Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes

Taku Uchiyama[1], Takashi Abe[2,3], Toshimichi Ikemura[2] & Kazuya Watanabe[1]

Recent awareness that most microorganisms in the environment are resistant to cultivation has prompted scientists to directly clone useful genes from environmental metagenomes[1]. Two screening methods are currently available for the metagenome approach, namely, nucleotide sequence–based screening[2] and enzyme activity–based screening[3]. Here we have introduced and optimized a third option for the isolation of novel catabolic operons, that is, substrate-induced gene expression screening (SIGEX). This method is based on the knowledge that catabolic-gene expression is generally induced by relevant substrates and, in many cases, controlled by regulatory elements situated proximate to catabolic genes. For SIGEX to be high throughput, we constructed an operon-trap *gfp*-expression vector available for shotgun cloning that allows for the selection of positive clones in liquid cultures by fluorescence-activated cell sorting. The utility of SIGEX was demonstrated by the cloning of aromatic hydrocarbon–induced genes from a groundwater metagenome library and subsequent genome-informatics analysis.

Current metagenome approaches use either nucleotide sequence–based screening or enzyme activity-based screening for isolation of catabolic genes[1]. Nucleotide sequence–based screening is highly efficient when PCR is used for direct retrieval of target DNA fragments[2]. However, because PCR primers have been designed by comparing a limited number of catabolic genes that were cloned from easily culturable laboratory isolates, PCR screening can access only homologs of known genes in a diverse, environmental, catabolic-gene pool. This is also the case for hybridization-based methods. In enzyme activity–based screening, a shotgun library is constructed and used to screen for clones expressing a desired enzymatic activity[3]. It has, however, been observed that some enzymes are difficult to express as active forms in cloning hosts[4] (e.g., in *Escherichia coli*). In addition, this method is generally laborious; a handful of clones expressing a desired enzymatic activity has to be selected from a vast quantity of colonies in a metagenome library (e.g., 4 from 930,000 (ref. 3)). Considering the limitations associated with the current screening methods, we have introduced SIGEX for the screening of

environmental metagenome libraries for isolating novel catabolic genes (see **Fig. 1**).

SIGEX is based on the knowledge that catabolic gene expression is generally induced by substrates and metabolites of catabolic enzymes and, in many cases, is controlled by regulatory elements situated proximate to catabolic genes. Thus, catabolic genes expressed constitutively and those with transcriptional regulators that are distantly situated cannot be obtained. We expect that a wide range of substrates that can migrate into the cytoplasm of a cloning host are available for SIGEX screening. In addition, it has been reported that transcriptional regulators of a broad range of bacteria, including *Bacillus subtilis*[5] and *Agrobacterium tumefaciens*[6], can work in concert with transcriptional machineries in the cloning host (that is, *E. coli*), which led us to believe that catabolic operons of a broad range of bacteria may be obtained by SIGEX.

To make SIGEX a high-throughput process, we constructed an operon-trap *gfp*-expression vector (designated p18GFP) suitable for shotgun cloning and used this in combination with fluorescence-activated cell sorting (FACS) for high-throughput selection of positive clones in liquid cultures. The total procedure of the SIGEX scheme is presented in **Figure 1**, which comprises four steps: (i) construction of a metagenomic library in liquid culture, (ii) removal of clones containing self-ligation plasmids and those expressing green fluorescent protein (GFP) constitutively, (iii) selection by sorting of clones expressing GFP in the presence of a target substrate, and (iv) colony isolation of the sorted cells on agar plates. This scheme obviates the need for forming a large number of colonies for a primary metagenome library.

A promoter-trap system has previously been harnessed for isolating genes of a pathogenic strain expressed within host cells[7]. SIGEX, on the other hand, uses the *lac* promoter for eliminating clones harboring self-ligation plasmids by FACS after induction with isopropyl-beta-D-thiogalactopyranoside (IPTG). This procedure can remove false positives and improve the efficiency for selecting positive clones in step iii. DNA fragments obtained by the SIGEX method are expected to contain a catabolic operon fragment relevant to the target compounds, that is, open reading frames (ORFs) coding for transcriptional regulator(s) and catabolic enzyme(s). When a partial operon is

---

[1]Laboratory of Applied Microbiology, Marine Biotechnology Institute, 3-75-1 Heita, Kamaishi, Iwate 026-0001, Japan. [2]Department of Population Genetics, National Institute of Genetics and the Graduate University for Advanced Studies, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. [3]Xanagen Inc., 2-7-14 Higashinakano, Nakano, Tokyo 164-0003, Japan. Correspondence should be addressed to K.W. (kazuya.watanabe@mbio.jp).
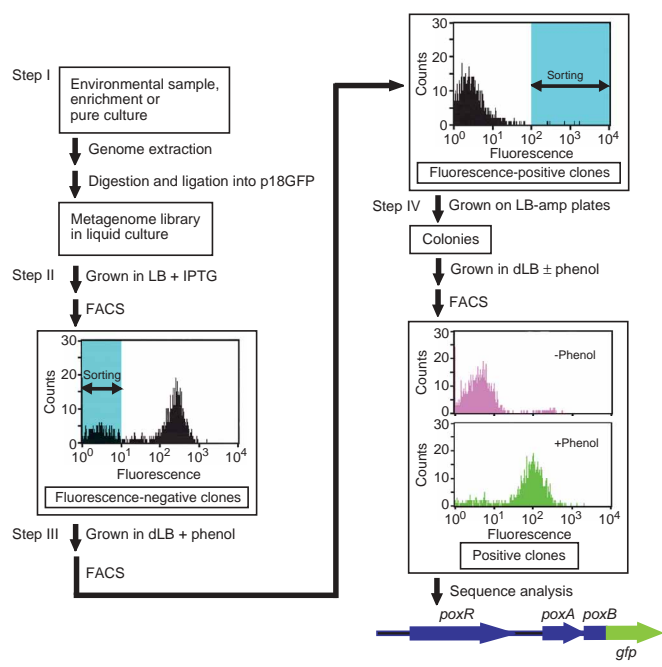
Figure 1 Schematic representation of the SIGEX scheme with an example of cloning of a phenol-degradative operon fragment from *R. eutropha* strain E2. Phenol was used as an induction substrate.

obtained, the whole operon can subsequently be retrieved by standard procedures, such as inverse PCR[8].

To develop the SIGEX method and to examine its utility, we first attempted to isolate a known phenol-degradative operon (the *pox* operon[9], NCBI accession no. AF026065) from a *Ralstonia eutropha* E2 genomic library (**Fig. 1**). In this experiment, we used a diluted Luria-Bertani (LB)-type medium in steps iii and iv because we have found that the *pox*-controlled GFP expression in *E. coli* harboring p18GFP-pox was very weak in the standard LB medium[10] supplemented with phenol. Experiments with several different types of media, including

LB-type media containing different concentrations of organic nutrients and M9 media[10] containing single carbon sources, have shown that the dLB medium plus phenol worked best to express GFP fluorescence from the *pox* transcriptional unit. Similar expression inhibitions in nutrient-rich media (exemplified by LB and nutrient broth) have been observed for other catabolic operons[11,12], suggesting the wide utility of this diluted medium in the SIGEX scheme. Among the positive clones (those expressing GFP fluorescence in the presence but not the absence of phenol) obtained in step iv, two clones were subjected to sequence analysis. We found that they contained the same *pox* fragment comprised of *poxRAB*[9] (genes for a positive transcriptional regulator and two subunits of phenol hydroxylase), demonstrating the utility of the SIGEX scheme for isolating a target catabolic-operon fragment.

An environmental metagenome library (containing approximately 152,000 clones with an average insertion size of 7 kb) was constructed from groundwater microbial flora[13]. For isolating aromatic-hydrocarbon catabolic operon fragments, the library was subjected to SIGEX screening with induction substrates of benzoate and naphthalene (benzoate-negative clones in step iii was subjected to induction with naphthalene). Benzoate was primarily used because this compound is a key intermediate metabolite in aromatic-hydrocarbon catabolic pathways and acts as an inducer for the expression of upper[14] and lower[15] operons. In addition, some benzoate derivatives, such as salicylate, are industrially important. Screening by SIGEX with the two substrates was accomplished within 4 d and yielded 62 positive clones (58 for benzoate and 4 for naphthalene).

In the screening with benzoate, the sorting ratio (a ratio of the number of fluorescence-positive cells sorted to total cells subjected to the flow sorting in step iii) was 2.3×10^{-4}, the viability (a ratio of the number of colonies formed to the fluorescence-positive cells sorted in step iv) was 0.14, whereas the positive ratio (a ratio of the number of positive clones to colonies analyzed in step iv) was 0.33. Restriction fragment length polymorphism (RFLP) analysis was conducted to determine redundancy among the positive clones, showing that benzoate-positive clones were divided into 33 types, whereas naphthalene-positive clones were divided into 2 types. A large number of RFLP types was obtained with benzoate, probably because it is a

Figure 2 RFLP types obtained by SIGEX. BZO, RFLP types obtained with benzoate; NAP, RFLP types obtained with naphthalene. Arrows indicate GFP (green), catabolic enzyme (blue), transcriptional regulator (pink), transporter protein (orange), transposase (pale blue) and other or unknown (black). Phylogenetic affiliation was analyzed by the SOM analysis[22], in which a percentage represents the confidence level; BP, *Betaproteobacteria*; GP, *Gammaproteobacteria*; AC, *Actinobacteria*; AP, *Alphaproteobacteria*; DP, *Deltaproteobacteria*. Induction efficiency was calculated by dividing GFP fluorescence of the *E. coli* clone under substrate induction by GFP fluorescence under no induction.

Name	Operon structure	Phylogenetic affiliation (%)			Induction efficiency (fold)
		1st	2nd	3rd	
BZO23		BP(96)	GP(4)		40
BZO26		GP(100)			302
BZO32		AC(96)	BP(4)		6
BZO47		GP(100)			152
BZO62		GP(66)	BP(23)		13
BZO70		AP(91)	GP(7)	BP(2)	7
BZO71		GP(87)	AP(9)	BP(4)	160
BZO135		DP(81)	AP(14)	BP(5)	11
NAP1		GP(97)	AP(1)	AC(1)	12
NAP3		BP(94)	BP(6)		53

Table 1 Results of homology search

ORF		Homologous protein (function identified)				
Name	Length (aa)	Name and function	Accession no.	Length (aa)	Host organism	Identity (aa %)
Bzo23-1	137	Probable transcriptional regulator, MarR family	AAQ60396	147	*Chromobacterium violaceum*	63
Bzo23-2	109	Dissimilatory sulfite reductase, gamma subunit	ZP_00287929	102	*Magnetococcus sp.* MC-1	42
Bzo23-3	172	Repressor of drug resistance pump, MarR family	NP_417169	176	*E. coli* K12	36
Bzo23-4	480	Fusaric acid–resistance protein FusA precursor	P24126	530	*Burkholderia cepacia*	30
Bzo23-5	423	EmrA; multidrug-resistance protein A	NP_439059	390	*Haemophilus influenzae*	39
Bzo23-6	123[a]	EmrB; multidrug-resistance protein B	NP_439058	510	*H. influenzae*	62
Bzo26-1	328	ABC-type transporter sulfonate-binding protein	BAC00972	321	*P. putida*	31
Bzo26-2	315	ABC-type transporter sulfonate-binding protein	BAC00972	321	*P. putida*	28
Bzo26-3	299	Acetyl-xylan esterase	AAC06115	335	*Streptomyces lividans*	25
Bzo26-4	295	CatR; catBC operon transcriptional regulator	P20667	289	*P. putida*	71
Bzo26-5	382	CatB; muconate lactonizing enzyme I	1307186A	375	*P. putida*	74
Bzo26-6	120	CatC; muconolactone delta-isomerase (MIase)	P00948	96	*P. putida*	77
Bzo26-7	309	CatA; catechol 1,2-dioxygenase	BAA07037	311	*P. putida*	65
Bzo26-8	324	BenR; benABC operon transcriptional activator	AAF63447	318	*P. putida*	68
Bzo26-9	298[a]	BenA; benzoate dioxigenase, alpha subunit	AAF63448	452	*P. putida*	85
Bzo32-1	242[a]	NADPH:quinone reductase	ZP_00292163	328	*Thermobifida fusca*	45
Bzo32-2	385	CshA; citrate synthase	AAM92490	492	*Dictyostelium discoideum*	54
Bzo32-3	421	Decarboxylase	BAC79026	538	*Streptomyces sp.* AW-7161	45
Bzo32-4	313	Hydrogen peroxide–inducible genes activator	AAL53818	301	*Brucella melitensis* 16M	30
Bzo32-5	443	PlsC; transmembrane phospholipid biosynthesis	NP_856155	580	*Mycobacterium bovis*	26
Bzo32-6	369[a]	AcoA; aconitase	AAN61439	398	*Emericella nidulans*	51
Bzo47-1	281[a]	Ifc3; fumarate reductase flavoprotein subunit	CAB37062	588	*Shewanella fridimarina*	23
Bzo47-2	316	LysR family transcriptional regulator	CAD44482	307	*P. stutzeri*	35
Bzo47-3	324	Electron transfer flavoprotein alpha polypeptide	NP_000117	333	*Homo sapiens*	36
Bzo47-4	262	Electron-transferring flavoprotein beta subunit	AAC31169	270	*Megasphaera elsdenii*	26
Bzo47-5	459	Ifc3; fumarate reductase flavoprotein subunit	CAB37062	588	*Sh. fridimarina*	30
Bzo47-6	278	Hydantoin racemase	Q00924	249	*Pseudomonas* SP-NS671	30
Bzo47-7	312	GstR; LysR family transcriptional regulator	AAC17939	342	*Bradyrhizobium japonicam*	36
Bzo62-1	139[a]	ATPase components of ABC transporter	ZP_00152430	559	*Dechloromonas aromatica*	92
Bzo62-2	641	Sensor histidine kinase with a receiver domain	CAE27122	863	*Rhodopseudomonas palustris*	27
Bzo62-3	315	Probable mdcF malonate transporter	NP_102819	316	*Mesorhizobium loti*	28
Bzo62-4	361	Tartrate dehydrogenase (TDH)	Q51945	365	*P. putida*	74
Bzo62-5	303	Transcriptional regulator LysR-type	BAB35931	314	*E. coli* O157:H7	54

Table 1 continued on following page

central intermediate metabolite in a variety of degradative pathways. It is also conceivable that gene-expression systems in the host may have affected the number of benzoate-positive clones, because *E. coli* is known to have some benzoate-responsive elements, for example, *merR*[16].

To examine if SIGEX-selected metagenome fragments actually contained relevant catabolic genes, we analyzed nucleotide sequences of 10 RFLP types (8 benzoate-positive and 2 naphthalene-positive RFLP types; see **Fig. 2** and **Table 1**). We found that BZO26 contained ORFs homologous to genes in benzoate-degradative[17] and catechol-degradative[18] operons. Other ORF sequences that were relevant to aromatic-hydrocarbon transformation included Bzo70-5 (related to xenobiotic reductase for nitroaromatics[19]), Bzo135-6 (related to tyrosine phenol lyase[20]) and Nap3-6 (related to hydroxylase for a ubiquinone precursor[21]). We also found ORFs homologous to genes encoding catabolic enzymes for nonaromatic substrates, for example, Bzo47-6, Bzo62-4 and Bzo71-8. These sequence analyses showed that

putative catabolic genes were enriched in the SIGEX-selected metagenome fragments; some ORFs related to genes coding for transporter proteins were also found.

Recently, a method for phylogenetic affiliation of a genome fragment by analyzing the pattern of oligonucleotide appearance has been developed[22]. We applied this method to analyze the metagenome fragments obtained by SIGEX (**Fig. 2**), and showed that they originated from various Gram-negative and Gram-positive bacteria. **Figure 2** also presents induction efficiencies ([fluorescence after induction with the substrate]/[fluorescence without induction] in step IV) for the SIGEX-selected fragments. We found that some RFLP types that originated from *Gammaproteobacteria* exhibited high induction efficiencies; this is likely, because gammaproteobacterial *E. coli* was used as the host. SIGEX could also isolate RFLP types with relatively low induction efficiencies (they mostly originated from organisms distantly related to *E. coli*). These data suggest that the diversity of metagenome fragments selected by SIGEX depends on the

Table 1 Continued

| ORF		Homologous protein (function identified)				
Name	Length (aa)	Name and function	Accession no.	Length (aa)	Host organism	Identity (aa %)
Bzo70-1	336	Permeases of the drug/metabolite transporter	ZP_00005499	307	*Rhodobacter sphaeroides*	27
Bzo70-2	359	N-acetyl quinovosamine synthesis protein	AAQ93037	309	*Rhizobium etli*	24
Bzo70-3	129	HTH-type transcriptional regulator hmrR	Q9X5U4	129	*Rhi. leguminosarum*	40
Bzo70-4	197	AcrR; regulatory protein	AAL32124	214	*Proteus minabilis*	41
Bzo70-5	366	Xenobiotic reductase B	AAF02539	349	*P. putida*	51
Bzo70-6	262	FabG; 3-oxoacyl-[acyl-carrier-protein] reductase	NP_280196	255	*Halobacterium* sp. NRC-1	33
Bzo70-7	67*	Quinone oxidoreductase	NP_241801	322	*Bacillus halodurans*	43
Bzo71-1	209	PsrA; transcriptional regulator TetR family	CAC17801	237	*P. putida*	29
Bzo71-2	254	Bcepa02002564; hypothetical protein	ZP_00216133	150	*Burkholderia cepacia*	38
Bzo71-3	391	FadAx; 3-ketoacyl-CoA thiolase	AAK18171	397	*P. putida*	52
Bzo71-4	332	Cyclase	CAA07767	257	*S. argillaceus*	47
Bzo71-5	370	Outer membrane protein (porin)	ZP_00212955	358	*Bu. cepacia*	36
Bzo71-6	243	PsrA; transcriptional regulator TetR family	CAC17801	237	*P. putida*	25
Bzo71-7	310	PobR; transcriptional regulator AraC family	CAB64665	292	*P. putida*	38
Bzo71-8	390	P450terp	AAA25996	428	*Pseudomonas* sp.	29
Bzo71-9	91[a]	Putidaredoxin	BAA00414	107	*P. putida*	51
Bzo135-1	250	ToxA; methyltransferase	BAA92862	245	*Burkholderia glumae*	38
Bzo135-2	236	Uncharacterized protein conserved in bacteria	ZP_00270060	183	*Rhodospirillum rubrum*	29
Bzo135-3	327	Paraquant-inducible protein B	ZP_00130522	343	*Desulfovibrio desulfricans* G20	28
Bzo135-4	285	ABC-type transporter, ATPase component	ZP_00130521	257	*D. desulfricans* G20	58
Bzo135-5	405	ABC-type transporter, permease component	ZP_00130520	373	*D. desulfricans* G20	42
Bzo135-6	480	Tyrosine phenol-lyase	Q08897	458	*Symbiobacterium thermophilum*	46
Bzo135-7	123	Hypothetical protein	ZP_00136757	121	*P. aeruginosa*	57
Nap1-1	271[a]	Transposase	CAB54057	355	*P. putida*	50
Nap1-2	226	BarB; transcriptional regulator, TetR family	BAA23612	216	*S. virginiae*	29
Nap1-3	254	DdsA; decaprenyl diphosphate synthase	CAD24417	333	*Paracoccus zeaxanthinifaciens*	31
Nap1-4	202	IcaR; transcriptional regulator, TetR family	AAC06121	185	*Staphylococcus epidermidis*	30
Nap1-5	217	TolA protein	P19934	421	*E. coli*	37
Nap1-6	148	Molecular chaperone	ZP_00321150	187	*Haemophilus influenzae*	30
Nap3-1	413	Raeut022065; hypothetical protein	ZP_00170803	336	*Ralstonia eutropha* JMP134	37
Nap3-2	168	Methyl-accepting chemotaxis protein	ZP_00301462	39	*Geobacter metallireducens*	39
Nap3-3	216	LanK; transcriptional regulator, TetR family	AAD13556	192	*S. cyanogenus*	41
Nap3-4	186	TolA protein	P19334	421	*E. coli*	43
Nap3-5	253	NE0179; hypothetical protein	NP_840273	228	*Nitrosomonas europaea*	47
Nap3-6	580	UbiB; 2-octaprenylphenol hydroxylase	NP_418279	546	*E. coli*	30

[a]Partial.

gate setting in FACS in step iii. Low gate setting may increase the chance to obtain genome fragments from diverse bacteria, although this may also increase the ratio of false positives. Alternatively, use of another bacterium (e.g., *B. subtilis*) as a cloning host may be useful for obtaining different fragments.

Among ORFs found in the SIGEX-selected fragments, we were interested in characterizing the putative P450 enzyme encoded by *bzo*71-8, because Bzo71-8 showed substantial homology (71%) only to putative P450 (accession no. NP_946360) found in the genome-sequenced *Rhodopseudomonas palustris*[23]. Bzo71-8 also showed moderate homology (29%) to monoterpene-transforming P450 of *Pseudomonas putida*[24] (**Table 1**), whereas the relevance to aromatic-hydrocarbon transformation was not apparent in the sequence analysis. *E. coli* harboring p18GFP-BZO71 was first subjected to the transformation assay with a variety of benzoate derivatives; however, we could not detect any transformation products. In addition, the typical P450 absorption spectrum was not observed in its cell-free

extract. We therefore subcloned *bzo*71-8 into an expression plasmid (pET28a) and attempted to express P450 under the control of the T7 promoter. The resultant cell-free extract exhibited typical absorption spectra of P450 (ref. 24), confirming that Bzo71-8 was really P450 (**Fig. 3a**).

Next, a spectrophotometric substrate-binding assay[25] was done with benzoate relatives; shifts of the absorption peak from 416 nm to around 395 nm due to changes in the spin state of the heme[25] were observed with several compounds, including 4-hydroxybenzoate (4-HB), but not with benzoate or 2- and 3-hydroxylated derivatives (**Fig. 3b**). To identify transformation products, Bzo71-8 was mixed with recombinant electron-transfer proteins (that is, putidaredoxin (PdX) and putidaredoxin reductase (PdR) of *P. putida*[26]), and the mixture was used for the transformation assay. The electron transfer between these proteins was confirmed by a spectrophotometric analysis (**Fig. 3c**). Among the compounds bound to Bzo71-8, gas chromatography–mass spectrometry (GC-MS) analysis detected

Figure 3 Characterization of the Bzo71-8 P450.
(**a**) Absorption spectra of Bzo71-8. The oxidized form exhibited a typical Soret peak at 416 nm, whereas the peak of CO-binding Bzo71-8 (reduced with sodium hydrosulfite) was shifted to 446 nm. (**b**) Substrate binding assay as assessed by a shift of the Soret peak of the oxidized form. By adding 4-HB, the peak was shifted from 416 nm to 396 nm. (**c**) Electron transfer to Bzo71-8. The absorption peak of CO-binding P450 appeared when Bzo71-8 was incubated with PdR and PdX, indicating that the PdR/PdX system can reduce Bzo71-8. In this panel, the transformation reaction (4-HB to protocatechuate) is also shown.
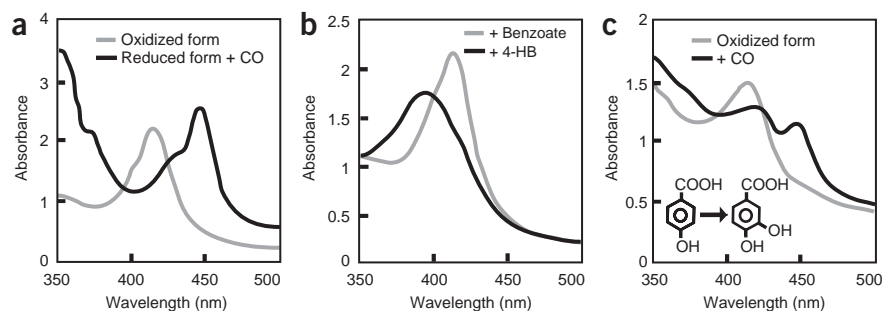
products only from 4-HB and identified it as protocatechuate (**Fig. 3c**). To our knowledge, Bzo71-8 is the first P450 that specifically catalyzes 3-hydroxylation of 4-HB.

In conclusion, this study demonstrated the utility of the SIGEX scheme for the isolation of novel catabolic genes from a metagenome library. As shown in the characterization of the Bzo71-8 P450, SIGEX facilitated isolation of a catabolic gene whose enzymatic activity was not easily expressed and analyzed in a cloning host (e.g., *E. coli*). Another advantage of SIGEX is that substrates for an unknown enzyme found in a metagenome fragment could be deduced from the induction substrate used in SIGEX screening, although possible promiscuity of this linkage should be considered. Based on the results of the present study, SIGEX can be used as an approach to facilitate access to novel biocatalysts in natural genetic resources.

METHODS

Strains and plasmid. *E. coli* JM109 was used as the host organism, and p18GFP was used as a vector for constructing a genomic library. *E. coli* JM109 carrying p18GFP or its derivatives was grown in Luria-Bertani (LB) medium[10] containing 100 μg/ml ampicillin and 200 μM IPTG at 37 °C or the dLB medium containing ampicillin at 30 °C. dLB was composed per liter of 0.5 g of Bacto yeast extract, 1 g of Bacto tryptone, 1 g of NaCl, 10 mM MgSO₄ and 0.2% maltose. *R. eutropha* E2 was grown at 30 °C in dLB medium containing 100 μg/ml ampicillin and 50 μg/ml kanamycin. Genomic DNA of strain E2 was extracted by the method described elsewhere[27].

For constructing p18GFP, a *gfp* gene was amplified by PCR from pET-GFPer[28] using restriction-site containing primers: 5′-CCGGATCCTAATTAAT-TAAGAAGGAGATAT-3′ (forward) and 5′-GCTTGCATGCTTAGTATAGTT-CATCCATGC-3′ (reverse). The forward primer also contained sequences for stop codons in three frames. The amplicon was digested with *Bam*HI and *Sph*I and ligated into pUC18 to construct p18GFP.

Groundwater sample and DNA extraction. Groundwater was obtained from the TK101 underground crude-oil storage cavity at Kuji in Iwate, Japan, in June 2003. Characteristics of this groundwater have been reported elsewhere[13]. Microorganisms in the groundwater were collected by filtration on a 0.22-μm pore size membrane (type GV, Millipore) and recovered by washing the filter with TE buffer[10]. The groundwater metagenome was extracted from the collected microorganisms as described[27].

Construction of genomic library. Genomic DNA of strain E2 or groundwater microorganisms was partially digested with *Sau*3AI, and DNA fragments of 5–10 kb were recovered by agarose gel electrophoresis and the RECOCHIP kit (Takara). The DNA fragments were ligated with *Bam*HI-digested alkalipho-sphatase-treated p18GFP and used to transform *E. coli* JM109 competent cells (Takara) by electroporation. The number of clones harboring inserts and the average insert sizes were analyzed by plating a small fraction of the library on LB- ampicillin plates and analyzing by restriction-endonuclease

digestion. A part of the library in liquid culture was stored at –80 °C in the presence of 15% glycerol.

Flow cytometry and FACS. A FACSVantage SE (Becton Dickinson) machine equipped with an argon laser was used according to the manufacturer's instructions. The 488-nm emission line of the laser with an output power of 1.0 W was used for measuring the forward scatter (488-nm band pass filter for detection) and fluorescence intensity (detector FL1 with a 530-nm band pass filter). Fluorebrite beads (0.5 μm; Polysciences) were used to check the instrument performance before sample analysis. The CELLQuest software (Becton Dickinson) was used for data acquisition and analysis. For preparing cells to be subjected to FACS, the dLB- ampicillin medium supplemented with an induction substrate (2 mM) or IPTG (0.5 mM) was inoculated with cells from a genomic library and cells were grown at 30 °C.

RFLP and sequence analyses. For RFLP analysis of metagenomic fragments, restriction enzymes *Eco*RI and *Pst*I were used. Nucleotide sequences were determined by primer walking using a 3700 DNA sequencer (Applied Biosys-tems). ORFs were analyzed using the ORF finder program in the NCBI homepage (http://www.ncbi.nlm.nih.gov/), and homology search was conducted using the Blast program in GenBank (http://www.ncbi.nlm.nih.gov/BLAST). Phylogenetic analysis of metagenome fragments was conducted by the self-organization map algorithm as described elsewhere[22].

Expression and analyses of BZO71-8 P450. The *bzo71-8* gene was recovered by PCR with restriction-site containing primers 5′-CGTCTCCCATGTTCAGTTTT GACCCCTATT-3′ (forward) and 5′-GGATCCTTAATTAGGATGCTGGCAGGC CAG-3′ (reverse) and ligated into pET28a (Novagen). *E. coli* BL21 (DE3) pLysS (Novagen) was transformed with the resultant plasmid. For checking the expression of P450, 100 ml of LB medium containing kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) was inoculated with 1 ml of the overnight culture in the same medium and shaken at 37 °C. When the optical density at 600 nm (A_{600}) was reached at 0.8–1.0, 1 mM α-aminolevulic acid and 0.5 mM FeCl₃ were added to the culture. After further incubation for 30 min at 30 °C, expression was induced by adding 0.5 mM IPTG, and cells were grown for a further 20 h at 25 °C. Cells were collected by centrifugation, resuspended in 20 mM potassium phosphate buffer (pH 7.4) containing 20% glycerol and broken by a French press (Ohtake works). The extract was centrifuged at 12,000g for 20 min, and the resultant supernatant was used for spectro-photometric analyses.

A coding region for putidaredoxin and putidaredoxin reductase was recov-ered from the *P. putida* genome[26] by PCR using restriction-site containing primers (5′-CGTCTCCCATGTCTAAAGTAGTGTATGTGT-3′ [forward] and 5′-CGTCTCATCGATTACCATTGCCTATCGGGA-3′ [reverse] for the PdX gene; 5′-CGTCTCCCATGAACGCAAACGACAACGTGG-3′ [forward] and 5′-CGTCTCATCGATCAGGCACTACTCAGTTCA-3′ [reverse] for the PdR gene). Each of these PCR products was ligated into pET28a to generate the expression vector, pETPdX or pETPdR. *E. coli* Rosetta (DE3) was transformed with pETPdX, whereas *E. coli* BL21 (DE3) pLysS was transformed with pETPdR. Cells were grown at 37 °C in TB medium[10] supplemented with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) for BL21[pETPdR] or

kanamycin (50 μg/ml) for Rosseta[pETPdX] up to A_{600} of approximately 1.0. The temperature was then lowered to 30 °C, and cells were grown for 20 h after adding 0.5 mM IPTG. For a transformation assay, cells expressing Bzo71-8, PdX and PdR were mixed in 20 mM potassium phosphate buffer (pH 7.4) containing 20% glycerol and disrupted using the French press. A P450 reaction was initiated by adding NADH (1 mM) and a substrate (1 mM). After incubation at 30 °C for 12 h, the reaction mixture was acidified with HCl (pH 2.0) and extracted with ethyl acetate. The organic phase was recovered, dried and treated with the trimethylsilyl reagent as described elsewhere[29]. This sample was analyzed by using a QP5050 gas chromatography–mass spectrometry machine (Shimadzu) equipped with a DB-5 column (30 m in length, 0.25 mm in diameter; J&W Scientific) as described elsewhere[29].

Nucleotide sequence accession number. The sequences reported in this paper have been deposited in GenBank/EMBL/DDBJ under accession no. AB186499 to AB186504 and AB190317 to AB190320.

1. Daniel, R. The soil metagenome—a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* **15**, 199–204 (2004).
2. Okuta, A., Ohnishi, K. & Harayama, S. PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. *Gene* **212**, 221–228 (1998).
3. Henne, A., Daniel, R., Schmitz, R.A. & Gottschalk, G. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl. Environ. Microbiol.* **65**, 3901–3907 (1999).
4. Teramoto, M., Futamata, H., Harayama, S. & Watanabe, K. Characterization of a high-affinity phenol hydroxylase from *Comamonas testosteroni* R5 by gene cloning, and expression in *Pseudomonas aeruginosa* PAO1c. *Mol. Gen. Genet.* **262**, 552–558 (1999).
5. Debarbouille, M., Martin-Verstraete, I., Klier, A. & Rapoport, G. The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to both sigma[54]- and phosphotransferase system-dependent regulators. *Proc. Natl. Acad. Sci. USA* **88**, 2212–2216.
6. Rogowsky, P.M. *et al.* Molecular characterization of the vir regulon of *Agrobacterium tumefaciens*: complete nucleotide sequence and gene organization of the 2863-kbp regulon cloned as a single unit. *Plasmid* **23**, 85–106 (1990).
7. Valdivia, R.H. & Falkow, S. Fluorescence-based isolation of bacterial genes expressed within host cells. *Science* **277**, 2007–2011 (1997).
8. Ochman, H., Gerber, A.S. & Hartl, D.L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623 (1988).
9. Hino, S., Watanabe, K. & Takahashi, N. Phenol hydroxylase cloned from *Ralstonia eutropha* strain E2 exhibits novel kinetic properties. *Microbiology* **144**, 1765–1772 (1998).
10. Sambrook, J. & Russell, D.W.. *Molecular Cloning: A Laboratory Manual*, edn. 3 (Cold Spring Harbor Press, Cold Spring Harbor, New York, 2001).
11. Sze, C.C., Moore, T. & Shingler, V. Growth phase-dependent transcription of the sigma(54)-dependent Po promoter controlling the *Pseudomonas*-derived (methyl)phenol dmp operon of pVI150. *J. Bacteriol.* **178**, 3727–3735 (1996).
12. Yuste, L., Canosa, I. & Rojo, F. Carbon-source-dependent expression of the PalkB promoter from the *Pseudomonas oleovorans* alkane degradation pathway. *J. Bacteriol.* **180**, 5218–5226 (1998).
13. Watanabe, K. *et al.* Molecular characterization of bacterial populations in petroleum-contaminated groundwater discharged from underground crude-oil-storage cavities. *Appl. Environ. Microbiol.* **66**, 4803–4809 (2000).
14. Marques, S. & Ramos, J.L. Transcriptional control of the *Pseudomonas putida* TOL plasmid catabolic pathways. *Mol. Microbiol.* **9**, 923–929 (1993).
15. Parsek, M.R., Shinabarger, D.L., Rothmel, R.K. & Chakrabarty, A.M. Roles of CatR and cis,cis-muconate in activation of the catBC operon, which is involved in benzoate degradation in *Pseudomonas putida*. *J. Bacteriol.* **174**, 7798–7806 (1992).
16. Seoane, A.S. & Levy, S.B. Characterization of MarR, the repressor of the multiple antibiotic resistance (*mar*) operon in *Eschericia coli*. *J. Bacteriol.* **177**, 3414–3419 (1995).
17. Cowles, C.E., Nichols, N.N. & Harwood, C.S. BenR, a XylS homolog, regulates three different pathways of aromatic acid degradation in *Pseudomonas putida*. *J. Bacteriol.* **182**, 6339–6346 (2000).
18. Rothmel, R.K. *et al.* Nucleotide sequencing and characterization of *Pseudomonas putida catR*: a positive regulator of the *catBC* operon is a member of the LysR family. *J. Bacteriol.* **172**, 922–931 (1990).
19. Blehert, D.S., Fox, B.G. & Chambliss, G.H. Cloning and sequence analysis of two *Pseudomonas* flavoprotein xenobiotic reductases. *J. Bacteriol.* **181**, 6254–6263 (1999).
20. Kumagai, H., Yamada, H., Matsui, H., Ohkishi, H. & Ogata, K. Tyrosine phenol lyase I. Purification, crystallization, and properties. *J. Biol. Chem.* **245**, 1767–1772 (1970).
21. Poon, W.W. *et al.* Identification of *Escherichia coli ubiB*, a gene required for the first monooxygenase step in ubiquinone biosynthesis. *J. Bacteriol.* **182**, 5139–5146 (2000).
22. Abe, T. *et al.* Informatics for unveiling hidden genome signatures. *Genome Res.* **13**, 693–702 (2003).
23. Larimer, F.W. *et al.* Complete genome sequence of the metabolically versatile photo-synthetic bacterium *Rhodopseudomonas palustris*. *Nat. Biotechnol.* **22**, 55–61 (2004).
24. Peterson, J. *et al.* Cytochrome P-450terp. Isolation and purification of the protein and cloning and sequencing of its operon. *J. Biol. Chem.* **267**, 14193–14203 (1992).
25. George, P., Bettlestene, J. & Griffith, J.S. in *Haematin Enzymes*, vol. 1 (eds. Falk, J.E., Lemberg, R., & Morton, R.K.) 105–141 (Pergamon Press Ltd, Oxford 1961).
26. Peterson, J.A., Lorence, M.C. & Amarneh, B. Putidaredoxin reductase and putidaredoxin. *J. Biol. Chem.* **265**, 6066–6073 (1990).
27. Silhavy, T.J., Berman, M.L. & Enquist, L.W. *Experiments With Gene Fusions* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1984).
28. Sakikawa, C., Taguchi, H., Makino, Y. & Yoshida, M. On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *J. Biol. Chem.* **274**, 21251–21256 (1999).
29. Peng, X., Misawa, N. & Harayama, S. Isolation and characterization of thermophilic Bacilli degrading cinnamic, 4-coumaric, and ferulic acids. *Appl. Environ. Microbiol.* **69**, 1417–1427 (2003).