RESEARCH ARTICLE

Bacterial community changes during bioremediation of aliphatic hydrocarbon-contaminated soil

Cécile Militon[1,2], Delphine Boucher[1,2], Cédric Vachelard[3], Geoffrey Perchet[3], Vincent Barra[4,5], Julien Troquet[3], Eric Peyretaillade[1,2] & Pierre Peyret[1,2]

[1]Clermont Université, Université d'Auvergne, Laboratoire: Microorganismes Génome et Environnement, Clermont-Ferrand, France; [2]CNRS, UMR 6023, Laboratoire: Microorganismes Génome et Environnement (LMGE), Aubiere, France; [3]Biobasic Environnement, Biopôle Clermont-Limagne, Saint-Beauzire, France; [4]Clermont Université, Université Blaise Pascal, LIMOS, Clermont-Ferrand, France; and [5]CNRS, UMR 6158, LIMOS, Aubière, France

Correspondence: Pierre Peyret, CNRS, UMR 6023, Laboratoire: Microorganismes Génome et Environnement (LMGE), Campus des cézeaux, 24 Avenue des Landais, 63177 Aubière, France. Tel.: +33 4 7340 7604; fax: +33 4 7340 7670; e-mail: pierre.peyret@univ-bpclermont.fr

Received 21 April 2010; revised 11 September 2010; accepted 22 September 2010.
Final version published online November 2010.

DOI:10.1111/j.1574-6941.2010.00982.x

Editor: Max Häggblom

Keywords
bacterial diversity; 16S rRNA gene; aliphatic hydrocarbons; bioremediation.

Abstract

The microbial community response during the oxygen biostimulation process of aged oil-polluted soils is poorly documented and there is no reference for the long-term monitoring of the unsaturated zone. To assess the potential effect of air supply on hydrocarbon fate and microbial community structure, two treatments (0 and 0.056 mol h^{-1} molar flow rate of oxygen) were performed in fixed bed reactors containing oil-polluted soil. Microbial activity was monitored continuously over 2 years throughout the oxygen biostimulation process. Microbial community structure before and after treatment for 12 and 24 months was determined using a dual rRNA/rRNA gene approach, allowing us to characterize bacteria that were presumably metabolically active and therefore responsible for the functionality of the community in this polluted soil. Clone library analysis revealed that the microbial community contained many rare phylotypes. These have never been observed in other studied ecosystems. The bacterial community shifted from *Gammaproteobacteria* to *Actinobacteria* during the treatment. Without aeration, the samples were dominated by a phylotype linked to the *Streptomyces*. Members belonging to eight dominant phylotypes were well adapted to the aeration process. Aeration stimulated an *Actinobacteria* phylotype that might be involved in restoring the ecosystem studied. Phylogenetic analyses suggested that this phylotype is a novel, deep-branching member of the *Actinobacteria* related to the well-studied genus *Acidimicrobium*.

Introduction

Hydrocarbons are widespread in soil subsurfaces as a result of industrial activity and accidental leakages during polluting transport and storage. Because of their impacts on human and environmental health (Hutcheson *et al.*, 1996), it has been necessary to develop site-restoring processes that can destroy or render harmless pollutants (Khan *et al.*, 2004). Petroleum is a complex mixture of thousands of organic compounds including a volatile aromatic fraction and a less volatile fraction (i.e. the aliphatic fraction). The aliphatic fraction consists of a wide range of hydrocarbons, ranging from light gases (C_1–C_4) to heavy residues (C_{35}–C_{40}). Spilled petroleum derivatives (mainly aliphatic hydrocarbons) are persistent in soil, and cause a significant deterioration in soil physical properties (Luthy *et al.*, 1997) and important changes in microorganism populations (Aislabie *et al.*, 2004). Aliphatic hydrocarbons exhibit a low chemical reactivity owing to their lack of functional groups. However, microorganisms have developed efficient strategies to degrade alkanes and to transform the hydrocarbons to more easily metabolizable substrates (Wentzel *et al.*, 2007). The cell contact with hydrophobic substrate is crucial because the initial step in aliphatic and aromatic hydrocarbon degradation is often mediated by oxidation reactions catalyzed by membrane-bound oxygenases (van Beilen & Funhoff, 2007). Interactions with the mineral and organic fractions may also result in a reduction in biological

FEMS MICROBIOLOGY ECOLOGY

availability and allow these contaminants to persist in soil. The degree of these physical interactions increases with time, and has been termed 'aging' (Hatzinger & Alexander, 2002).

Hydrocarbon degraders are classified in Bacteria (*Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, *Chlamydiae* and *Deinococcus-Thermus*), Archaea (*Halobacteriales*), Fungi and algae (Prince *et al.*, 2010). Although the fate of hydrocarbons and the extent to which microorganisms contribute to their biodegradation seem to be dependent on local environmental variables (Dibble & Bartha, 1979; Leahy & Colwell, 1990), some trends have emerged, such as the predominance of *Gammaproteobacteria* following hydrocarbon contamination in soils (Pucci *et al.*, 2000; Mills *et al.*, 2003; Popp *et al.*, 2006). Thus, the abilities of microorganisms to degrade complex substrates such as hydrocarbons have been studied for a few decades (Atlas & Bartha, 1972) and experiments have been conducted to use them to develop bioremediation strategies. One of the *in situ* bioremediation technologies relies on air sparging (Bass *et al.*, 2000; Yang *et al.*, 2005). Although primarily a physical treatment technology, the subsurface air flow induced can potentially increase oxygen supply and promote aerobic biodegradation. However, biodegradation is not always stimulated (Baldwin *et al.*, 2009). Despite the wide application of this technique to enhance bioremediation, there is little knowledge of the stimulated biocatalysts and the overall microbiological characteristics of the process (Zucchi *et al.*, 2003; Cavalca *et al.*, 2004; Kabelitz *et al.*, 2009; Nyyssönen *et al.*, 2009). The main limit of these investigations is the difficulty in identifying all microbial populations and studying their changes in complex environments (e.g. soils).

To gain more insights into the changes taking place in natural microbial communities under strong selective pressure, soil that was highly contaminated with aliphatic hydrocarbons and at different stages of an enhanced aerobic bioremediation was analyzed. In this paper, we report the monitoring of the hydrocarbon loss and the diversity of soil bacterial community during a 2-year pilot-scale bioremediation process for an alluvial soil contaminated with various aliphatic hydrocarbons. The results presented show a continuous depletion of total organic pollutants due to oxygen amendment associated with changes in the autochthonous bacterial communities. The changes observed in the active bacterial communities could be associated with the successful remediation of the contaminated soil. The microbial community changes were followed by the construction of both nearly full-length 16S rRNA genes and 16S rRNA sequence libraries, which offer a unique opportunity for an in-depth phylogenetic analysis to describe the complexity of diversity within various major microbial phyla encountered in soil.

Materials and methods

Soil samples

Samples were collected from an industrial site in September 2005 (the location cannot be indicated due to a confidentiality agreement). The contamination of this site has resulted from the progressive leakage of a storage tank following 20 years of industrial activities. This alluvial soil was slightly acidic (pH 6.6) and showed a low moisture content (12%). The soil was analyzed by Wessling Laboratories (France). The total organic matter was 0.1% and the total N was 500 mg kg^{-1} of dry matter. Ammonium, chloride, nitrate, phosphate, sulfate, calcium, copper, magnesium and phosphorus concentrations were 26, 310, 1, 0.009, 440, 9600, 200, 7200 and 900 mg kg^{-1} of dry matter, respectively. The microbial population of this contaminated soil was indigenous and no microorganisms were added before or during the experiment.

Soil bioremediation

The polluted soil was collected and maintained as a column for use in the reactors. Fixed bed reactors (height: 45 cm) with an 8-L working volume were designed and built to enable control of the liquid-phase composition (water quantity: 1.6 L, wet soil weight: 13.6 kg, soil moisture: 12%) and the air flow rate (volumetric flow rate: 100 mL min^{-1}, molar flow rate: 0.268 mol h^{-1}, molar flow rate of oxygen: 0.056 mol h^{-1}, oxygen concentration at saturation C*: 2.6E-04 mol O_2 L^{-1} water, superficial velocity: 0.6 mL min^{-1} cm^{-2}), in order to study at a pilot plant scale the influence of the different operating variables on the kinetics of the degradation (Fig. 1). Experiments were carried out at room temperature. Ambient carbon dioxide (CO_2) present in the injected air (340 p.p.m.) was first eliminated by passage through a cartridge of concentrated potassium hydroxide. The gaseous exhaust of each bioreactor was bubbled through a 0.5 M potassium hydroxide solution and the carbonate content of the solution connected to each reactor was then determined regularly by titrimetric analysis. CO_2 was considered an indicator of the biodegradation of organic pollutants by microorganisms, enabling an indirect measurement of the decreasing pollution load. The decreasing rate of pollution was then determined according to the approach developed by Troquet *et al.* (2003).

The total organic pollution load of the soil samples was measured with Soxhlet extraction of the hydrocarbon fraction present in the ground. Forty grams of soil was weighed in a cellulose cartridge, which was then placed in the Soxhlet extractor. The extraction of the organic fraction was performed with 100 mL of cyclohexane for 4 h. The weight of the organic extract was then measured after solvent

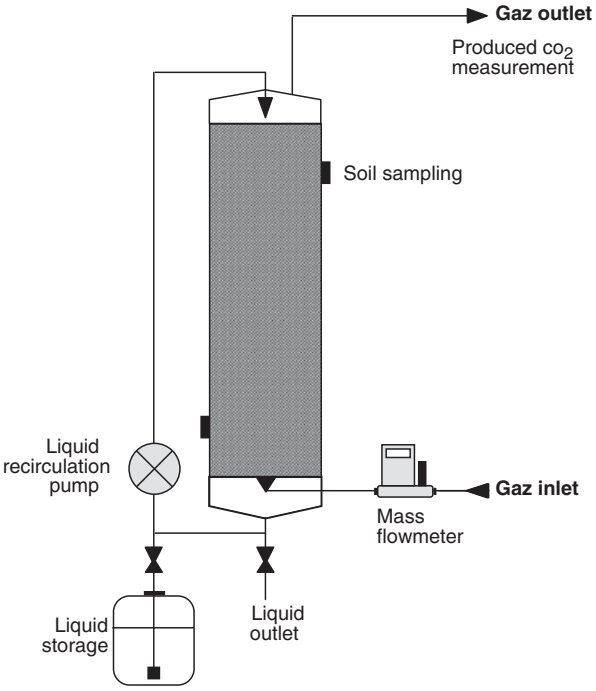

Fig. 1. General configuration of the bioremediation column reactors. Biobasic Environnement™.

evaporation. Results are given in mg hydrocarbons kg^{-1} dry soil (p.p.m.).

The extracted pollutants were diluted in dichloromethane and then filtered through a 0.45-μm filter. These samples were analyzed using a GC/MS apparatus (Agilent Technologies 6890-plus, 5973 Network). The analysis was performed using a Supelco Equity-5 GC column (30 m long, 0.25 mm inner diameter, 0.25 μm thickness). The initial temperature was 40 °C, later increased to 270 °C at 10 °C min^{-1}, and the final temperature was held for 15 min. The carrier gas was helium. Data acquisition and analysis were performed using the ENHANCED CHEMSTATION C1701CA software (Agilent Technologies). Total petroleum hydrocarbons (TPH) were analyzed using GC coupled with a flame ionization detector (GC–FID 7890, Agilent Technologies) according to ISO 16703 standards. An accredited analysis laboratory (Wessling Laboratories) performed the analysis.

Nucleic acids' recovery

The assessment of microbial community structure using molecular tools requires a sufficient number of sequencing efforts and an appropriate sampling strategy, taking into account microbial diversity and the heterogeneous distribution of microorganisms. A sampling aliquot of soil ≥ 1 g is required to obtain robust and reproducible fingerprinting analysis of the genetic structure of the microbial communities (Ranjard *et al.*, 2003). To minimize this heterogeneity, we have knowingly recovered DNA and RNA with 5 and 10 g of soil, respectively. The samples used for total nucleic acid recovery were not destructively taken from bioremediation column reactors (Biobasic Environnement™). Because of the difficulty of simultaneously recovering DNA/RNA of good quality and quantity from the polluted soil studied, different protocols were used to recover these nucleic acids. DNA was extracted and purified from 5 g of soil according to the protocol of Zhou *et al.* (1996). This DNA extraction protocol was slightly modified by the addition of a phenol and phenol/chloroform/isoamyl alcohol (IAA) (25 : 24 : 1) purification step before the chloroform/IAA purification. Total RNA was extracted and purified from 10 g of soil that had been incubated for 30 min at 37 °C with 30 mL NLS buffer [1% *N*-lauroyl sarcosine (w/v); 1% cetyltrimethylammonium bromide (w/v); 0.5 M EDTA], 100 U RNase inhibitor (Sigma) and 5 g proteinase K (Sigma). This mix was then agitated twice for 30 s at 30 Hz in an MM301 bead-beater (Retsch) with 15 g of glass and zirconium beads (106 μM, 2 mm and 1 cm, respectively) before a 10-min incubation at 60 °C. The solution was centrifuged for 10 min at 8500 ***g*** at 4 °C and the supernatant was purified using phenol, phenol/chloroform (1 : 1) and chloroform extraction steps. RNA was pelleted, washed with 70% ethanol and concentrated using the RNeasy Mini Kit as recommended by the manufacturer (Qiagen GmbH, Hilden, Germany). Residual DNA in the RNA samples was destroyed using 2 U RNase-free DNase I (AmpGrade, Invitrogen). DNA and RNA were analyzed on the 2100 Bioanalyzer as recommended by the manufacturer. The DNA and RNA solutions were stored at −20 °C for later use. For each observational time (i.e. 0, 12 and 24 months), only a single extraction was performed because we optimized the protocols and checked the quality and quantity of extracted nucleic acids.

Reverse transcription of 16S rRNA

Denatured 16S rRNA (125 ng) was reverse transcribed using the 1406r primer (5′-ACGGGCGGTGWGTRCAA-3′). This primer targets most of microorganisms belonging to Bacteria and Archaea (Baker *et al.*, 2003). Reactions (20 μL) containing 0.625 μM reverse primers, 100 U Superscript III reverse transcriptase (RT) (Invitrogen), 1 × RT buffer, 0.1 M dithiothreitol, 0.25 mM dNTP (Invitrogen), 1 U RNasin+ (Promega) and 19 μg T4 gene 32 Protein (Roche) were performed for 2 h at 42 °C. After reverse transcription, RT products (especially RNA strands) were degraded by alkaline hydrolysis (0.25 M NaOH for 15 min at 65 °C) and analyzed on the 2100 Bioanalyzer using the RNA 6000 Nano and Pico Kits.

PCR amplification of 16S rRNA genes

PCR amplifications were performed using the primers 27f (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1406r as

described previously (Suzuki & Giovannoni, 1996; Baker *et al.*, 2003). The 50-μL reaction mixture contained 50 ng of soil DNA or RT products, 1 U GoTaq DNA polymerase (Promega), 1 × PCR buffer containing $MgCl_2$ (1.5 mM final concentration), 0.2 μM of each primer and 250 μM of each deoxynucleotide. The reactions were performed in an iCycler thermocycler (Bio-Rad) using a program with an initial denaturation step for 5 min at 95 °C, followed by 35 cycles of three steps (95 °C for 1 min, 59 °C for 30 s and 72 °C for 1 min) and a final extension step for 7 min at 72 °C. The presence and sizes of the amplification products were determined by agarose gel electrophoresis. Bands with the expected sizes were excised and purified using the Qiaquick Gel Elution Kit as suggested by the manufacturer (Qiagen GmbH).

Library construction, screening and sequencing

Agarose-purified PCR products were ligated into the pCR®II-TOPO® vector (TOPO TA Cloning® Kit Dual Promoter, Invitrogen) and transformed into *Escherichia coli* One Shot® TOP10 cells (Invitrogen) according to the manufacturer's instructions. Plasmid DNA was extracted using the Montage Plasmid $Miniprep_{HTS}$ 96 Kit (Millipore). Each clone was digested with the EcoRI restriction enzyme, analyzed by electrophoresis on a 1% w/v agarose gel and stained with ethidium bromide. Positive, purified plasmids were sequenced using the Sanger method of MWG Biotech (Ebersberg, Germany) with both SP6 and T7 vector primers. All the sequences were deposited in the EMBL-Bank with accession numbers AM934704–AM936916 and FM209057–FM209176.

Sequence treatment and statistical analysis

Sequence treatment and joining were performed using the pregap4 and gap4 tools of the STADEN PACKAGE program (Staden, 1996). The sequences were sense-oriented using the OrientationChecker tool (Ashelford *et al.*, 2006) and aligned using the CLUSTAL W (1.83) and MULTIPLE SEQUENCE ALIGNMENT program (Thompson *et al.*, 1994). Sequences were checked for chimeric PCR artifacts using MALLARD software (Ashelford *et al.*, 2006) by inspection of sequence alignments. Potential composite sequences were excluded. The nonchimeric sequences were aligned using the NAST algorithm (DeSantis *et al.*, 2006) of the Greengenes Web interface (greengenes.lbl.gov/NAST). With the Greengenes and Ribosomal Database Project (Cole *et al.*, 2009) tools, it is possible to classify sequences and create distance matrices. In the same way, the reference sequence from each operational taxonomic unit (OTU) was aligned with complete sequences from an ARB database using the latter's automatic alignment tool (http://www.arb-home.de) (Ludwig *et al.*, 2004). This alignment was checked and corrected manually, taking into consideration the secondary structure of the rRNA molecule. Sequences were inserted into an optimized tree according to maximum parsimony criteria without allowing changes to existing tree topology. Bootstrap analysis was performed using the ARB parsimony bootstrap algorithm.

The OTUs that bring together sequences sharing > 97% identity were then characterized using the furthest neighbor clustering algorithm of the DOTUR program (Schloss & Handelsman, 2005). Shannon diversity indices (H') and the probable numbers of phylotypes (S_{Chao1}), along with the lower and upper bounds of the confidence intervals, named 95% lci and 95% uci, respectively, were also calculated using this tool. The formulas used by the software are described in the web site of DOTUR (http://www.plantpath.wisc.edu/fac/joh/dotur.html). The Good's (C) and equitability (J) indices were autonomously calculated using the following formulas:

$$C = [1 - (n_1/N)] \times 100$$

where n_1 is the number of singletons (OTUs with a single sequence) and N is the number of individuals sampled;

$$J = H'/H'_{max}$$

where H' is the Shannon index and H'_{max} is the H' index when each OTU is represented by a single sequence.

The rarefaction calculations were performed using the ANALYTIC RAREFACTION (1.3) program (Raup, 1972) of the Stratigraphy Lab, University of Georgia (http://www.uga.edu/~strata/software/Software.html). Finally, comparisons of microbial community memberships and structures were evaluated with SONS (shared OTUs and similarity), a computer program using nonparametric estimators (Schloss & Handelsman, 2006).

Results and discussion

Bioremediation process

Over 2 years, the alluvial soil of an industrial site contaminated by various aliphatic hydrocarbons (10 400 p.p.m. of C_{10}–C_{40} compounds) was exposed to (1) an oxygen biostimulation process or to (2) an absence of aeration (control) in two different reactors. Using the monitoring of CO_2 production, a continuous and significant hydrocarbon loss was observed during the oxygen biostimulation process (Fig. 2). Different phases were observed, in particular a significant removal of hydrocarbons between 12 and 24 months (Fig. 2). Quantification of the pollutants during the bioremediation process showed a decrease in the pollutant after 12 months (1600 p.p.m.) and 24 months (4500 p.p.m.) (Table 1). These results (15% and 43%

decrease in the pollutant, respectively) were similar to those calculated through indirect monitoring and gas balance calculations. TPH analysis by GC–FID showed a main fraction of hydrocarbons characterized by long chain-length C_{21}–C_{35} molecules (74% of the TPH). The major degradation occurred for medium chain-length C_{12}–C_{20} hydrocarbons after 24 months (80%), compared with heavier hydrocarbons (38% and 44% for C_{21}–C_{35} and C_{35}–C_{40} molecules, respectively).

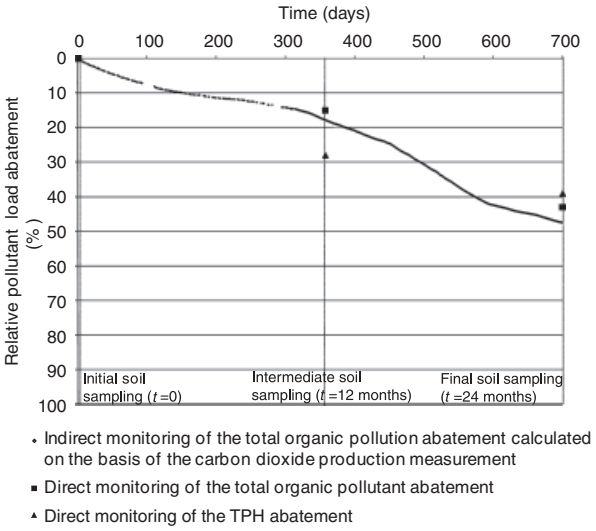

Fig. 2. Monitoring of pollutant abatement in column reactors.

The results for the total organic pollutant load were in agreement with the results obtained for the TPH analysis (± 4%) (Table 1). Moreover, the assessment of the bioremediation process through indirect gas balance calculations seemed to be a sound approach (± 10%) (Fig. 2). No qualitative or quantitative loss of TPH was observed without aeration (control) after 24 months, suggesting a negligible role of biotic and abiotic losses without aeration (data not shown). However, the absence of a killed control prevents us from excluding any abiotic degradation during the bioremediation process, namely by air stripping. Thus, these results highlight that the oxygen biostimulation process is crucial for restoring the studied soil. This study suggests that aeration stimulates microbial community activities, including those involved in hydrocarbon mineralization, and/or may facilitate the interactions between the pollutants and the microbial community. To assess bacterial community changes during the biodegradation process, we described the diversity of the polluted ecosystem by analyzing libraries from small subunit (SSU) rRNA genes and SSU rRNA.

Table 1. Global organic pollutant and TPH concentrations during the bioremediation process

Time point	Unit	Initial	t = 12 months	t = 24 months
Total organic pollutant load*	p.p.m.	10 500	8904	6038
Relative SD	%	2,5	4,6	11,5
Total organic pollutant abatement	p.p.m.	0	1596	4463
Total organic pollutant abatement	%	0	15	43
Total petroleum hydrocarbons†	p.p.m.	4600	3300	2800
Hydrocarbons (C10–C12)	p.p.m.	100	50	20
Hydrocarbons (C12–C16)	p.p.m.	100	50	20
Hydrocarbons (C16–C21)	p.p.m.	700	200	590
Hydrocarbons (C21–C35)	p.p.m.	3400	2400	2100
Hydrocarbons (C35–C40)	p.p.m.	480	410	270
Hydrocarbons (C10–C12) abatement	p.p.m.	0	50	80
Hydrocarbons (C12–C16) abatement	p.p.m.	0	50	80
Hydrocarbons (C16–C21) abatement	p.p.m.	0	500	− 390‡
Hydrocarbons (C21–C35) abatement	p.p.m.	0	1000	1300
Hydrocarbons (C35–C40) abatement	p.p.m.	0	70	210
Hydrocarbons (C10–C12) abatement	%	0	50	80
Hydrocarbons (C12–C16) abatement	%	0	50	80
Hydrocarbons (C16–C21) abatement	%	0	71	–
Hydrocarbons (C21–C35) abatement	%	0	29	38
Hydrocarbons (C35–C40) abatement	%	0	15	44
TPH abatement	p.p.m.	0	1300	1800
TPH abatement	%	0	28	39

*Results are the average of five replicates.

†Results are given from only one replicate according to the small sample quantity available.

‡This abnormal value is due to the heterogeneity of the sample.

Total pollutant load was determined by extraction into dichloromethane (DCM) and then weighing the residue.

The TPH was determined by GC–FID analysis.

General description of the 16S rRNA genes and 16S rRNA libraries

The microbial diversity of the polluted soil was studied before treatment (t_0) and during the bioremediation process after 12 and 24 months (t_{+12} and t_{+24}, respectively). To gain an expanded view of the microbial diversity in this soil, we used a dual rRNA/rRNA gene approach, allowing us to characterize bacteria that were presumably metabolically active and therefore responsible for the functionality of the community in this polluted soil. The natural shift of the microbial community without aeration (control) was also investigated at t_{+24}. We did not build an rRNA library for this control because degradation of the pollutants was not observed.

From the seven libraries, 2333 nearly full-length 16S rRNA genes and 16S rRNA sequences were investigated. The present study constitutes an extensive survey of polluted soil. A subdivision into OTUs sharing > 97% identity was performed ($OTU_{0.03}$ definition). Using this combined approach, 388 different OTUs were detected (Supporting Information, Table S1). More than half of the phylotypes were only distantly related to previously identified species (< 97% of sequence similarity). This situation was especially marked for the *Actinobacteria*, where > 70% of the phylotypes were only distantly related (< 97% of sequence similarity) to known phylotypes (Fig. 3). Bacterial communities in complex ecosystems such as soils typically exhibit a distribution pattern in which most bacterial species are present in low abundance (Pedros-Alio, 2006, 2007; Sogin *et al.*, 2006; Ashby *et al.*, 2007; Elshahed *et al.*, 2008). Our study showed a high bacterial diversity. Fourteen bacterial phyla were observed: *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chlorobi*, *Chloroflexi*, *Gemmatimonadetes*, *Nitrospirae*, *OP10*, *Planctomycetes*, *Proteobacteria* (*Alpha*, *Beta*, *Delta*- and *Gamma* subdivisions), *TM6*, *TM7*, *Verrucomicrobia* and *WS3*. Bacterial diversity in soil has been investigated extensively using culture-independent 16S rRNA gene-based analysis (Youssef & Elshahed, 2009). The absolute majority of soil sequences belong to nine major bacterial phyla: the *Proteobacteria* (mainly the *Alpha*, *Beta*, *Delta* subdivisions), *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Bacteroidetes*, *Firmicutes*, *Planctomycetes*, *Verrucomicrobia* and *Gemmatimonadetes* (Janssen, 2006). The phyla detected in the polluted environment are characteristic of a soil ecosystem. Surprisingly, no sequences belonging to *Firmicutes* were identified, but their weak presence in a hydrocarbon-polluted soil has been observed (Daane *et al.*, 2001; Popp *et al.*, 2006).

Several bacterial divisions present only in DNA-based libraries were detected by comparison of the six different libraries for each observational time during the bioremediation process: *Bacteroidetes*, *OP10*, *Planctomycetes*, *TM7* and *Verrucomicrobia* at t_0, *Chlorobi* and *Chloroflexi* at t_{+12}, and *OP10* and *Verrucomicrobia* at t_{+24} (Table S1). However, a relatively small number of phylotypes and sequences belonged to these DNA-specific bacterial divisions. The 16S rRNA libraries were also characterized by an 'RNA library-specific' bacterial group: the *TM6* division (Table S1). A comparison at the phylotype level revealed that only 32 phylotypes (58.4% of the sequences) were found in both libraries at t_0, 45 (61.0% of the sequences) at t_{+12}, and 44 (53.7% of the sequences) at t_{+24}, indicating the differential recognition between both types of molecules. 16S rRNA genes and 16S rRNA sequences mirrored, respectively, the 'total' microbial community (active/quiescent/dead) and the 'active' one. Bacterial divisions present only in DNA libraries could theoretically provide a glimpse of quiescent bacteria present in the polluted soil or extracellular DNA released from dead cells. The unexpected contrary situation could be explained by a small, but very active (high number of ribosomes), number of cells, or by methodological biases, such as the random clone selection for sequencing or a differential cell lysis efficiency during DNA and RNA extractions (Suzuki & Giovannoni, 1996; von Wintzingerode *et al.*, 1997).

Rarefaction analyses were performed to assess the representativeness of the libraries (Fig. 4). For each library, the resulting curve did not attain its asymptotic phase, an indication of the incompletely described diversity of the microbial communities studied. Nevertheless, the values of Good's indices (*C*) showed that the sizes of the libraries were sufficient to cover 74–88% of the bacterial communities. To estimate similarity between communities based on membership and structure, we used SONS, a computer program using nonparametric estimators (Table 2). The J_{abund} values of RNA libraries were high, but significantly different from 1.0. These estimates suggested that although low-abundance members might not have been shared between communities; the most abundant members were shared in RNA libraries. J_{abund} values of the DNA libraries were quite low, suggesting that only the highly dominant phylotypes might be shared. The very low J_{abund} values of the DNA control library indicate significant differences in bacterial communities. To compare the structures of the communities from the seven libraries, the community similarity index, θ, was calculated (Table 2). Although the community structures were not identical, a good conservation was observed for the active microorganisms described by the RNA libraries with a closer proximity before (t_0) and after the first sampling (t_{+12}) of the bioremediation process. A more pronounced change in the microbial community structure was observed during the bioremediation process (t_{+12} to t_{+24}), which could be correlated to the change in the degradation rate of the pollutants at 24 months. A good similarity was also observed between DNA t_0 and DNA t_{+12}. The θ values for the comparisons between the DNA control library and the

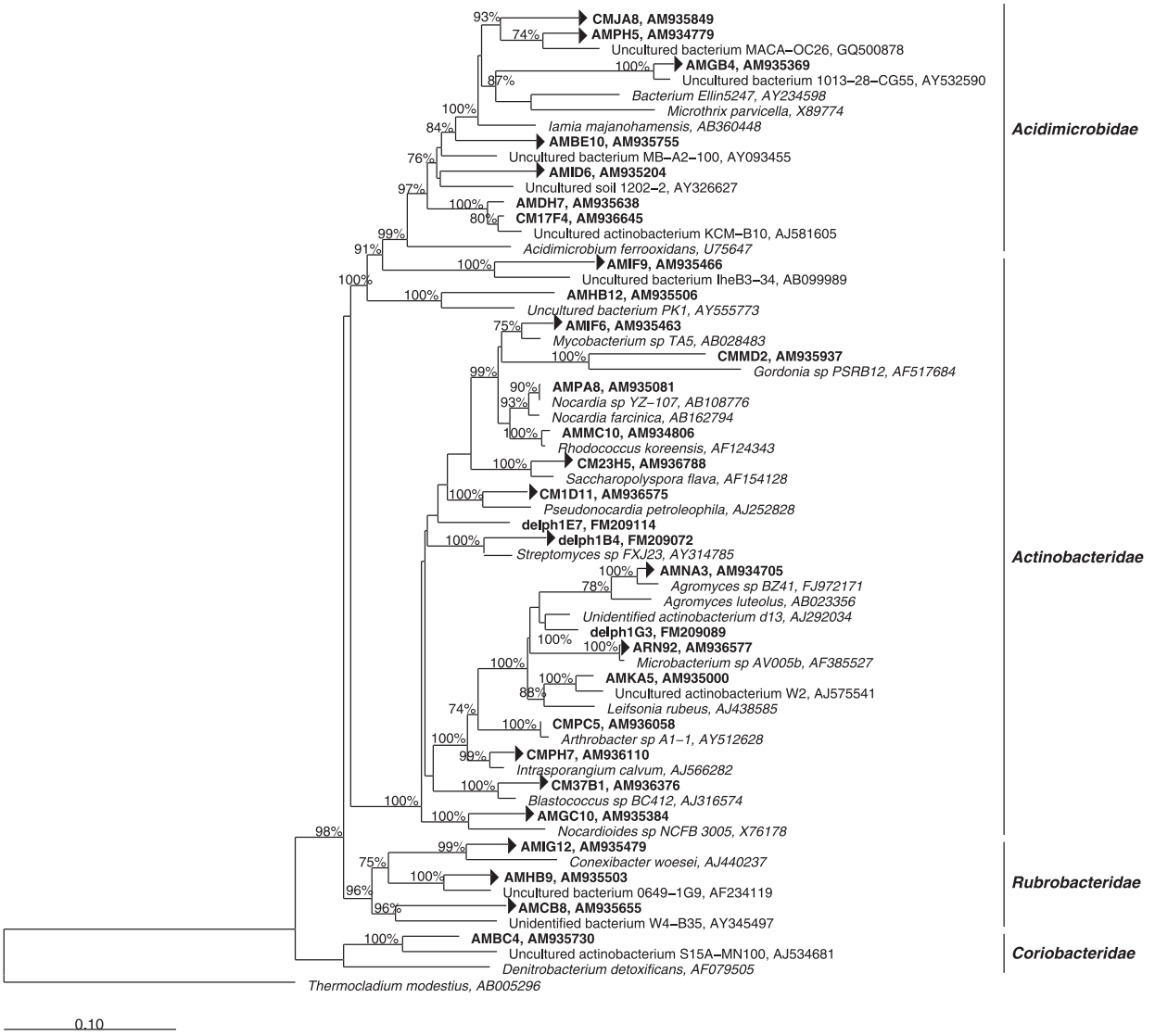

Fig. 3. Distance phylogram of *Actinobacteria* OTU sequences based on aligned near-full-length SSU rRNA genes and rRNA sequences from hydrocarbon-polluted soil libraries (names in bold) as well as representative sequences from the ARB library. The tree was rooted with the Archaea *Thermocladium modistius* (AB005296). Accession numbers are also indicated. Bootstrap values of $> 70\%$ (for 1000 iterations) are shown. The scale bars represent a distance of 10%.

six other libraries were very low. These results confirmed that a significant modification of the microbial diversity was induced by the restoring system.

Phylogenetic analysis of the polluted soil before the bioremediation process

The accidental release of laminated oil into the studied soil has probably induced a shift of the community structure in response to perturbation by contaminants, but the lack of uncontaminated soil prevents us from formally describing the autochthonous microbial community from the pristine soil and its evolution over time after the oil spill up to the beginning of the bioremediation process (t_0). The initial bacterial community structure in a soil is a function of the geographical location, environmental conditions and the edaphic parameters, mainly soil pH, which could largely explain the diversity and richness of soil bacterial communities (Fierer & Jackson, 2006). In this study, a high level of community diversity ($H' = 4.476$) was observed. The subdivision in OTUs sharing $> 97\%$ identity ($OTU_{0.03}$ definition) highlighted > 150 different phylotypes. Only half of these phylotypes were active in the polluted soil before the bioremediation process (Table S1). The two largest and most diverse divisions were the *Alphaproteobacteria* and the *Gammaproteobacteria*, especially for the 'active' bacterial

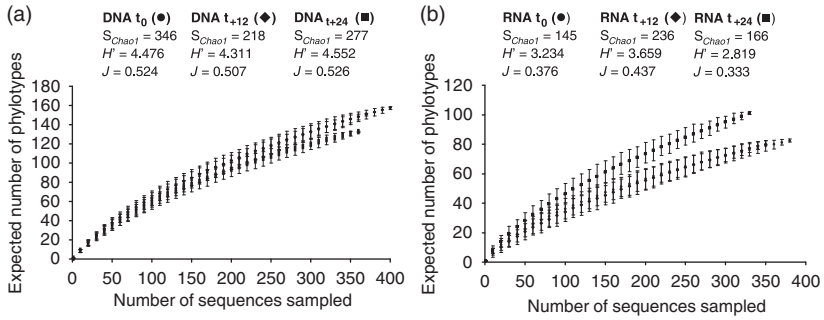

Fig. 4. Statistical analyses of libraries obtained from 16S rRNA genes (DNA) (a) and 16S rRNA (RNA) (b) throughout the pilot-scale bioremediation process of the hydrocarbon-polluted soil. Rarefaction plots and calculation of the estimated numbers of phylotypes (S_{Chao1}), Shannon's diversity indices (H') and equitability indices (J) have been performed on libraries obtained before the beginning of the process (t_0 ●), after 12 months (t_{+12} ◆) and after 24 months (t_{+24} ■). Phylotypes are shown at the 3% genetic distance level. Error bars represent the 95% confidence interval.

Table 2. Similarity assessment between communities based on membership and structure using SONS, a computer program that uses nonparametric estimators

	J_{abund}	θ
RNA t_0–t_{+12}	0.74, SE = 0.11	0.79, SE = 0.04
RNA t_{+12}–t_{+24}	0.71, SE = 0.20	0.53, SE = 0.06
RNA t_0–t_{+24}	0.54, SE = 0.18	0.44, SE = 0.05
DNA t_0–t_{+12}	0.52, SE = 0.07	0.66, SE = 0.06
DNA t_{+12}–t_{+24}	0.43, SE = 0.08	0.27, SE = 0.04
DNA t_0–t_{+24}	0.36, SE = 0.06	0.27, SE = 0.05
Control – DNA t_0	0.42, SE = 0.12	0.17, SE = 0.05
Control – DNA t_{+12}	0.34, SE = 0.09	0.22, SE = 0.06
Control – DNA t_{+24}	0.45, SE = 0.15	0.16, SE = 0.05
Control – RNA t_0	0.35, SE = 0.12	0.17, SE = 0.05
Control – RNA t_{+12}	0.37, SE = 0.11	0.19, SE = 0.06
Control – RNA t_{+24}	0.42, SE = 0.17	0.09, SE = 0.04

The abundance-based Jaccard (J_{abund}) similarity index and the community similarity index, θ, are shown for the different DNA and RNA libraries throughout the pilot-scale bioremediation process of the hydrocarbon-polluted soil (t_0, before the beginning of the process; t_{+12}, after 12 months; t_{+24}, after 24 months). Control is a DNA library obtained from the polluted soil without biostimulation after 24 months.

community (Table S1). The *Betaproteobacteria* and *Deltaproteobacteria* divisions showed a lower level of diversity and abundance. The phylum of *Proteobacteria* is the most abundant in soil libraries (Janssen, 2006). The *Proteobacteria* encompass enormous morphological, physiological and metabolic diversity, and are of great importance to global carbon, nitrogen and sulfur cycles (Kersters *et al.*, 2006). In nonpolluted soils, *Alphaproteobacteria* is the most abundant class and *Gammaproteobacteria* is often the least abundant (Spain *et al.*, 2009). Our results are in agreement with most studies showing the importance of the *Proteobacteria*, especially the *Gamma* division, in hydrocarbon-polluted soil microbial communities (Margesin *et al.*, 2003; Hamamura *et al.*, 2006; Popp *et al.*, 2006) or natural asphalts (Kim & Crowley, 2007). This so-called 'gamma-shift' seems to be characteristic of soils polluted by hydrocarbons (Kabelitz *et al.*, 2009). Moreover, these results agree with previous observations that all proteobacterial divisions, with the possible exception of *Epsilonproteobacteria*, respond positively to the input of hydrocarbons (Greer *et al.*, 2010). The third and fourth most diverse and abundant phyla are *Actinobacteria* and *Acidobacteria*. The use of 16S rRNA sequences revealed a surprisingly high abundance of *Actinobacteria* in the 'active community'. *Actinobacteria* are high G+C, Gram-positive bacteria, comprising a wide range of physiologically diverse species. On a 3-year clean-up site, Kabelitz *et al.* (2009) also observed a predominance of sequence types affiliated with the *Actinobacteria*. In our study, *Gemmatimonadetes*, *Chloroflexi*, *Bacteroidetes* and *Planctomycetes* were composed of few sequences, but were distributed to a large number of phylotypes (Table S1). Finally, *Chlorobi*, *Nitrospirae*, *OP10*, *TM6*, *TM7*, *Verrucomicrobia* and *WS3* could be considered rare phyla, with only one to seven sequences (Table S1).

Major bacterial phylotypes before the bioremediation process

Although the *Gamma* class was less diversified than *Alphaproteobacteria*, it was the most abundant bacterial division (Table S1). Five phylotypes (AMNA4, AMNG8, AMNA6, AMNC6 and AMNB8) represented 77% of the *Gammaproteobacteria* in the DNA library. Typical sequences of these dominant phylotypes are distantly related by ARB affiliation to *Microbulbifer salipaludis* (GenBank accession no. AF479688: *Alteromonadales*; *Alteromonadaceae*; 91% identity), *Lysobacter brunescens* (AB161360: *Xanthomonadales*; *Xanthomonadaceae*; 96% identity), *Hydrocarboniphaga effusa* (AY363244: *Xanthomonadales*; *Sinobacteraceae*; 86% identity), *Acidithiobacillus caldus* (AB023405: *Acidithiobacillales*; *Acidithiobacillaceae*; 87% identity); and *Thiorhodovibrio winogradsky* (AB016986: *Chromatiales*; *Chromatiaceae*; 89% identity), respectively. The five dominant phylotypes described previously in the DNA library were uncovered in

the RNA library for four of them (AMNA4, AMNG8, AMNC6 and AMNB8) and constituted 86% of the *Gammaproteobacteria* sequences. The phylotype related to AMNA6 was not identified. We noticed that the phylotype related to *T. winogradsky* was more active in the RNA library than in the DNA library.

The analysis of 16S rRNA sequences revealed an active phylotype CMJC10 (AM935867) distantly related to proteobacterium BHI8073 (AJ431227: 80% identity) for the *Deltaproteobacteria* phylum and the importance of *Actinobacteria* with one dominant phylotype (74% of the actinobacterial sequences) characterized by the representative sequence CMJA8 (AM935849). CMJA8 displayed the closest sequence similarity to *Acidimicrobium ferrooxidans* (U75647: *Acidimicrobidae*; *Acidimicrobiales*; *Acidimicrobineae*; *Acidimicrobiaceae*, 88% identity). The *Acidimicrobidae*, a subclass of the *Actinobacteria* (Stackebrandt *et al.*, 1997; Zhi *et al.*, 2009), is currently composed of a single genus (*Acidimicrobium*) and a single species (*A. ferrooxidans*). *Acidimicrobium* are found in warm, acidic, sulfur- or mineral sulfide-rich environments (Clark & Norris, 1996). Interestingly, molecular environmental studies allowed the discovery of several novel 16S rRNA sequence clusters with a probable worldwide distribution in different terrestrial environments, forming several lineages deeply rooted in the *Acidimicrobiales* order (Rheims *et al.*, 1999). CMJA8 was equally closely related to the new isolate strain *Iamia majanohamensis* (AB360448: *Acidimicrobidae*; *Acidimicrobiales*; *Acidimicrobineae*; *Iamiaceae*, 92% identity) (Kurahashi *et al.*, 2009) and the filamentous bacteria *Microthrix parvicella* (X89774: *Candidatus* Microthrix, 87% identity) (Rossetti *et al.*, 2005). CMJA8 should be a novel, deep-rooting lineage within the class *Actinobacteria* and clusters with uncultured bacteria and the well-studied *A. ferrooxidans*.

These results of the RNA library were used to identify metabolically active bacteria that are responsible for the functionality of the community in this polluted soil. The changes in the bacterial community were monitored after 12 and 24 months of treatment.

Modification of bacterial community during the pilot-scale bioremediation process

The bacterial community changed at each sampling date during the bioremediation process (Fig. 5) in correlation with the depletion of the pollutants. *Proteobacteria* still dominated the microbial community after 12 months of bioremediation, but active proteobacterial phylotypes were less diversified for *Alpha-*, *Beta-* and *Deltaproteobacteria* (Table S1). Because of the high diversity of these divisions with no dominant phylotypes, the side effects of the sampling method on infrequent populations could explain these observations. We cannot exclude that some phylotypes

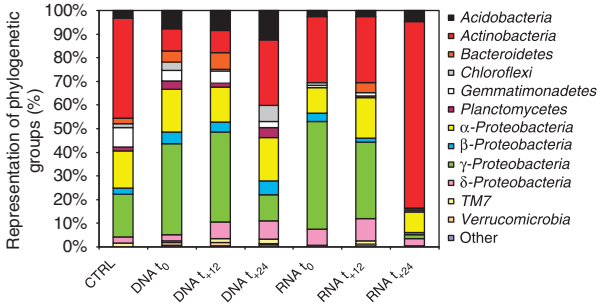

Fig. 5. Distribution of phylogenetic groups observed during the bioremediation process of aliphatic hydrocarbon-polluted soil. Phylotypes have been identified in libraries obtained from 16S rRNA genes (DNA) and 16S rRNA (RNA) throughout the pilot-scale bioremediation process of the hydrocarbon-polluted soil (t_0, before the beginning of the process; t_{+12}, after 12 months; t_{+24}, after 24 months). CTRL is a DNA library obtained from the polluted soil without biostimulation after 24 months. The percentages of the phylogenetically classified sequences are given. Phylogenetic groups accounting for < 1% of the sequences (*Chlorobi*, *Nitrospirae*, *OP10*, *TM6* and *WS3*) are summarized in the artificial group 'Other'.

could be more sensitive or could not adapt to or resist this new environment during the bioremediation treatment. During the last stage of this treatment, a significant change in the bacterial community occurred. This shift was particularly important for the 'active' populations. While *Proteobacteria* dominate the 'total' bacterial community (43%), relatively few proteobacterial sequences were detected in the 'active' bacterial community (14%). A strong reduction of abundance was especially observed for the *Gamma* division (Table S1; Fig. 5). Only the *Alphaproteobacteria* still represent a significant part of the diversity without dominant phylotypes. The typical soil lineages exhibited a large difference in the relative abundances between different soils, with the exception of *Alphaproteobacteria*, where the relative abundance in various soils exhibited remarkable stability in spite of the changes in diversity rank (Spain *et al.*, 2009). The same observation was made during the bioremediation process with minor fluctuations in the 'total' bacterial community, whereas a noticeable fluctuation in the 'active' fraction was noticed (Table S1). Surprisingly, the RNA libraries' analysis highlighted that *Actinobacteria* were predominant (77.5%) in the 'active' bacterial community particularly after 24 months of oxygen biostimulation (Fig. 5). Initially, a bacterial community dominated by the Gram-negative bacteria (*Gammaproteobacteria*, *Alphaproteobacteria*) was present in the polluted soil (t_0); however, after 24 months, *Actinobacteria* were dominant, accompanied by lower proportions of *Gammaproteobacteria* (Table S1). *Actinobacteria* are known to be common in soil environments, and play a key role in the decomposition of organic matter (Prince *et al.*, 2010). One bacterial phylotype (FM209072), belonging to *Streptomyces*, dominated the

'total' bacterial community inhabiting the nontreated soil. This Gram-positive *Actinomycetales* (nondetected in the initial bacterial community) has been identified previously in moisture-limited environments contaminated with hydrocarbons (Pucci *et al.*, 2000).

Changes in the dominant phylotypes during the pilot-scale bioremediation process

During the bioremediation process, aeration induced the selection of active actinobacterial phylotypes (Fig. 6). The CMJA8 phylotype described previously was the most abundant phylotype during the last stage of the bioremediation process. In this active bacterial community, *Actinobacteria* increased from nearly 30% to approximately 80%. Furthermore, the actinobacterial phylotype ARN92 (AM936577) belonging to *Microbacterium* was the second most active phylotype at the end of the treatment. These observations mirrored results from culture-based approaches suggesting that *Actinobacteria* (*Actinomycetales*) might play a role in the bioremediation of alkane-contaminated moisture-limited soils (Pucci *et al.*, 2000). *Microbacterium* currently contains hydrocarbon degraders of high-molecular-weight polycyclic aromatic hydrocarbons (Gauthier *et al.*, 2003). Thus, the enhanced aerobic bioremediation stimulated CMJA8 and ARN92 phylotypes, which might be involved in the restoration of the ecosystem studied. As already mentioned, CMJA8 is a novel, deep-rooting lineage within the class *Actinobacteria*.

Bacterial core-set characterization

The study of bacterial communities during the entire bioremediation process highlighted the presence of a bacterial 'core-set' (OTUs present at every stage of the process), characterized by 10 phylotypes (representative sequences at t_0: the *Gammaproteobacteria* AMNG8, AMNA4, AMNA6, AMLA2, the *Betaproteobacteria* AMNB2, the *Alphaproteobacteria* AMNH8, the *Chloroflexi* AMLB4, AMKH1, and the *Actinobacteria* AMPH5, AMNA3), corresponding to 21% of the sequences obtained from the three DNA libraries. The first three phylotypes correspond to the dominant phylotypes described previously. We did not find sequences showing an identity of at least 97% (the threshold used to define OTU) to AMLA2 and AMLB4 by BLAST search. The best identities were found with sequences EU287122 (94% identity with AMLA2) and EF157231 (93% identity with AMLB4). AMNH8 (98% identity with AY625147), AMNA3 (98% identity with FJ972171) and AMNA4 (98% similarity to AY622251) were related to few sequences (4, 1 and 3 sequences, respectively) in public databases exclusively found in polluted environments (groundwater and soil contaminated by chlorinated compounds, wastewater, soils contaminated by hydrocarbons, heavy metals or uranium). AMKH1 (99% similarity to HM299006) and AMPH5 (97% identity with GQ500878) were also related to few sequences in the public databases found in various environments (soils, air, human skin). Finally, AMNG8 (97% identity with AB473904), AMNA6 (99% identity with AY955091) and AMNB2 (99% identity with AJ867658) were more frequently related to other sequences in public databases. AMNB2 was closely related to the recently genome-sequenced species *Methylibium petroleiphilum* (Kane *et al.*, 2007), a methylotroph able to degrade various hydrocarbons. The core-set was limited to three phylotypes (representative sequences at t_0: CMJA6, CMMB4, CMJA8) for the metabolically active communities, corresponding to 30% of the sequences obtained from the three RNA libraries. Phylotypes linked to CMJA6 (*Gammaproteobacteria*) and CMMB4 (*Alphaproteobacteria*) were described previously with the representative sequences of DNA libraries (AMLA2 and AMNH8, respectively). CMJA8 is the representative sequence of the active dominant phylotype already described.

Conclusions

Given the nature of hydrocarbon contamination of soils and the importance of bioremediation strategies, understanding the fate and behavior of aliphatic hydrocarbons is imperative, particularly microorganism–contaminant interactions (Stroud *et al.*, 2007). Biodegradation by microorganisms is the key removal process of hydrocarbons in soils, which is

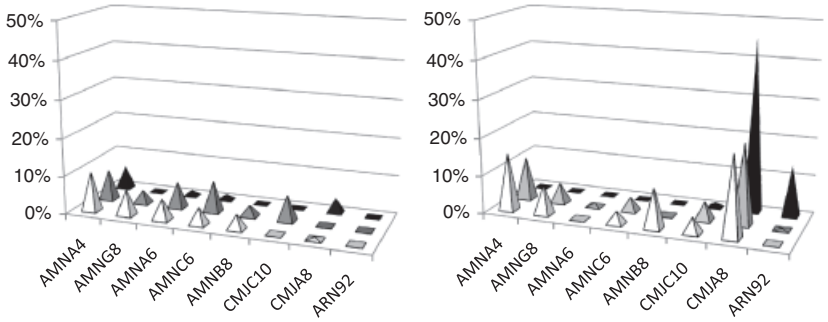

Fig. 6. Proportion of dominant phylotypes and changes during the bioremediation process of aliphatic hydrocarbon-polluted soil. Proportions of dominant phylotypes have been characterized in libraries obtained from 16S rRNA genes (a) and 16S rRNA (b) throughout the pilot-scale bioremediation process of the hydrocarbon-polluted soil (t_0, white; t_{+12}, gray; t_{+24}, dark gray).

controlled by hydrocarbon physicochemistry, environmental conditions, bioavailability and the presence of catabolically active microorganisms. Aeration induced a constant degradation of the pollutants, probably due to a combination of physical actions and biotic actions by the biostimulation of the microbial communities. We demonstrated that in a highly polluted ecosystem, the bacterial community is diverse and mainly composed of rare phylotypes. According to ecological and evolutionary theory, diverse communities provide a larger contribution to ecosystem functions and service compared with less diverse counterparts (Bell *et al*., 2005). A small number of dominant phylotypes (five *Gammaproteobacteria*, one *Deltaproteobacteria* and three *Actinobacteria*) were identified in the polluted environment. Within these, the CMJA8 and ARN92 phylotypes were highly active during the last stage of the enhanced aerobic bioremediation process, suggesting a role in restoring the ecosystem studied. These culture-independent studies provide a first look at potential hydrocarbon-degrading candidates inhabiting the studied polluted soil. However, an explicit link between bacterial populations and their role in hydrocarbon biodegradation has not been established.

Acknowledgements

This research was supported by the ACI Non Pollution-Depollution Oxygen project. C.M. was supported by a PhD fellowship from the 'Ministère de l'éducation, de la recherche et de la technologie' of France.

References

Aislabie JM, Balks MR, Foght JM & Waterhouse EJ (2004) Hydrocarbon spills on Antarctic soils: effects and management. *Environ Sci Technol* **38**: 1265–1274.

Ashby MN, Rine J, Mongodin EF, Nelson KE & Dimster-Denk D (2007) Serial analysis of rRNA genes and the unexpected dominance of rare members of microbial communities. *Appl Environ Microb* **73**: 4532–4542.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ & Weightman AJ (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microb* **72**: 5734–5741.

Atlas RM & Bartha R (1972) Degradation and mineralization of petroleum by two bacteria isolated from coastal waters. *Biotechnol Bioeng* **14**: 297–308.

Baker GC, Smith JJ & Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Meth* **55**: 541–555.

Baldwin BR, Nakatsu CH, Nebe J, Wickham GS, Parks C & Nies L (2009) Enumeration of aromatic oxygenase genes to evaluate biodegradation during multi-phase extraction at a gasoline-contaminated site. *J Hazard Mater* **163**: 524–530.

Bass DH, Hastings NA & Brown RA (2000) Performance of air sparging systems: a review of case studies. *J Hazard Mater* **72**: 101–119.

Bell T, Newman JA, Silverman BW, Turner SL & Lilley AK (2005) The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157–1160.

Cavalca L, Dell'Amico E & Andreoni V (2004) Intrinsic bioremediability of an aromatic hydrocarbon-polluted groundwater: diversity of bacterial population and toluene monoxygenase genes. *Appl Microbiol Biot* **64**: 576–587.

Clark DA & Norris PR (1996) *Acidimicrobium ferrooxidans* gen. nov., sp. nov.: mixed-culture ferrous iron oxidation with *Sulfobacillus* species. *Microbiology* **142**: 785–790.

Cole JR, Wang Q, Cardenas E *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Daane LL, Harjono I, Zylstra GJ & Haggblom MM (2001) Isolation and characterization of polycyclic aromatic hydrocarbon-degrading bacteria associated with the rhizosphere of salt marsh plants. *Appl Environ Microb* **67**: 2683–2691.

DeSantis TZ Jr, Hugenholtz P, Keller K *et al.* (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34**: W394–W399.

Dibble JT & Bartha R (1979) Effect of environmental parameters on the biodegradation of oil sludge. *Appl Environ Microb* **37**: 729–739.

Elshahed MS, Youssef NH, Spain AM *et al.* (2008) Novelty and uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microb* **74**: 5422–5428.

Fierer N & Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *P Natl Acad Sci USA* **103**: 626–631.

Gauthier E, Deziel E, Villemur R, Juteau P, Lepine F & Beaudet R (2003) Initial characterization of new bacteria degrading high-molecular weight polycyclic aromatic hydrocarbons isolated from a 2-year enrichment in a two-liquid-phase culture system. *J Appl Microbiol* **94**: 301–311.

Greer CW, Whyte LG & Niederberger TD (2010) Microbial communities in hydrocarbon-contaminated temperate, tropical, alpine, and polar soils. *Handbook of Hydrocarbon and Lipid Microbiology, Vol. 22* (Timmis KN, ed), pp. 2313–2328. Springer-Verlag, Berlin.

Hamamura N, Olson SH, Ward DM & Inskeep WP (2006) Microbial population dynamics associated with crude-oil biodegradation in diverse soils. *Appl Environ Microb* **72**: 6316–6324.

Hatzinger PB & Alexander M (2002) Effect of aging of chemicals in soil on their biodegradability and extractability. *Environ Sci Technol* **29**: 537–545.

Hutcheson MS, Pedersen D, Anastas ND, Fitzgerald J & Silverman D (1996) Beyond TPH: health-based evaluation of petroleum hydrocarbon exposures. *Regul Toxicol Pharm* **24**: 85–101.

Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microb* **72**: 1719–1728.

Kabelitz N, Machackova J, Imfeld G, Brennerova M, Pieper DH, Heipieper HJ & Junca H (2009) Enhancement of the microbial community biomass and diversity during air sparging bioremediation of a soil highly contaminated with kerosene and BTEX. *Appl Microbiol Biot* **82**: 565–577.

Kane SR, Chakicherla AY, Chain PS *et al.* (2007) Whole-genome analysis of the methyl tert-butyl ether-degrading beta-proteobacterium *Methylibium petroleiphilum* PM1. *J Bacteriol* **189**: 1931–1945.

Kersters K, De Vos P, Gillis M, Swings J, Vandamme P & Stackebrant E (2006) Introduction to the *Proteobacteria*. *The Prokaryotes*, Vol. 5, 3rd edn (Dworkin M, Falkow S, Rosenberg E, Schleifer KH & Stackebrandt E, eds), pp. 4–37. Springer, New York.

Khan FI, Husain T & Hejazi R (2004) An overview and analysis of site remediation technologies. *J Environ Manage* **71**: 95–122.

Kim JS & Crowley DE (2007) Microbial diversity in natural asphalts of the Rancho La Brea Tar Pits. *Appl Environ Microb* **73**: 4579–4591.

Kurahashi M, Fukunaga Y, Sakiyama Y, Harayama S & Yokota A (2009) *Iamia majanohamensis* gen. nov., sp. nov., an actinobacterium isolated from sea cucumber *Holothuria edulis*, and proposal of *Iamiaceae* fam. nov. *Int J Syst Evol Micr* **59**: 869–873.

Leahy JG & Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. *Microbiol Rev* **54**: 305–315.

Ludwig W, Strunk O, Westram R *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Luthy RG, Aiken GR, Brusseau ML *et al.* (1997) Sequestration of hydrophobic organic contaminants by geosorbents. *Environ Sci Technol* **31**: 3341–3347.

Margesin R, Labbe D, Schinner F, Greer CW & Whyte LG (2003) Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils. *Appl Environ Microb* **69**: 3085–3092.

Mills DK, Fitzgerald K, Litchfield CD & Gillevet PM (2003) A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum-contaminated soils. *J Microbiol Meth* **54**: 57–74.

Nyyssönen M, Kapanen A, Piskonen R, Lukkari T & Itävaara M (2009) Functional genes reveal the intrinsic PAH biodegradation potential in creosote-contaminated groundwater following *in situ* biostimulation. *Appl Microbiol Biot* **84**: 169–182.

Pedros-Alio C (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.

Pedros-Alio C (2007) Ecology. Dipping into the rare biosphere. *Science* **315**: 192–193.

Popp N, Schlomann M & Mau M (2006) Bacterial diversity in the active stage of a bioremediation system for mineral oil hydrocarbon-contaminated soils. *Microbiology* **152**: 3291–3304.

Prince RC, Gramain A & McGenity TJ (2010) Prokaryotic hydrocarbon degraders. *Handbook of Hydrocarbon and Lipid Microbiology*, *Vol. 19* (Timmis KN, ed), pp. 1669–1692. Springer-Verlag, Berlin.

Pucci OH, Bak MA, Peressutti SR, Klein I, Härtig C, Alvarez HM & Wünsche L (2000) Influence of crude oil contamination on the bacterial community of semiarid soils of Patagonia (Argentina). *Acta Biotechnol* **20**: 129–146.

Ranjard L, Lejon DP, Mougel C, Schehrer L, Merdinoglu D & Chaussod R (2003) Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environ Microbiol* **5**: 1111–1120.

Raup DM (1972) Taxonomic diversity during the Phanerozoic. *Science* **177**: 1065–1071.

Rheims H, Felske A, Seufert S & Stackebrandt E (1999) Molecular monitoring of an uncultured group of the class *Actinobacteria* in two terrestrial environments. *J Microbiol Meth* **36**: 65–75.

Rossetti S, Tomei MC, Nielsen PH & Tandoi V (2005) '*Microthrix parvicella*', a filamentous bacterium causing bulking and foaming in activated sludge systems: a review of current knowledge. *FEMS Microbiol Rev* **29**: 49–64.

Schloss PD & Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microb* **71**: 1501–1506.

Schloss PD & Handelsman J (2006) Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microb* **72**: 6773–6779.

Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *P Natl Acad Sci USA* **103**: 12115–12120.

Spain AM, Krumholz LR & Elshahed MS (2009) Abundance, composition, diversity and novelty of soil Proteobacteria. *ISME J* **3**: 992–1000.

Stackebrandt E, Rainey FA & Ward-Rainey NL (1997) Proposal for a new hierarchic classification system, Actinobacteria classis nov. *Int J Syst Bacteriol* **47**: 479–491.

Staden R (1996) The Staden sequence analysis package. *Mol Biotechnol* **5**: 233–241.

Stroud JL, Paton GI & Semple KT (2007) Microbe–aliphatic hydrocarbon interactions in soil: implications for biodegradation and bioremediation. *J Appl Microbiol* **102**: 1239–1253.

Suzuki MT & Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microb* **62**: 625–630.

Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Troquet J, Larroche C & Dussap C-G (2003) Evidence for the occurrence of an oxygen limitation during soil bioremediation by solid-state fermentation. *Biochem Eng J* **13**: 103–112.

van Beilen JB & Funhoff EG (2007) Alkane hydroxylases involved in microbial alkane degradation. *Appl Microbiol Biot* **74**: 13–21.

von Wintzingerode F, Gobel UB & Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213–229.

Wentzel A, Ellingsen TE, Kotlar HK, Zotchev SB & Throne-Holst M (2007) Bacterial metabolism of long-chain *n*-alkanes. *Appl Microbiol Biot* **76**: 1209–1221.

Yang X, Beckmann D, Fiorenza S & Niedermeier C (2005) Field study of pulsed air sparging for remediation of petroleum hydrocarbon contaminated soil and groundwater. *Environ Sci Technol* **39**: 7279–7286.

Youssef NH & Elshahed MS (2009) Diversity rankings among bacterial lineages in soil. *ISME J* **3**: 305–313.

Zhi XY, Li WJ & Stackebrandt E (2009) An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int J Syst Evol Micr* **59**: 589–608.

Zhou J, Bruns MA & Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microb* **62**: 316–322.

Zucchi M, Angiolini L, Borin S *et al*. (2003) Response of bacterial community during bioremediation of an oil-polluted soil. *J Appl Microbiol* **94**: 248–257.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Bacterial diversity observed during the bioremediation process of the aliphatic hydrocarbon-polluted soil.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.