Journal of Hazardous Materials 177 (2010) 593–601

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Molecular analysis of the catechol-degrading bacterial community in a coal wasteland heavily contaminated with PAHs

Najoi El Azhari*, Marion Devers-Lamrani, Guillaume Chatagnier, Nadine Rouard, Fabrice Martin-Laurent

INRA/Université de Bourgogne, Soil and Environmental Microbiology, 17 rue sully, BP 86510, 21065 Dijon Cedex, France

ARTICLE INFO

Article history:
Received 18 June 2009
Received in revised form
15 December 2009
Accepted 17 December 2009
Available online 7 January 2010

Keywords:
Catechol
Soil DNA
PCR
Biodiversity
catA sequence

ABSTRACT

A PCR-based molecular tool was developed to estimate the diversity of the catechol-degrading bacterial community in a coal wasteland heavily contaminated with PAHS. A degenerate primer pair specific to *catA* sequences was designed by multiple alignment of known sequences coding a key intermediate of the β-ketoadiapate pathway degrading catechol, namely catechol 1,2-dioxygenase. The specificity of this primer pair was assessed in 21 pure strains by PCR and sequencing. Comparison of the 16S rDNA and *catA* phylogenies revealed an absence of congruence between these two genes. The primer set was able to amplify *catA* sequences in DNA extracts from an industrial soil highly contaminated with heavy metals and polycyclic aromatic hydrocarbons (PAHs). RFLP screening of the *catA* library (95 clones) yielded 32 RFLP families. All of the 43 clone sequences obtained exhibited 86% identity on average to known CatA. Phylogenetic analysis revealed that these CatA sequences were related to Actinobacteria, α-, β- and γ-Proteobacteria phyla and confirmed the absence of congruence with 16S rDNA sequences, which implies horizontal gene transfer of the cat gene cluster between soil microbiota. Our results suggest that the diversity of the *catA* bacterial community is maintained in highly contaminated soil.

1. Introduction

As a result of the co-evolution of bacteria and plants, the soil microflora has adapted to make use of plant-derived aromatic compounds such as lignin [1,2]. Industrial activities and intensive agriculture have polluted the environment with numerous aromatic compounds such as polycyclic aromatic hydrocarbons (PAHs) and pesticides. These compounds are known to accumulate in the food chain and have adverse effects on animal and human health, and also to be toxic to living organisms and affect functioning of the overall ecosystem [3–6]. The bacterial populations involved in degrading these natural compounds have subsequently adapted to xenobiotic degradation through processes involving gene fluxes and mutations [7–9]. They are thus of major significance as regards soil ecosystem services [10]. Some of the enzymes involved in the biodegradation of phenolic compounds have been extensively characterised in several cultivable microbial strains [11] but the ecological importance of these strains regarding aromatics biodegradation in soil is not yet well established. Bacterial communities possessing the sets of genes which code these enzymes almost invariably transform such toxic compounds into protocatechuate and catechol [12,13]. In aerobic environments the main route of degradation of these molecules is the β-ketoadipate pathway [14,15].

The initial step of the catechol branch is operated by the catechol 1,2-dioxygenase (1,2-CTD) which catalyses the ring fission [15]. The gene encoding the 1,2-CTD, namely *catA*, is widely dispersed within soil microbiota [15]. Despite the important gene sequence polymorphism in phylogenetically divergent microorganisms, the functional redundancy leads to conserved region within the active site [16–18]. Thus the catechol-degrading bacterial community plays a key role in carbon cycling as well as in pollutant degradation in the environment [4,19].

However, the ecology of this microbial community has remained obscure in terms of the diversity of this community, in response to PAHs contamination. Most studies have focused on detailed description of the processes responsible for the specific activity of this enzyme or on characterisation of the *cat* operon in pure strains [12,20,21]. The recent development of new molecular tools, based on direct amplification of soil DNA extracts, provides a welcome alternative to the Pasteurian approaches as these are known to be biased, due to the poor culturability of most soil microorganisms. Several studies report the development of primer

* Corresponding author. Present address: Welience Agro-Environment, 8 Avenue Jean Bertin, BP 66917, 21065 Dijon Cedex, France. Tel.: +33 3 80 69 35 93; fax: +33 3 80 69 32 24.
E-mail addresses: najoi.el-azhari@welience.com, najoi.el-azhari@dijon.inra.fr (N. El Azhari).

0304-3894/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2009.12.074

Table 1
Twenty-one bacterial strains (set A) and culture media used for testing the specificity of the primer set. Lines in bold font correspond to strains harbouring genes coding both intradiol dioxygenases of the β-ketoadipate pathway (i.e. *catA* and *pcaH*).

Strain	Phylum	Source	Medium	pcaH	catA	% identity
Rhodococcus globerulus PWD1	Actinobacteria	University of Wales, UK	LB	–	–	
Geodermatophilus sp. B1	Actinobacteria	Bayer Crop Science, France	LB	+	–	
Corynebacterium equi ATCC14887	Actinobacteria	University of Witwatersrand, South Africa	LB	+	–	
Streptomyces sp. 2065	Actinobacteria	University of British Columbia, Canada	ISP4	+	–	
Streptomyces argentoleus B15	Actinobacteria	Bayer Crop Science, France	LB	–	–	
Streptomyces argilaceus	Actinobacteria	University of Oviedo, Spain	LB	–	–	
***Streptomyces verticillium* B4**	**Actinobacteria**	Bayer Crop Science, France	**LB**	+	+	**87% *Burkholderia* sp.**
Streptomyces griseus B24	Actinobacteria	Bayer Crop Science, France	LB	–	–	
Streptomyces griseolus B22	Actinobacteria	Bayer Crop Science, France	LB	–	–	
***Sphingomonas* sp. RW1**	**α-Proteobacteria**	**GBF, Germany**	**M9 + benzofuran**	+	+	**99% *Pseudomonas putida* KT4220**
Ralstonia sp. JS705 (CLCA)	β-Proteobacteria	AFRL/MLQL, Florida	Cardiff	–	–	
***Ralstonia eutropha* JMP134**	**β-Proteobacteria**	**IOEZ, Germany**	**Cardiff**	+	+	**88% *Pseudomonas putida***
Variovorax paradoxus C1Cl	β-Proteobacteria	INRA-Dijon, France	LB	+	–	
Variovorax sp. S1lq	β-Proteobacteria	INRA-Dijon, France	LB	–	–	
Ralstonia sp. TCF	β-Proteobacteria	INRA-Dijon, France	Cardiff	+	–	
Ralstonia sp. JD6	β-Proteobacteria	INRA-Dijon, France	Cardiff	+	–	
Alcaligenes eutrophus JMP131	β-Proteobacteria	INRA-Dijon, France	Cardiff	+	–	
Pseudomonas azelaïca	γ-Proteobacteria	EAWAG, Switzerland	LB	–	+	84% *Pseudomonas putida*
***Pseudomonas* sp. 6978**	**γ-Proteobacteria**	**University of Minnesota**	**Brunner Medium + 3-chlorobenzoate**	+	+	**84% *Pseudomonas putida***
***Pseudomonasputida* ATCC23975**	**γ-Proteobacteria**	**Lispcomb Lab, Minnesota**	**LB**	+	+	**73% *Pseudomonas putida***
***Pseudomonasputida* DSM2112**	**γ-Proteobacteria**	**DSMZ, Germany**	**LB + streptomycin**	+	+	**99% *Pseudomonas putida* KT4220**

Medium references: ISP4 [39], M9 + benzofuran [40], Cardiff [41], Brunner Medium + 3-chlorobenzoate [42].

Table 2
Sequences used for the comparison of phylogenies. Catechol 1,2-dioxygenase (1,2-CTD) accession numbers in bold correspond to amino acid sequences derived from complete genome, chromosome or plasmid sequences (i.e. sequences not from partial genomes such as operons or amplicons). CCD signifies chlorocatechol dioxygenase. Genomic location on plasmid or chromosome is indicated, ND means non-determined (i.e. sequences from complete genomes).

Strain	Phylum	Catechol 1,2-dioxygenase Acc. nos.	Genomic location	16S Acc. nos.
Streptomyces verticilium B4	Actinobacteria	EU169382	ND	FJ227307
Rhodococcus opacus strain 1CP	Actinobacteria	DQ146627 (CCD)	ND	Y11893
Mycobacterium gilvum PYR-GCK	Actinobacteria	**YP_001132628.1 (1,2-CTD)**	Chromosome	NC_009338
Arthrobacter sp. FB24 c	Actinobacteria	**YP 831415.1 (1,2-CTD)**	Chromosome	NC_008541
Xanthobacter autotrphicus Py2	α-Protobacteria	**ABS66382.1 (1,2-CTD)**	ND	NC_009720
Sphingomonas sp. RW1	α-Protobacteria	EU169377	ND	CP000699
Burkholderia pseudomallei	β-Protobacteria	**YP_001076597.1 (1,2-CTD)**	Chromosome	NC_009078
Ralstonia eutropha JMP134	β-Protobacteria	EU169383	ND	NC_007347
Bordetella petrii DSMZ 12804	β-Protobacteria	**YP_001630005.1 (1,2-CTD)**	ND	NC_010170
Achromobacter denitrificans	β-Protobacteria	**NP_990899.1 (CCD)**	Plasmid	EU622534
Pseudomonas putida ATCC23975	γ-Protobacteria	EU169379	ND	FJ227304
Pseudomonas sp. DSMZ 6978	γ-Protobacteria	EU169380	ND	FJ227306
Pseudomonas azelaica	γ-Protobacteria	EU169381	ND	FJ227303
Pseudomonas putida DSMZ 2112	γ-Protobacteria	EU169378	ND	FJ227305
Pseudomonas arvilla	γ-Protobacteria	D37783 (1,2-CTD)	ND	AB236664
Pseudomonas aeruginosa J5-2	γ-Protobacteria	EF111021 (CCD)	ND	EF530572
Acinetobacter baumannii AYE	γ-Protobacteria	**YP 001713609.1 (1,2-CTD)**	Chromosome	NC_010410
Pseudomonas putida F1				
a	γ-Protobacteria	**ABQ78195.1 (1,2-CTD)**	ND	NC_009512
b		**ABQ78684.1 (1,2-CTD)**	ND	
Pseudomonas aeruginosa PAO1	γ-Protobacteria	**NP 251197.1 (1,2-CTD)**	ND	NC_002516
Pseudomonas nitroreducens Jin-1	γ-Protobacteria	**ABL07613.1 (CCD)**	ND	EF534986

sets aiming at amplifying genes related to PAHs degradation on pure strains [22,23]. To our knowledge, only a few number of studies relates the use of *catA* PCR quantification on seawater but the identity of the amplicons was not verified [22,24]. Molecular approaches based on amplification of functional genes from environmental DNA will provide new insight in catechol-degrading community playing a key role in soil ecosystemic services. In addition, it will offer a new tool to monitor this bacterial community in response to disturbances of the soil environment. Therefore, our aim was to estimate the diversity of the catechol-degrading bacterial community in soil by designing a degenerate primer set which targets the *catA*. The specificity of this primer pair was assessed on DNA extracted from pure strains known to degrade aromatic compounds. This new molecular tool, which involved PCR amplification, clone library screening and sequence analyses, was then used to describe the diversity of bacterial *catA* in a coal wasteland heavily contaminated with PAHs.

2. Materials and methods

2.1. Bacterial strains—DNA extraction

DNA from 21 bacterial strains (set A), belonging to Actinobacteria, α-, β- and γ-Proteobacteria and harbouring various oxygenase genes, was used as template to check the specificity of the primer set (Table 1). The bacterial strains were grown in appropriate liquid media (Table 1) at 28 °C for 16 h at 150 rpm. DNA was prepared from 1 mL of culture aliquot using a proteinase K treatment and thermal shock as described previously [25], then stored at −20 °C until used.

2.2. Phylogenetic analysis of the catA gene and PCR primers design

16S rDNA and *catA* phylogenies were established and compared to estimate possible congruence. Twenty sequences of 16S rDNA (set B1) from Actinobacteria, α-, β- and γ-Proteobacteria phyla were aligned to construct a phylogenetic tree which was then compared with the corresponding *catA* and/or *ccd* (chlorocatechol 1,2-dioxygenase) gene (set B2) sequence-based phylogenetic tree (Table 2).

Gene sequences were retrieved from the GenBank database or from this work. Missing 16S rDNA sequences from *Pseudomonas azelaica*, *Pseudomonas putida* ATCC23975, *Pseudomonas putida* DSMZ 2112, *Pseudomonas* sp. DSMZ 6978 and *Streptomyces verticillium* B4 were obtained by amplifying, cloning and sequencing amplicons using the universal primer set 27_f 5′-AGAGTTTGATCMTGGCTCAG-3′ and 1492_r 5′-GGYTACCTTGTTACGACTT-3′. Alignments were obtained by using the ClustalX program (ftp://ftp.ebi.ac.uk/pub/software/clustalw2/2.0.10/) and neighbour-joining trees were generated by applying the NJPlot program (http://pbil.univ-lyon1.fr/software/njplot.html). The gram+ sequences were used to form the outgroup of the 16S rDNA tree and Arthrobacter sp. FB24, *Mycobacterium gilvum* PYR-GCK, *Rhodococcus opacus* 1CP, *Achromobacter denitrificans* L2, *Pseudomonas nitroreducens* Jin1 and *Pseudomonas aeruginosa* J5-2 sequences to form the outgroup of the *catA* tree. An applet was used for the pairwise comparison of alternative phylogenies http://www.mas.ncl.ac.uk/~ntmwn/phylo_comparison/pairwise.html.

Ten 1,2-CTD (CatA) protein sequences (set C) from γ-Proteobacteria and Actinobacteria phyla were aligned using the ClustalX program (GenBank accession number: O33538) (*R. opacus*); O33950 (*Acinetobacter lwoffii*); P07773 (*Acinetobacter* sp. ADP1); P31019 (*Pseudomonas* sp. EST1001); Q43984 (*P. putida*); Q51433 (*P. putida*); Q51960 (*P. putida*); Q52041 (*P. putida*); Q9Z9Y0 (*Frateuria* sp: ANA-18); Q9Z9Y4 (*Frateuria* sp. ANA-18). PCR primers were designed on the basis of this subset of *catA* sequences (C, Fig. 1), using the Primer3 program available online (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). This primer set was tested *in silico* on known *catA* sequences.

2.3. catA PCR amplification conditions

catA PCR products were amplified from 12.5 ng of bacterial DNA in a final volume of 25 μL containing 1.5 U of *Taq* DNA polymerase (Q-biogene, France), 50 pmol of $CATA_f$ 5′-ACV CCV CGH ACC ATY

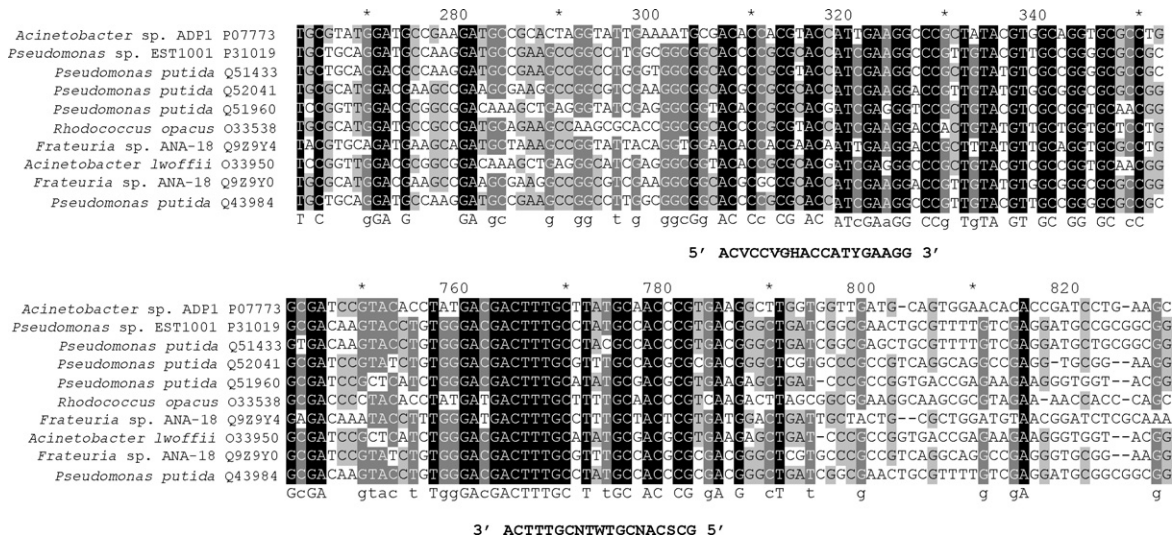

Fig. 1. Primer set design based on the multiple alignment of 10 *catA* sequences (set C, GenBank accession numbers: O33538; O33950; P07773; P31019; Q43984; Q51433; Q51960; Q51041; Q9Z9Y0; Q9Z9Y4). The CATAf/r degenerate primer is indicated beneath the alignment (H: A/C/T; N: A/C/G/T; S: G/C; W: A/C/G; Y: C/T).

GAA GG-3′ (GC content: 60.8% ca.; hypothetical T_m: 61.7 °C ca.), and 50 pmol of $CATA_r$. 5′-CGS GTN GCA WAN GCA AAGT-3′ (GC content: 52.6% ca.; hypothetical T_m: 56.7 °C ca.). Touchdown PCR amplification was performed with a MasterCycler (Eppendorf, Germany) under the following conditions: 94 °C for 5 min, 8 cycles of 94 °C for 1 min, 62 °C for 1 min decreasing by 0.5 °C per cycle until the annealing temperature reached 58 °C, elongation step at 72 °C for 1 min. An additional 30 cycles were performed at an annealing temperature of 58 °C followed by a final elongation at 72 °C for 7 min.

2.4. Cloning and sequencing of catA amplicons

The *catA* PCR products in the agarose gel were purified using the MinElute Gel Extraction Kit according to manufacturer's instructions (Qiagen, Germany). The PCR fragments were ligated into the pGEM-T easy® Vector (Promega, WI, USA) following the protocol provided by the manufacturer. Recombinant clones were screened by SP6/T7 PCR. The positive clones were sequenced with a CEQ-8000 automatic sequencer using the DTCS-1 kit according to the supplier's recommendations (Beckman Coulter, CA, USA). The *catA* and 16S rDNA sequences were deposited at Genbank under the accession numbers indicated in Table 2.

2.5. Soil samples

Soil was collected from a coal wasteland at Neuves-Maisons (Lorraine, France) in accordance with ISO 10381-6 recommendations (Soil quality – Sampling – Part 6: Guidance on the collection, handling and storage of soil for the assessment of aerobic microbial processes in the laboratory). A composite sample of the Ap horizon (0–10 cm) was collected from each experimental field [26]. The soil samples were immediately sieved (5-mm mesh) to remove plant debris but retain a maximum number of aerobic microorganisms within the aggregates. 250-mg aliquots were promptly frozen in liquid nitrogen and stored at −80 °C until soil DNA extraction.

The soil physicochemical properties were as follows: clay (12.6%), silt (26.1%), sand (61.3%), organic carbon (7.06%), total nitrogen (0.27%), C/N (25.8), organic matter (12.2%) and pH 7.13. The soil concentrations of PAHs and heavy metals are shown in Table 3. The total heavy metals and PAH measurements were processed using the standardised methods NF EN ISO 1 1885 and XP-X-33-012, respectively.

2.6. Soil DNA extraction

DNA was extracted from three independent 250 mg replicates of soil sample, according to Martin-Laurent et al. [27]. The DNA extracts were purified using a polyvinylpolypyrrolidone spin column and a Sepharose 4B spin column, then checked for integrity by electrophoresis on a 1% agarose gel and quantified on gel using standard concentrations of calf thymus.

2.7. catA library screening and sequencing

catA PCR fragments generated from the three soil replicates were pooled, cloned and screened as described above. PCR products of the recombinant clones (5 μL) were digested overnight at 37 °C with MspI restriction enzyme. The choice of restriction enzyme was based on *in silico* digestion of *catA* sequences from seven pure strains in our collection (cf. §2.4). Ninety-five clones were subjected to RFLP screening. At least one clone was sequenced for each RFLP type identified. Sequencing was carried out as described above.

Amino acid sequences which reveals the specific signature of CatA enzyme was used to verify the identity of the isolated PCR fragments. Amino acid sequences were deduced using the Expasy – Translate Tool (http://www.expasy.ch/tools/dna.html). Closest relatives were obtained using the Basic Local Alignment Search Tool (BLAST) P (http://www.ncbi.nlm.nih.gov/BLAST/). Sequences were aligned with ClustalX software. A neighbour-joining tree was computed with the NJPlot software package (i.e. § 2.2).

2.8. Diversity analysis

A rarefaction curve was obtained by applying the freeware program aRarefactWin.exe (http://www.uga.edu/~strata/AnRareReadme.html). The diversity of the *catA* bacterial community was assessed using (i) the Shannon–Weiner index calculated as: $H' = -\sum p_i \ln p_i$, where $p_i = \frac{n_i}{N}$ and n_i is the number of clones in each phylotype and N the total number of clones and (ii) the Simpson Diversity index D calculated as: $D = \sum(p_i^2)$.

Table 3
Amounts of PAHs and heavy metal contaminants measured in the Neuves-Maisons soil.

	Neuves-Maisons (NM)	Permissible maximum level of heavy metals in soils according to Council Directive 86/278 of June 1986
Trace metal elements		
Copper (mg kg^{-1})	95.5	100
Zinc (mg kg^{-1})	2770	300
Lead (mg kg^{-1})	683	500
Cadmium (mg kg^{-1})	2.37	2
Mercury (mg kg^{-1})	3.38	1
PAH concentration (mg kg^{-1})		
Fluorene	34	
Phenanthrene	120	
Anthracene	51	
Fluoranthene	210	
Pyrene	170	
Benz[*a*]anthracene	105	
Chrysene	95	
Benzo[*b*]fluoranthene + benzo[*k*] fluoranthene	144	
Benzo[*a*]pyrene	92	
Benzo[*a*,*h*]anthracene	10	
Benzo[*g*,*h*,*i*]perylene	58	
Total 16 PAHs (mg kg^{-1})	1219	

3. Results

3.1. Phylogenetic analysis

An *in silico* study was performed to establish the level of congruence between the *catA* and 16S rDNA phylogenies. Two data sets (B1 and B2, Table 2) were used to infer phylogenetic trees, firstly from a ca. 750-nt alignment of Actinobacteria, α-, β- and γ-Proteobacteria *16S rDNA* sequences and secondly from a 470-nt alignment of *catA* sequences from the same taxa. The sequences were selected from the available *16S rDNA* and *catA* sequences identified in these same microorganisms. The tree based on alignment of the *16S rDNA* from 20 sequences of ca. 750 bp exhibited two main clusters (Fig. 2a), one of Actinobacteria, the other of Proteobacteria. As expected, the second cluster contained three sub-groups, consisting of α-, β- and γ-Proteobacteria, respectively. The *catA* tree, in contrast, showed two main groups with the sequences spread along the tree in different clusters (Fig. 2b). The first group consisted of Actinobacteria, β- and γ-Proteobacteria and the second of Actinobacteria and α-, β- and γ-Proteobacteria.

The phylogeny inferred from the *catA* gene sequence alignment (set B2) was arrayed against the phylogeny obtained from *16S rDNA* gene (set B1) analysis (Fig. 2a and b). The structure of the *16S rDNA* phylogenetic tree was consistent with the expected taxonomy. The

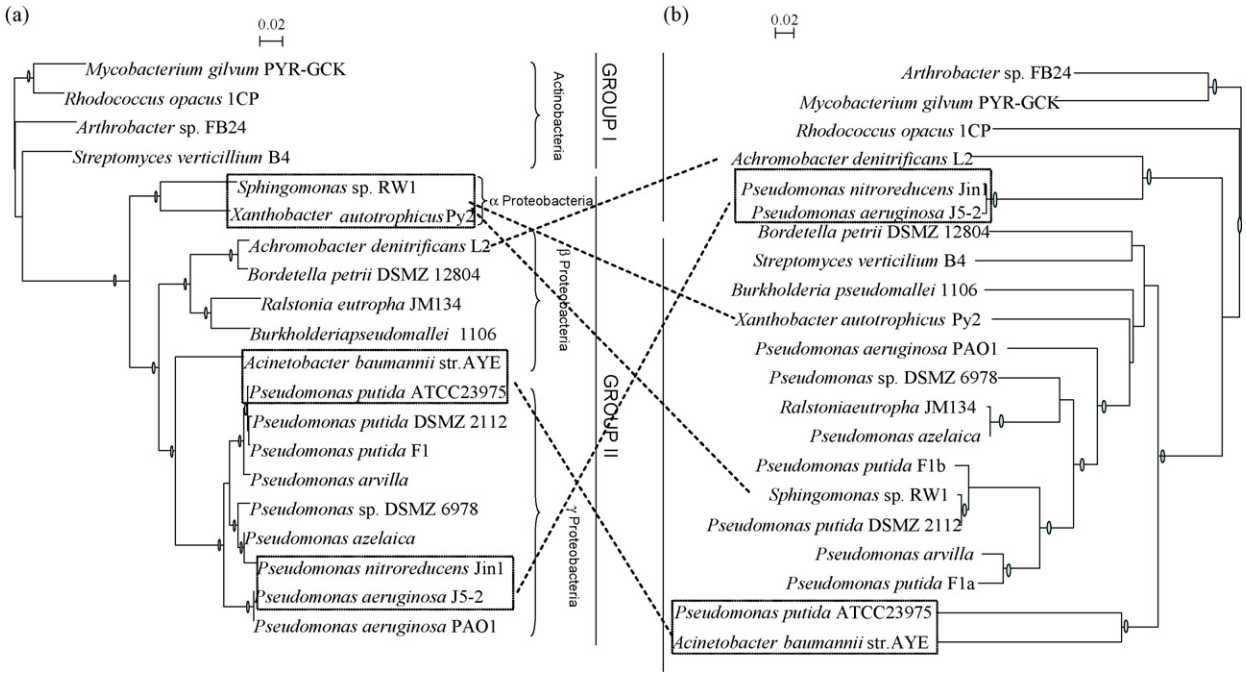

Fig. 2. Comparison of 16SrDNA and *catA* phylogenies (sets B1 and B2). (a) Phylogenetic tree of 16S rDNA sequences based on the alignment of ca. 750-nt fragments: the two clusters are indicated as Group I and Group II. The different sub-clusters gather different phyla as indicated. Bootstrap values greater than 700 iterations (of 1000) are marked at branch nodes. (b) Phylogenetic tree of *catA* sequences based on the alignment of ca. 470-nt fragments: the two clusters are indicated as Group I and Group II. Bootstrap values greater than 700 iterations (out of 1000) are marked at branch nodes. Discontinuous lines between the two trees represent examples of non-congruence in the phylogenies.

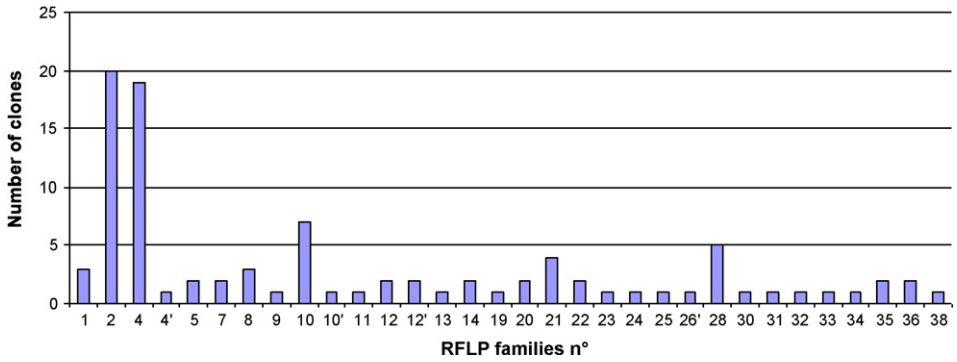

Fig. 3. Distribution of *catA* RFLP families.

two main clusters (i.e. group I and group II) segregated into Actinobacteria and Proteobacteria phyla with a bootstrap support of 100% (Fig. 2a) and the Proteobacteria correctly divided into α-, β- and γ-Proteobacteria classes. In contrast, the two groups inferred from the *16S rDNA* phylogenies (i.e. group I and group II) were not preserved in the *catA* tree. Although two clusters (i.e. groups I and II, Fig. 2b) were indeed present, the taxonomical structure was not retrieved suggesting non-congruence of the two phylogenies. For example, in the *16S rDNA* phylogenetic tree, *Pseudomonas* spp. were grouped within the same cluster in group II, but were spread over both groups in the *catA* tree, with a bootstrap support of 100% (Fig. 2a and b). Pairwise comparison of the two trees revealed a likelihood of 44.5% and statistically confirmed this observation (data not shown).

3.2. catA primers design

Sixty dioxygenase sequences (i.e. PcaH and CatA) were subjected to multiple alignment in order to reveal conserved nucleotide regions in which degenerated primers could be designed. The tremendous variability of the dioxygenase amino acid sequences was confirmed (data not shown, i.e. Supplementary data 1). A subset of 10 *catA* sequences (set C) from γ-Proteobacteria was aligned and used to design a degenerate primer pair: CATAf 5′-ACV CCV CGH ACC ATY GAA GG-3′ and CATAr 5′-CGS GTN GCA WAN GCA AAGT-3′ (Fig. 1) which amplified a 470 bp *catA* fragment. The specificity of this primer pair was tested on DNA extracts from 21 bacterial strains (set A) in the Actinobacteria, α-, β- and γ-Proteobacteria phyla, known to harbour oxygenase activity. Seven strains in this bacterial collection exhibited a single *catA* PCR fragment of the expected size: *Pseudomonas azelaïca*, *Pseudomonas* sp. DSMZ 6978, *P. putida* ATCC 23975, *P. putida* DSMZ 2112, *Ralstonia eutropha* JMP 134, *Sphingomonas* sp. RW1 and *S. verticillium* B4. Fourteen strains did not exhibit any *catA* amplification (Table 1). The identities of the deduced amino acid sequences to known CatA sequences, ranged from 87 to 100%.

3.3. catA sequence diversity in soil

3.3.1. RFLP family diversity

catA sequences were amplified from Neuves-Maisons soil (Table 3). This coal wasteland can be considered as heavily contaminated with PAHs since most heavy metal concentrations are above the EU legislation limits (Council Directive 86/278 of June 1986, Table 3) and the PAH concentration is high (i.e. 1 219 mg per kg of soil, Table 3).

PCR amplification yielded a single amplicon of approximately 470 bp. The *catA* amplicons were used to construct a clone library. Ninety-five clones were screened by PCR-RFLP analysis using MspI restriction enzyme and clustered according to pattern into RFLP families. Thirty-two *catA* RFLP families were identified. The RFLP profile distribution revealed two dominant profiles (i.e. RFLP profile nos. 2 and 4, Fig. 3), each representing approximately 20% of the total screened clones. Numerous poorly represented RFLP profiles, representing approximately 60% of the total screened clones, were also detected. This co-dominance was corroborated by the Simpson index value of $D = 0.11$.

The species richness of the *catA* sequences recovered from the Neuves-Maisons soil DNA was then assessed by rarefaction curve analysis in which the number of RFLP profiles was represented in relation to the number of screened clones (Fig. 4). The asymptote of the rarefaction curve was not attained which indicated that the total diversity of the *catA* sequences had not been recovered. This was corroborated by the Shannon–Weiner index of $H' = 2.99$.

3.3.2. Phylogenetic diversity

The diversity of the *catA* sequences was evaluated by sequencing 43 clones from the 32 RFLP types in the *catA* library (1–3 clones per RFLP family). The clones exhibited a slight length polymorphism ranging from 464 to 473 bp. Comparison of the amino acid sequences deduced from the *catA* nucleotidic sequences with CatA sequences derived from known bacteria revealed identities ranging from 64 to 100%. The CatA amino acid sequences were aligned and a phylogenetic tree was constructed (Fig. 5). The CatA sequences branched with bacterial sequences belonging to Actinobacteria, α-, β- and γ-Proteobacteria. Two distinct clusters were identified at 5% similarity.

Cluster 1 consisted of the Actinobacterium *S. verticillium* B4 and grouped 6 CatA sequences representing 4 RFLP profiles (i.e. RFLP profile nos. 22, 28, 30 and 31). Cluster 2 grouped 37 CatA sequences and could be divided into two main sub-clusters with 10% similar-

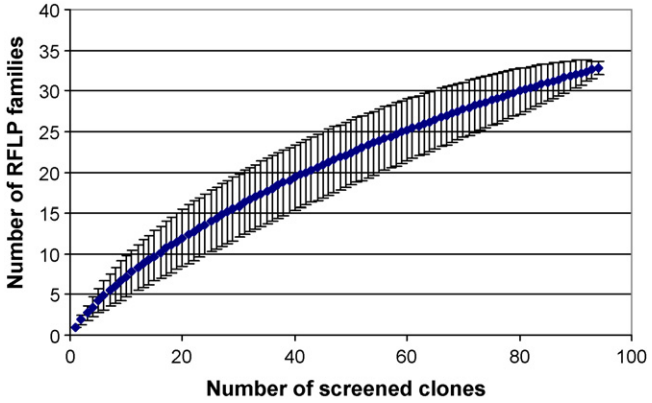

Fig. 4. Rarefaction curve of the observed diversity of *catA* RFLP families in the Neuves-Maisons soil. The error bars are 95% confidence intervals calculated from the variance of the number of phylotypes.

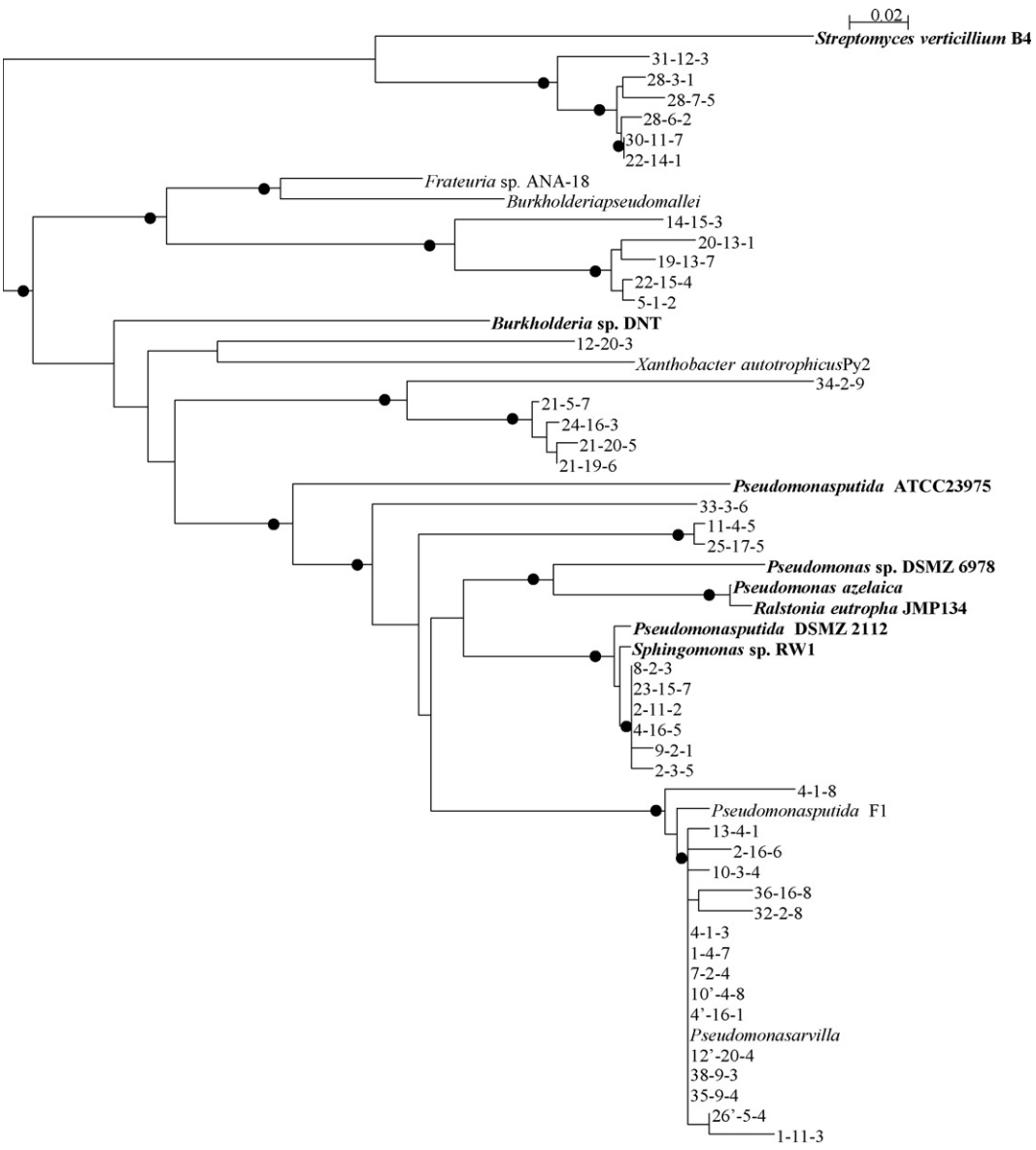

Fig. 5. Phylogenetic tree of CatA sequences derived from Neuves-Maisons soil. The tree is based on the ClustalX multiple alignment of deduced amino acid sequences of CatA. Strains indicated in bold font correspond to the CatA sequences described in this study. Bootstrap values greater than 700 iterations (out of 1000) are marked at branch nodes. The clones are named as follows: RFLP family number–clone number (composed of two figures).

ity. The first sub-cluster grouped five CatA sequences representing five different RFLP profiles (nos. 5, 15, 19, 20 and 22), branching with α-, β- and γ-Proteobacteria CatA amino acid sequences. Sub-cluster 2 consisted mainly of γ-Proteobacteria CatA sequences and grouped 31 clone sequences representing 24 RFLP profiles. Three of these profiles (i.e. RFLP profiles 21, 24 and 34) did not branch significantly with any known CatA sequence. Sub-cluster 2 also enclosed CatA sequences belonging to RFLP profiles 2, 4, 8, 9, 11, 23, 25 and 33, which showed significant identity to *Pseudomonas*-, *Sphingomonas*- and *Ralstonia*-like CatA. Finally, RFLP profiles 1, 2, 4, 4′, 7, 10, 10′, 12′, 13, 26′, 32, 35 and 38 exhibited significant identity to *Pseudomonas* CatA.

4. Discussion

Microbial oxygenases as determinants of the fate of aromatic compounds in the environment, notably synthetic aromatic contaminants, are of major interest [15]. Dioxygenases of the β-ketoadipate (*ortho*-) pathway are known to be widespread in bacteria. Catechol dioxygenases, in particular, play a key role in the biodegradation of xenobiotics [4]. Our aim here was to develop a molecular tool to estimate the diversity of the catechol-degrading bacterial community in a soil heavily contaminated with PAHs.

The developed primer set amplified *catA* gene fragments from seven pure strains in the screened bacterial collection (set A), known for their oxygenase activity or involvement in aromatics biodegradation. Thirteen of the 21 strains tested here (set A) exhibited the *pcaH* gene [28], seven the *catA* gene and six exhibited both (Table 1). Based on this analysis, we could conclude that at least 14 strains degrade PAHs through the β-ketoadipate pathway. One could hypothesise that the seven remaining strains either (i) possessed *catA* or *pcaH* genes presenting sequence polymorphisms, such as oxygenases from modified *ortho*-pathways, which prevented amplification with the CATA or PCAH primer pairs or (ii) used another pathway (i.e. catechol 2,3-dioxygenase) [15]. Thus the β-ketoadipate pathway, which plays a central role in the degradation of naturally occurring aromatic compounds derived from plant components as well as that of environmental pollu-

tants such as PAHs, is widespread in microorganisms [15]. This further justifies the need for a molecular tool to monitor this functional community in soils. Unexpectedly, amplicons were not only obtained from the γ-Proteobacteria phylum. Although the primer set design was based on γ-Proteobacterial *catA* sequences, an amplicon corresponding to the CatA sequence was also exhibited by *Sphingomonas* sp. RW1 (α-Proteobacterium), *R. eutropha* JMP 134 (β-Proteobacterium) and *S. verticillium* B4 (Actinobacterium). These results suggest that the CATA primer set based on the alignment of γ-Proteobacteria is also able to amplify *catA* sequences from other phyla and implies that similar *catA* genes might be retrieved from different phyla. We therefore checked the *16S rDNA* and *catA* phylogenies for congruence (sets B1 and B2). Sequences from known microorganisms were retrieved from the Genbank database and used to construct phylogenetic trees. In both cases, although two clusters were present (i.e. groups I and II) the taxonomical structure of the trees could not be retrieved which suggested that the two phylogenies were not congruent.

This suggested either (i) that the *16S rDNA* and *catA* genes evolved independently or (ii) that horizontal gene transfer had occurred [29,30]. This second hypothesis is supported by the fact that *cat* genes of the *ortho*-pathway (i.e. β-ketoadipate pathway) as well as the equivalent of the modified *ortho*-pathway (i.e. *ccd* or *clc* genes for the degradation of chlorocatechol) are often plasmid borne (i.e. marked as (Pl) in the *catA* tree) [21,31]. On the basis of this hypothesis, *A. denitrificans* could be a potential recipient of the plasmid-borne *cat* operon since *A. denitrificans* L2 branches into group II in the *16S rDNA* based tree, whereas it is located in group I in the *catA* tree.

The second step in the validation of this molecular tool consisted of testing the primer set on nucleic acids from a coal wasteland soil heavily contaminated with heavy metals and polycyclic aromatic hydrocarbons (PAHs) (Table 3). *catA* sequences were successfully amplified from this soil. The CATA primer set was also used to amplify soil DNA extracts from five agricultural soils in which no amplicon was detected (data not shown). The successful amplification of 16S rDNA from these different soils DNA extracts and amplification of *pcaH* by quantitative PCR (i.e. same DNA templates) excluded the possibility of PCR inhibition [32]. In addition, the same soil DNA extracts were submitted to inhibition test in another study showing no or slight inhibition in qPCR assays [32]. The absence of amplicons in these five soils DNA extracts suggested that the microbial population harbouring the gene was too poorly abundant to be detectable under these PCR conditions.

The detection of *catA* sequences in the heavily PAH-contaminated Neuves-Maisons soil was in accordance with the previous studies involving traditional culture-dependent methods which reported the enrichment of contaminated soils in PAH-degrading bacterial populations [18]. This was recently confirmed by qPCR assays using PCR primers that specifically target the functional PAH-degrading genes. An increase in the biodegradation potential of a PAH-degrading bacterial consortium was observed in contaminated environments [33]. Our results seem to indicate that the CatA catalysing the ring fission of catechol, a key intermediate in the β-ketoadipate pathway involved in the initial reactions of PAHs transformation, would be a relevant biomarker for monitoring PAH contamination in industrial soils.

The diversity of the *catA* community was assessed by clone library analysis and the abundance of this community by rarefaction analysis. The non-attainment of the asymptote on the rarefaction curve suggested that the full diversity of *catA* sequences amplified from the Neuves-Maisons soil DNA was not recovered and that the primer set could amplify a still broader diversity. This was corroborated by the diversity indexes. Altogether, these results indicated the ability of the CATA primer set to highlight the broad diversity of the *cat* bacterial community in the Neuves-Maisons soil. Moreover, similar diversity indices determined from the global community structure and assessed by DGGE fingerprinting, were observed in soils contaminated with PAHs and heavy metals [34].

Sequence analysis of the *catA* clone library revealed identities ranging from 64 to 100%. As the lowest shared identity observed between CatA sequences from different bacterial strains can be as low as 55%, it could be concluded that the *catA* sequences retrieved from soil corresponded to the expected gene [15]. Moreover, alignment of the amino acid sequences revealed a specific signature consisting of one tyrosyl and two histidyl residues involved in Fe^{3+} binding (data not shown, i.e. Supplementary data 2) [35]. Only one clone in the RFLP profile no. 1 (i.e. 1-11-3) was lacking in one histidyl residue. This clone could be a mutant coding for a non-functional enzyme or, possibly, a PCR mismatch or sequencing error. These results further confirmed that the *catA* sequences amplified from Neuves-Maisons soil were mainly affiliated with γ-Proteobacteria-like *catA* and some with other phyla. It has to be noticed that two clones from RFLP group 22 are segregated in the clusters one and two. This observation highlights the limit of RFLP screening which does not reveal all the polymorphisms existing in a gene sequence.

Some of the known sequences related to the amplified *catA* sequences are harboured by bacterial strains recognised to successfully degrade certain aromatic compounds. Indeed, *Burkholderia* sp. DNT is able to degrade a wide range of nitroaromatic compounds *via* nitrocatechol an analogue of catechol, the substrate of CatA [36]. Some sequences also branched with *R. eutropha* JMP 134, a Gram-negative bacterial strain able to degrade an impressive list of chloroaromatic compounds [37,38].

5. Conclusions

Despite the considerable divergence of known *catA* gene sequences, we were successful in designing a degenerate primer pair specific to *catA* sequences. We also showed that these *catA* sequences could not be related to known bacterial taxa due to the lack of congruence with 16S *rDNA* sequences. This primer set constitutes an interesting molecular tool and provides a unique insight into the biodiversity of the *catA* bacterial community in PAHs-contaminated soils even if the diversity is not fully recovered. Future work will be aimed at exploring the pertinence of this new functional biomarker in monitoring PAH contamination in soils.

Acknowledgements

The authors would like to thank Antonio Bispo (ADEME) for helpful discussions and Diana Warwick for editing this manuscript. We thank Corinne Leyval and Thierry Beguiristain (LIMOS, Nancy, France) for providing soil samples of Neuves-Maisons. We also would like to thank the Service of Sequencing and of Genotyping (SSG) of the SERCOBIO (INRA/Burgundy University) for sequencing facilities and Karen Boucherot and Céline Gouwie for technical support. The PhD work of Najoi El Azhari was funded by an ADEME/Région Bourgogne grant no. TE204-37. The work was funded by an INRA/Région Bourgogne contract (nos. B04828 and B04829).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhazmat.2009.12.074.

References

[1] C.L. Wang, S. Takenaka, S. Murakami, K. Aoki, Isolation of a benzoate-utilizing *Pseudomonas* strain from soil and production of catechol from benzoate by transpositional mutants, Microbiol. Res. 156 (2001) 151–158.

[2] S.-C. Tsai, L.-D. Tsai, Y.-K. Li, An isolated Candida albicans TL3 capable of degrading phenol at large concentration, Biosci. Biotechnol. Biochem. 69 (2005) 2358–2367.
[3] B. Bukowska, S. Kowalska, The presence and toxicity of phenol derivatives—their effect on human erythrocytes, Curr. Top. Biophys. 27 (2003) 43–51.
[4] A. Mrozik, Z. Piotrowska-Seget, S. Labuzek, Bacterial degradation and bioremediation of polycyclic aromatic hydrocarbons, Pol. J. Environ. Stud. 12 (2003) 15–25.
[5] A.J. Sweetman, M. Dalla Valle, K. Prevedouros, K.C. Jones, The role of soil organic carbon in the global cycling of persistent organic pollutants (POPs): interpreting and modelling field data, Chemosphere 60 (2005) 959–972.
[6] Y. Du, M. Zhou, L. Lei, Role of the intermediates in the degradation of phenolic compounds by Fenton-like process, J. Hazard. Mater. 3 (2006).
[7] A.W. Warhurst, C.A. Fewson, Biotransformations catalyzed by the genus Rhodococcus, Crit. Rev. Biotechnol. 15 (1994) 29–73.
[8] G.A. Kowalchuk, G.B. Hartnett, A. Benson, J.E. Houghton, K.L. Ngai, L.N. Ornston, Contrasting patterns of evolutionary divergence within the *Acinetobacter calcoaceticus* pca operon, Gene 146 (1994) 23–30.
[9] S. Copley, Evolution of a metabolic pathway for the degradation of a toxic xenobiotic: the patchwork approach, Trends Biochem. Sci. 25 (2000) 261–265.
[10] L. Fishbein, An overview of Environ. and toxicological aspects of aromatic hydrocarbons. III. Xylene, Sci. Total Environ. 43 (1985) 165–183.
[11] M.T. Garcia, A. Ventosa, E. Mellado, Catabolic versatility of aromatic compound-degrading halophilic bacteria, FEMS Microbiol. Ecol. 54 (2005) 97–109.
[12] P.D. Strachan, A.A. Freer, C.A. Fewson, Purification and characterization of catechol 1,2-dioxygenase from *Rhodococcus rhodochrous* NCIMB 13259 and cloning and sequencing of its catA gene, Biochem. J. 333 (1998) 741–747.
[13] S.I. Kim, S.-Y. Song, K.-W. Kim, E.-M. Ho, K.-H. Oh, Proteomic analysis of the benzoate degradation pathway in Acinetobacter sp. KS-1, Res. Microbiol. 154 (2003) 697–703.
[14] R.N. Patel, C.T. Hou, A. Felix, M.O. Lillard, Catechol 1,2-dioxygenase from Acinetobacter calcoaceticus: purification and properties, J. Bacteriol. 127 (1976) 536–544.
[15] C.S. Harwood, R.E. Parales, The β-ketoadipate pathway and the biology of self-identity, Ann. Rev. Microbiol. 50 (1996) 553–590.
[16] R. Yamahara, S. Ogo, Y. Watanabe, T. Funabiki, T. Jitsukawa, H. Masuda, H. Einaga, (Catecholato)iron (III) complexes with tetradnetate tripodal ligands containing substituted phenol and pyridine units as structural and functional model complexes for the catechol-bound intermediate of intradiol-cleaving catechol dioxygenases, Inorg. Chim. Acta 300–302 (2000) 587–596.
[17] M. Velusamy, R. Mayilmurugan, M. Palaniandavar, Functional models for catechol dioxygenases: Iron(III) complexes of *cis*-facially coordinating linear 3N ligands, J. Inorg. Biochem. 99 (2005) 1032–1042.
[18] C.E. Cerniglia, Biodegradation of polycyclic aromatic hydrocarbons, Biodegradation 3 (1992) 351–368.
[19] E. Matsumura, S. Oii, S. Murakami, S. Takenaka, K. Aoki, Constitutive synthesis, purification, and characterization of catechol 1,2-dioxygenase from the aniline-assimilating bacterium *Rhodococcus* sp. AN-22, J. Biosci. Bioeng. 98 (2004) 71–76.
[20] C.A. Earhart, M.W. Vetting, R. Gosu, I. Michaud-Soret, L. Que, D.H. Ohlendorf, Structure of catechol 1,2-dioxygenase from *Pseudomonas arvilla*, Biochem. Biophys. Res. Commun. 338 (2005) 198–205.
[21] H. Habe, J.-S. Chung, A. Ishida, K. Kusaga, K. Ide, T. Takemura, H. Nojiri, H. Yamane, T. Omori, The fluorene catabolic linear plasmid in *Terrabacter* sp. strain DBF63 carries the β-ketoadipate pathway genes *pcaRHGBDCFIJ*, also found in proteobacteria, Microbiology 151 (2005) 3713–3722.
[22] K. Sei, K.-I. Asano, N. Tateishi, K. Mori, M. Ike, M. Fujita, Design of PCR primers and gene probes for the general detection of bacterial populations capable of degrading aromatic compounds via catechol cleavage pathways, J. Biosci. Bioeng. 88 (1999) 542–550.
[23] A.D. Laurie, G. Jones, Quantification of phnAc and nahAc in contaminated New Zealand soils by competitive PCR, Appl. Environ. Microbiol. 66 (2000) 14–17.
[24] K. Sei, D. Inoue, K. Wada, K. Mori, M. Ike, T. Kohno, M. Fujita, Monitoring behaviour of catabolic genes and change of microbial community structures in seawater microcosms during aromatic compound degradation, Water Res. 38 (2004) 4405–4414.
[25] D. Chèneby, S. Hallet, M. Mondon, F. Martin-Laurent, J.-C. Germon, L. Philippot, Genetic characterization of the nitrate reducing community based on *narG* nucleotide sequence analysis, Microb. Ecol. 46 (2003) 113–121.
[26] V. Stolbovoy, L. Montanarella, N. Filippi, S. Selvaradjou, P. Panagos, J. Gallego, Soil Sampling Protocol to Certify the Changes of Organic Carbon Stock in Mineral Soils of European Union. In L. Office for Official Publications of the European Communities (Ed.), vol. EUR 21576 EN., 2005.
[27] F. Martin-Laurent, L. Philippot, S. Hallet, R. Chaussod, J.-C. Germon, G. Soulas, G. Catroux, DNA extraction from soils: old bias for new microbial diversity analysis methods, Appl. Environ. Microbiol. 67 (2001) 2354–2359.
[28] N. El Azhari, S. Chabaud, A. Percept, D. Bru, F. Martin-Laurent, *pcaH*, a molecular marker for estimating the diversity of the protocatechuate-degrading bacterial community in the soil environment, Pest Manag. Sci. 63 (2007) 459–467.
[29] A. Buchan, E.L. Neidle, M.A. Moran, Diversity of the ring-cleaving dioxygenase gene *pcaH* in a salt marsh bacterial community, Appl. Environ. Microbiol. 67 (2001) 5801–5809.
[30] D.H. Clayton, S.E. Bush, K.P. Johnson, Ecology of congruence: past meets present, Syst. Biol. 53 (2004) 165–173.
[31] B. Frantz, K.-L. Ngai, D.K. Chatterjee, L.N. Ornston, A.M. Chakrabarty, Nucleotide sequence and expression of *clcD*, a plasmid-borne dienelactone hydrolase gene from *Pseudomonas* sp. strain B13, J. Bacteriol. 169 (1987) 704–709.
[32] N. El Azhari, D. Bru, A. Sarr, F. Martin-Laurent, Estimation of the density of the protocatechuate-degrading bacterial community in soil by real-time PCR, Eur. J. Soil Sci. 59 (2008) 665–673.
[33] A. Cébron, M.-P. Norini, T. Beguiristain, C. Leyval, Real-time PCR quantification of PAH-ring hydroxylating dioxygenase (PAH-RHDα) genes from Gram positive and Gram negative bacteria in soil and sediment samples, J. Microbiol. Meth. 73 (2008) 148–159.
[34] A. Vivas, B. Moreno, C. del Val, C. Macci, G. Masciandaroc, E. Benitez, Metabolic and bacterial diversity in soils historically contaminated by heavy metals and hydrocarbons, J. Environ. Monit. 10 (2008) 1241–1372.
[35] M.W. Vetting, D.H. Ohlendorf, The 1.8 Å crystal structure of catechol 1,2-dioxygenase reveals a novel hydrophobic helical zipper as a subunit linker, Structure 8 (2000) 429–440.
[36] T. Leungsakul, G.R. Johnson, T.K. Wood, Protein engineering of the 4-methyl-5-nitrocatechol monooxygenase from *Burkholderia* sp. strain DNT for enhanced degradation of nitroaromatics, Appl. Environ. Microbiol. 72 (2006) 3933–3939.
[37] J.H. Leveau, A.J.B. Zehnder, J.R. van der Meer, The *tfdKtfdK* gene product facilitates uptake of 2,4-dichlorophenoxyacetate by *Ralstonia eutropha* JMP134(pJP4), J. Bacteriol. 108 (1998) 2237–2243.
[38] I. Plumeier, D. Pérez-Pantoa, S. Heim, B. Gonzàlez, D.H. Pieper, Importance of different tfd genes for degradation of chloroaromatics by *Ralstonia eutropha* JMP134, J. Bacteriol. 184 (2002) 4054–4064.
[39] S.G. Iwagami, K. Yang, J. Davies, Characterization of the protocatechuic acid gene cluster from *Streptomyces* sp. strain 2065, Appl. Environ. Microbiol. 66 (2000) 1499–1508.
[40] P. Fortnagel, H. Harms, R.-M. Wittich, S. Krohn, H. Meyer, V. Sinnwell, H. Wilkes, W. Francke, Metabolism of dibenzofuran by *Pseudomonas* sp. strain HH69 and the mixed culture HH27, Appl. Environ. Microbiol. 56 (1990) 1148–1156.
[41] L. Cavalca, A. Hartmann, N. Rouard, G. Soulas, Diversity of *tfdC* genes: distribution and polymorphism among 2,4-dichlorophenoxyactic acid degrading soil bacteria, FEMS Microbiol. Ecol. 29 (1999) 45–58.
[42] P.A. Williams, K. Murray, Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (*arvilla*) mt-2: evidence for the existence of a TOL plasmid, J. Bacteriol. 120 (1974) 416–423.