Environmental Microbiology (2007) **9**(9), 2289–2297 doi:10.1111/j.1462-2920.2007.01342.x

Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds

Hikaru Suenaga,[1] Tsutomu Ohnuki[1] and Kentaro Miyazaki[1,2]*

[1]*Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan.*

[2]*Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan.*

Summary

A metagenomic approach was taken to retrieve catabolic operons for aromatic compounds from activated sludge used to treat coke plant wastewater. Metagenomic DNA extracted from the sludge was cloned into fosmids and the resulting *Escherichia coli* library was screened for extradiol dioxygenases (EDOs) using catechol as a substrate, yielding 91 EDO-positive clones. Based on their substrate specificity for various catecholic compounds, 38 clones were subjected to sequence analysis. Each insert contained at least one EDO gene, and a total of 43 EDO genes were identified. More than half of these belonged to new EDO subfamilies: I.1.C (2 clones), I.2.G (20 clones), I.3.M (2 clones) and I.3.N (1 clone). The fact that novel I.2.G family genes were over-represented in these clones suggested that these genes play a specific role in environmental aromatic degradation. The I.2.G clones were further classified into six groups based on single-nucleotide polymorphisms (SNPs). Based on the combination of the SNPs, the evolutionary lineage of the genes was reconstructed; further, taking the activities of the clones into account, potential adaptive mutations were identified. The metagenomic approach was thus used to retrieve novel EDO genes as well as to gain insights into the gene evolution of EDOs.

Received 15 February, 2007; accepted 12 April, 2007. *For correspondence. E-mail miyazaki-kentaro@aist.go.jp; Tel. (+81) 29 861 6060; Fax (+81) 29 861 6413.

Introduction

Bacteria that degrade aromatic compounds are widely distributed in the environment and are important for breaking down both natural and xenobiotic compounds. Such bacteria also play critical roles in the global carbon cycle. Aerobic bacterial aromatic compound degradation pathways commonly have catecholic compounds as intermediates (Smith, 1990). Extradiol dioxygenases (EDOs), which catalyse the ring cleavage of catecholic compounds to produce meta-cleavage yellow compounds, play a key role in determining the specificity of these catabolic pathways (Hirose *et al*., 1994; Vaillancourt *et al*., 2004; Fortin *et al*., 2005). On the basis of amino acid sequence similarities, EDOs are classified into two evolutionarily independent families, types I and II, which have totally different primary sequences (Eltis and Bolin, 1996). Microbial screening has been used to isolate EDOs, as well as entire degradation pathways, for use in bioremediation (Furukawa *et al*., 2004; Fortin *et al*., 2005). The majority of known EDOs are type I: subfamily I.2.A, which contains catechol 2,3-dioxygenase, and subfamily I.3.A, which contains 2,3-dihydroxybiphenyl dioxygenase (Eltis and Bolin, 1996; Mesarch *et al*., 2000; Junca *et al*., 2004).

More than 99% of microorganisms are thought to be unculturable or difficult to culture in a laboratory using standard cultivation methods; it seems likely that these uncultured microorganisms include diverse microbes that are distantly related to cultured microorganisms (Amann *et al*., 1995). A metagenomic approach, in which all genome sequences in an environment are non-selectivity cloned into a single library, allows the study of genomes from organisms that are not easily cultured (Handelsman, 2004). One advantage of this approach is that it can, in theory, accelerate the identification of novel genes with industrial value. For example, efforts have been made to retrieve genes such as chitinase, dehydrogenase, oxidoreductase, amylase, esterase, endoglucanase and cyclodextrinase from metagenomic libraries by functional screening (10^4–10^5; Cottrell *et al*., 1999; Henne *et al*., 2000; Knietsch *et al*., 2003; Yun *et al*., 2004; Ferrer *et al*., 2005).

Coke plant wastewater contains various organic pollutants, such as phenol, mono- and polycyclic

nitrogen-containing aromatics, oxygen- and sulfur-containing heterocyclic compounds, and polycyclic aromatic hydrocarbons (Stamoudis and Luthy, 1980). Activated sludge, which is widely used for wastewater treatment, is a complex microbial community that contains various microorganisms responsible for decomposing such compounds (Wagner and Loy, 2002). Metagenomic libraries from activated sludge should therefore provide a great opportunity to study novel aspects of the microbiological activity in this complex system as well as to identify novel genes that are difficult to isolate through a conventional cultivation-based approach.

In this study, we used activated sludge as a metagenomic DNA resource and developed a high throughput screening system using a 96-well plate format. Numerous EDOs were retrieved, sequenced and characterized. Based on gene sequence analysis of members of a new subfamily, 1.2.G, we also discuss the evolution of these genes.

Results and discussion

Construction of a metagenomic library

Prior to construction of a metagenomic library, the diversity of microorganisms present in the activated sludge was evaluated by 16S rDNA sequence analysis. The rDNA genes were amplified with a set of bacterial universal polymerase chain reaction primers using the metagenomic DNA as a template. Amplicons were subcloned into a plasmid, a total of 90 *Escherichia coli* clones were picked randomly, and the cloned 16S rDNA sequences were analysed. The metagenomic DNA comprised highly diverse phylotypes: *Proteobacteria* (41.5%), *Actinobacteria* (28.1%), *Bacteroidetes* (16.9%) and 'unclassified' (13.5%). Overall, the composition appeared unbiased and similar to DNA from other activated sludge samples (Bond *et al*., 1995; Snaidr *et al*., 1997; Daims *et al*., 2001; Juretschko *et al*., 2002; Wagner and Loy, 2002).

The metagenomic *E. coli* library was created using a fosmid as a vector and an average insert size of 33 kb. The library comprised 96 000 clones covering 3.2 Gb of metagenomic DNA; these clones were divided into 10 96-well plates (100 clones well^{-1}), which reduced screening time. An advantage to using cosmid or fosmids is that they can accommodate longer inserts (30–40 kb) than can common plasmid vectors (< 5 kb) (Entcheva *et al*., 2001; Daniel, 2005). As catabolic operons (gene clusters) for aromatic compounds often range from 10 to 30 kb, we used a fosmid so that we could clone whole-gene clusters.

Functional screening of a metagenomic fosmid library

We screened the metagenomic library for EDO activity using catechol as a substrate. The entire screening process involved growing fosmid-containing *E. coli* in liquid medium, extracting proteins from the cells, mixing the proteins with catechol and detecting enzymatic activity by monitoring the yellow colour of the product, 2-hydroxymuconate semialdehyde. This screen identified 85 positive wells. Although the 'catechol spray method' is convenient and therefore frequently used in microbial screening for aromatic degradation pathways, this method did not work efficiently in our metagenomic library screening.

A secondary screen was carried out to isolate positive clone(s) from the each positive well containing 100 mixed clones. From each well, an aliquot was taken and cultured on agar plates. Several hundred colonies were then randomly picked and screened for EDO activity again using the 96-well format. If multiple clones were obtained from a single well, the identities of each clone were checked by restriction fragment length polymorphism analysis. Out of 85 wells, six wells (2C5, 4C2, 4E12, 5B2, 8G4 and 10D8) had two different positive clones, and the rest had one positive clone each; thus a total of 91 positive fosmid clones were obtained.

Substrate specificities of the 91 positive fosmids (i.e. whether the inserts functioned as a whole operon) were tested using catechol and its derivatives (3-methylcatechol, 4-methylcatechol, 4-chlorocatechol and 2,3-dihydroxybiphenyl) (Fig. 1). All clones developed yellow colour with catechol, the substrate used in the initial screening. Three clones (2H2, 5F10 and 9B1) showed activity with all substrates, and most clones had different activities with different derivatives. In the case of catechol, most of the fosmid clones had a yellow colour after 1 h, which subsequently disappeared as the yellow product was degraded by the action of a subsequent enzymatic transformation. On the other hand, most of the positive fosmid clones transformed 2,3-dihydroxybiphenyl to the corresponding meta-cleavage yellow compound (2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid) after prolonged incubation (16 h). Based on substrate specificity, 38 fosmids were randomly selected and completely sequenced using the shotgun method.

Sequence analysis of the metagenome EDO genes

Based on their amino acid sequences, Eltis and Bolin (1996) classified EDOs into two families, type I and II. The 35 known EDOs obtained by microbial screening (as of 1996) were classified this way. Type I enzymes are homologous to each other; the majority of them are classified into subfamilies I.2.A and I.3.A, which include catechol 2,3-dioxygenase and 2,3-dihydroxybiphenyl dioxygenase respectively. There are fewer type II enzymes; they lack sequence similarity to each other and to type I enzymes. DNA sequencing revealed at least one

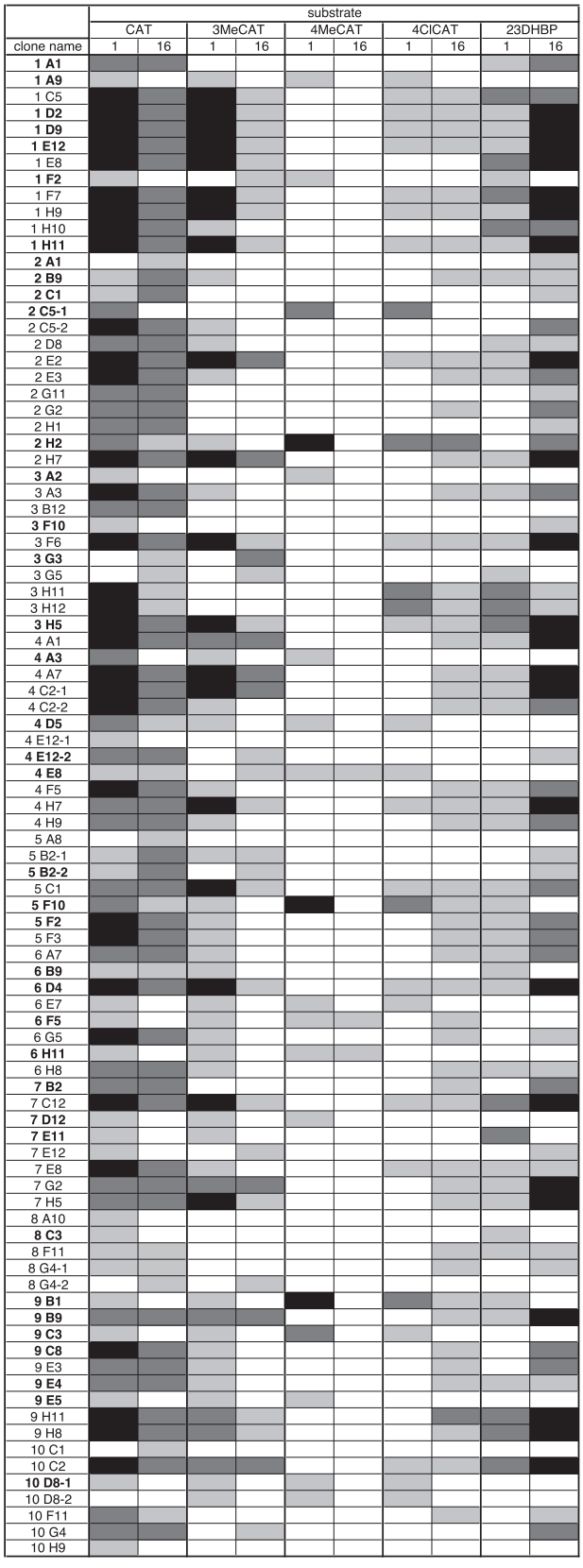

Fig. 1. Formation of meta-cleavage yellow compounds from catecholic compounds by EDOs expressed in *E. coli*. A metagenomic library of fosmid clones was screened using catechol as a substrate. The activity level is shown in grey scale: no colour, no yellow compound formation; light grey, OD < 0.5; medium grey, OD 0.5–1.0; dark grey, OD > 1.0 at the corresponding absorption maximum. Values represent the average of three independent experiments. Thirty-eight fosmids used for subsequent DNA sequencing are shown in bold. CAT, catechol; 3MeCAT, 3-methylcatechol; 4MeCAT, 4-methylcatechol; 4ClCAT, 4-chlorocatechol; 23DHBP, 2,3-dihydroxybiphnyl.

type I EDO gene in each fosmid (Table 1). Other genes, such as phenol hydroxylases, were found in the flanking regions. Five fosmids, S-2C5, S-3F10, S-5F10, S-6B9 and S-9E4, contained two EDO genes, to yield a total of 43 genes (40 full-length and 3 partial) from the 38 sequenced clones.

Many rhodococcal strains have multiple EDO genes, which may account for the unique ability of this bacterial group to degrade various aromatic compounds (Asturias and Timmis, 1993; Maeda *et al*., 1995). The metagenomic library used here was likely to contain DNA from rhodococci, as demonstrated by the 16S rDNA sequencing results (classified as *Actinobacteria*); thus, the 2C5, 3F10, 5F10, 6B9 and 9E4 clones may be derived from rhodococci. Our failure to identify any type II EDO genes probably reflects their scarcity in the environment, as reported by Eltis and Bolin (1996). Further sequence analysis suggested that some of the clones we isolated contained microbial DNA from unculturable (or yet to be cultured) species, and this evidence is presented below.

The phylogenetic relationship of 45 metagenomic and several representative EDOs obtained by microbial screening is shown in Fig. 2. Eight EDOs (3F-10, 9E4-2 partial, 9B1, 6B9-2, 2H2, 5F10-2, 3A2 and 2C5-2) belonging to subfamily I.2.A were identified in our screen; this subfamily counts as members most of the EDOs previously identified using agar plate-based screening. However, our metagenomic approach also identified 37 additional EDOs belonging to other subfamilies, including EDOs which belong to four newly defined subfamilies: I.1.C, I.2.G, 1.3.M and I.3.N. Approximately half of these EDOs (20) were in subfamily I.2.G, which branched at a deep point in the lineage. These families had not previously been identified using cultivation-based screening and were thus missing on the phylogenetic tree, supporting the idea that the metagenomic library used here contained genes from a wide variety of microorganisms that may be difficult to cultivate and which have not been well studied.

Functional analyses of metagenomic EDOs

Next, functional analysis of the metagenomic EDOs was conducted. Clones carrying full-length EDO genes that

Table 1. List of predicted EDOs obtained by metagenomic library screening.

Name	EDO subfamily	Length (aa)	Most similar protein	Host organism	AA identity (%)	Bacterial division
S-1A1	I.2.G	329	Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-1A9	I.2.B	309	Catechol 2,3-dioxygenase	*Sphingomonas* sp. A8AN3	176/310 (56)	α-*Proteobacteria*
S-1D2	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-1D9	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-1E12	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-1F2	I.3.N	297	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bordetella avium* 197N	169/292 (57)	β-*Proteobacteria*
S-1H11	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-2A1	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-2B9	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	157/308 (51)	*Firmicutes*
S-2C1	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	157/308 (51)	*Firmicutes*
S-2C5-1		165[a]	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	92/167 (55)	*Firmicutes*
S-2C5-2	I.2.A	307	Catechol 2,3-dioxygenase	*Acinetobacter* sp. YAA	213/307 (69)	*Actinobacteria*
S-2H2	I.2.A	308	Catechol 2,3-dioxygenase	*Ralstonia* sp. KN1	182/310 (59)	β-*Proteobacteria*
S-3A2	I.2.A	308	Catechol 2,3-dioxygenase	*Ralstonia* sp. KN1	195/311 (63)	β-*Proteobacteria*
S-3F10-1	I.2.A	309	Catechol 2,3-dioxygenase	*Delftia tsuruhatensis*	305/308 (99)	β-*Proteobacteria*
S-3F10-2	I.2.C	315	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida*	295/314 (94)	γ-*Proteobacteria*
S-3G3		324[a]	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-3H5	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-4A3	I.1.C	176	Putative dioxygenase	*Bordetella avium* 197N	61/134 (45)	β-*Proteobacteria*
S-4D5	I.2.B	309	Catechol 2,3-dioxygenase	*Sphingomonas* sp. A8AN3	176/310 (57)	α-*Proteobacteria*
S-4E8	I.2.C	315	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida*	294/314 (94)	γ-*Proteobacteria*
S-4E12	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-5B2	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-5F2	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-5F10-1	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-5F10-2	I.2.A	308	Catechol 2,3-dioxygenase	*Ralstonia* sp. KN1	182/310 (59)	β-*Proteobacteria*
S-6B9-1	I.2.A	308	Catechol 2,3-dioxygenase	*Ralstonia* sp. KN1	182/310 (59)	β-*Proteobacteria*
S-6B9-2	I.1.C	176	Putative dioxygenase	*Bordetella avium* 197N	61/134 (45)	β-*Proteobacteria*
S-6D4	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-6F5	I.2.B	309	Catechol 2,3-dioxygenase	*Sphingomonas* sp. A8AN3	176/310 (57)	α-*Proteobacteria*
S-6H11	I.2.C	315	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida*	294/314 (94)	γ-*Proteobacteria*
S-7B2	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-7D12	I.2.C	315	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida*	295/314 (94)	γ-*Proteobacteria*
S-7E11	I.3.M	314	Dihydroxybiphenyl dioxygenase	*Terrabacter* sp. DPO360	126/294 (42)	*Actinobacteria*
S-8C3	I.3.M	314	Dihydroxybiphenyl dioxygenase	*Terrabacter* sp. DPO360	126/294 (42)	*Actinobacteria*
S-9B1	I.2.A	308	Catechol 2,3-dioxygenase	*Ralstonia* sp. KN1	182/310 (59)	β-*Proteobacteria*
S-9B9	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-9C3	I.2.C	315	Catechol 2,3-dioxygenase	*Pseudomonas* sp. K82	288/314 (92)	γ-*Proteobacteria*
S-9C8	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-9E4-1	I.2.G	329	2,3-Dihydroxybiphenyl 1,2-dioxygenase	*Bacillus* sp. JF8	156/308 (51)	*Firmicutes*
S-9E4-2		219[a]	Catechol 2,3-dioxygenase	*Ralstonia* sp. KN1	59/79 (75)	β-*Proteobacteria*
S-9E5	I.2.C	315	3-Methylcatechol 2,3-dioxygenase	*Pseudomonas putida*	295/314 (94)	γ-*Proteobacteria*
S-10D8	I.2.B	309	Catechol 2,3-dioxygenase	*Sphingomonas* sp. A8AN3	176/310 (57)	α-*Proteobacteria*

a. Partial sequence; the complete EDO gene sequence was not obtained by shotgun sequencing.
aa, amino acid.

were previously generated for DNA shotgun sequencing were utilized (Table 2). Because 3-methylcatechol dioxygenase (TodE), a well-characterized I.3.B enzyme from *Pseudomonas putida* F1 (Zylstra and Gibson, 1989), cleaves all five substrates tested, the substrate specificities of the metagenomic EDOs were compared with TodE activity. Extradiol dioxygenases in family I.2 are known to preferentially cleave monocyclic aromatic compounds such as catechol (Eltis and Bolin, 1996). As for the metagenomic clones (S-2C5-3, S-2H2 and S-3A2 in I.2.A; S-1A9 and S-10D8 in I.2.B; S-6H11 and S-9C3 in I.2.C), they also preferentially reacted with monocyclic compounds. Exceptions to this were the enzymes in subfamily I.2.G, which displayed nearly equal reactivity to catechol and 2,3-dihydroxybiphenyl. The preference (3-methylcatechol > catechol > 2,3-dihydroxybiphenyl > 4-methylcatechol > 4-chlorocatechol) was different from that of their most similar homologue, 2,3-dihydroxybiphenyl 1,2-dioxygenase from *Bacillus* sp. JF8 in subfamily I.2.F (Hatta *et al*., 2003), which prefers substrates in the following order: 2,3-dihydroxybiphenyl > 3-methylcatechol > catechol > 4-methylcatechol > 4-chlorocatechol.

Enzymes in family I.3 prefer polycyclic aromatic compounds (Eltis and Bolin, 1996). EDO S-7E11, belonging to subfamily I.3.M, exhibited specificity similar to that of 2,3-dihydroxybiphenyl dioxygenase from *Pseudomonas pseudoalcaligenes* KF707 (Taira *et al*., 1992) in subfamily I.3.A: both reacted with 2,3-dihydroxybiphenyl, but were

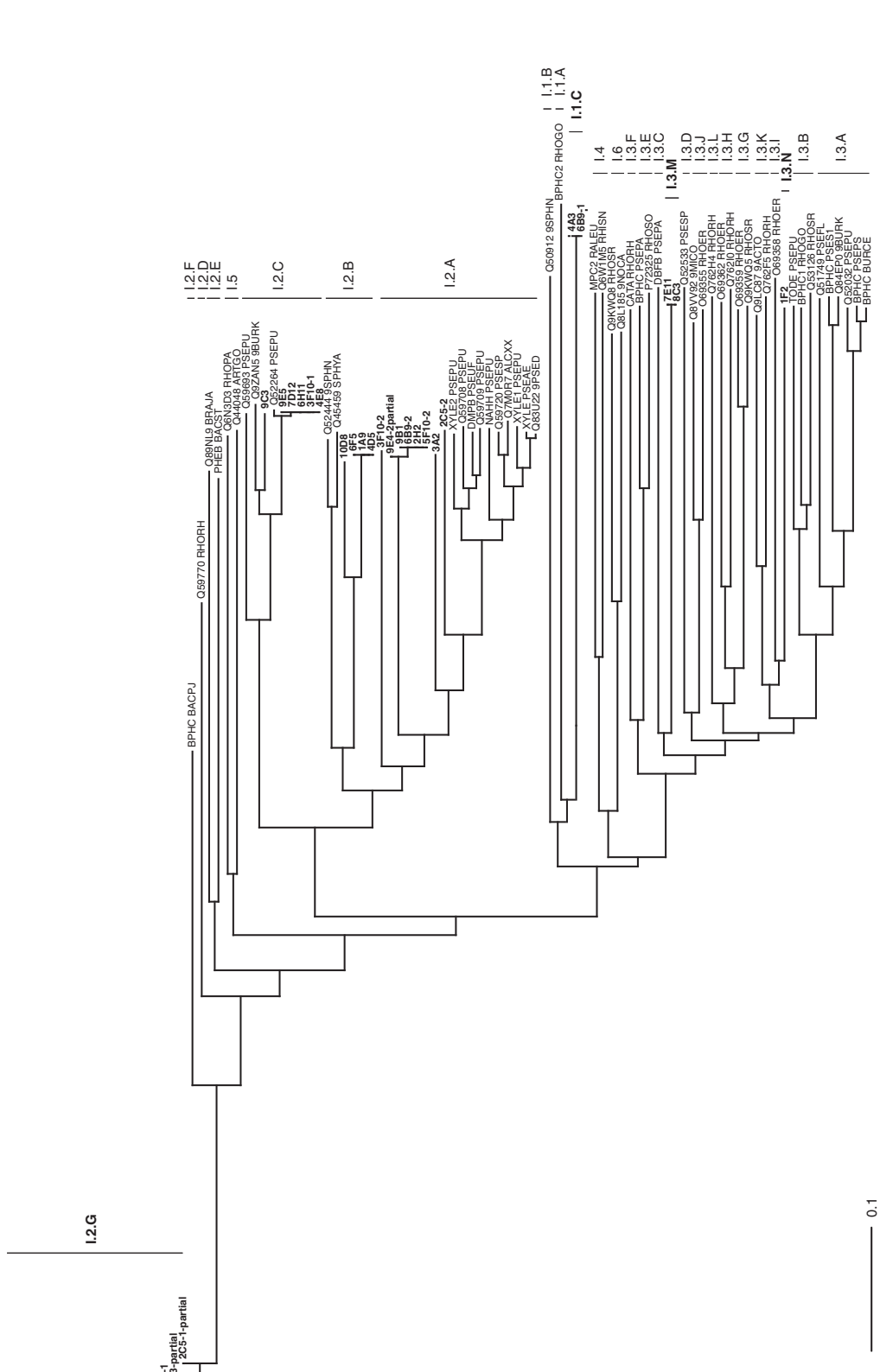

Fig. 2. A phylogenetic tree showing both metagenomic EDOs and previously identified type I EDOs. The metagenomic clones isolated in this study are indicated in bold. Enzymes belonging to same subfamily share > 54% amino acid sequence identity in pairwise alignments (Eltis and Bolin, 1996). The accession numbers of the previously identified EDOs are as follows: BPHC BACPJ, Q8GR45; Q59770 RHORH, Q59770; PHEB BACST, P31003; Q89NL9 BRAJA, Q89NL9; Q59693 PSEPU, Q59693; Q9ZAN5 9BURK, Q9ZAN5; Q52264 PSEPU, Q52264; Q52444 9SPHN, Q52444; Q45459 SPHYA, Q45459; XYLE2 PSEPU, Q04285; Q59708 PSEPU, Q59708; DMPB PSEUF, P17262; Q59709 PSEPU, Q59709; NAHH PSEPU, P08127; Q59720 PSESP, Q59720; Q7M0R7 ALCXX, Q7M0R7; XYLE1 PSEPU, P06622; XYLE PSEAE, P27887; Q83U22 9PSED, Q83U22; Q6N3D3 RHOPA, CGA009; Q44048 ARTGO, Q44048; Q50912 9SPHN, Q50912; MPC2 RALEU, P17296; Q6W1M5 RHISN, Q6W1M5; Q9KWQ8 RHOSR, Q9KWQ8; Q8L185 9NOCA, Q8L185; BPHC2 RHOGO, P47232; DBFB PSEPA, P47243; CATA RHORH, Q53034; BPHC PSEPA, P11122; P72325 RHOSO, P72325; Q52533 PSESP, Q52533; Q8VV92 9MICO, Q8VV92; O69355 RHOER, O69355; Q762H4 RHORH, Q762H4; O69362 RHOER, O69362; Q76210 RHORH, Q76210; O69359 RHOER, O69359; Q9KWQ5 RHOSR, Q9KWQ5; Q9LC87 9ACTO, Q9LC87; Q762F5 RHORH, Q762F5; O69358 RHOER, O69358; TODE PSEPU, P13453; BPHC1 RHOGO, P47231; Q53126 RHOSR, Q53126; Q51749 PSEFL, Q51749; BPHC PSES1, P17297; Q84EP0 9BURK, Q84EP0; Q52032 PSEPU, Q52032; BPHC PSEPS, P08695; BPHC BURCE, P47228.

Table 2. Activities of metagenomic EDOs with various catecholic substrates.

		Relative activity (%)[a]				
Clone	Subfamily	CAT	3MeCAT	4MeCAT	4ClCAT	23DHBP
TodE	I.3.B	100	210	13	19	63
S-4A3	I.1.C	12	6.2	1.4	< 1.0	< 1.0
S-2C5-3	I.2.A	228	30	40	30	21
S-2H2	I.2.A	387	145	201	195	25
S-3A2	I.2.A	2.6	1.5	1.3	1.2	< 1.0
S-1A9	I.2.B	147	71	97	85	6.7
S-10D8	I.2.B	99	50	81	77	3.0
S-6H11	I.2.C	31	19	6.3	7.1	1.8
S-9C3	I.2.C	439	322	169	121	15
S-1A1	I.2.G	22	36	2.7	1.9	8.6
S-1D2	I.2.G	58	95	8.3	4.5	18
S-2A1	I.2.G	20	33	1.7	< 1.0	7.4
S-7B2	I.2.G	16	25	1.2	1.7	6.3
S-7E11	I.3.M	16	19	5.3	4.1	140
S-1F2	I.3.N	74	65	40	37	134

a. Activities relative to TodE activity for catechol (144.89 U per gram of protein) are shown for cloned EDOs. Values are the average from three independent experiments. CAT, catechol; 3MeCAT, 3-methylcatechol; 4MeCAT, 4-methylcatechol; 4ClCAT, 4-chlorocatechol; 23DHBP, 2,3-dihydroxybiphenyl.

nearly inactive with the other substrates tested. S-1F2, belonging to subfamily I.3.N, showed activity with all substrates tested, similar to TodE. Subfamily I.1 consists of short EDOs that are approximately half the size of other EDOs. These EDOs include BphC (I.1.B, Q50912_9SPIN in Fig. 2) from *Sphingomonas xenophaga* BN6 (Heiss *et al*., 1995); BphC2 (I.1.A, BPHC2_RHOGO in Fig. 2) and BphC3 (I.1.A, not shown in Fig. 2) from *Rhodococcus globerulus* P6 (Asturias and Timmis, 1993); and BphC2 (I.1.A, not shown in Fig. 2) from *Rhodococcus erythropolis* TA421 (Maeda *et al*., 1995). Although these enzymes all prefer 2,3-dihydroxybiphenyl, metagenomic EDO S-4A3 (I.1.C) hardly reacted with the substrate, indicating that this enzyme evolved differently from the other short EDOs and acquired distinct substrate specificity.

Gene evolution of I.2.G EDOs

It is known that minor changes in primary EDO sequence can result in marked changes in catalytic activity (Ramos *et al*., 1987). As demonstrated by directed evolution experiments and rational protein design, single-amino-acid substitutions can drastically change the substrate- and regio-specificities of aromatic hydrocarbon dioxygenases (Suenaga *et al*., 2002; Furukawa *et al*., 2004). Naturally occurring as well as engineered single-amino-acid changes have also been found to influence catalytic properties. For example, an EDO belonging to subfamily I.2.A, which has Tyr at amino acid position 218, has a low turnover number and high affinity for catechol substrates. The enzyme is found in both slightly and highly contaminated sites. In contrast, a variant having His at position 218 has a high turnover number and low affinity, and is found only in highly contaminated sites. These observations imply that specific EDO genes have been positively selected in response to aromatic compound contamination (Junca *et al*., 2004).

In this study, the I.2.G subfamily was highly represented, and detailed analysis of these enzymes shed light on EDO gene evolution. The genes in this subfamily were not identical to each other; 18 genes (not counting the two partial genes) carried eight single-nucleotide polymorphisms (SNPs) and were further classified into six groups. Based on the combinations of the SNPs as well as on their flanking sequences, an evolutionary lineage was constructed (Fig. 3). It is evident that these genes evolved from a common ancestor (group 1) and diverged through accumulations of single-base mutations (groups 2–6) and deletion of the 3′-terminal region of a 3′-flanking gene (S-1D2, S-1D9, S-1E12, S-1H11, S-3H5, S-6D4 and S-9B9 in group 2).

This analysis gave us a ‘snapshot’ of dynamic gene evolution in an ecosystem, and is the first time such a result has been obtained using a metagenomic approach. By carefully comparing the lineage and the enzymatic activities of some selected I.2.G clones (Table 2), we found that the residue at amino acid position 241 is key for determining protein activity. S-1D2 (group 2), which had Thr241, exhibited higher activity than rest of the enzymes (S-1A1 in group 1, S-2A1 in group 5 and S-7B2 in group 3) that had Ala at position 241. Therefore, the Ala241 to Thr241 mutation most likely occurred during adaptation to contamination conditions. The crystal structure of a homologous homoprotocatechuate 2,3-dioxygenase (from *Arthrobacter globiformis*; MndD) has been solved (Vetting *et al*., 2004). The amino acid sequence of MndD

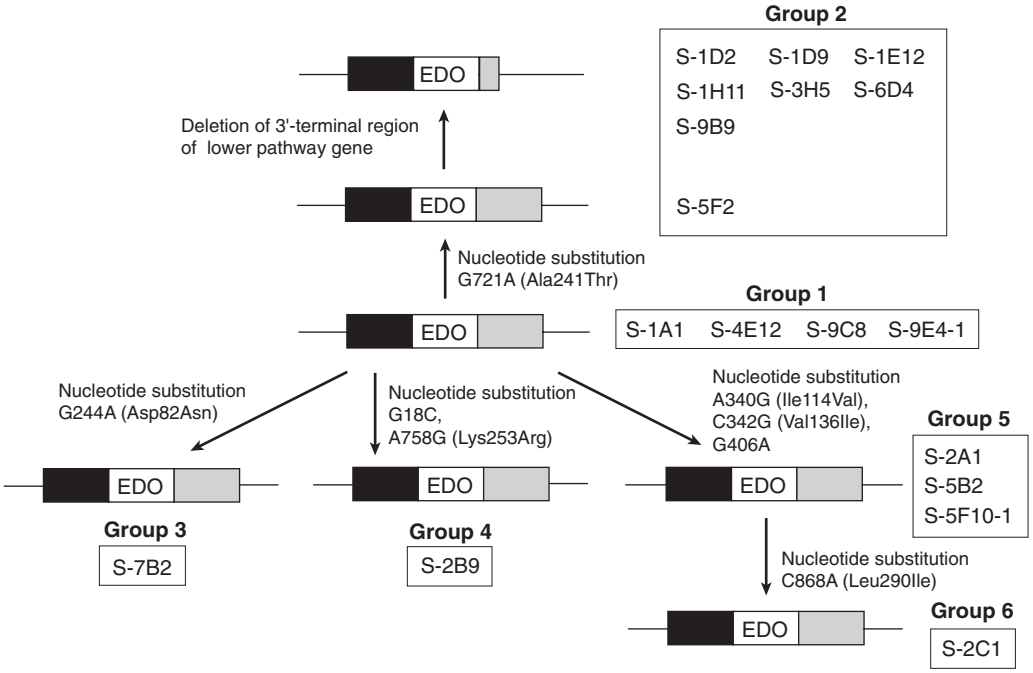

Fig. 3. Proposed evolutionary lineage of metagenomic I.2.G EDOs. The black squares represent genes for upper pathways (encoding the putative hydroxylase component) and the grey squares represent genes for lower pathways (encoding the putative hydrolase).

is about 30% identical to S-1D2. In MndD, Gly233 corresponds to Thr241 of S-1D2, and is located on the surface of the protein, far from the active site. Thus, it is not clear why a residue at this position would influence the EDO's activity.

In summary, by conducting a metagenomic screen for catechol degradation pathways, we identified 91 EDO genes. Of these, 38 clones were sequenced and we identified 43 EDO genes, including 25 belonging to four new subfamilies. In addition, we captured a snapshot of gene evolution. These results suggest that this approach may prove helpful for analysing and developing bioremediation processes.

Experimental procedures

Activated sludge and DNA extraction

Activated sludge was collected from the aeration tank of a coke plant wastewater treatment plant in Japan. The sludge sample (50 mg) was re-suspended in 1 ml of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), followed by dispersion of the microbial aggregates by sonification with the Handy Sonic UR-20P (Tomy Seiko, Tokyo, Japan) at maximum power for 15 s. Total genomic DNA was purified as described previously (Gabor *et al*., 2003).

Cloning and DNA sequencing of 16S rRNA gene

Polymerase chain reaction amplification of the 16S rRNA genes was performed using purified metagenomic DNA as a template and bacterial universal primers (Lane, 1991). Polymerase chain reaction products were ligated into a pGEM-Teasy vector (Promega) transformed into competent *E. coli* DH5α cells. A total of 90 clones were randomly selected and sequenced. Sequencing was performed with primer 907R (5′-CCGYCAATTCMTTTRAGTTT-3′) where Y is C or T, M is A or C and R is A or G, a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems), and an ABI3730XL (Applied Biosystems). All sequences (500–600 nucleotides) were compared with those in the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST) using the BLAST program. To detect and omit chimeric DNAs, the CHECK-CHIMERA program of the Ribosomal Database Project (Maidak *et al*., 1997) was used.

Construction of the metagenomic fosmid library

A metagenomic library was constructed using a Copy-Control Fosmid Library Production Kit (Epicentre). Metagenomic DNA was first separated by pulsed field gel electrophoresis and 33–48 kb DNA fragments were excised, purified, blunt-ended and ligated into the pCC1FOS fosmid included in the kit. The ligation mixture was then packaged into lambda phage using MaxPlax Lambda Packaging Extracts. *Escherichia coli* EPI-300 T1^R was then transformed with the packaged library, and transformants were selected on LB agar plates containing 12.5 μg ml^{-1} chloramphenicol (LB/Cm). Colonies were picked and re-suspended in separated wells of 96-well plates containing 200 μl of LB/Cm; each well contained 100 independent clones. A total of 96 000 colonies were used to construct a library in 10 96-well plates.

Screening the metagenomic fosmid library for EDO activity

A small aliquot of cells from each well of the metagenomic library was transferred to a second set of 96-well plates containing 1 ml of 2xYT plus 12.5 μg ml^{-1} chloramphenicol (2xYT/Cm) and incubated at 37°C for 18 h without agitation. Subsequently, 200 μl of the overnight culture was transferred to fresh 96-well plates containing 800 μl of 2xYT/Cm, grown at 37°C for 30 min with mild agitation at 250 r.p.m. (Taitec, BioShaker M-BR-024), and 1 μl of CopyControl Induction Solution (Epicentre) was added. The solution increases the copy number of fosmids in *E. coli*, which may promote protein expression in the cells. Cells were grown with vigorous agitation (1200 r.p.m.) at 37°C for 2 h, harvested by centrifugation (3000 r.p.m., 15 min, 4°C), re-suspended in 50 mM phosphate buffer (pH 7.5) and disrupted by addition of 150 μl of BugBuster Plus Benzonase Nuclease (Novagen). For each sample, cell debris was removed by centrifugation (3000 r.p.m., 15 min, 4°C) and the supernatant transferred to a clean plate. Substrate (5 μl of catechol; 0.5 mM final) was added to 100 μl of each cell extract, and the reaction plates were incubated with mild agitation (250 r.p.m.) at 25°C. Positive clones were identified after incubation for 1 h or 16 h.

Isolation of single positive fosmid clones from positive wells

As each well comprised 100 different fosmid clones, a second screen was carried out to isolate single positive fosmid(s) from the positive wells. A small sample of culture was taken from the positive wells, diluted and cultured on LB/Cm agar plates. Several hundred colonies were then screened for EDO activity in the same manner as for the first screen. When several colonies were picked from the same wells, they were further subjected to restriction fragment length polymorphism analysis using EcoRI and PstI.

DNA sequencing and gene annotation

Shotgun DNA sequencing of positive fosmid clones was performed at Dragon Genomics Center, Takara Bio. The resultant nucleotide sequences were used for similarity searches with BLASTX software and the NCBI database using the default parameters. Annotation was carried out based on the BLASTX results. Multiple sequence alignment was carried out using CLUSTALX software, and a phylogenic tree was generated by the neighbour-joining method (Saitou and Nei, 1987). In the classification system for type I EDOs, enzymes belonging to the same subfamily exhibit greater than 54% amino acid identity in pairwise comparisons (Eltis and Bolin, 1996).

Enzyme activity assays

Cells extracts were prepared from *E. coli* EPI-300 $T1^R$ cells carrying positive fosmids as described above. Enzyme activities were determined by the formation of ring meta-cleavage products using 0.5 mM catechol or related compounds. After 1 h or 16 h, absorbance, as a measure of enzymatic activity, was measured on a Molecular Devices plate reader (VERSA max). Plasmids containing various EDO genes in pUC118 were obtained in shotgun sequencing experiments, and were introduced into *E. coli* JM109. Recombinant *E. coli* were then grown in 96-well plates containing 1 ml of LB medium, 50 μg ml^{-1} ampicillin and 0.1 mM isopropyl β-D-thiogalactoside. Enzymes were extracted from the cells in the same manner as for fosmid-carrying *E. coli*, and activities were determined using 0.5 mM catechol and related compounds. The relative ring-cleavage activities were determined from the absorption coefficients of the ring-cleavage product formed from the following substrate: catechol, 375 nm ($\varepsilon = 33\,000\ M^{-1}cm^{-1}$); 3-methyl catechol, 388 nm ($\varepsilon = 13\,400\ M^{-1}cm^{-1}$); 4-methyl catechol, 379 nm ($\varepsilon = 28\,100\ M^{-1}cm^{-1}$); 4-chloro catechol, 379 nm ($\varepsilon = 36\,100\ M^{-1}cm^{-1}$); and 2,3-dihydroxy biphenyl, 434 nm ($\varepsilon = 13\,200\ M^{-1}cm^{-1}$) respectively (Hirose *et al*., 1994). One unit of enzyme activity was defined as the amount of enzyme that catalysed the formation of 1 μmol of the product per minute. Protein concentration was determined using Quick Start Bradford Dye Reagent (Bio-Rad) and bovine gamma globulin for the standard curve.

Nucleotide sequences accession number

The 38-nucleotide sequences reported in this article have been deposited in GenBank/EMBL/DDBJ under Accession No. AB266111 to AB266148.

References

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143–169.

Asturias, J.A., and Timmis, K.N. (1993) Three different 2,3-dihydroxybiphenyl 1,2-dioxygenase genes in the gram-positive polychlorobiphenyl-degrading bacterium *Rhodococcus globerulus* P6. *J Bacteriol* **175:** 4631–4640.

Bond, P.L., Hugenholtz, P., Keller, J., and Blackall, L.L. (1995) Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl Environ Microbiol* **61:** 1910–1916.

Cottrell, M.T., Moore, J.A., and Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* **65:** 2553–2557.

Daims, H., Nielsen, J.L., Nielsen, P.H., Schleifer, K.H., and Wagner, M. (2001) *In situ* characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* **67:** 5273–5284.

Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* **3:** 470–478.

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

Entcheva, P., Liebl, W., Johann, A., Hartsch, T., and Streit, W.R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* **67:** 89–99.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A., *et al.* (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* **7:** 1996–2010.

Fortin, P.D., MacPherson, I., Neau, D.B., Bolin, J.T., and Eltis, L.D. (2005) Directed evolution of a ring-cleaving dioxygenase for polychlorinated biphenyl degradation. *J Biol Chem* **280:** 42307–42314.

Furukawa, K., Suenaga, H., and Goto, M. (2004) Biphenyl dioxygenase: functional versatilities and directed evolution. *J Bacteriol* **186:** 5189–5196.

Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol Ecol* **44:** 153–163.

Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68:** 669–685.

Hatta, T., Mukerjee-Dhar, G., Damborsky, J., Kiyohara, H., and Kimbara, K. (2003) Characterization of a novel thermostable Mn(II)-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase from a polychlorinated biphenyl- and naphthalene-degrading *Bacillus* sp. JF8. *J Biol Chem* **278:** 21483–21492.

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.

Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66:** 3113–3116.

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 Microbiol Lett* **118:** 273–278.

Junca, H., Plumeier, I., Hecht, H.J., and Pieper, D.H. (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. *Microbiol* **150:** 4181–4187.

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

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and Daniel, R. (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* **69:** 1408–1416.

Lane, D.J. (1991) *Nucleic Acid Techniques in Bacterial Systematics.* New York, NY, USA: John Wiley & Sons.

Maeda, M., Chung, S.Y., Song, E., and Kudo, T. (1995) Multiple genes encoding 2,3-dihydroxybiphenyl 1,2-dioxygenase in the gram-positive polychlorinated biphenyl-degrading bacterium *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem. *Appl Environ Microbiol* **61:** 549–555.

Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J., and Woese, C.R. (1997) The RDP (Ribosomal Database Project). *Nucleic Acids Res* **25:** 109–111.

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

Ramos, J.L., Wasserfallen, A., Rose, K., and Timmis, K.N. (1987) Redesigning metabolic routes: manipulation of TOL plasmid pathway for catabolism of alkylbenzoates. *Science* **235:** 593–596.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4:** 406–425.

Smith, M.R. (1990) The biodegradation of aromatic hydrocarbons by bacteria. *Biodegradation* **1:** 191–206.

Snaidr, J., Amann, R., Huber, I., Ludwig, W., and Schleifer, K.-H. (1997) Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. *Appl Environ Microbiol* **63:** 2884–2896.

Stamoudis, V.C., and Luthy, R.G. (1980) Determination of biological removal of organic constituents in quench waters from high-BTU coal-gasification pilot plants. *Water Res* **14:** 1143–1156.

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.

Taira, K., Hirose, J., Hayashida, S., and Furukawa, K. (1992) Analysis of *bph* operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenses* KF707. *J Biol Chem* **267:** 4844–4853.

Vaillancourt, F.H., Bolin, J.T., and Eltis, L.D. (2004) Ring-cleavage dioxygenases, In *Pseudomonas*, Vol. 3. Ramos, J.-L. (ed.). New York, NY, USA: Kluwer Academic/Plenum Publishers, pp. 359–395.

Vetting, M.H., Wackett, L.P., Que, J., Jr, Lipscomb, J.D., and Ohlendorf, D.H. (2004) Crystallographic comparison of manganese- and iron-dependent homoprotocatechuate 2,3-dioxygenases. *J Bacteriol* **186:** 1945–1958.

Wagner, M., and Loy, A. (2002) Bacterial community composition and function in sewage treatment systems. *Curr Opin Biotechnol* **13:** 218–227.

Yun, J., Kang, S., Park, S., Yoon, H., Kim, M.J., Heu, S., and Ryu, S. (2004) Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl Environ Microbiol* **70:** 7229–7235.

Zylstra, G.J., and Gibson, D.T. (1989) Toluene degradation by *Pseudomonas putida* F1: nucleotide sequence of the todC1C2BADE genes and their expression in *Escherichia coli*. *J Biol Chem* **264:** 14940–14946.

