This is a postprint of an article published in

Junca, H., Pieper, D.H.

**Functional gene diversity analysis in BTEX contaminated soils by means of
PCR-SSCP DNA fingerprinting: Comparative diversity assessment against
bacterial isolates and PCR-DNA clone libraries
(2004) Environmental Microbiology, 6 (2), pp. 95-110.**

Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries

Howard Junca and Dietmar H. Pieper*

Department of Environmental Microbiology, GBF-German Research Centre for Biotechnology, Mascheroder Weg 1, D-38124 Braunschweig, Germany.

*For correspondence. E-mail: dpi@gbf.de, Tel. (+49) 531 6181 467, Fax (+49) 531 6181 411

Keywords

PCR-SSCP molecular fingerprint, BTEX contamination, catechol 2,3-dioxygenase, gene diversity, catabolic potential, biodegradation, natural microbial communities.

Summary

Developments in molecular biology based techniques have led to rapid and reliable tools to characterize microbial community structures and to monitor their dynamics under in-situ conditions. However, there has been a distinct lack of emphasis on monitoring the functional diversity in the environment. Genes encoding catechol 2,3-dioxygenases (C23O), as key enzymes of various aerobic aromatic degradation pathways, were used as functional targets to assess the catabolic gene diversity in differentially BTEX contaminated environments by polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP). Site specific PCR-SSCP fingerprints were obtained, showing that gene diversity experienced shifts correlated to temporal changes and levels of contamination. PCR-SSCP enabled the recovery of predominant gene polymorphs, and results closely matched with the information retrieved from random sequencing of PCR-DNA clone libraries. A new method for isolating strains capable of growing on BTEX compounds was developed to diminish pre-selection or enrichment bias and to assess the function of predominant gene polymorphs. C23O abundance in isolates correlated with the levels of BTEX pollution in the soil samples analysed. Isolates harbouring C23O genes, identical to the gene polymorph predominant in all contaminated sites analysed, showed an unexpected benzene but not toluene mineralising phenotype whereas isolates harbouring a C23O gene variant differing by a single point mutation and observed in highly polluted sites only, were capable, among some other isolates, to mineralise benzene and toluene, indicating a catabolically determined sharing of carbon sources on-site. The PCR-SSCP technique is thus a powerful tool for assessing the diversity of functional genes and the identification of predominant gene polymorphs in environmental samples as a prerequisite to understand the functioning of microbial communities.

1 Introduction

2
3 There is a growing interest in understanding how microbial communities react to challenging
4 conditions, e.g., how they adapt to increasing levels of pollution. Genetic techniques such as
5 Polymerase Chain Reaction (PCR) amplification using universal primers targeting 16S rRNA genes,
6 allow the cultivation-independent examination of microbial community structures and their
7 responses under stressful conditions. Many previous studies have focused on the taxonomic
8 description of natural microbial communities, their changes and dominant key players (Torsvik and
9 Ovreas, 2002). However, various functional genes such as catabolic genes are often localized on
10 mobile genetic elements susceptible to high mutation rates and horizontal transfer (Lawrence, 2002;
11 Williams *et al.*, 2002). Therefore, a certain level of metabolic potential or activity is not necessarily
12 reflected by the taxonomic affiliation of organisms or by the microbial community structure. PCR-
13 based methods have been successfully applied for detecting catabolic genes in the environment and
14 quantitative PCR methods been demonstrated to be powerful tools for monitoring bioremediation
15 processes (Stapleton and Sayler, 1998; Widada *et al.*, 2002). Such methods, however, have focused
16 mainly on the characterisation of the presence, or abundance of a family of catabolic genes (Sotsky
17 *et al.*, 1994; Meyer *et al.*, 1999; Mesarch *et al.*, 2000; Beller *et al.*, 2002; Marlowe *et al.*, 2002; Baldwin
18 *et al.*, 2003) without assessing the finer levels of sequence diversity. The diversity of catabolic gene
19 sequences often reflects differences in substrate specificity or affinity (Cerdan *et al.*, 1994; Beil *et al.*,
20 1998; Parales *et al.*, 2000; Suenaga *et al.*, 2002), and single amino acid differences have been reported
21 to drastically change substrate specificity of toluene dioxygenase (Beil *et al.*, 1998), regioselectivity of
22 naphthalene dioxygenase (Parales *et al.*, 2000) and toluene 2-monooxygenase (Pikus *et al.*, 1997), or
23 activity of muconate cycloisomerase (Kaulmann *et al.*, 2001). A more detailed picture of the
24 catabolic gene structure and sequence diversity in environmental samples will, thus, significantly
25 increase our knowledge of the functional potential of microbial communities. Moreover, shifts in
26 catabolic gene structure will allow the deduction of the evolutionary fitness of catabolic genes,
27 operons and their respective hosts. A variety of molecular fingerprinting approaches previously
28 developed to assess community structure, via the analysis of 16S rDNA or rRNA diversity, may be
29 applied to define functional gene structure (Widada *et al.*, 2002).

30 Benzene, toluene, ethylbenzene and xylenes, commonly referred to as BTEX, are important
31 monoaromatic, environmental contaminants world-wide. Their aerobic degradation can proceed by
32 various different pathways, involving either an activation of the aromatic ring by dioxygenation or
33 monooxygenation reactions, or through a processing of the side chain(s) (Williams and Sayers,
34 1994). Whereas a relatively broad diversity of activation mechanisms is possible, these pathways
35 usually converge in the formation of (substituted) catechols as central intermediates (Reineke, 1998)
36 which, in the majority of already described pathways, are subject to extradiol cleavage. Extradiol
37 dioxygenases are, thus, key enzymes in the degradation of aromatic compounds and many of such
38 proteins and their coding sequences have been described, purified and characterised in the last

decades. Examination of their evolutionary relationships (Eltis and Bolin, 1996) showed that the majority of extradiol dioxygenases with preferences for monocyclic substrates were, at the protein level, phylogenetically closely related, constituting the so-called family of I.2 extradiol dioxygenases. Subfamily I.2.A seems to be of particular importance for the degradation of monocyclic aromatic compounds (Eltis and Bolin, 1996). Members of this enzyme subfamily were reported to differ significantly in their substrate specificity (Kitayama *et al.*, 1996) and, thus, the metabolic network in environmental communities will be significantly interconnected with the C23O diversity. Due to their central role in aromatic catabolism, the catechol 2,3-dioxygenases (C23O) subfamily I.2.A genes have been analysed in studies of diverse environments and their presence has been observed in various contaminated sites (Chatfield and Williams, 1986; Wikstrom *et al.*, 1996; Daly *et al.*, 1997; Okuta *et al.*, 1998; Meyer *et al.*, 1999; Cavalca *et al.*, 2000; Mesarch *et al.*, 2000) however, no fingerprinting methods to generate reliable and characteristic profiles from different environmental samples and to effectively extract predominant or specific gene sequences have been reported. In this study we introduce the application of a modified polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) analysis with the single-strand removal method previously adapted for molecular fingerprinting of microbial community structures (Schwieger and Tebbe, 1998), for determining the characteristic sequence types of C23O genes as an indicator of catabolic potential. These PCR-SSCP results are assessed by comparing sequence data from PCR-DNA clone libraries and C23O sequences and metabolic performance of microorganisms exhibiting C23O activity. PCR-SSCP is demonstrated to be a reliable and rational tool to rapidly determine sequence diversity within a catabolic gene family in environmental samples.

Results and Discussion

Optimisation of PCR-SSCP for generation of C23O fingerprints

In order to analyse the sequence diversity of C23O genes in soils differing in contamination level, the PCR-SSCP technique was optimised to generate fingerprints of C23O genes from total DNA extracts. Even though (Fig. 1A) it was possible to resolve, e.g., the different 238 nucleotide C23O gene single-strand conformations of *P. putida* mt-2 and *P. putida* HS1 fragments, which share a DNA sequence similarity of 95%, PCR-SSCP resolution of single-strands from 13 reference strain PCR-amplifications was low, and higher concentrations (>300 ng) of single-strands resulted in multiple conformations of the same sequence (Fig. 1A and supplementary online material 3), a phenomenon previously reported for single-strand DNA molecules (Atha *et al.*, 2001; Schmalenberger *et al.*, 2001). When the PCR primer set to amplify this 238-bp fragment was applied to soil DNA, amplification products were obtained from only 3 out of 5 contaminated soil samples (A1Y, A3Y and 1Y). ssDNA of the PCR amplicons obtained from these soil samples did not

1 generate clear banding patterns, indicating the presence of overlapping polymorphic sequences with
2 similar electrophoretic mobilities, as observed in ssDNA artificial polymorphic mixtures of this
3 fragment (supplementary online material 3).

4 An improved resolution was achieved using ssDNA obtained from 583-bp C23O fragments (Fig.
5 1B), allowing a discrimination of the main polymorphic ssDNA mobility of each different reference
6 ssDNA. However, multiple conformations of the same sequence types were still observed, and trials
7 to amplify the 583-bp from soil samples were successful only in the same 3 contaminated samples as
8 it was observed for the PCR-amplification of the 238-bp fragment (1Y, A1Y and 3Y). In contrast,
9 PCR amplifications of the 527-bp C23O fragments generated positive signals from all 5
10 contaminated samples, and even a slight positive signal from the non-contaminated soil. Thus, the
11 PCR primers developed in the course of this study were more effective than the primers previously
12 used to amplify C23O fragments from the environmental samples tested.

13 PCR-SSCP analysis of the 527-bp C23O fragments showed clearly different mobilities of highly
14 similar C23O fragments and one predominant single-strand conformation per reference strain (Fig.
15 1C). Fragment mixtures from environmental samples (Fig. 1D) produced strands of clearly
16 differential mobilities, demonstrating the potential of the PCR-SSCP technique to retrieve
17 polymorphic patterns of functional genes from the environment.

18 19 *Diversity of C23O fingerprints and genetic information recovered from PCR-SSCP gels* 20

21 To identify the predominant products by DNA sequencing, different C23O gene single-strands
22 (“bands”), designated according to their electrophoretic mobilities as A – M (Fig. 1D) and
23 representing at least 85% of the total signal intensity of each profile, were extracted from the gels
24 and reamplified by PCR. All predominant bands observed in the profiles were composed of
25 unambiguous single-strand species. This contrasts recent literature on profiling the 16S rDNA
26 diversity by temperature gradient gel electrophoresis (TGGE), denaturing gradient gel
27 electrophoresis (DGGE), or single-strand conformation polymorphism (SSCP), which showed that
28 single bands may contain more than one sequence (Bruns *et al.*, 2001; Kisand and Wikner, 2003;
29 Schmalenberger and Tebbe, 2003), a fact of specific importance if samples of high diversity are
30 profiled. The profiling of a defined catabolic gene family with limited diversity, obviously lower the
31 possibility of different single-strands having identical mobility. However, sequencing of low
32 intensity bands showed some ambiguous bases in the sequence. Such ambiguous positions could be
33 due to overlapping ssDNA with few sequence differences and present at lower concentrations, as
34 reported in PCR-SSCP 16S rDNA studies (Schmalenberger and Tebbe, 2003). For this reason, only
35 an unambiguous 301-bp block of the sequences recovered from the ssDNA bands, corresponding to
36 nucleotides 314 to 614 of *P. putida* mt-2 C23O gene (protein positions 106 to 204) was used for
37 phylogenetic analyses (Fig. 1E-F). The majority of bands with high relative intensities and identical
38 mobility, like band G, present in all profiles, contained identical single-strands. In a few cases (e.g.

bands G, M, H, I and L in profile 1Y), identical 301-bp sequences were recovered from different bands. Most probably, these different single-strand mobilities are due to sequence differences actually found in regions outside of the 301 base block analysed. This hypothesis is supported by cloned PCR-amplified C23O gene sequences wherein single base differences between closely related sequences were observed in these regions (see below). The treatment of bands of identical sequence and mobility in different profiles as the same unique character, of identical sequences but different mobilities as different characters, and the treatment of different sequences exhibiting identical mobility as unique and independent characters, allowed the calculation of a dendrogram showing the similarities between the PCR-SSCP profiles (Fig. 1E). High similarities were observed between PCR-SSCP profiles of samples from identical sites and horizons sampled at different time points, whereas low PCR-SSCP profile similarities were evident from samples differing in contamination level or taken from different soil horizons. These observations indicated that the catabolic gene composition might depend on the environmental conditions and is probably driven by the contamination level. The sequence of band G, predominant in fingerprints of all contaminated samples is closely related to the *P. stutzeri* AN10 C23O gene sequence (Fig. 1F). Sequences closely related to the *Alcaligenes* sp. KF711 C23O gene were observed frequently in DNA extracted from capillary fringe soil samples. The sequence of a very intense and distinct ssDNA band, specific for the highly contaminated site A1Y, clustered within the isolated branch of the C23O gene of *P. putida* G7. Additionally, differences between highly-contaminated and low-contaminated sites were detectable. A sequence type identical to that retrieved from band E (97% similarity to *P. aeruginosa* JI104-1 C23O putative protein) was observed in highly contaminated samples only, whereas band A (closest relative *P. aeruginosa* ZD 4-3 C23O gene) was observed only in slightly contaminated samples. Thus, PCR-SSCP targeting the C23O gene was used successfully to identify subsets of C23O genes which are predominant throughout the whole contaminated area analysed, as well as subsets of genes which are predominant only at specific sites differing in contamination level or soil horizon.

Diversity of C23O sequences recovered from PCR clone libraries

The same 527-bp C23O gene PCR amplification products, from 1Y and 3Y total soil DNA, used to produce PCR-SSCP profiles were also used to generate clone libraries for determination of the diversity of C23O gene sequence types, a commonly used technique to assess diversity in a mixed PCR product (Yeates *et al.*, 2000; Hamelin *et al.*, 2002; Norton *et al.*, 2002). From the inserts, an unambiguous sequence length of 515-bp (corresponding to positions 193 to 707 or amino acids 65 to 235 of the *P. putida* mt-2 C23O gene) could be determined for each selected clone. The results (Fig. 2A) were in close agreement with the information recovered from PCR-SSCP. A major clustering of sequence types closely related with *P. stutzeri* AN10 C23O – *P. aeruginosa* JI104 C23O was observed (Fig. 2A). The 18 distinct DNA sequences converge to only 5 putative protein

sequence types indicating that the amino acid sequence integrity of the enzyme seems to be very important.

Isolation and characterisation of strains exhibiting C23O activity

The fact that C23O genes related to *Pseudomonas stutzeri* AN10 were observed to be dominant in all contaminated sites analysed raised the question on their importance for BTEX degradation. We thus aimed to isolate from the samples a broad collection of strains harbouring C23O activity. Analysis of reference strains revealed, that low levels of constitutive expression of C23O on R2A agar plates was sufficient to observe transformation of catechol into the yellow coloured 2-hydroxymuconic semialdehyde. Dilutions of soil samples were spread on R2A agar plates and all colonies were analysed for C23O activity by spraying with catechol. An extensive collection of strains exhibiting C23O activity could be isolated. Whereas there was no significant difference in the number of colony forming units from the differently contaminated soils tested (approximately $5 \pm 3 \times 10^6$ CFU/g of soil), the percentages of colonies exhibiting *meta*-cleavage activity varied significantly, ranging from <3%, 10% and >70% of total CFU/g growing on R2A plates obtained from 0Y, 3Y and 1Y soil samples, respectively, correlating with the contamination levels. The environmental conditions and, probably, the contamination level were selective for strains possessing C23O activity. As total CFUs did not differ drastically it can be suggested that nutrients other than available carbon are limiting in the environment analysed. However, as the fraction of bacteria cultured here comprises only a subset of the total microflora, such assumptions remain speculative.

A high proportion (37 of 46 isolates) of the strains expressing *meta*-cleavage activity could use at least one of the BTEX compounds as a sole carbon source (Table 1), showing a high correlation between those characteristics. The DNAs of the isolates were used as templates in order to amplify C23O subfamily I.2.A gene fragments. Primer sets amplifying 238-bp, 527-bp and 583-bp fragments all showed successful amplification in the majority (26 out of 37) of isolates capable to grow on BTEX compounds. Some of the isolates exhibiting *meta*-cleavage activity and the ability to grow on BTEX compounds (11 isolates) were not amplifiable with C23O specific primers. As, besides members of C23O subfamily I.2.A, enzymes of subfamily I.2.C and I.3.B also have been shown to be involved in the degradation of BTEX (Eltis and Bolin, 1996) it is highly probable, that this minor fraction of isolates contain genes of these subfamilies. A high proportion of 14 isolates, 13 of them carrying a C23O subfamily I.2.A gene, was capable, under the conditions tested, to grow only on benzene as a sole source of carbon, and microorganisms with such a phenotype have rarely been reported.

A selection of 30 isolates, preferentially those showing catabolic capabilities against BTEX or an amplifiable C23O fragment, were analysed by ARDRA, and three distinct *AluI* ARDRA profiles were observed (Fig. 2B and Table 1). 16S rDNA sequencing of 13 representative strains revealed highest

similarities to ribosomal sequences of *Pseudomonas* spp.: 6 strains exhibiting ARDRA profile PS1, 99.0 % to *P. fluorescens* DSM 50108, 6 strains exhibiting ARDRA profile PS2, 98.4 % to *P. brassicacearum* str. NFM421, and 1 strain exhibiting ARDRA profile PS3, 98.2 % to *P. corrugata* ATCC 17484.

Diversity of C23O genes from isolates

A primer pair annealing at the start and the stop codons of C23O I.2.A genes was designed and it showed amplification efficiency similar to that of primer pairs amplifying gene fragments. Use of this primer pair allowed not only to determine the whole gene sequence but a very convenient direct cloning and overexpression of C23O genes from isolates. The estimated phylogeny of the proteins deduced from PCR-amplified C23O gene sequences (Fig. 2C) from 19 isolates, comprising different morphotypes and BTEX degradation profiles, showed five distinct groups of sequence types. The predominant C23O DNA sequence type observed, identical in 11 different isolates (Fig. 2C), designated A1, differed from the *P. stutzeri* AN10 sequence (Bosch *et al.*, 2000) by only a single nucleotide, which represents one amino acid difference in position 218 (*Tyr* instead of *His*) and from the *P. aeruginosa* JI104-1 C23O sequence by six nucleotides, corresponding to only two amino acid differences in the putative protein sequence (*Ala* instead of *Arg* at position 60, and *Tyr* instead of *His* at position 218). This sequence type was identical to sequence recovered from the most intense bands of the SSCP profiles (bands G, Fig. 1F), and to the most abundant sequence retrieved from the C23O PCR clone libraries (Fig. 2A, group 1). A sequence identical to *P. stutzeri* AN10 C23O, designated type A2 (Fig. 2C) was observed in two isolates from site 1Y only as well as in two clones from the clone library of sample 1Y (Fig. 2A, group 5). The sequence difference between a type A1 and a type A2 sequence is not included in the 301 base block used for phylogenetic analysis of PCR-SSCP bands. A detailed sequence analysis upstream of this block, however, gave no indications for the massive presence of an A2-type sequence in any of the analysed bands. A second C23O sequence type, retrieved from 4 isolates, was related to the isolated branch of the *P. putida* G7 C23O gene in the I.2.A subfamily. This sequence type could also be detected by PCR-SSCP or in the clone library (Fig. 1F and 2A) Two isolates contained C23O sequences, closely related to the *P. putida* mt53 (only found by this method) or the *P. putida* H C23O gene, respectively (also found in 3 of 27 clones from 3Y PCR library). The general phylogenetic distribution of the sequences from the isolates showed a clear predominant sequence polymorphism in organisms from soil samples 1Y and 3Y, as well as the coexistence of C23O genes of more distant evolutionary origins in isolates recovered from the same environment.

10 of 11 isolates that harboured a C23O type A1 use only benzene as a sole carbon source,-whereas both isolates harbouring a C23O type A2, were capable to grow on benzene, toluene and ethylbenzene. Isolate A1YC1 (C23O type D) can use only toluene, *m*-xylene and *p*-xylene as a sole

carbon source. Its C23O sequence is closely related to xylE genes of TOL plasmids pWW0 and pWW53, and the catabolic phenotype of the isolate hints on the degradation of toluene and xylenes via the respective benzoates and thus a localization of the gene in an operon similar to the classical TOL pathway. Whereas there is a strong correlation between the degradative phenotype and the C23O type observed, there is no such correlation between the exact phylogenetic position of the isolates and the C23O gene variant they contain. As an example, strains 1YdBTEX2 and 1YB3 had exactly the same 16S rDNA sequence, but differed in growth phenotype and C23O gene comprised. On the other hand, 1YdBTEX2 and 3YdBTEX1 harboured the same C23O gene and shared the same growth phenotype, but differed in their 16S rDNA sequences and probably belong to different *Pseudomonas* species.

As there is a strong relationship between type of C23O present and growth phenotype, the question should be answered if, and in which way this gene and gene product or the operon it is located in, influences the growth phenotype. It could be reasoned that observation of C23O is just fortuitous and that growth on benzene only is due to an induction of a catechol *ortho*-cleavage route, a pathway reported not to be suited for the degradation of methylaromatics including toluene (Knackmuss *et al.*, 1976). However, control experiments using five isolates with different growth phenotype and C23O genes (1YC2 and 1YB2, type A1; 1Yxyl1, 1YdBTEX2, type A2; 1XB2, type B) showed that all those isolates exhibited high C23O activity (1000 – 4000 U/g protein), but negligible catechol 1,2-dioxygenase activity (< 50 U/g protein) during growth on benzene and thus use a *meta*-cleavage pathway. Two isolates, 1YB2 and 1YdBTEX2, representing type A1 and type A2 hosts, were selected for further analysis. Both strains showed a high constitutive expression of catechol 2,3-dioxygenase of 650 and 2900 U/g protein, respectively, only slightly lower than the activity during growth of benzene (Table 2). The C23O genes of both strains were overexpressed in *E.coli* and the gene products compared to those expressed during growth on glucose or benzene. Despite the difference in only one amino-acid residue, the substrate specificities of the overexpressed type A1 and type A2 enzymes differed significantly (Table 2). Whereas the type A2 His218 variant of 1YdBTEX2 showed, among the substrates tested, the highest activity with catechol, the type A1 Tyr218 variant of 1YB2 exhibited highest activity with 4-methylcatechol and a relatively elevated activity with 3-methylcatechol. An influence of this amino acid on substrate specificity was not expected from the reported crystal structure of the C23O of *P. putida* mt-2 (Kita *et al.*, 1999). Partially purified enzymes from glucose grown cells of either 1YB2 or 1YdBTEX2 exhibited substrate specificities identical to the respective overexpressed proteins and showed identical retention behaviour during anion exchange chromatography (elution at 0.34 M NaCl). Similarly, only one C23O activity was observed when analysing proteins of benzene grown cells of either 1YB2 or 1YdBTEX2 by anion exchange chromatography. To further evidence the involvement of the respective type A1 or type A2 C23O in degradation of benzene, aliquots of highly active fractions obtained by anion exchange chromatography were subjected to SDS-PAGE, blotted onto a polyvinylidene difluoride membrane and major protein bands of a molecular mass of 35 ± 2 Kda

1 analysed by N-terminal sequencing. The amino-terminal sequences of the proteins from 1YB2,
2 1YdBTEX2, JM109(pA1Tyr) and JM109(pA2His) were identical and identified as
3 MNKGIMRPGHVQL; this sequence is identical to the sequence of the C23O proteins of *P. stutzeri*
4 AN10 and *P. aeruginosa* JI104. Inner sequences WGVNEVNPEAWPR and
5 FQAPSGHHFELYADKEYTGK obtained from both the 1YB2 and 1YdBTEX2 derived proteins were
6 identical to those deduced from the type A1 and type A2 C23O gene sequences. Thus, the C23O
7 activity constitutively expressed in glucose-grown cells is due to a type A1 C23O in 1YB2 and to a
8 type A2 C23O in 1YdBTEX2.

9 However, the observed differences in substrate specificity do not explain the difference in growth
10 behaviour, as the type A1 C23O of benzene degraders is highly active with methylsubstituted
11 catechols. It was therefore analysed if strain 1YB2 contains an active 2-hydroxymuconic
12 semialdehyde hydrolase (xylF). XylF is necessary for degradation of toluene via 3-methylcatechol
13 and 2-hydroxy-6-oxo-hepta-2,4-dieneoate, but dispensable for growth on benzene via catechol and
14 2-hydroxymuconic semialdehyde, as 2-hydroxymuconic semialdehyde can be degraded via the
15 alternative branch of the meta-cleavage pathway involving a 2-hydroxymuconic semialdehyde
16 dehydrogenase, xylG (Sala-Trepat *et al.*, 1972). Intriguingly 1YB2 as well as 1YdBTEX2 showed
17 constitutive expression of high levels of a 2-hydroxymuconic semialdehyde hydrolase of 170 ± 20
18 U/g protein against 2-hydroxy-6-oxo-hepta-2,4-dieneoate and of 70 ± 10 U/g protein against 2-
19 hydroxymuconic semialdehyde and of a 2-hydroxymuconic semialdehyde dehydrogenase (20 ± 5
20 U/g and 65 ± 10 U/g protein in cell extracts of 1YB2 or 1YdBTEX2, respectively). It is thus evident,
21 that the difference in growth phenotype of 1YB2 and 1YdBTEX2 is not due to substrate specificity
22 differences of C23O enzymes, nor to the absence of a functional 2-hydroxymuconic semialdehyde
23 hydrolase.

24 Nevertheless, as the presence of a A1 type C23O gene is directly correlated with a benzene, and the
25 presence of a A2 type C23O gene with a benzene/toluene degrading phenotype, it can be assumed
26 that these C23O genes are linked, possibly in an operon, with other catabolic pathway genes, which
27 either, by their substrate specificity or by controlling metabolic flux through the pathway, determine
28 the range of substrates that can be mineralized. Specifically, the accumulation of catechols formed
29 during metabolism is known to prevent growth of single organisms on aromatic compounds
30 (Perez-Pantoja *et al.*, 2003).

31 32 33 Conclusions

34
35 The alignment of a common putative 99 amino acid stretch derived from C23O gene sequence
36 information obtained from bacterial isolates, PCR clone libraries and extracted PCR-SSCP bands,
37 allowed a comparison of the three different methods applied (Fig. 3), specifically for soil samples 1Y
38 and 3Y. All the methods showed a predominance of sequences related to the *P. stutzeri* AN10 C23O

1 gene. The diversity of C23O gene sequences from isolates was, as expected, lower. As an example,
2 the second most commonly observed C23O gene sequence observed in the clone library and the
3 PCR-SSCP profile, related to the *P. putida* CF600 C23O gene, was not observed in any of the
4 isolates. As 20 and 27 cloned PCR-amplified C23O genes were sequenced for sites 3Y and 1Y
5 respectively, the sequence types detected at higher numbers can be assumed to be more prevalent in
6 the samples. Those sequences were also predominant in the PCR-SSCP profiles. Thus, PCR-SSCP
7 profiling has provided a reliable overview of the relative abundance and diversity of C23O genes
8 within different environmental conditions. We have shown a successful example of adaptation of
9 PCR-SSCP to generate fingerprints of functional genes, which makes very likely further
10 developments, optimisations and applications of this technique to target a number of gene families
11 playing functional roles in the environments. Together with promising methods to characterise
12 functional gene expression in environmental samples (Wawer *et al.*, 1997; Miskin *et al.*, 1999;
13 Koizumi *et al.*, 2002; Manefield *et al.*, 2002; Marlowe *et al.*, 2002), this technique would help future
14 trends to characterize and compare catabolic gene diversity and structure, as well as active operons
15 and pathways under changing environmental conditions.

16 Even though the expression of those genes upon BTEX stress has not been proven directly in the
17 environmental samples, and could be constitutive as in the isolates, it appears evident that specific
18 C23O genes and operons have been positively selected during the adaptation of soils to BTEX
19 contamination. Moreover, from studies on isolates we could show that the presence of a C23O type
20 A1 gene seems to be related to a benzene degrading phenotype. Those genes were dominant in all
21 samples analysed and can be related to the fact that benzene was the major pollutant in all those
22 samples. The isolates represent a new, not previously reported growth phenotype of obvious
23 selective advantage under environmental conditions. Whereas organisms harbouring a C23O type
24 A1 containing pathway are highly specialized, organisms harbouring C23O type A2 genes, from
25 studies of isolates, can be related to a benzene/toluene degrading phenotype, and were detected only
26 in highly contaminated samples, which had been stressed with a significant amount of toluene in
27 addition to benzene. If such kind of pathway actually evolved from a pathway comprising a C23O
28 type A1 remains to be elucidated. Nevertheless, the application of a multifaceted approach by a new
29 PCR-SSCP of functional genes, its validation by clone libraries and a non-biased isolation of BTEX
30 degraders and their biochemical characterization allowed a better understanding on the complexity
31 of BTEX degradation in the environment and gives a metabolic explanation to the fact that complex
32 communities are involved in mineralisation. Currently microcosm experiments and on site studies
33 are performed to analyse shifts in catabolic gene composition in more detail.

Experimental Procedures

Microorganisms, samples, isolation and culture conditions

Reference strains used in this study were *P. stutzeri* AN10 (Bosch *et al.*, 2000), *P. putida* CF600 (Bartilson and Shingler, 1989), *P. putida* G7 (Ghosal *et al.*, 1987), *P. putida* H (Herrmann *et al.*, 1995), *Pseudomonas sp.* IC (Carrington *et al.*, 1994), *P. aeruginosa* JI104 (Kitayama *et al.*, 1996), *Sphingomonas sp.* KF711 (Moon *et al.*, 1996), *P. putida* mt-2 (Nakai *et al.*, 1983), *P. putida* mt53 (Keil *et al.*, 1985), *P. stutzeri* OM1 (Ouchiyama *et al.*, 1998), *P. stutzeri* OX1 (Arengi *et al.*, 2001), *P. putida* HS1 (Benjamin *et al.*, 1991) and *P. putida* 3,5X (Hopper and Kemp, 1980). Bacterial strains were cultured on R2A agar at 30°C for 3 days, and stored at -70°C in glycerol stocks (Sambrook *et al.*, 1989).

Soil samples were collected from the unsaturated (X) and the capillary fringe (Y) horizons from an aquifer located in the Czech Republic, which carries an underground BTEX plume. The sampling sites were designated 0 (non polluted soil), 3 (slightly BTEX contaminated soil), and 1 (highly BTEX contaminated soil). In two successive sampling campaigns 2001 - 2002, the BTEX concentrations in the groundwater were determined (Aquatest a.s. Prague) ranging from 320, 97, 6 and 13 mg/l of benzene, toluene, ethylbenzene and xylenes respectively (site 1), to 55, 2, 0.007 and 0.012 mg/l of benzene, toluene, ethylbenzene and xylenes respectively (site 3). Samples were designated according to sampling site (0, 1 or 3) and soil horizon (X or Y). To differentiate between samples times, samples collected in 2001 were additionally designated by an A.

Microorganisms with C23O activity were isolated by plating 0.1 ml of appropriate dilutions of soil samples onto R2A agar plates incubated in the absence or presence of aromatic compounds (benzene (B), *m*-xylene (Xyl), an equimolar mixture of BTEX (BTEX) or a 1:100 dilution of the equimolar mixture of BTEX (dBTEX) dissolved in heptamethylnonane). In each case, 20 µl of substrate were supplemented via the vapour phase by placing them into eppendorf pipette tips inside the petri dishes. After 3 days of incubation at 30°C, plates were screened for the presence of colonies exhibiting C23O activity by spraying with an aqueous solution of catechol (100 mM). A subset of colonies with different colony morphotypes and exhibiting yellow coloration upon spraying were selected and purified. The utilization by the isolates of single BTEX compounds as a sole carbon and energy source was determined by growth on mineral medium plates (Dorn *et al.*, 1974) solidified with SeaKem LE Agarose (FMC Bioproducts) and supplemented via the vapour phase with 20 µl of benzene, toluene, ethylbenzene or *o*-, *m*-, or *p*-xylene. To discard false positives, the growth results were assessed by comparison with the growth characteristics of the isolates on mineral medium agarose plates.

Primer design for amplification of C23O subfamily I.2.A gene fragments

In the literature, different primer sets, have been reported. Out of previously described primers pairs specific to C23O subfamily I.2.A genes (Wikstrom *et al.*, 1996; Okuta *et al.*, 1998; Meyer *et al.*, 1999; Mesarch *et al.*, 2000), a set targeting a 238-bp fragment, reported to be designed based on a multiple DNA sequence alignment of 8 reference was tested (Mesarch *et al.*, 2000). This primer pair amplified the expected PCR product size from all 13 reference strains, including *P. putida* G7 C23O (data not shown). Multiple DNA and protein alignments (supplementary online material 1) of 21 subfamily I.2.A C23O sequences were used to localize further conserved regions for primer design. Three new primer sets were found suitable to amplify the expected PCR product sizes of 527-bp, 583-bp and 934-bp from DNA of all 13 reference strains (example of 527-bp amplifications in supplementary online material 2). When DNA of isolates was used as template, similar positive PCR signals were obtained with each primer set (data not shown). Thus, the primer set producing the longest fragment of 934-bp was selected to amplify, sequence, and obtain recombinant expression of genes from the isolates.

DNA extraction and PCR

DNA from soil samples was extracted and purified with the Fast Prep Soil DNA Extraction Kit (BIO101). To extract DNA from bacterial strains, single colonies were boiled and centrifuged (Kanakaraj *et al.*, 1998), 4 µl of the supernatants were used as a PCR template in a final volume of 50 µl of a mixture containing 1X PCR Buffer (Promega) supplemented with 1.5 mM MgCl₂, 200 µM dNTP's, 0.5 µM of each primer (synthesized by Invitrogen) and 0.3 U/µl Taq DNA Polymerase (Promega).

To amplify and sequence catechol 2,3-dioxygenase gene fragments the following primers were used: 23CAT-F 5' CGA CCT GAT CTC CAT GAC CGA 3' and DEG-R 5' TYA GGT CAK MAC GGT CA 3' to amplify a 238-bp C23O gene fragment (Mesarch *et al.* 2000) comprising positions 687 to 924 of *P. putida* mt-2 xylE CDS, forward 5' ATG GAT TTY ATG GSB TTC AAG G 3' and reverse 5' TCG ATV GAK GTR TCG GTC ATG 3' to amplify a 527-bp C23O gene fragment corresponding to nucleotide positions 193 to 719 of *P. putida* mt-2 xylE gene (this study), forward 5' TCT AYC TSA AGG CYT GGA 3' and reverse 5' TCG GTC ATG GAG ATC AGG TCG 3' to amplify a 583-bp C23O gene fragment corresponding to nucleotide positions 125 to 707 of *P. putida* mt-2 xylE gene (this study), forward 5'AGG TGW CGT SAT GAA MAA AGG 3' and reverse 5' TYA GGT SAK MAC GGT CAK GAA 3' to amplify the complete C23O gene plus 10 bases upstream producing a fragment of an expected size of 934-bp, corresponding to nucleotide positions -10 to 924 of *P. putida* mt-2 xylE gene (this study). To determine optimal annealing temperatures for the primer sets, gradient PCR was performed with annealing temperatures between 50.4 to 60°C. All PCR

programs comprised an initial step at 95°C for 3 min, followed by 35 cycles of denaturation at 94°C for 45 s, annealing for 45 s, and elongation at 72°C (the duration of the elongation step was varied according to the targeted PCR product size, using 10 sec per 100-bp) and a final elongation step at 72°C for 8 min. When the PCR products were intended to be applied on PCR-SSCP gels, annealing temperature of 52°C was applied. When the PCR products were amplified for direct sequencing, a touchdown PCR approach was used for 10 cycles, with an initial annealing temperature of 60°C followed by a decrease in annealing temperature of 1°C per cycle, followed by 25 PCR cycles at a constant annealing temperature of 55°C. To evaluate PCR product sizes, 4 µl of the PCR reactions were analysed by agarose gel electrophoresis (1.5% agarose, 10 cm length, 1X TAE running buffer, 1 hour at 95 volts). DNA bands were visualized by ethidium bromide staining (Sambrook *et al.*, 1989). To compare amplification efficiency among the primer sets, DNA extracted from *P. putida* mt-2 or a mixture of equal amounts of *P. putida* mt-2, *Pseudomonas* sp. IC, *P. putida* G7, and *P. aeruginosa* J1104 DNA were used as template, and serial dilutions were applied in PCR reactions with each primer set. Purification of PCR products was done with either the Qiaquick PCR Cleaning Kit or the Gel Extraction Kit (Qiagen) according to the manufacturer instructions. PCR amplification of 16S rDNA fragments from isolates was performed as previously described (Karlson *et al.*, 1993). 5 µl of the purified PCR products were subjected to amplified ribosomal DNA restriction analysis (ARDRA) using the restriction enzyme *AluI* (Mocali *et al.*, 2003) and representatives of distinctive restriction patterns were fully sequenced.

The 934-bp C23O PCR fragments of 1YB2 (type A1) and 1YdBTEX2 (type A2) were cloned into pGEM-T by using the pGEM-T Easy Kit (Promega) to give pA1Tyr and pA2His, respectively. These plasmids were transformed into *Escherichia coli* JM109 competent cells (Promega). *E. coli* cells were cultured on LB medium at 37°C, supplemented with ampicillin at a final concentration of 100 µg/ml.

PCR-SSCP

Single-stranded DNA from PCR products was obtained as previously described (Schwieger and Tebbe, 1998). Briefly, PCR as performed with one of the primers being 5'phosphorylated, PCR products were eluted from agarose gels, and the phosphorylated strands were digested by lambda exonuclease (NEB). The remaining single-strands were purified with Qiaquick PCR Cleaning Kit (Qiagen), dried by vacuum centrifugation, resuspended in 6 µl of loading buffer (95% formamide, 0.25% bromophenol blue and 0.25% xylene cyanol), and denatured for 5 minutes at 94 °C, followed by instant cooling on water ice bath for 3 minutes. The separation conditions were standardised in a DCode System for PCR-SSCP (Bio-Rad) coupled to a cooling bath device (Lauda E100).

Optimised running parameters were 120V (10 mA) for 18 h at a constant temperature of 26°C on 20 cm x 20 cm x 0,75 mm 0.6X MDE gels in 0.7X TBE (Sambrook *et al.*, 1989) as a running buffer.

Optimal results were obtained when ssDNA obtained from 100 - 400 ng dsDNA was loaded onto the gels and a slightly enhanced resolution was achieved when the amplified single-strands of the reverse primer were subjected to PCR-SSCP analysis. For nucleic acid detection, gels were silver stained as reported previously (Bassam *et al.*, 1991). Single-strand electrophoretic mobilities corresponding to different conformations were excised from dried gels, and DNA extracted by the "Crush and Soak" method (Sambrook *et al.*, 1989). PCR reamplification of the excised and eluted single-strands was made with the same primers used to generate the original dsDNA fragment. PCR-SSCP bands of different mobilities were quantified by using Scion Image Beta 4.0.2 software (Scion Corporation) and Imagequant Software (Molecular Dynamics). The similarity matrix of PCR-SSCP fingerprints was calculated using Bio1D v.99.02 Software (Vilber Lourmat) either with Nei et Li (Dice) or Jaccard coefficients 5% confidence applied to both compared bands. A dendrogram of similarity values in the matrix was calculated using the UPGMA algorithm.

DNA sequencing and phylogenetic analyses

Nucleotide sequencing of PCR fragments or plasmids with cloned inserts was carried out on both strands using Taq dye-deoxy terminator in an ABI 373A automatic DNA sequencer (Perkin-Elmer Applied Biosystems) using M13 forward and reverse primers (Sambrook *et al.*, 1989). Primers used for sequence reactions were the same as those used in the original PCR. Contigs were assembled with Sequencher software (V 4.0.5). Alignments were generated with CLUSTAL X 1.8 windows interface of CLUSTAL W program using default values (Thompson *et al.*, 1997). Their results were edited and translated using GeneDoc (version 2.6.001). The Sequence Match program (Cole *et al.*, 2003) was used to find the closest relative to the 16S rDNA sequences obtained. Phylogenetic trees were obtained with the option available on CLUSTAL program through Neighbor-Joining (N-J) algorithm method. Distances were generated using the Kimura Matrix, and tree stability was supported through Bootstrap analysis, values higher than 50% of 1000 replications (seed value 111) are shown on respective nodes. Trees were visualized with NJplot (Perriere and Gouy, 1996). To visualize relationships between the sequences retrieved by all the methods, the program Treecon for Windows (1.3b) (Van de Peer and De Wachter, 1993) was used to estimate distances using a Kimura matrix with 500 bootstrap samples and to infer tree topology by UPGMA clustering method. Sequence data reported in this study have been submitted to the GenBank/EMBL//DBBJ databases and are available under accession numbers AJ544921 to AJ545012, AY228547, AY364085, AY364086 and AY364087.

Preparation of cell extracts

Isolates were routinely grown in mineral salts medium (Dorn *et al.*, 1974) containing 50 mM phosphate buffer (pH 7.4), supplemented with 4 mM of benzene (supplied via the vapour phase) or 1 % glucose. The baffled Erlenmeyer flasks were sealed with Teflon coated screw caps and incubated at 30°C on a rotary shaker (150 rpm). *E. coli* JM109 containing C23O genes were grown at 37°C in Luria broth medium containing 1 mM IPTG and 0.1 mg/ml ampicillin. Cells harvested from 100 – 200 ml of medium were resuspended in potassium phosphate buffer (50 mM, pH 7.5), and disrupted using a French press (Aminco, Silver Spring, MD, USA). Cell debris was removed by centrifugation at 100,000 x g for 1 h at 4°C.

Enzyme assays

Catechol 2,3-dioxygenase activity was determined in 50 mM K/Na-phosphate (pH 7.5) with 0.1 mM of catechol, 3-methylcatechol, 4-methylcatechol or 4-chlorocatechol as substrates, using extinction coefficients of reaction products previously described (catechol, $\epsilon_{375\text{nm}} = 36,000 \text{ M}^{-1} \text{ cm}^{-1}$; 3-methylcatechol, $\epsilon_{388\text{nm}} = 13,800 \text{ M}^{-1} \text{ cm}^{-1}$; 4-methylcatechol, $\epsilon_{382\text{nm}} = 28,100 \text{ M}^{-1} \text{ cm}^{-1}$, , 4-chlorocatechol, $\epsilon_{379\text{nm}} = 39,600 \text{ M}^{-1} \text{ cm}^{-1}$ (Hirose *et al.*, 1994; Heiss *et al.*, 1995). Catechol 1,2-dioxygenase was analyzed as previously described (Dorn and Knackmuss, 1978). Kinetic measurements were recorded on an UV 2100 spectrophotometer (Shimadzu Corporation, Kyoto, Japan). 2-Hydroxymuconic semialdehyde hydrolase was measured by determining the NAD-independent decrease in concentration of 2-hydroxymuconic semialdehyde (0.05 mM) at 375 nm or of 2-hydroxy-6-oxo-hepta-2,4-dieneoate (0.05 mM) at 388 nm (Sala-Trepat *et al.*, 1972). 2-Hydroxymuconic semialdehyde dehydrogenase was measured by determining the NAD-dependent decrease in concentration of 2-hydroxymuconic semialdehyd (0.05 mM) at 375 nm. NAD was added to a final concentration of 0.5 mM. 2-Hydroxymuconic semialdehyde dehydrogenase activity was calculated by subtracting the 2-hydroxymuconic semialdehyde hydrolase activity from the overall rate of 2-hydroxymuconic semialdehyde disappearance. 2-Hydroxymuconic semialdehyde and 2-hydroxy-6-oxo-hepta-2,4-dieneoate were prepared *in situ* by incubation of 0.05 mM solutions of catechol or 3-methylcatechol in phosphate buffer (50 mM, pH 7.4) with partially purified C23O from *E. coli* JM109 (pA2His) (corresponding to 0.1 – 0.2 U/ml). Protein concentrations were determined by the method of Bradford (Bradford, 1976) with bovine serum albumin as the standard.

Enzyme purification and sequencing

Catechol 2,3-dioxygenases were partially purified using a Fast Protein Liquid Chromatography system (Amersham Biosciences, Freiburg, Germany). Cell extracts of *E. coli* JM109 (pA2His), *E. coli* JM109 (pA1Tyr), strain 1YB2 or strain 1YdBTEX2 were separately applied to a MonoQ HR 5/5 column and eluted by a linear gradient of NaCl (0 to 0.5 M over 25 ml) in Tris HCl (50 mM, pH 7.5) (flow rate 1 ml/min; fraction volume 0,5 ml). (SDS-PAGE) was performed on a Bio-Rad Miniprotein II essentially as described previously (Laemmli, 1970). The acrylamide concentrations for concentrating and separating gels were 5 and 10% (wt/vol), respectively. The proteins were stained by Coomassie brilliant blue R250. Proteins were electroblotted onto a polyvinylidene difluoride membrane, and the membrane was stained with Coomassie brilliant blue R250. N-terminal amino acid sequencing was performed with an Applied Biosystems model 494A Procise HT sequencer. In-gel trypsin digests of proteins were analysed with a matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) Reflex apparatus (Bruker) as previously described (Heim *et al.*, 2002). Respective peptide fingerprints obtained were compared by BLAST tool (Tatusova and Madden, 1999) against protein databases or putative proteins of gene or gene fragments sequenced in the course of this study.

Acknowledgements

We wish to thank Prof. P.A Williams, Dr A. Kitayama, Dr. R. Bosch, Dr. T. Omori, Prof. Y. Kim, Prof. V. Shingler, Prof. H. Herrmann, Prof. N. Ornston, Dr. D. Young, Prof. D. Kunz and Prof. D.J. Hopper, for kindly providing the reference strains used in this work and Barbara Hendrickx, Dr Winnie Dejonghe, Dr Miroslav Cerník and Dr Tomáš Lederer for soil sampling. We also thank Iris Plumeier and Annette Krueger for excellent technical assistance, Dr. Daniela Regenhart for guidance in MALDI-TOF technique, and Dr. Lotte H.E. Gabriel-Jürgens, Dr. Dirk F. Wenderoth and Dr. Edward R.B. Moore for useful comments on the manuscript. This work was supported by grant QLK3-CT-2000-00731 from the European Community.

References

- Arenghi, F.L., Berlanda, D., Galli, E., Sello, G., and Barbieri, P. (2001) Organization and regulation of *meta* cleavage pathway genes for toluene and *o*-xylene derivative degradation in *Pseudomonas stutzeri* OX1. *Appl Environ Microbiol* **67**: 3304-3308.
- Atha, D.H., Kasprzak, W., O'Connell, C.D., and Shapiro, B.A. (2001) Prediction of DNA single-strand conformation polymorphism: analysis by capillary electrophoresis and computerized DNA modeling. *Nucleic Acids Res* **29**: 4643-4653.
- Baldwin, B.R., Nakatsu, C.H., and Nies, L. (2003) Detection and enumeration of aromatic oxygenase genes by multiplex and real-time PCR. *Appl Environ Microbiol* **69**: 3350-3358.
- Bartilson, M., and Shingler, V. (1989) Nucleotide sequence and expression of the catechol 2,3-dioxygenase-encoding gene of phenol-catabolizing *Pseudomonas* CF600. *Gene* **85**: 233-238.
- Bassam, B.J., Caetano-Anolles, G., and Gresshoff, P.M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* **196**: 80-83.
- Beil, S., Mason, J.R., Timmis, K.N., and Pieper, D.H. (1998) Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene. *J Bacteriol* **180**: 5520-5528.
- Beller, H.R., Kane, S.R., Legler, T.C., and Alvarez, P.J. (2002) A real-time polymerase chain reaction method for monitoring anaerobic, hydrocarbon-degrading bacteria based on a catabolic gene. *Environ Sci Technol* **36**: 3977-3984.
- Benjamin, R.C., Voss, J.A., and Kunz, D.A. (1991) Nucleotide sequence of *xylE* from the TOL pDK1 plasmid and structural comparison with isofunctional catechol-2,3-dioxygenase genes from TOL, pWW0 and NAH7. *J Bacteriol* **173**: 2724-2728.
- Bosch, R., Garcia-Valdes, E., and Moore, E.R. (2000) Complete nucleotide sequence and evolutionary significance of a chromosomally encoded naphthalene-degradation lower pathway from *Pseudomonas stutzeri* AN10. *Gene* **245**: 65-74.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.
- Bruns, M.A., Hanson, J.R., Mefford, J., and Scow, K.M. (2001) Isolate PM1 populations are dominant and novel methyl tert-butyl ether-degrading bacteria in compost biofilter enrichments. *Environ Microbiol* **3**: 220-225.
- Carrington, B., Lowe, A., Shaw, L.E., and Williams, P.A. (1994) The lower pathway operon for benzoate catabolism in biphenyl-utilizing *Pseudomonas* sp. strain IC and the nucleotide sequence of the *bphE* gene for catechol 2,3-dioxygenase. *Microbiology* **140**: 499-508.
- Cavalca, L., Di Gennaro, P., Colombo, M., Andreoni, V., Bernasconi, S., Ronco, I., and Bestetti, G. (2000) Distribution of catabolic pathways in some hydrocarbon-degrading bacteria from a subsurface polluted soil. *Res Microbiol* **151**: 877-887.

Cerdan, P., Wasserfallen, A., Rekik, M., Timmis, K.N., and Harayama, S. (1994) Substrate specificity of catechol 2,3-dioxygenase encoded by TOL plasmid pWW0 of *Pseudomonas putida* and its relationship to cell growth. *J Bacteriol* **176**: 6074-6081.

Chatfield, L.K., and Williams, P.A. (1986) Naturally occurring TOL plasmids in *Pseudomonas* strains carry either two homologous or two nonhomologous catechol 2,3-oxygenase genes. *J Bacteriol* **168**: 878-885.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A. et al. (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* **31**: 442-443.

Daly, K., Dixon, A.C., Swannell, R.P., Lepo, J.E., and Head, I.M. (1997) Diversity among aromatic hydrocarbon-degrading bacteria and their *meta*-cleavage genes. *J Appl Microbiol* **83**: 421-429.

Dorn, E., and Knackmuss, H.J. (1978) Chemical structure and biodegradability of halogenated aromatic compounds. Two catechol 1,2-dioxygenases from a 3-chlorobenzoate-grown pseudomonad. *Biochem J* **174**: 73-84.

Dorn, E., Hellwig, M., Reineke, W., and Knackmuss, H.J. (1974) Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. *Arch Microbiol* **99**: 61-70.

Eltis, L.D., and Bolin, J.T. (1996) Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* **178**: 5930-5937.

Ghosal, D., You, I.S., and Gunsalus, I.C. (1987) Nucleotide sequence and expression of gene *nahH* of plasmid NAH7 and homology with gene *xyIE* of TOL pWWO. *Gene* **55**: 19-28.

Hamelin, J., Fromin, N., Tarnawski, S., Teyssier-Cuvelles, S., and Aragno, M. (2002) *nifH* gene diversity in the bacterial community associated with the rhizosphere of *Molinia coerulea*, an oligonitrophilic perennial grass. *Environ Microbiol* **4**: 477-481.

Heim, S., Del Mar Lleo, M., Bonato, B., Guzman, C.A., and Canepari, P. (2002) The viable but nonculturable state and starvation are different stress responses of *Enterococcus faecalis*, as determined by proteome analysis. *J Bacteriol* **184**: 6739-6745.

Heiss, G., Stolz, A., Kuhm, A.E., Muller, C., Klein, J., Altenbuchner, J., and Knackmuss, H.J. (1995) Characterization of a 2,3-dihydroxybiphenyl dioxygenase from the naphthalenesulfonate-degrading bacterium strain BN6. *J Bacteriol* **177**: 5865-5871.

Herrmann, H., Muller, C., Schmidt, I., Mahnke, J., Petruschka, L., and Hahnke, K. (1995) Localization and organization of phenol degradation genes of *Pseudomonas putida* strain H. *Mol Gen Genet* **247**: 240-246.

Hirose, J., Kimura, N., Suyama, A., Kobayashi, A., Hayashida, S., and Furukawa, K. (1994) Functional and structural relationship of various extradiol aromatic ring-cleavage dioxygenases of *Pseudomonas* origin. *FEMS Microbiology Letters* **118**: 273-277.

Hopper, D.J., and Kemp, P.D. (1980) Regulation of enzymes of the 3,5-xyleneol-degradative pathway in *Pseudomonas putida*: evidence for a plasmid. *J Bacteriol* **142**: 21-26.

Kanakaraj, R., Harris, D.L., Songer, J.G., and Bosworth, B. (1998) Multiplex PCR assay for detection of *Clostridium perfringens* in feces and intestinal contents of pigs and in swine feed. *Vet Microbiol* **63**: 29-38.

Karlson, U., Dwyer, D.F., Hooper, S.W., Moore, E.R., Timmis, K.N., and Eltis, L.D. (1993) Two independently regulated cytochromes P-450 in a *Rhodococcus rhodochrous* strain that degrades 2-ethoxyphenol and 4-methoxybenzoate. *J Bacteriol* **175**: 1467-1474.

Kaulmann, U., Kaschabek, S.R., and Schlomann, M. (2001) Mechanism of chloride elimination from 3-chloro- and 2,4-dichloro-cis,cis-muconate: new insight obtained from analysis of muconate cycloisomerase variant CatB-K169A. *J Bacteriol* **183**: 4551-4561.

Keil, H., Keil, S., Pickup, R.W., and Williams, P.A. (1985) Evolutionary conservation of genes coding for *meta* pathway enzymes within TOL plasmids pWW0 and pWW53. *J Bacteriol* **164**: 887-895.

Kisand, V., and Wikner, J. (2003) Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *Journal of Microbiological Methods* **54**: 183-191.

Kita, A., Kita, S., Fujisawa, I., Inaka, K., Ishida, T., Horiike, K. et al. (1999) An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure Fold Des* **7**: 25-34.

Kitayama, A., Achioku, T., Yanagawa, T., Kanou, K., Kikuchi, M., Ueda, H. et al. (1996) Cloning and characterization of extradiol aromatic ring-cleavage dioxygenases of *Pseudomonas aeruginosa* J1104. *J Ferment Bioeng* **82**: 217-223.

Knackmuss, H.-J., Hellwig, M., Lackner, H., and Otting, W. (1976) Cometabolism of 3-methylbenzoate and methylcatechols by a 3-chlorobenzoate utilizing *Pseudomonas*: Accumulation of (+)-2,5-dihydro-4-methyl- and (+)-2,5-dihydro-2-methyl-5-oxo-furan-2-acetic acid. *European Journal of Applied Microbiology* **2**: 267 - 276.

Koizumi, Y., Kelly, J.J., Nakagawa, T., Urakawa, H., El-Fantroussi, S., Al-Muzaini, S. et al. (2002) Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization, and DNA microarray technology. *Appl Environ Microbiol* **68**: 3215-3225.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lawrence, J.G. (2002) Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**: 449-460.

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

Marlowe, E.M., Wang, J.M., Pepper, I.L., and Maier, R.M. (2002) Application of a reverse transcription-PCR assay to monitor regulation of the catabolic *nahAc* gene during phenanthrene degradation. *Biodegradation* **13**: 251-260.

1 Mesarch, M.B., Nakatsu, C.H., and Nies, L. (2000) Development of catechol 2,3-dioxygenase-specific
2 primers for monitoring bioremediation by competitive quantitative PCR. *Appl Environ Microbiol*
3 **66**: 678-683.

4 Meyer, S., Moser, R., Neef, A., Stahl, U., and Kampfer, P. (1999) Differential detection of key
5 enzymes of polyaromatic-hydrocarbon-degrading bacteria using PCR and gene probes. *Microbiology*
6 **145**: 1731-1741.

7 Miskin, I.P., Farrimond, P., and Head, I.M. (1999) Identification of novel bacterial lineages as active
8 members of microbial populations in a freshwater sediment using a rapid RNA extraction procedure
9 and RT-PCR. *Microbiology* **145**: 1977-1987.

10 Mocali, S., Bertelli, E., Di Cello, F., Mengoni, A., Sfalanga, A., Viliani, F. et al. (2003) Fluctuation of
11 bacteria isolated from elm tissues during different seasons and from different plant organs. *Res*
12 *Microbiol* **154**: 105-114.

13 Moon, J., Min, K.R., Kim, C.K., Min, K.H., and Kim, Y. (1996) Characterization of the gene
14 encoding catechol 2,3-dioxygenase of *Alcaligenes* sp. KF711: overexpression, enzyme purification,
15 and nucleotide sequencing. *Arch Biochem Biophys* **332**: 248-254.

16 Nakai, C., Kagamiyama, H., Nozaki, M., Nakazawa, T., Inouye, S., Ebina, Y., and Nakazawa, A.
17 (1983) Complete nucleotide sequence of the metapyrocatechase gene on the TOL plasmid of
18 *Pseudomonas putida* mt-2. *J Biol Chem* **258**: 2923-2928.

19 Norton, J.M., Alzerreca, J.J., Suwa, Y., and Klotz, M.G. (2002) Diversity of ammonia
20 monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch Microbiol* **177**: 139-149.

21 Okuta, A., Ohnishi, K., and Harayama, S. (1998) PCR isolation of catechol 2,3-dioxygenase gene
22 fragments from environmental samples and their assembly into functional genes. *Gene* **212**: 221-
23 228.

24 Ouchiya, N., Miyachi, S., and Omori, T. (1998) Cloning and nucleotide sequence of carbazole
25 catabolic genes from *Pseudomonas stutzeri* strain OM1, isolated from activated sludge. *J Gen Appl*
26 *Microbiol* **44**: 57-63.

27 Parales, R.E., Lee, K., Resnick, S.M., Jiang, H., Lessner, D.J., and Gibson, D.T. (2000) Substrate
28 specificity of naphthalene dioxygenase: effect of specific amino acids at the active site of the enzyme.
29 *J Bacteriol* **182**: 1641-1649.

30 Perez-Pantoja, D., Ledger, T., Pieper, D.H., and Gonzalez, B. (2003) Efficient turnover of
31 chlorocatechols is essential for growth of *Ralstonia eutropha* JMP134(pJP4) in 3-chlorobenzoic acid.
32 *J Bacteriol* **185**: 1534-1542.

33 Perriere, G., and Gouy, M. (1996) WWW-query: an on-line retrieval system for biological sequence
34 banks. *Biochimie* **78**: 364-369.

35 Pikus, J.D., Studts, J.M., McClay, K., Steffan, R.J., and Fox, B.G. (1997) Changes in the
36 regiospecificity of aromatic hydroxylation produced by active site engineering in the diiron enzyme
37 toluene 4-monooxygenase. *Biochemistry* **36**: 9283-9289.

Reineke, W. (1998) Development of hybrid strains for the mineralization of chloroaromatics by patchwork assembly. *Annu Rev Microbiol* **52**: 287-331.

Sala-Trepat, J.M., Murray, K., and Williams, P.A. (1972) The metabolic divergence in the meta cleavage of catechols by *Pseudomonas putida* NCIB 10015. Physiological significance and evolutionary implications. *Eur J Biochem* **28**: 347-356.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Schmalenberger, A., and Tebbe, C.C. (2003) Bacterial diversity in maize rhizospheres: conclusions on the use of genetic profiles based on PCR-amplified partial small subunit rRNA genes in ecological studies. *Mol Ecol* **12**: 251-262.

Schmalenberger, A., Schwieger, F., and Tebbe, C.C. (2001) Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. *Appl Environ Microbiol* **67**: 3557-3563.

Schwieger, F., and Tebbe, C.C. (1998) A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl Environ Microbiol* **64**: 4870-4876.

Sotsky, J.B., Greer, C.W., and Atlas, R.M. (1994) Frequency of genes in aromatic and aliphatic hydrocarbon biodegradation pathways within bacterial populations from Alaskan sediments. *Can J Microbiol* **40**: 981-985.

Stapleton, R.D., and Sayler, G.S. (1998) Assessment of the microbiological potential for the natural attenuation of petroleum hydrocarbons in a shallow aquifer system. *Microb Ecol* **36**: 349-361.

Suenaga, H., Watanabe, T., Sato, M., Ngadiman, and Furukawa, K. (2002) Alteration of regiospecificity in biphenyl dioxygenase by active-site engineering. *J Bacteriol* **184**: 3682-3688.

Tatusova, T.A., and Madden, T.L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**: 247-250.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876-4882.

Torsvik, V., and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240-245.

Van de Peer, Y., and De Wachter, R. (1993) TREECON: a software package for the construction and drawing of evolutionary trees. *Comput Appl Biosci* **9**: 177-182.

Wawer, C., Jetten, M.S., and Muyzer, G. (1997) Genetic diversity and expression of the [NiFe] hydrogenase large-subunit gene of *Desulfovibrio* spp. in environmental samples. *Appl Environ Microbiol* **63**: 4360-4369.

Widada, J., Nojiri, H., and Omori, T. (2002) Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic genes in bioremediation. *Appl Microbiol Biotechnol* **60**: 45-59.

- 1 Wikstrom, P., Wiklund, A., Andersson, A.C., and Forsman, M. (1996) DNA recovery and PCR
2 quantification of catechol 2,3-dioxygenase genes from different soil types. *J Biotechnol* **52**: 107-120.
- 3 Williams, P.A., and Sayers, J.R. (1994) The evolution of pathways for aromatic hydrocarbon
4 oxidation in *Pseudomonas*. *Biodegradation* **5**: 195-217.
- 5 Williams, P.A., Jones, R.M., and Shaw, L.E. (2002) A third transposable element, ISPpu12, from the
6 toluene-xylene catabolic plasmid pWW0 of *Pseudomonas putida* mt-2. *J Bacteriol* **184**: 6572-6580.
- 7 Yeates, C., Holmes, A.J., and Gillings, M.R. (2000) Novel forms of ring-hydroxylating dioxygenases
8 are widespread in pristine and contaminated soils. *Environ Microbiol* **2**: 644-653.

Table 1. Utilization of single BTEX compounds as sole carbon sources by bacterial isolates exhibiting *meta*-cleavage activity.

Isolate Name	B	T	E	m-X	p-X	o-X	C23O	ARDRA type
A1YC1	-	+	-	+	+	-	+	PS2*
A1YC2	-	+	-	+	+	-	+	PS2*
1Y C1	+	+	+	-	-	-	+	PS1
1Y C2	+	-	-	-	-	-	+	PS1*
1Y C3	+	-	-	-	-	-	+	PS1
1Y Xyl1	+	+	+	-	-	-	+	PS1
1Y Xyl2	+	+	+	-	-	-	+	PS1
1Y Xyl3	+	+	+	-	-	-	+	PS1
1Y dBTEX1 -	-	-	-	-	-	+	ND	
1Y dBTEX2 +	+	+	-	-	-	+	PS1*	
1Y dBTEX3 +	-	-	-	-	-	+	PS1	
1Y BTEX1	+	-	-	-	-	-	+	PS1
1Y BTEX3	+	-	-	-	-	-	+	PS1
1Y B1	+	-	-	-	-	-	+	PS1
1Y B2	+	-	-	-	-	-	+	PS1*
1Y B3	+	-	-	-	-	-	+	PS1*
3Y C1b	+	+	-	-	-	-	-	ND
3Y C1	-	-	-	-	-	-	-	ND
3Y C2	+	+	+	-	-	-	-	ND
3Y C3	+	-	-	-	-	-	-	ND
3Y C4	+	+	+	-	-	-	-	ND
3Y C5	+	+	-	-	-	-	+	PS2
3Y Xyl2	+	+	-	-	-	-	-	PS2*
3Y Xyl1	+	+	-	-	-	-	+	PS2
3Y Xyl2b	+	-	-	-	-	-	+	PS2
3Y Xyl3	+	-	-	-	-	-	+	PS2
3Y dBTEX1 -	-	-	-	-	-	-	PS2*	
3Y dBTEX2 +	+	+	-	-	-	+	ND	
3Y dBTEX1b+	-	-	-	-	-	+	PS2*	
3Y BTEX2b +	+	+	-	-	-	-	PS2*	
3Y B1	-	-	-	-	-	-	-	ND
3Y B1b	-	-	-	-	-	-	-	ND
3Y B2	-	-	-	-	-	-	-	ND
1X C1	+	+	+	-	-	-	-	ND
1X C2	+	+	-	-	-	-	+	PS2
1X Xyl1	+	+	+	-	-	-	-	ND
1X Xyl1b	+	-	-	-	-	-	-	ND
1X Xyl2	-	-	-	-	-	-	+	PS1
1X dBTEX1 -	-	-	-	-	-	+	PS1*	
1X BTEX1	+	+	+	-	-	-	+	PS1
1X BTEX3	+	+	+	-	-	-	-	PS1
1X BTEX4	+	+	+	-	-	-	+	PS1
1X BTEX1b +	+	+	-	-	-	-	ND	
1X B1	+	-	-	-	-	-	+	PS3*
1X B1b	-	-	-	-	-	-	+	ND
1X B2	+	+	+	-	-	-	+	PS1*

Strains were designated according to sample origin and aromatic cosubstrates included during isolation on R2A plates prior as described in the experimental procedures. Growth was determined on solid mineral medium supplemented with single aromatic substrates as sole carbon and energy source. Aromatic substrate used were benzene (B), toluene (T), ethylbenzene (E), *meta*-xylene (m-X), *para*-xylene (p-X) and *ortho*-xylene (o-X). All isolates were analysed for the presence of C23O subfamily I.2.A amplifiable gene. (+), PCR product of the expected length by all 4 primers, (-) no PCR product by any of the primers. A selection of isolates were analysed by ARDRA and the respective banding profile (ARDRA type PS1, PS2 or PS3, respectively, Fig. 2B) is indicated. (*) denote those isolates further analysed by full sequencing of the 16S rDNA PCR fragment. ND, not determined.

Table 2. Substrate specificity and activity of catechol 2,3-dioxygenases of strains 1YB2, 1YdBTEX2, *E.coli* JM109 (pA1Tyr) and *E.coli* JM109 (pA2His).

Substrate	Strain			
	1YB2	1YdBTEX2	JM109(pA1Tyr)	JM109(pA2His)
Catechol	100	100	100	100
		(830 ± 40) (4.200 ± 1800)	(1.300 ± 80) (4.800 ± 200)	
3-Methylcatechol	45 ± 5	20 ± 5	45 ± 5	20 ± 5
4-Methylcatechol	120 ± 10 50 ± 10	140 ± 20	50 ± 10	
4-Chlorocatechol	40 ± 5	35 ± 5	45 ± 10	35 ± 5

Activities were determined at substrate concentrations of 0.1 mM and are given relative to that with catechol (100%). Activities in cell extracts (U/g protein) are given in parentheses.

Figure legends

Fig. 1.

Analysis of C23O gene fragments from reference strains or environmental DNA by PCR-SSCP with the single-strand removal method.

A. Mobility of 238b C23O ssDNA fragments amplified from *P. putida* mt-2 (lane 1-4) or *P. putida* HS1 (lane 5-8). A total of 520 ng (lane 1), 208 ng (lane 2), 104 ng (lane 3-4), 600 ng (lane 5), 240 ng (lane 6) or 120 ng (lane 7-8) was digested by lambda exonuclease for 3 h (lane 1-3 and 5-7) or 16 h (lane 4 and 8). B. Mobility of 583b C23O ssDNA fragments amplified from *P. stutzeri* AN10 (lane 1), *P. putida* CF600 (lane 2), *P. putida* G7 (lane 3), *Pseudomonas* sp. IC (lane 4), *P. putida* HS1 (lane 5), *P. putida* 3,5X (lane 6), and of a mixture of those amplification products (lane 7). The respective dominant bands are indicated by small letters. C. Mobility of 527b C23O ssDNA fragments amplified from *P. stutzeri* AN10 (lane 1), *P. putida* CF600 (lane 2), *P. putida* G7 (lane 3), *Pseudomonas* sp. IC (lane 4), *P. putida* HS1 (lane 5) and *P. putida* 3,5X (lane 6). M, 75 ng DNA Molecular Weight Marker X (Roche). D. Mobility of 527b C23O ssDNA fragments amplified from soil DNA, that has been extracted from samples 3Y (lane 1), 1Y (lane 2), A3Y (lane 3), A1Y (lane 4), 3X (lane 5) and 0Y (lane 6), respectively. Approximately 400 ng of dsDNA PCR product were digested and loaded. M, 75 ng of DNA Molecular Weight Marker X (Roche). Bands (A to M) subjected to sequence analysis are indicated by small letters. E. Dendrogram showing the degree of similarity of C23O gene structures as determined in D of samples 3Y, 1Y, A3Y, A1Y, 3X and 0Y (Dice coefficient, 5% confidence). Scale representing percentage of similarity between the profiles is shown at the left side. F. Phylogenetic tree of deduced C23O protein sequences from sequences recovered of PCR-SSCP bands, indicated as (SS), and designated according to sample origin and mobility (alignment of 99 amino acids, bar represents 5 amino acid changes per 100 amino acids). The sequences obtained were aligned with known members of the C23O subfamily I.2.A (indicated by their DDBJ/EMBL/GenBank accession number, organism of origin and strain designation). The C23O protein sequence of *S. yanoikuyae* B1, a member of the subfamily I.2.B, was used as an outgroup. Alignments were performed with CLUSTAL W using default values. N-J phylogenetic trees were generated using the option available in CLUSTAL W program. Bootstrap values above 50% from 1000 neighbor-joining trees are indicated to the left of the nodes.

Fig. 2.

Phylogenetic trees of deduced amino acid sequences of C23O gene fragments from PCR clone libraries and isolates and characteristic ARDRA profiles of isolates. Sequences were aligned with known members of the C23O subfamily I.2.A (indicated by their DDBJ/EMBL/GenBank accession number, organism of origin and strain designation). The C23O protein sequence of *S. yanoikuyae* B1, a member of the Subfamily I.2.B, was used as an outgroup. Alignments were performed with CLUSTAL W using default values. Phylogenetic trees were generated using the option available in CLUSTAL W program. Bootstrap values above 50% from 1000 neighbor-joining trees are indicated to the left of the nodes.

A. Phylogenetic tree of deduced C23O protein sequences from PCR clone library sequences, indicated as (CL) and designated according to the site of origin (alignment of 171 amino acids, bar represents 5 amino acid changes per 100 amino acids). The insert shows the phylogenetic tree

from the DNA alignment of PCR clone sequences closely related to *P. stutzeri* AN10 C23O (alignment of 513 nucleotides, bar represents 5 nucleotide changes per 1000 nucleotides). Triangles represent the 5 groups of identical putative protein sequences originating from 18 different DNA sequences. B. ARDRA types of isolates. C. Phylogenetic tree of deduced C23O protein sequences from bacterial isolates, indicated as (IS), from sites 3Y, 1Y, and 1X. (alignment of 268 amino acids, bar represents 2 amino acid changes per 100 amino acids). Triangles at the left side indicate groups of sequences (A to D) found in the strain collection.

Fig. 3.

Cumulative joint UPGMA clustering representation of related sequences from PCR-SSCP bands (SS), clone libraries (CL) and isolates (IS) (99 amino acids, scale represents similarity coefficient). The sequences obtained by all methods were aligned with known members of the C23O subfamily I.2.A (indicated by their DDBJ/EMBL/GenBank accession number, organism of origin and strain designation). The C23O protein sequence of *S. yanoikuyae* B1, a member of the subfamily I.2.B, was used as outgroup. The site of origin of the sequence retrieved is indicated at the right border together with a graphical representation of relative abundance. The abundance of a specific C23O sequence from isolates (IS) is given as number of isolates showing this sequence/number of total isolates from that site, and additionally by (■) indicating the number of isolates with the respective sequence, and (■) indicating the number of other isolates from that site. The abundance of C23O sequence from clone libraries (CL) is given as ratio of clones showing this sequence/number of total clones from that site, and additionally represented by (/) indicating the number of isolates showing this sequence, and (/) indicating the number of other clones from that site. The abundance of C23O sequence from PCR-SSCP bands (the respective mobilities are indicated) is given as percentage of total signal intensity, which is additionally represented by (f) in front of the sequence. Total of black and grey left curly braces correspond to 100%.

Fig. 1

Fig. 2

Additional material to be published on the World Wide Web Site

Supplementary online material 1

Protein (A) and DNA (B) multiple alignments of Exradiol Dioxygenases Subfamily I.2.A genes used for primer design. Sequences names correspond to DDBJ/EMBL/GenBank accession number, organism of origin and strain designation.

A

B

Supplementary online material 2

PCR products obtained from reference strains, isolates and soil DNA using a primer set targeting a 527-bp fragment of C23O. Lanes 2 to 11, 16 to 25, and 27 contain respectively amplifications from *P. stutzeri* AN10, *P. putida* CF600, *P. putida* G7, *Pseudomonas* sp. IC, *P. putida* H (PG320), *P. putida* KT2440 plus pPGH11 (PG125 *P. putida* H), *E. coli* JM109 plus pCNU401 (*Sphingomonas* sp. KF711), *P. aeruginosa* JI104, *P. putida* mt-2, *P. putida* mt53, *P. stutzeri* OM1, *P. stutzeri* OX1, *P. putida* HS1, *P. putida* 35X from Shingler Lab, *P. putida* 3,5X from Hopper lab, Isolate A1YC1, Isolate A2YC1, Soil 1Y, Soil A3Y, Soil A1Y, and *P. putida* KT2440 TOL+. Any amplification was obtained in triplicates of negative controls (data not shown). Lanes 1, 12, 15 and 26 were loaded with 200 ng of molecular weight marker V (Roche), lanes 13, 14, and 28 were loaded with 250 ng of O'rangeRuler 50bp DNA ladder (MBI Fermentas).

Supplementary online material 3

SSCP profiles of a 238b C23O fragment PCR amplified using as template DNA extracted from (lanes 1 to18) *P. putida* AN10, *P. putida* CF600, *P. putida* G7, *Pseudomonas* sp. IC, *P. aeruginosa* JI104, *Sphingomonas* sp. KF711, *P. putida* mt2, *P. putida* mt53, *P. stutzeri* OM1, *P. stutzeri* OX1, *P. putida* H, *P. putida* HS1, *P. putida* 3,5X, Isolate A1YC1, Isolate A2YC1, equivalent mixture of all the previous fragments, Soil A1Y and Soil A2Y. Lanes M were loaded with 40 ng of molecular weight marker X (Roche)