteria*, which included the orders *Rhodobacterales*, *Caulobacterales*, *Rhodospirillales*, *Parvularculales*, and *Rhizobiales* (Fig. 1B). In this case, the majority of the clones were related to the order *Parvularculales*, but clone identification at higher taxonomic ranks was not possible by comparison of 16S rRNA gene sequences with databases and subsequent phylogenetic analyses, suggesting that the corresponding organisms may represent new taxa (Fig. 3). Literature data about *Parvularculales* are scarce, and the only valid species described to date is *Parvularcula bermudensis*, a bacterium isolated from a seawater sample [9]. In the class *Betaproteobacteria*, the order *Rhodocyclales* was the most abundant, mainly represented by *Thauera* spp. (Fig. 3). The phylum *Chloroflexi* was the second more abundant in the CAS library (Fig. 1A), accounting for 14% of total clones. Filamentous representatives of this phylum have

been commonly found in activated sludge wastewater treatment plants, and they have occasionally been associated with bulking incidences [3, 4]. Members of *Chloroflexi* are ecologically significant in the MBR process, since they are responsible for degradation of carbohydrates and cellular materials, which consequently reduces membrane fouling potential [38].

The remaining phyla, such as *Acidobacteria* and *Chlorobi* for the CAS library, and *Bacteroidetes*, *Firmicutes*, TM7, and *Planctomycetes* for the MBR library, showed low abundance in the bacterial communities under study (Fig. 4 and 5). Schloss and Handelsman [47] suggested that these poorly sampled members represent rare members in microbial communities, but may play significant functional roles in the microbial community. Other speculative explanation for the low abundance of such taxa would be PCR bias introduced by the use of bacterial primers designed based on rRNA sequences from cultured organisms, consequently amplifying environmental sequences similar to those already existing in the databases [45].

Richness Analyses

Collector's curves, also called species accumulation curves, plot the cumulative number of species recorded as a function of sampling effort [13]. Collector's curves illustrate the rate at which new species are found. The complete coverage of data set is expected to result in a plateau-shaped curve. In this study, collector's curves at the estimated phylum level (distance=0.20) reached clear saturation for the CAS library, suggesting that the sampling effort was sufficient to reveal all phyla present in this sample. For the MBR library, the curve tends to an asymptote, but additional sampling is necessary for saturation. At the species level (distance=0.03), the weakly curvilinear plots showed that phylotype richness in both wastewater libraries was underestimated, suggesting that further sampling from both libraries may have revealed more diversity at the species level (Fig. 6). These results were not expected, since the sampled refinery wastewater treatment plants consist of closed systems that receive large amounts of pollutant compounds and therefore tend to select specific populations capable of utilizing and/or degrading such compounds, thus reducing diversity. Other microbial population studies corroborate our results in the sense that the number of sampled clones was insufficient to reach saturation, when the OTU definition was set at 3% [20, 23]. Hughes *et al.* [25] argued that microbial diversity cannot be fully estimated because of the high bacterial diversity in environmental samples, small sample sizes, or both, resulting in linear or close to linear accumulation curves. On the other hand, exhaustive inventories of microbial communities still remain impractical for most laboratories of microbial ecology.

Nonparametric richness estimators, such as ACE (abundance-based coverage estimators) and Chao1, were
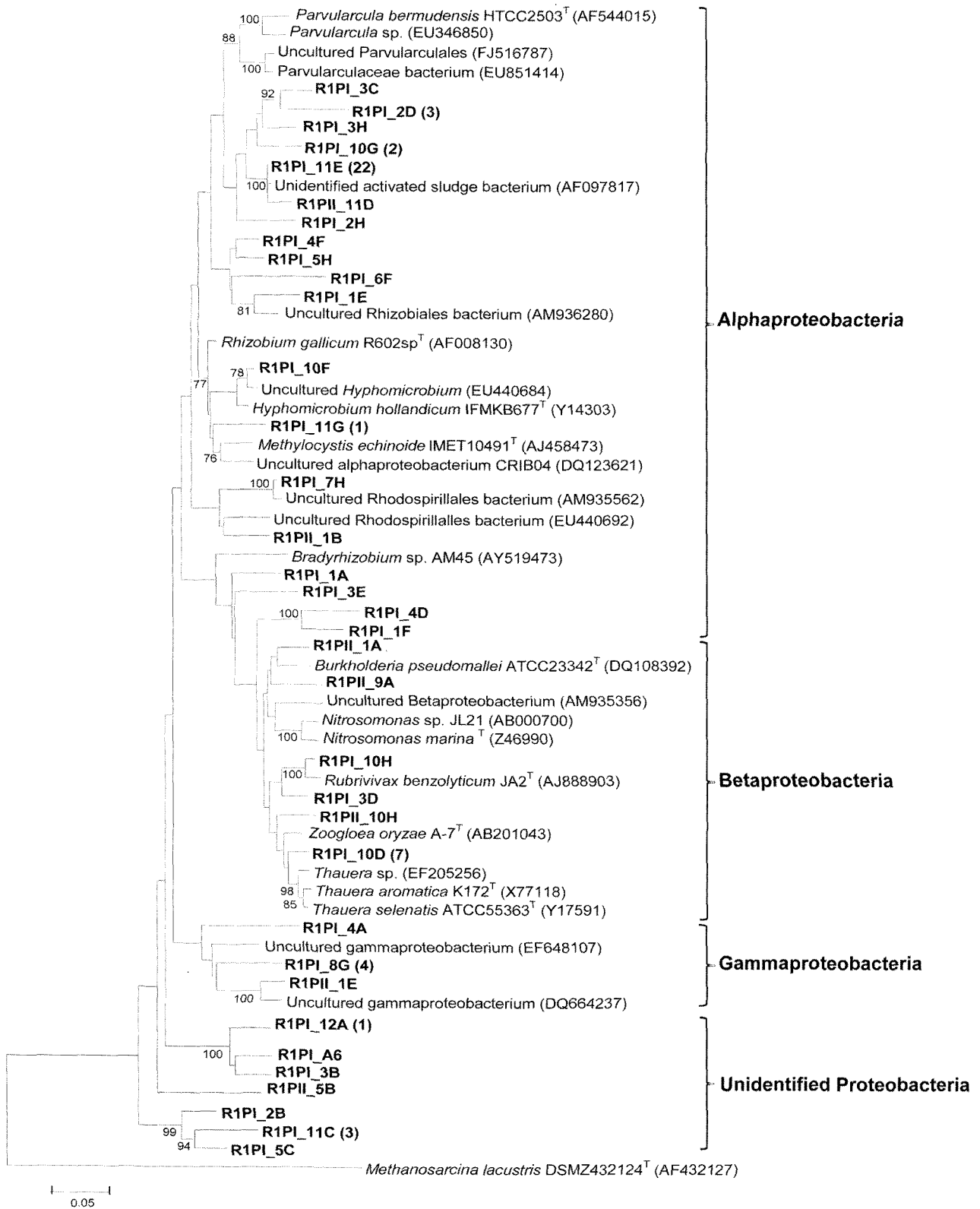
Fig. 3. Phylogenetic analysis based on partial 16S rRNA sequences from the CAS clone library (R1) representing members of phylum *Proteobacteria* and related species.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanosarcina lacustris* was used as the outgroup.
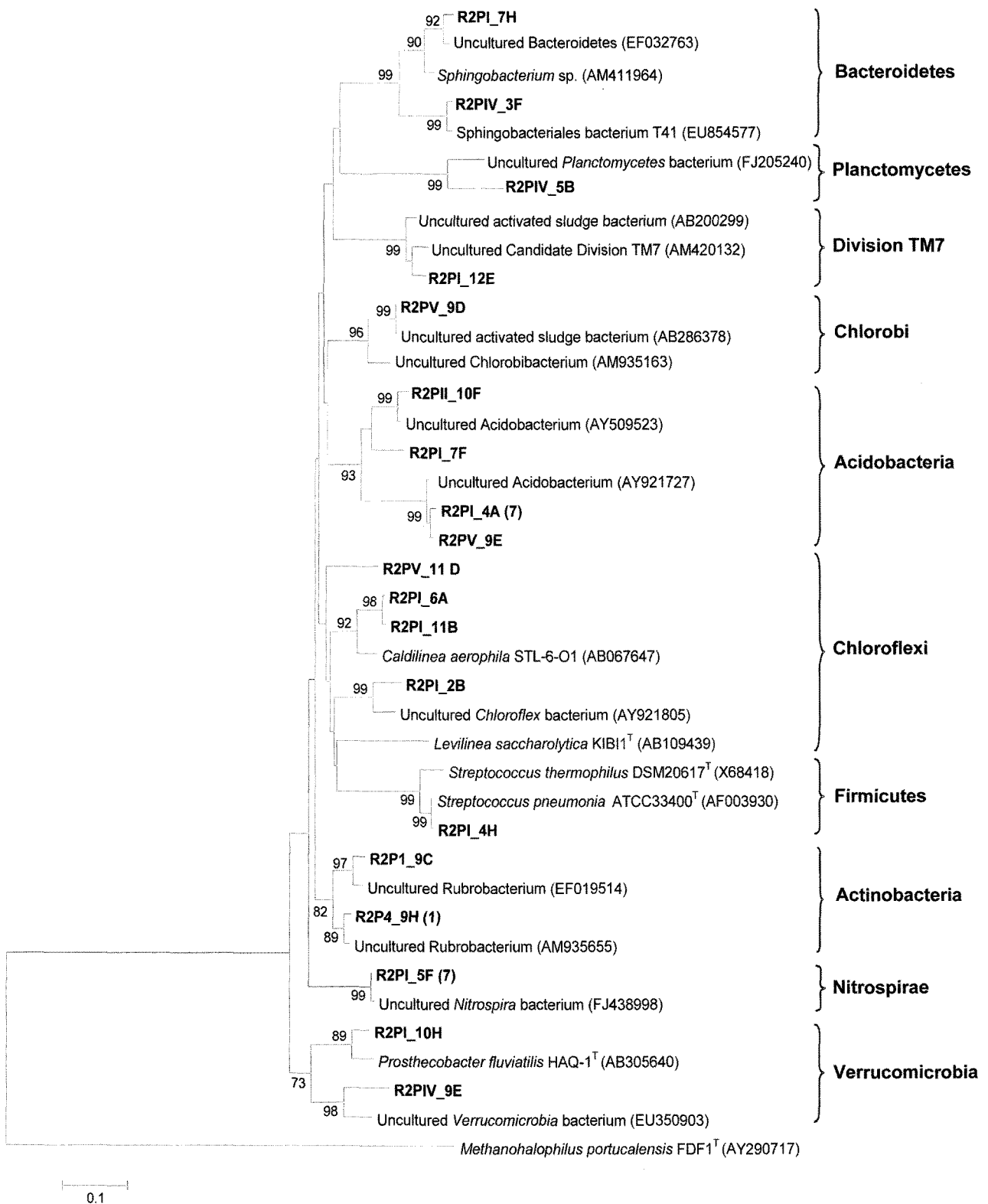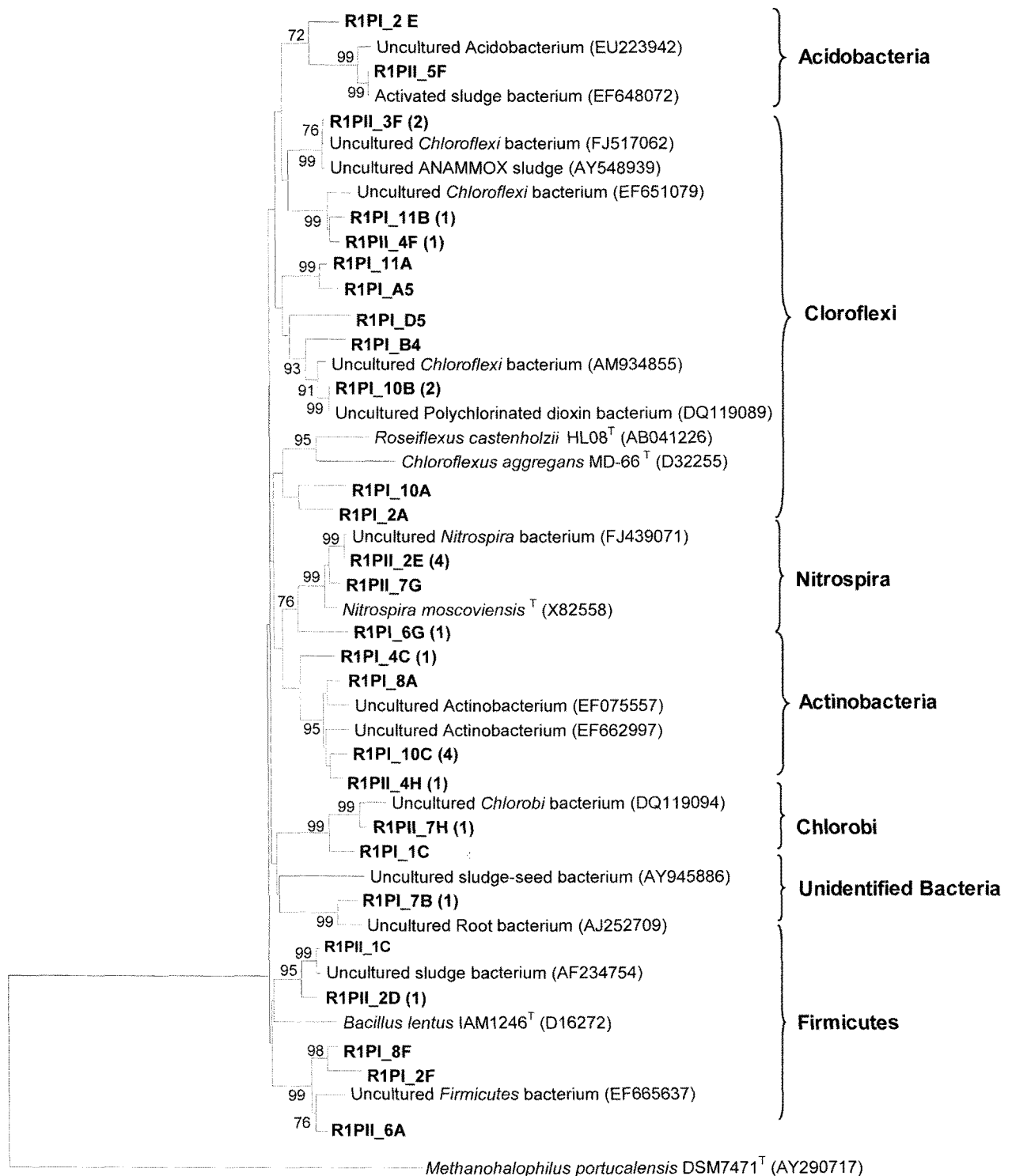
Fig. 4. Phylogenetic analysis based on partial 16S rRNA sequences from MBR clones (R2) representing members of major phyla in Bacteria Domain, except *Proteobacteria*, and related species.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanohalophilus portucalensis* was used as the outgroup.

Fig. 5. Phylogenetic analysis based on partial 16S rRNA sequences from CAS clones (R1) representing members of major phyla in Bacteria Domain, except *Proteobacteria*, and related species.

Evolutionary distances were based on the Kimura 2p model and tree reconstruction on the neighbor-joining method. Bootstrap values (1,000 replicate runs, shown as %) greater than 70% are listed. GenBank accession numbers are listed after species names. *Methanohalophilus portucalensis* was used as the outgroup.
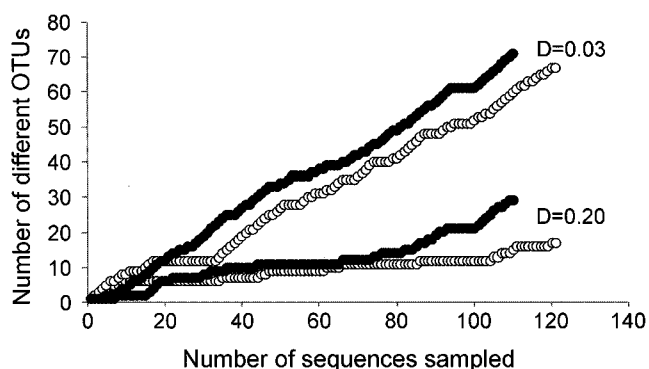
Fig. 6. Collector's curve of observed OTU richness in MBR (filled circle) and CAS (opened circle) sludge samples at evolutionary distances (D) of 0.03 and 0.20.

described to estimate total phylotype richness from the frequency of phylotypes in a clone library [25]. These indices were used in this study to estimate the total number of OTUs in bacterial communities of the sludge samples. In both libraries, the estimated number of phylotypes at the species level was higher than the observed numbers (Table 2). On the other hand, the estimated phylotype richness at the phylum level was similar to the observed one for both MBR and CAS libraries, suggesting that all phyla present in the sludges were sampled in this study. The combined data of richness analyses led us to conclude that the diversity that was observed in both libraries at the species level is an underestimate of the total bacterial diversity present in the environment, and further sampling would yield an increased observed diversity.

Diversity Analyses

Diversity indices serve as a valuable tool to quantify diversity in a community and describe its numerical structure, combining richness and evenness components. Two diversity indices, the Shannon index and the reciprocal of Simpson's index (1/D), were used to assess the bacterial community diversity. Simpson's index is heavily weighted towards the most

abundant species in the sample, while being less sensible to species richness. The Shannon index is positively correlated with species richness and evenness and gives more weight per individual to rare than common species, being sensitive to sample size [34]. Higher numbers indicate greater diversity. Both diversity indices were different between the libraries and revealed greater bacterial diversity for the MBR library, considering an evolutionary distance of 0.03 (Table 2). Magurran [34] described that the value of the Shannon index obtained from empirical data usually falls between 1.5 and 3.5 and rarely surpasses 4. Based on diversity index values found in the literature for sludge [32], soil [20], and marine sponge samples [39], the Shannon diversity indices of the CAS and MBR libraries were considered high (distance=0.03). According to previous data [60], the value of the reciprocal Simpson's index (1/D) below 50 found for CAS (Table 2) might indicate a dominance profile in this community. These results could be explained by the high abundance of the sequence types related to the order *Parvularculales* in the CAS library (Fig. 1B). Several researchers agree that the microbial population dynamics of MBR systems is different from that of the conventional activated sludge process because of the long sludge retention (>SRT), low food-to-microorganism ratio, complete biomass retention, and lack of sludge recirculation [11, 59]. These conditions allow for adaptation of microorganisms in general and of potentially slow growing specialist bacteria in particular, which establish a more diverse microbial community with broader physical capabilities in the system [58].

The computer program LIBSHUFF was used to compare the genetic coverage obtained in the MBR and CAS 16S rRNA libraries. Two gene libraries are assumed to be different if their homologous and heterologous coverage curves differ significantly [51]. These analyses showed statistically significant differences between the genetic coverage obtained in libraries MBR and CAS, with low P values for both analyses (P=0.001) (Fig. 7A and 7B). More information on these differences was obtained by examination

Table 2. Richness and diversity estimates for bacterial 16S rRNA gene clone libraries from MBR and CAS processes (sequence classification based on the cutoff value determined by DOTUR).

Source (n[a])	Distance[b]	Richness[c]	ACE[d]	Chao1[e]	Shannon[f]	1/Simpson[g]
MBR (110)	0.20	29	34	30	2.7	14
	0.03	71	145	167	4.0	97
CAS (121)	0.20	17	18	17	2.3	9
	0.03	67	179	150	3.6	30

[a]n, Number of gene sequences analyzed.
[b]80% Identity was estimated as the phylum-level distance (D=0.20), and 97% identity was estimated as the species-level distance (D=0.03).
[c]Richness is based on observed unique OTUs.
[d]Nonparametric statistical prediction of total richness of different OTUs based on distribution of abundant (10) and rare (10) OTUs.
[e]Nonparametric statistical predictions of total richness of OTUs based on distribution of singletons and doubletons.
[f]Shannon diversity index. A higher number represents more diversity.
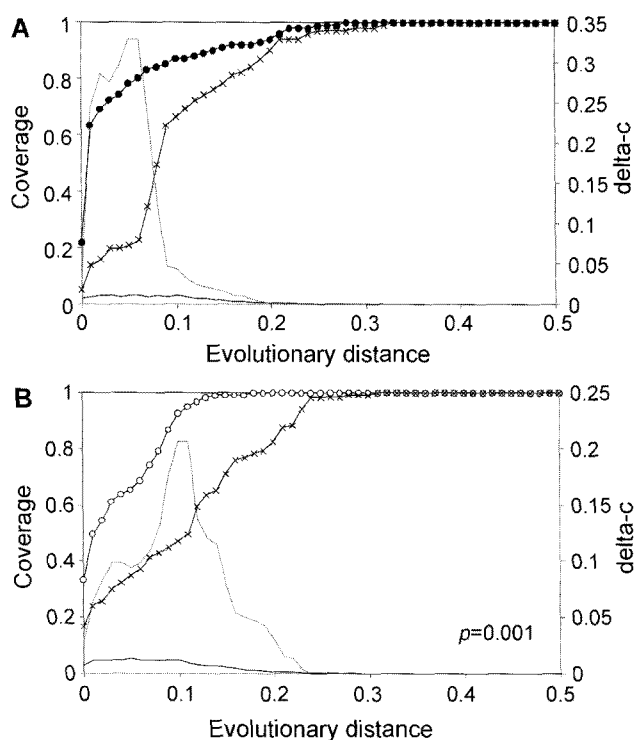[g]Reciprocal of Simpson's diversity index. A higher number represents more diversity.

Fig. 7. Genetic coverage comparison between libraries MBR and CAS.

A. MBR homologous coverage (filled circle). B. CAS homologous coverage (opened circle). Heterologous coverages are represented by crosses. Broken lines indicate the value of $(C_X - C_{XY})^2$ for the original samples and solid lines show $P = 0.05$ value of $(C_X - C_{XY})^2$ for the randomized samples.

differences in microbial community structure between the processes, which were clearly demonstrated in this work. Diversity calculations revealed high Shannon and Simpson indices for both libraries. Communities highly diverse offer the advantage of flexibility to adapt to changes in environments, what could explain the stable performance of both CAS and MBR processes. This work was one of a few that described, qualitatively and quantitatively, the structure of bacterial communities in MBR and CAS processes of wastewater treatment systems from petroleum refineries. A thorough understanding of the microbial community composition provides a basis for future application of methods for *in situ* analyses of the function of specific microorganisms, and will thus help to establish the links between microbial biodiversity and process stability in wastewater treatment plants.

Acknowledgments

of the distribution of $(Cx - Cxy)^2$ with D, which revealed that the values calculated for the real data exceeded the values calculated by using randomized data (at a P value of 0.05) at taxonomic distances lower than 0.23. These results showed that the CAS and MBR libraries differed greatly at the level of closely related sequences, but shared most deep taxa. Homologous coverage curves indicated that the MBR library had high coverage estimates at taxonomic distances correspondent to deeply branched taxa, such as phylum (D≥0.20). The CAS library, on the other hand, had high coverage estimates at D≥0.10, suggesting that higher taxa such as orders and classes have been fully sampled in the library. These results are consistent with the nonparametric statistical calculations of total richness (ACE and Chao), which yielded greater values for the MBR library at D=0.20.

Analysis of bacterial diversity in MBR and CAS samples originated from wastewater treatments of petroleum refineries revealed that the composition of bacterial communities was significantly different between the two processes. To date, a lot of literature have been published regarding the organic removal performance, sludge characteristics, operational conditions, and so on, and illustrated the differences between MBRs and the CAS system. These characteristics obviously contribute to yield

REFERENCES

1. Amann, R. I., W. Ludwig, and K. Schleifer. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59:** 143–169.
2. Barbosa, V. L., S. D. Atkins, V. P. Barbosa, J. E. Burgess, and R. M. Stuetz. 2006. Characterization of *Thiobacillus thioparus* isolated from a activated sludge bioreactor used for hydrogen sulde treatment. *J. Appl. Microbiol.* **101:** 1269–1281.
3. Beer, M., E. M. Seviour, Y. Kong, M. Cunningham, L. L. Blackall, and R. J. Seviour. 2002. Phylogeny of the filamentous bacterium Eikelboom Type 1851, and design and application of a 16S rRNA targeted oligonucleotide probe for its fluorescence *in situ* identification in activated sludge. *FEMS Microbiol. Lett.* **207:** 179–183.
4. Bjornsson, L., P. Hugenholtz, G. W. Tyson, and L. L. Blackall. 2002. Filamentous *Chloroflexi* (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology* **148:** 2309–2318.
5. Bodour, A. A., J.-M. Wang, M. L. Brusseau, and R. M. Maier. 2003. Temporal change in culturable phenanthrene degraders in response to long-term exposure to phenanthrene in a soil column system. *Environ. Microbiol.* **5:** 888–895.
6. Braile, P. M. 1979. *Manual of Industrial Wastewater Treatment*, 1st Ed. São Paulo, CETESB.
7. Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11:** 783–791.
8. Chao, A., M.-C. Ma, and M. C. K. Yang. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80:** 193–201.

9. Cho, J. C. and S. J. Giovannoni. 2003. *Parvularcula bermudensis* gen. nov., sp. nov., a marine bacterium that forms a deep branch in the α-Proteobacteria. *Intl. J. Syst. Evol. Microbiol.* **53:** 1031–1036.

10. Cho, K.-S., M. Hirai, and M. Shoda. 1991. Degradation characteristics of hydrogen sulfide, ethanethiol, dimethyl sulfide and dimethyl disulfide by *Thiobacillus thioparus* DW 44 isolated from peat biofilter. *J. Ferment. Bioeng.* **71:** 384–389.

11. Choi, J.-H., S. H. Lee, K. Fukushi, and K. Yamamoto. 2007. Comparison of sludge characteristics and PCR–DGGE based microbial diversity of nanofiltration and microfiltration membrane bioreactor. *Chemosphere* **67:** 1543–1550.

12. Clescerl, L. G., A. E. Greenberg, and A. D. Eaton. 1998. *Standard Methods for the Examination of Water and Wastewater.* APHA, AWWA, WEF, Washington, U.S.A.

13. Colwell, R. K. and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* **345:** 101–118.

14. Duncan, K. E., K. L. Sublette, P. A. Rider, A. Stepp, R. R. Beitle, J. A. Conner, and R. Kolhatkar. 2001. Analysis of a microbial community oxidizing inorganic sulfide and mercaptans. *Biotechnol. Prog.* **17:** 768–774.

15. Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.

16. Felsenstein, J. 1989. PHYLIP – phylogeny inference package (version 3.2). *Cladistics* **5:** 164–166.

17. Ferreira, D. F. 2000. Análises estatísticas por meio do Sisvar para Windows versão 4.0. *In:* 45ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, pp. 255–258. UFSCar, São Carlos, SP.

18. Good, I. J. 1953. The population frequency of species and the estimation of population parameters. *Biometrika* **40:** 237–264.

19. Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res.* **8:** 195–202.

20. Gremion, F., A. Chatzinotas, and H. Harms. 2003. Comparative 16S rDNA and 16S rRNA sequence analysis indicates that *Actinobacteria* might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. *Environ. Microbiol.* **5:** 896–907.

21. Grobkopf, R., S. Stubner, and W. Liesack. 1998. Novel Euryarchaeota lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms. *Appl. Environ. Microbiol.* **64:** 960–969.

22. Henze, M., P. Harremoes, J. la Cour Jansen, and E. Arvin. 1997. Wastewater treatment. *In* U. Forstner, R. J. Murphy, and W. H. Rulkens (eds.). Environmental Engineering. 2nd Ed. Springer, Berlin.

23. Horn, M. A., H. L. Drake, and A. Schramm. 2006. Nitrous oxide reductase genes (*nosZ*) of denitrifying microbial populations in soil and the earthworm gut are phylogenetically similar. *Appl. Environ. Microbiol.* **72:** 1019–1026.

24. Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180:** 4765–4774.

25. Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan. 2001. Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67:** 4399–4406.

26. Judd, S. The status of membrane bioreactor technology. 2008. *Trends Biotechnol.* **26:** 109–116.

27. Juretschko, S., L. Loy, A. Lehner, and M. Wagner. 2002. The microbial community composition of a nitrifying–denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. *System. Appl. Microbiol.* **25:** 84–99.

28. Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Mol. Biol. Evol.* **16:** 111–120.

29. Kraigher, B., T. Tina Kosjek, E. Heath, B. Kompare, and I. Mandic-Mulec. 2008. Influence of pharmaceutical residues on the structure of activated sludge bacterial communities in wastewater treatment bioreactors. *Water Res.* **42:** 4578–4588.

30. Lane, D. L., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* **82:** 6955–6959.

31. Lane, D. 1991. 16S/23S rRNA sequencing, pp. 115–175. *In* E. Stackebrandt and M. Goodfellow (eds.). *Nucleic Acid Techniques in Bacterial Systematics.* John Wiley & Sons, New York, N.Y.

32. Lin, S., S. Jin, L. Fu, C. Quan, and Y. S. Yang. 2009. Microbial community variation and functions to excess sludge reduction in a novel gravel contact oxidation reactor. *J. Hazard. Mater.* **165:** 1083–1090.

33. Loy, A., H. Daims, and M. Wagner. 2002. Activated sludge: Molecular techniques for determining community composition, pp. 26–43. *In* G. Bitton (ed.). *The Encyclopedia of Environmental Microbiology.* Wiley, New York.

34. Magurran, A. E. 2004. *Measuring Biological Diversity.* Blackwell Publishing, Oxford.

35. Manefield, M., A. S. Whiteley, R. I. Griffiths, and M. J. Bailey. 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl. Environ. Microbiol.* **68:** 5367–5373.

36. Mao, Y., X. Zhang, X. Yan, B. Liu, and L. Zhao. 2008. Development of group-specific PCR–DGGE fingerprinting for monitoring structural changes of *Thauera* spp. in an industrial wastewater treatment plant responding to operational perturbations. *J. Microbiol. Meth.* **75:** 231–236.

37. Mariano, J. B. 2001. Environmental impacts of petroleum refining. MSc. Dissertation, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

38. Miura, Y., Y. Watanabe, and S. Okabe. 2007. Significance of *Chloroflexi* in performance of submerged membrane bioreactor (MBR) treating municipal wastewater. *Environ. Sci. Technol.* **41:** 7787–7794.

39. Mohamed, N. M., J. J. Enticknap, J. E. Lohr, S. M. McIntosh, and R. T. Hill. 2008. Changes in bacterial communities of the marine sponge *Mycale laxissima* on transfer into aquaculture. *Appl. Environ Microbiol.* **74:** 1209–1222.

40. Monteiro, J. P., A. Magrini, A. S. Szklo, and R. T. S. Nunes. 2004. Management of water in the petroleum refinery and perspective of reuse. *In:* Ist Academic Congress of Environment and Development of Rio de Janeiro, RJ, Brazil. CADMA. FGV, Rio de Janeiro.

41. Mota, S. 1997. *Introduction to Environmental Engineering.* 1st Ed. Rio de Janeiro, Brazil, ABES.

42. Neria-Gonzáles, I., E. T. Wang, F. Ramírez, J. M. Romero, and C. Hernández-Rodríguez. 2006. Characterization of bacterial

community associated to biofilms of corroded oil pipelines from southeast of Mexico. *Anaerobe* **12:** 122–133.

43. Pereira, R. C. 2005. Development and application of a new methodology for the speciation analysis of selenium in aqueous effluents from petroleum refineries. MSc. Dissertation, PUC-Rio de Janeiro, RJ, Brazil.

44. Pettitt, A. N. 1982. Cramer–von Mises statistic, pp. 220–221. *In* S. Kotz and N. L. Johnson (eds.). *Encyclopedia of Statistical Sciences.* Wiley-Interscience, New York, N.Y.

45. Schloss, P. D. and J. Handelsman. 2004. Status of the Microbial Census. *Microbiol. Mol. Biol. Rev.* **68:** 686–691.

46. Schloss, P. D. and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71:** 1501–1506.

47. Schloss, P. D. and J. Handelsman. 2006. Toward a census of bacteria in soil. *PLoS Comp. Biol.* **2:** e92.

48. Shinoda, Y., J. Akagi, Y. Uchihashi, A. Hiráis, H. Yukawa, H. Yurimoto, Y. Sakai, and N. Kato. 2005. Anaerobic degradation of aromatic compounds by *Magnetospirillum* strains: Isolation and degradation genes. *Biosci. Biotechnol. Biochem.* **69:** 1483–1491.

49. Shokrollahzadeh, S., F. Azizmohseni, F. Golmohammad, H. Shokouhi, and F. Khademhaghighat. 2008. Biodegradation potential and bacterial diversity of a petrochemical wastewater treatment plant in Iran. *Bioresource Technol.* **99:** 6127–6133.

50. Silva, C. C., A. F. Viero, A. C. F. Dias, F. D. Andreote, E. J. Jesus, S. O. De Paula, A. P. R. Torres, V. M. J. Santiago, and V. M. Oliveira. Monitoring the bacterial community dynamics in a petroleum refinery wastewater membrane bioreactor fed with a high phenolic load. *J. Microbiol. Biotechnol.* Doi: 10.4014/jmb.0906.06001. [In press.]

51. Singleton, D. R., M. A. Furlong, S. L. Rathbun, and W. B. Whitman. 2001. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl. Environ. Microbiol.* **67:** 4374–4376.

52. Sipma, J., A. Svitelskaya, B. van der Marka, L. W. H. Pol, G. Lettinga, C. J. N. Buismana, and A. J. H. Janssena. 2004.

Potentials of biological oxidation processes for the treatment of spent sulfidic caustics containing thiols. *Water Res.* **38:** 4331–4340.

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

54. Tanji, Y., T. Kanagawa, and E. Mikami. 1989. Removal of dimethyl sulfide, methyl mercaptan, and hydrogen sulfide by immobilized *Thiobacillus thioparus* TK-m. *J. Ferment. Bioeng.* **68:** 280–285.

55. Tarlera, S. and E. B. Denner. 2003. *Sterolibacterium denitrificans* gen. nov., sp. nov., a novel cholesterol-oxidizing, denitrifying member of the β-*Proteobacteria. Int. J. Syst. Evol. Microbiol.* **53:** 1085–1091.

56. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24:** 4876–4882.

57. Viero, A. F., T. M. Melo, A. P. R. Torres, N. R. Ferreira, G. L. Sant'Anna Jr., C. P. Borges, and V. M. J. Santiago. 2008. The effects of long-term feeding of high organic loading in a submerged membrane bioreactor treating oil refinery wastewater. *J. Membrane Sci.* **319:** 223–230.

58. Wang, Z. and Z. Wu. 2009. Distribution and transformation of molecular weight of organic matters in membrane bioreactor and conventional activated sludge process. *Chem. Eng. J.* **150:** 396–402.

59. Wever, H., S. Weiss, T. Reemtsm, V. Vereecken, J. Muller, T. Knepper, *et al.* 2007. Comparison of sulfonated and other micropollutants removal in membrane bioreactor and conventional wastewater treatment. *Water Res.* **41:** 935–945.

60. Zhou, J., B. Xia, D. S. Treves, L.-Y. Wu, T. L. Marsh, R. V. O'Neill, A. V. Palumbo, and J. M. Tiedje. 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl. Environ. Microbiol.* **61:** 326–334.