Environmental Microbiology (2008) **10**(4), 978–987 doi:10.1111/j.1462-2920.2007.01518.x

Retrieval of entire genes from environmental DNA by inverse PCR with pre-amplification of target genes using primers containing locked nucleic acids

Kazutaka Yamada,[1]*† Takeshi Terahara,[2] Shinya Kurata,[1] Toyokazu Yokomaku,[1] Satoshi Tsuneda[3] and Shigeaki Harayama[4]

[1]*Technological Research Laboratory, Nippon Steel Kankyo Engineering Co., Ltd, 2-1-38 Shiohama, Kisarazu-shi, Chiba 292-0838, Japan.*

[2]*Department of Applied Chemistry, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan.*

[3]*Department of Life Science and Medical Bio-Science, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan.*

[4]*Biotechnology Development Center (NBRC), Department of Biotechnology, National Institute of Technology and Evaluation (NITE), 2-5-8 Kazusa-Kamatari, Kisarazu-shi, Chiba 292-0818, Japan.*

Summary

We had been unsuccessful to amplify desired nucleotide sequences from various environmental DNA samples by using the inverse polymerase chain reaction (IPCR) technique, most probably because the copy numbers of target DNA sequences had been quite low. To enrich the target DNA sequences prior to IPCR, a rolling-circle amplification was used with a site-specific primer containing locked nucleic acids (LNAs). This pre-amplified IPCR (PAI-PCR) method increased the sensitivity of PCR almost 10 000 times compared with the standard IPCR in model experiments using *Escherichia coli*. We then applied the PAI-PCR method to isolate glycosyl hydrolase genes from DNAs extracted from vermiform appendixes of horses and termite guts. The flanking sequences of the target genes were amplified and cloned successfully using PAI-PCR, whereas standard IPCR resulted in no amplification.

Received 4 April, 2007; accepted 31 October, 2007. *For correspondence. E-mail kyamada@nbrc.nite.go.jp; Tel. (+81) 438 20 5764; Fax (+81) 438 52 2314. †Present address: Biotechnology Development Center (NBRC), Department of Biotechnology, National Institute of Technology and Evaluation (NITE), 2-5-8 Kazusa-Kamatari, Kisarazu-shi, Chiba 292-0818, Japan.

Introduction

Many microorganisms in the natural environment have not yet been characterized because of difficulties in cultivating them by conventional cultivation techniques. Recently, efforts have been made to isolate genes of interest from environmental DNAs (collectively called metagenomes) by using metagenomic approaches that may provide access to novel genetic resources (Okuta *et al*., 1998; Rondon *et al*., 2000; Stokes *et al*., 2001; Sunna and Bergquist, 2003; Daniel, 2005; Ferrer *et al*., 2005; Hayashi *et al*., 2005; Lorenz and Eck, 2005; Uchiyama *et al*., 2005; Uchiyama and Watanabe, 2006). Several polymerase chain reaction (PCR)-based approaches are used to identify novel genes from metagenomes. One such method is 'cassette PCR' (Okuta *et al*., 1998), in which central parts of family genes are amplified by PCR, and the amplified DNAs are used to replace the central part of a previously cloned gene to form novel hybrid genes. Alternatively, inverse PCR (IPCR) (Ochman *et al*., 1988; Triglia *et al*., 1988) and vectorette PCR (Arnold and Hodgson, 1991; Kilstrup and Kristiansen, 2000) are the methods most widely used for cloning the upstream and downstream flanking regions of known sequences. However, few papers have been published that describe successful cloning of full-length genes from metagenomes by using gene-specific PCR (Eschenfeldt *et al*., 2001; Sunna and Bergquist, 2003; Hayashi *et al*., 2005; Tang *et al*., 2006; Uchiyama and Watanabe, 2006). It appears that improvements, such as pre-amplification of target genes, are necessary to overcome the difficulties associated with the isolation of full-length genes from metagenomes by IPCR or vectorette PCR.

When the amount of template DNA extracted from an environmental sample is small, pre-amplification of the template DNA is often carried out by isothermal DNA amplification, for example, by using phi29 DNA polymerase, on the basis of rolling-circle amplification (RCA) (Dean *et al*., 2001; Rector *et al*., 2004) or multiple displacement amplification (MDA) (Gonzalez *et al*., 2005;

Table 1. Oligonucleotides used in this study.

Name	Method	Sequence (5′–3′)	T_m (°C)[a]
Ecol-RC(DNA)	RCA	GTGAGTTG	25.0
Ecol-RC(LNA-5′)		G**T**G**A**GTTG	37.0
Ecol-RC(LNA-3′)		GTGAGT**T**G	30.0
Ecol-RC(LNA-Even)		G**T**G**AG**T**T**G	44.0
Ecol-iFw	IPCR	CCGGTACTGATGTGATGGCTGCTAGG	70.0
Ecol-iRv		GGGTTTATTTCTGGTGCGTTTCGTTGG	69.0
XYN-Fw	PCR	CATACKTTKGTTTGGCA	ND
XYN-Rv		TMGTTKACMACRTCCCA	ND
Xyn-RC(DNA)	RCA	CATACKTT	13.0–21.0
Xyn-RC(LNA-5′)		C**A**T**A**CKTT	21.0–29.0
Xyn-RC(LNA-3′)		CATACK**T**T	14.0–24.0
Xyn-RC(LNA-even)		C**A**T**AC**K**T**T	26.0–38.0
Xyn-0005LF	IPCR	TCTGGAAAAATATGTTACCGATGTT	60.0
Xyn-0005LR		CCAAGGACTTCTGTTTTACTGCC	64.0
Xyn-0007LF		TTTGTCACCGATGTAACGACGCATTT	67.0
Xyn-0007LR		GCGATAAAGAAACCGTCCTTGCC	67.0
Xyn-0022LF		ATATGCATGATGTATTCGAGTGGGCTGA	67.0
Xyn-0022LR		ACTCTCTGTCTGCTAACTCTCCGCTTGT	70.0
Xyn-0101LF		GTATCCGCGAGGTGTCGAGCC	70.0
Xyn-0101LR		AAAGGATGTCATCTTCTGCTG	60.0
Xyn-0102LF		TTTTTCCAGATTGCCAAGGACTTCT	64.0
Xyn-0102LR		ACTTATTTCCTGGGACGTGGTCAAC	67.0

a. Melting temperature values calculated by using the Exiqon T_m prediction tool, which is available online (http://lna-tm.com/).
LNA residues are depicted in boldface.
ND, not determined.

Jiang *et al*., 2005; Monstein *et al*., 2005; Raghunathan *et al*., 2005). The method is effective in whole-genome amplification of environmental samples (Abulencia *et al*., 2006; Yokouchi *et al*., 2006). Furthermore, selective amplification of a specific DNA sequence by RCA has also been carried out by using a set of primers, one being identical to the 5′-end and another being complementary to the 3′-end of the target sequence (Lizardi *et al*., 1998).

Because the specificity of the isothermal amplification is not high, especially when the reaction temperature is low or the initial amount of target DNA is small, selective amplification of a specific DNA region by the isothermal amplification is generally difficult. To reduce non-specific amplification derived from miss-priming, primers containing locked nucleic acids (LNAs) may be useful. Locked nucleic acids are DNA analogues in which the furanose ring in the sugar-phosphate backbone is chemically locked. Locked nucleic acids obey the Watson–Crick pairing rules, but have an increased specificity and a high affinity to complementary DNA (Sanjay *et al*., 1998; Vester and Wengel, 2004). Therefore, LNA oligonucleotides have been used as antisense molecules (Wahlestedt *et al*., 2000), as fluorescent *in situ* hybridization probes (Silahtaroglu *et al*., 2003), as primers for allele-specific PCR (Latorra *et al*., 2003), as probes for genotyping (Di Giusto and King, 2004), as inhibitors of undesired PCR amplification (Hummelshoj *et al*., 2005) and as sequencing primers (Levin *et al*., 2006).

In this study, we demonstrate that the sensitivity of IPCR dramatically increased when target DNA sequences were pre-amplified by RCA using LNA-containing site-specific primers. By using this pre-amplified inverse PCR (PAI-PCR) method, we successfully cloned novel genes encoding cellulolytic enzymes from environmental samples.

Results

Sensitivity of IPCR and PAI-PCR

The sensitivity of IPCR was examined for the amplification of a lysogenic lambda-phage sequence. The *Escherichia coli* (λ) DNA was digested with HindIII and self-circularized. The ligated DNA was purified, and serially diluted so that each tube contained 10^1–10^6 copies of each HindIII fragment. Inverse PCR was subsequently conducted using the Ecol-iFw and Ecol-iRv primers (Table 1) which amplify a 4880-bp-long fragment from the circularized lambda-phage DNA. A single band of approximately 5000 bp was amplified only when more than 10^6 copies of the DNA were present (Fig. 2).

To improve the sensitivity of IPCR, we tried to enrich the target DNA prior to IPCR by RCA using several target-specific primers (Table 1). After the RCA reaction, DNA was purified and used as a template for IPCR. The pre-amplification was effective to improve the sensitivity of IPCR. The effectiveness was different with different RCA primers: the sensitivity increased almost 100 times with the DNA primer, Ecol-RC(DNA), almost 1000 times with the LNA-containing primers, Ecol-RC(LNA-5′) and Ecol-

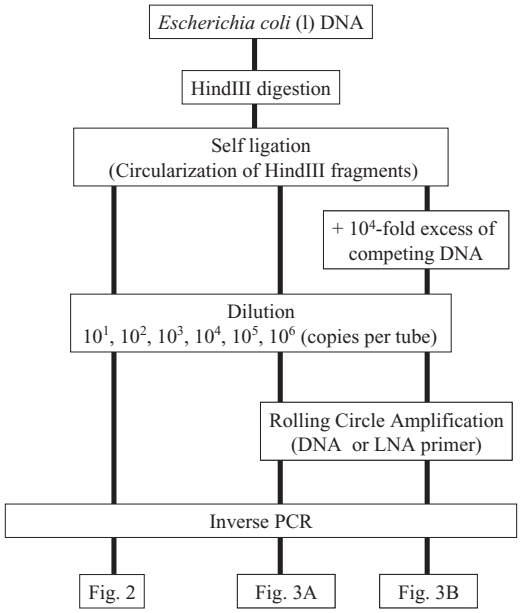

Fig. 1. Flow chart of the experiments for Figs 1 and 2. From the chromosomal DNA of *E. coli* (λ), a lambda DNA fragment was amplified by IPCR under a variety of conditions.

RC(LNA-3′), and more than 10^5 times with Ecol-RC(LNA-Even) (Fig. 3A). These results demonstrated that the RCA pre-amplification effectively increased the sensitivity of IPCR especially with LNA-containing primers. Next, we examined the sensitivity of PAI-PCR in the presence of excess competing DNA. The results with 10^4-fold excess of the competing DNA are shown in Fig. 3B. Under the conditions, the sensitivity decreased approximately 10-fold compared with that in the absence of the competing DNA. Even so, the amplification was possible using the primer Ecol-RC(LNA-even) when more than 10^2 copies of the target DNA existed as templates. These results indicated that target gene sequences from a gene of a bacterium constituting a small percentage (e.g. 0.01%) of the total microbial population could be recovered by PAI-PCR.

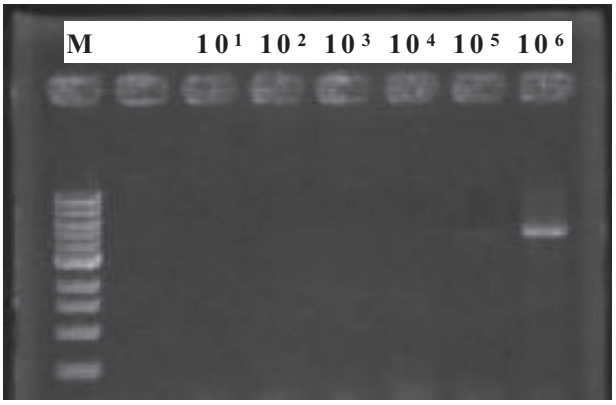

Fig. 2. Agarose gel electrophoresis of IPCR products amplified from DNA of *E. coli* (λ) without RCA. M indicates the 1.0 kb ladder lane, the intensity of the 3.0 kb band being higher than other bands.

Screening of genes encoding cellulolytic enzymes from environmental DNAs

DNAs extracted from vermiform appendixes of horses (Table 1) were used for PCR amplification of partial sequences of genes encoding glycosyl hydrolase family 10 (GHF10; http://cazy.org/fam/GH10.html). Polymerase chain reaction fragments (*c.* 160 bp) amplified from the DNA samples were ligated into a TA-cloning vector, and used to transform cells of *E. coli*. Transformants were randomly picked up and the sequences of cloned DNAs were determined. Eleven different DNA sequences were obtained, and the amino acid sequences translated from the sequences revealed 49–64% amino acid sequence similarities to known xylanases categorized as GHF10.

Based on the cloned sequences, IPCR primer sets for the amplification of the flanking regions of the targets were designed by using OLIGO software (TaKaRa-Bio). DNA extracted from the horse vermiform appendix was digested with several restriction enzymes and self-circularized. Then, PAI-PCR was performed, IPCR without RCA pre-

Table 2. Characterization of putative xylanase genes cloned from environmental samples.

Putative xylanase gene	Size			Homologue[c]		
	aa	MW[a]	Conserved domain[b]	Accession No.	Organism	Identity (%)
XynVA1	376	42 389	Glyco_hydro_10	AAL06078	Uncultured bacterium	40
XynVA2	396	44 154	Glyco_hydro_10	YP_001038252	*Clostridium thermocellum*	37
XynIO1	400	44 551	Glyco_hydro_10	BAA82143	*Clostridium stercorarium*	43
XynIO2	555	62 734	CBM4_9 + Glyco_hydro_10	CAA07173	*Bacillus* sp. Strain BP-23	39
XynIO4	381	42 533	Glyco_hydro_10	BAA82143	*Clostridium stercorarium*	35

a. Estimated using an ExPASy Proteomics Server tool (http://au.expasy.org/).
b. Identified by using the Conserved Domain Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) in the National Center for Biotehcnology Information (NCBI). Glyco_hydro_10, glycosyl hydrolases family 10 motif; CBM4_9, motif of carbohydrate-binding module.
c. XhomologProtein showing the highest similarity to the query detected by the Basic Local Alignment Search Tool (BLAST) search.
aa, amino acids; MW, molecular weight.

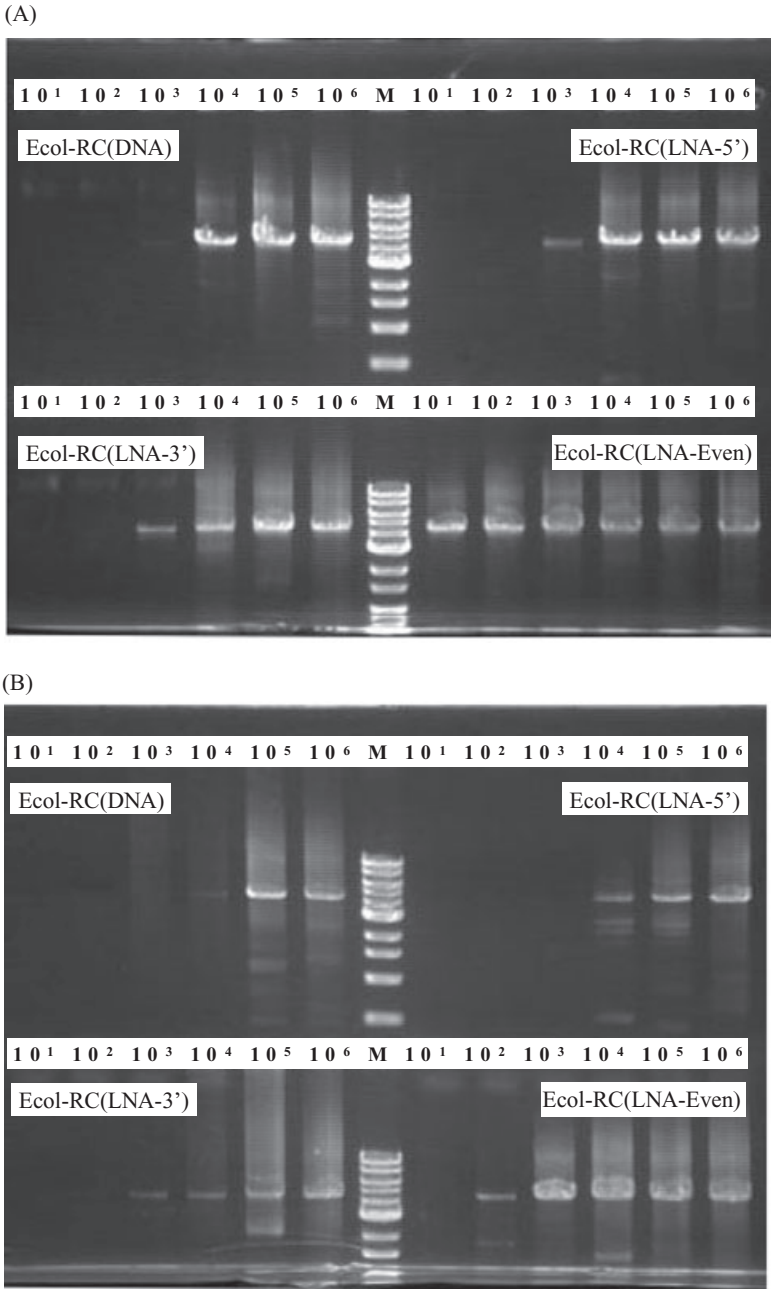

Fig. 3. Agarose gel electrophoresis of PAI-PCR products using (A) an *E. coli* (λ) DNA and (B) an *E. coli* (λ) DNA mixed with 10^4-fold excess of competing DNA. M indicates the 1.0 kb ladder lanes, the intensity of the 3.0 kb band being higher than other bands.

amplification being carried out as a control. The results with Xyn-RC(LNA-Even) as the RCA primer, and Xyn-0101LF plus Xyn-0101LR as the IPCR primers are shown in Fig. 4 (even-numbered lanes). A single amplification product was observed from all the five samples that had been digested by five different restriction enzymes. Pre-amplified inverse PCR amplification products were also recovered using other IPCR primer sets such as Xyn-0022LF and Xyn-0022LR (data no shown). However no amplification product was observed in any IPCR without RCA reaction (Fig. 4, odd-numbered lanes).

Sequencing analysis of the amplification products demonstrated that two putative novel xylanase genes were retrieved from horse vermiform appendixes. The genes were designated as *xynVA1* and *xynVA2* and their characteristics were described in Table 2. In both the *xynVA1* and *xynVA2* genes, a well-conserved catalytic domain, a motif of GHF10, was detected by using the NCBI Conserved Domain Search program, while a putative signal peptide in their N-terminal amino acid sequences was detected by using the SOSUI program (http://bpnuap.nagoya-u.ac.jp/sosui/). These findings suggest that both

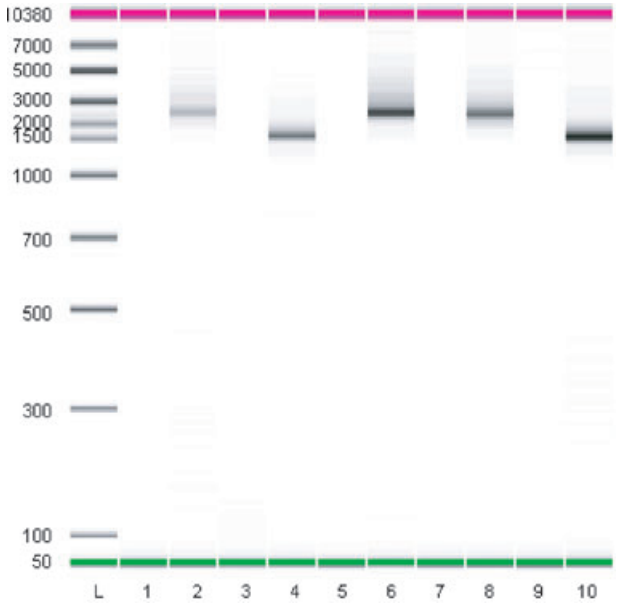

Fig. 4. Capillary electrophoreses of IPCR (odd-numbered lanes) and PAI-PCR (even-numbered lanes) products amplified from DNA extracted from horse vermiform appendixes. The electrophoreses followed by the analysis of the electrophoregrams were automatically performed by using an Agilent 2100 Bioanalyser with a DNA 7500 LabChip® kit (TaKaRa-Bio). The results are shown as a gel-like image. L is the ladder obtained with DNA fragment size markers (bp). The green and pink peaks are the lowest 50 bp and highest 10 380 bp markers respectively. The substrates used for IPCR or PAI-PCR were BamHI-digested DNA (lanes 1 and 2), EcoRI-digested DNA (lanes 3 and 4), HindIII-digested DNA (lanes 5 and 6), PstI-digested DNA (lanes 7 and 8) and XbaI-digested DNA (lanes 9 and 10).

the *xynVA1* and *xynVA2* genes are novel xylanase genes belonging to GHF10.

Similar experiments were carried out with 12 DNA samples extracted from termite guts (Table 1). Polymerase chain reaction fragments (*c.* 160 bp) were successfully amplified from all the 12 samples, then cloned and sequenced. When clones with sequence similarity higher than 90% were grouped, they were divided into 10 different groups. Between these 10 groups, the intergroup similarities ranged between 41% and 79%. A BLAST analysis demonstrated that the amino acid sequences translated from the DNA sequences revealed 40–74% amino acid sequence similarities to known xylanases categorized as GHF10. An IPCR primer set was designed for each of the 12 clones and PAI-PCR reactions were carried out. Pre-amplified inverse PCR amplification products were obtained from all the 12 DNA samples when an appropriate IPCR primer set was used together with either Xyn-RC(LNA-Even) or Xyn-RC(LNA-5′) as an RCA primer. When the RCA primer was Xyn-RC(LNA-Even), the sizes of the major products were generally larger than 2 kb suggesting successful PAI-PCR (Fig. 5A). To the contrary, the PAI-PCR products obtained with Xyn-RC(LNA-5′) contained both of large and small (< 0.5 kb) fragments, the smaller ones most probably being non-specific amplification products (Fig. 5B).

After the cloning of the larger fragments, the cloned DNAs were sequenced. From sequence data analysis, three putative xylanase genes were identified and designated as *xynIO1*, *xynIO2* and *xynIO4* (Table 2). Both the genes of *xynIO1* and *xynIO2* encoded polypeptides comprising a module of GHF10 and a putative signal peptide in their N-terminal amino acid sequences. Beside, the *xynIO2* product possessed a carbohydrate-binding module belonging to family 4_9 (CBM4_9) placed upstream of the GHF10 module. The *xynIO4* product had a module of GHF10 and two putative *trans*-membrane helices, one in the N-terminal and the other in the middle part of its gene product the polypeptide.

Discussion

Inverse PCR is a method for *in vitro* amplification of flanking regions of a known genomic sequence. However the low sensitivity of IPCR is well known and may be the result of a combination of at least two factors: imperfect efficiency of intramolecular ligation (Hayashi *et al.*, 1986), and a lower PCR efficiency with circularized DNA than with linear DNA (Mullis and Faloona, 1987; Laghi *et al.*, 2004). In fact, in a model experiment using the chromosomal DNA of a lambda lysogen of *E. coli*, 10^6 copies were required for successful IPCR amplification (Fig. 2). Because of this low sensitivity, the IPCR amplification technique is not readily applicable to environmental DNAs of high complexity. Recently, another method for improving the sensitivity of IPCR has been reported (Uchiyama and Watanabe, 2006). In this method, an affinity tag-labelled oligonucleotide is used in the first IPCR, followed by the affinity recovery of the IPCR products and nested IPCR with the recovered products.

In this study, we attempted to pre-amplify target DNAs prior to IPCR. Although RCA reaction is a popular and well-established method for pre-amplification, it was not effective in improving the yield of IPCR from a DNA primer specific to the target DNA (Fig. 3), most probably due to insufficient specificity of the RCA primers against a very low copy number of the target gene under isothermal amplification conditions. This result led us to use LNA-containing primer that binds to their complementary DNA sequences with increased affinity and a higher melting temperature (Vester and Wengel, 2004). We demonstrated that the use of LNA-containing primers in RCA reactions is effective in increasing the overall sensitivity of IPCR from 10 to 1000 times detecting as many as 10 copies of the target DNA in a sample (Fig. 3), and that the PAI-PCR method was effective even if the target chromo-

Fig. 5. Capillary electrophoreses of PAI-PCR products amplified from DNA extracted from 12 different termite gut samples. All DNA samples were digested with EcoRI, and an RCA primer [(A) Xyn-RC(LNA-Even); or (B) Xyn-RC(LNA-5′)] was used in the RCA reaction. Products were analysed with an Agilent 2100 Bioanalyser with a DNA 7500 LabChip® kit (TaKaRa-Bio).

some was diluted with 10^4-fold excess of competing DNA (Fig. 3B).

We demonstrated that the positions of LNA residues in a primer sequence significantly influence the effectiveness of RCA: the use of a primer with LNAs evenly distributed throughout its sequence (LNA-Even) offered a greater sensitivity than those with unevenly distributed LNAs (LNA-5′ and LNA-3′). Recently, an LNA-5′ primers was found to be significantly better than LNA-3′ and LNA-Even primers in a sequencing reaction (Levin *et al.*, 2006). On the other hand, an LNA-3′ primer showed the complete resistance against 3′→5′ exonuclease activities of proofreading DNA polymerases as well as of phi29 DNA polymerase (Di Giusto and King, 2004). The LNA-Even primers used in this study contained LNAs at the second, fourth, fifth and seventh positions from the 5′-ends (Table 1). We expect that these LNA-Even primers are tolerant to exonuclease digestion due to the 3′-end LNA while they show a greater efficacy of hybridization due to the 5′-end LNA.

We then used the PAI-PCR method to isolate novel genes from environmental DNAs. In this study, xylanase genes have been selected because of their applicability in many fields including industries of paper and pulp, food, environmental technologies and bio-fuelling. Xylanases catalyse the hydrolysis of the β-1,4-glycosidic bonds of xylan, which is the second most abundant component of plant cell walls after cellulose. In this experiment, we

Table 3. Environmental samples used for the isolation of glycosyl hydrolase genes.

Sample origin	Sampling site	Source of DNA	No. of DNA samples	No. of PCR-positive samples	No. of sequenced clones
Horses	Hayakita, Hokkaido	Vermiform appendix	6[a]	4[b]	70
Nasutitermes takasagoensis	Iriomote Island, Okinawa	Termite gut	12[c]	12	96

a. Sampled in December 2005.
b. Two among four samples were used for DNA sequencing.
c. Sampled in March 2004.

collected DNA from digestive tracts of two organisms capable of decomposing lignocelluloses, namely vermiform horse appendices and termite guts in which remarkable lignocellulolytic degradation is performed by prokaryotes. From both the horse vermiform appendixes and termite guts, we successively successfully amplified and cloned genes for GHF10 by using PAI-PCR while the standard IPCR without RCA reaction resulted in no amplification of the target sequences. Generally, primers containing evenly distributed LNAs provided better results in PAI-PCR in agreement with the *E. coli* (λ) experiments. However, primers of identical nucleotide sequences containing LNAs at different positions seem to have amplified different genes in the metagenomes: the *xynVA1*, *xynVA2* and *xynIO2* genes were obtained when an LNA-Even-type primer, Xyn-RC(LNA-Even), was used in an RCA reaction, whereas the *xynIO1* and *xynIO4* genes were obtained when an LNA-5′-type primer, Xyn-RC(LNA-5′), was used. This observation indicated that the specificity and sensitivity of LNA-containing primers may be also influenced by the position of LNAs, and more works are required to establish general design rules of LNA-containing primers for the RCA reaction. Anyway, the PAI-PCR method developed in this study strongly improved the efficacy of IPCR, and proved to be useful in the isolation of novel genes from environmental DNAs.

Experimental procedures

Preparation of DNA

Cell of *E. coli* strain W3110 harbouring a lambda prophage were grown at 37°C overnight in 100 ml of Luria–Bertani (LB) medium and harvested by centrifugation at 4000 *g* for 10 min. The genomic DNA of the cells was extracted by using DNeasy Tissue Kit (Qiagen GmbH, Hilden, Germany).

Horse vermiform appendixes were obtained from a meat-processing factory (Table 3), and DNA was extracted from each 5 g of the content fluid of the samples by using ISOIL (Nippongene, Tokyo, Japan) according to the manufacturer's instructions.

Termites, *Nasutitermes takasagoensis*, were collected from 12 independent nests in Iriomote Island, Japan (Table 3), and 100 digestive tracts were isolated from termites collected in each of the nests. The 100 digestive tracts from the same nest were combined, and genomic DNA was extracted by using ISOIL.

Oligonucleotide primers

Inverse PCR primers (Ecol-iFw and Ecol-iRv) and RCA primers [Ecol-RC(DNA), Ecol-RC(SD), Ecol-RC(LNA-5′), Ecol-RC(LNA-3′) and Ecol-RC(LNA-Even)] (Table 1) specific to the lambda-phage sequence were designed by using the 'OLIGO' software provided by TaKaRa-Bio (Otsu, Japan). For the amplification of partial fragments of genes encoding GHF10, PCR primers (XYN-FW and XYN-RV) described by Sunna and Bergquist (2003) were used. Locked nucleic acid-containing RCA primers [Xyn-RC(DNA), Xyn-RC(SD), Xyn-RC(LNA-5′), Xyn-RC(LNA-3′) and Xyn-RC(LNA-Even)] were designed based on the XYN-FW sequence. Inverse PCR primers (Xyn-0005LF and Xyn-0005LR; Xyn-0007LF and Xyn-0007LR; Xyn-0022LF and Xyn-0022LR; Xyn-0101LF and Xyn-0101LR; Xyn-0102LF and Xyn-0102LR) (Table 1) for the selective amplification of glycosyl hydrolase gene fragments were designed on the basis of the sequences of PCR-amplified glycosyl hydrolase gene fragments. DNA oligonucleotides were provided by Sigma-Aldrich Japan (Tokyo, Japan) and LNA oligonucleotides by Greiner (Frickenhausen, Germany).

IPCR of lambda-phage sequence

In the first series of experiments (Fig. 1), IPCR of a lambda-phage sequence was carried out without RCA. DNA (0.5 μg) extracted from *E. coli* cells was digested with 7.5 U of HindIII (TaKaRa-Bio) at 37°C for 5 h. The digested DNA was purified by using Microcon YM-100 (Millipore Corporation, Bedford, MA), and *c.* 200 ng of the DNA was self-circularized in 80 μl of the ligation buffer at 16°C overnight by using the Mighty Mix DNA ligation kit (TaKaRa-Bio). DNA in the reaction buffer was purified by using YM-100, and amounts equivalent to 1.0×10^6, 1.0×10^5, 1.0×10^4, 1.0×10^3, 1.0×10^2 and 1.0×10^1 copies of the purified DNA were used as templates for IPCR. Each of the templates was mixed with 0.2 mM dNTP, 0.2 μM each of primers Ecol-iFw and Ecol-iRv (Table 1), and 2.5 U of LA *Taq* DNA polymerase (TaKaRa-Bio) in 50 μl of GC Buffer I (TaKaRa-Bio), and IPCR was performed under the following conditions: 1 min of denaturation at 94°C, 50 cycles of 30 s at 94°C, 30 s at 60°C and 3 min at 72°C followed by 7 min of final extension at 72°C. Inverse PCR products were analysed by electrophoresis using 1.0% (w/v) of SeaKem GTG Agarose Gel (TaKaRa-Bio).

In the second series (Fig. 1), the lambda-phage sequence was pre-amplified by RCA prior to IPCR. To do it, the *E. coli* DNA was HindIII-digested and self-circularized as described above, and amounts equivalent to 1.0×10^6, 1.0×10^5, 1.0×10^4, 1.0×10^3, 1.0×10^2 and 1.0×10^1 copies were used as templates for RCA. Each of the templates was added in

40 μl of Phi29 DNA polymerase reaction buffer (New England Biolabs, Beverly, MA) containing 1.25 μM of one of the following primers, Ecol-RC(DNA), Ecol-RC(LNA-5′), Ecol-RC(LNA-3′) or Ecol-RC(LNA-Even) (Table 1). The mixture was heat-denatured at 95°C for 3 min then immediately cooled on ice. Subsequently, 10 μl of phi29 DNA polymerase reaction buffer containing 0.2 mM each of dNTP, 200 μg ml^{-1} BSA and 10 U of Phi29 DNA polymerase (New England Biolabs) was added, and the RCA reaction was performed at 30°C for 18 h. The enzyme was then subjected to thermal inactivation at 65°C for 10 min. DNA in the reaction mixture was purified by using YM-100 and subjected to IPCR. Inverse PCR and the analysis of IPCR products were carried out as described above.

In the third series (Fig. 1), RCA followed by IPCR of the lambda-phage sequence was carried out in the presence of excess competing DNA. The competing DNA was prepared by mixing the equal amounts of calf thymus DNA, herring sperm DNA and salmon sperm DNA (WAKO, Osaka, Japan). The *E. coli* DNA in an amount between 0.005 and 500 ng was mixed with 400 ng of the competing DNA, digested with HindIII and self-circularized as described above. Amounts equivalent to 1.0×10^6, 1.0×10^5, 1.0×10^4, 1.0×10^3, 1.0×10^2 and 1.0×10^1 copies of the lambda sequence were used for RCA followed by IPCR as described above. All the experiments were repeated three times.

Screening of genes encoding glycosyl hydrolase from environmental DNA

To amplify partial fragments of genes encoding GHF10, 10–250 ng of each DNA isolated from digestive tracts of horses or termites were added to 50 μl of EX *Taq* Buffer (TaKaTa-Bio) containing 0.2 mM each of dNTPs, 1 μM each of primers, XYN-Fw and XYN-Rv (Table 3), and 5 U of Ex *Taq* HotStart version (TaKaRa-Bio). Polymerase chain reaction was performed with the following cycling programme: 1 cycle of 30 s at 95°C; 30 cycles of 30 s at 95°C, 30 s at 55°C and 15 s at 72°C; and 7 min of final extension at 72°C. The PCR product was purified, ligated to the pCR®4-TOPO® Vector with a TOPO TA cloning kit (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol, and used to transform *ECOS*™-competent *E. coli* DH5α cells (Nippongene). Transformants were plated on LB agar containing 50 μg ml^{-1} ampicillin. Colonies were picked up, suspended in 1.3 ml of LB medium in the wells of a deep-well 96-well plate (Qiagen, Hilden, Germany) and grown at 37°C for 20 h. Cells in the cultures were harvested, and plasmid DNA was isolated by using a QIAprep® 96 Turbo BioRobot Kit (Qiagen) and a Qiagen Biorobot 8000 (Qiagen). DNA sequencing was performed by using M13-reverse and forward primers and a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), and analysed by using an ABI3130 Genetic Analyser. Sequence data were assembled by using ATGC software (Applied Biosystems).

Cloning of full-length genes for glycosyl hydrolases

The flanking sequences of genes for glycosyl hydrolase family 10 were cloned by the following procedures. Approximately 1 μg of DNA isolated from the digestive tracts was digested with 10 U of a BamHI, EcoRI, HindIII, PstI or XbaI restriction enzyme (TaKaRa-Bio) at 37°C for 5 h, then incubated at 70°C for 15 min to inactivate the restriction enzymes. The digested DNA was purified using YM-100 and self-circularized using a Mighty Mix DNA ligation kit (TaKaRa-Bio) at 16°C overnight in 100 μl of the ligation buffer. The self-circularized DNA was precipitated with ethanol and suspended in 500 μl of sterilized Milli-Q water. Next, 5 μl of the DNA solution (*c.* 10 ng) and LNA oligonucleotide [Xyn-RC(LNA-5′), Xyn-RC(LNA-3′) or Xyn-RC(LNA-Even)] (Table 1) were mixed with Phi29 DNA polymerase buffer to a final volume of 40 μl. Rolling-circle amplification reaction was carried out under the conditions described in the previous section. The RCA product was precipitated, suspended in 5 μl of sterilized Milli-Q water, and mixed with 45 μl of GC Buffer I containing 0.2 mM dNTP, 0.2 μM each of two IPCR primers (Xyn-0005LF plus Xyn-0005LR, Xyn-0007LF plus Xyn-0007LR, Xyn-0022LF plus Xyn-0022LR, Xyn-0101LF plus Xyn-0101LR or Xyn-0102LF plus Xyn-0102LR; Table 1) and 2.5 U of LA *Taq* DNA polymerase (TaKaRa-Bio). Inverse PCR was performed under the following conditions: 1 min of denaturation at 94°C, 30 cycles of 30 s at 94°C and 10 min at 60–68°C, and 7 min of final extension at 72°C. Products were analysed with an Agilent 2100 Bioanalyser with a DNA7500 LabChip® kit (TaKaRa-Bio). The resulting IPCR products were cloned, and plasmid DNAs containing inserts were sequenced as described above. For clones with longer inserts, complete sequencing was achieved by primer walking. Open reading frame (ORF) analysis was performed by using the GENETYX software package (Genetyx Corporation, Tokyo, Japan). Putative ORFs were translated and used as queries for BLASTP searches against the NCBI protein database.

Nucleotide sequence accession numbers

The nucleotide sequence data of the ORFs each encoding a putative xylanase, *xynVA1*, *xynVA2*, *xynIO1*, *xynIO2* or *xynIO4*, have been submitted to the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) and the GenBank® databases with the Accession No. AB269880 and AB359057 to AB359060 respectively.

Acknowledgements

This study was supported by the Grant No. 04000182-0 from the New Energy and Industrial Technology Development Organization (NEDO). We thank Dr Akira Sato for help in the sampling of termites and the isolation of their guts.

References

Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W., Chang, S.H., *et al.* (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* **72:** 3291–3301.

Arnold, C., and Hodgson, I.J. (1991) Vectorette PCR: a novel

approach to genomic walking. *PCR Methods Appl* **1:** 39–42.

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

Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11:** 1095–1099.

Di Giusto, D.A., and King, G.C. (2004) Strong positional preference in the interaction of LNA oligonucleotides with DNA polymerase and proofreading exonuclease activities: implications for genotyping assays. *Nucleic Acids Res* **32:** e32.

Eschenfeldt, W.H., Stols, L., Rosenbaum, H., Khambatta, Z.S., Quaite-Randall, E., Wu, S., *et al.* (2001) DNA from uncultured organisms as a source of 2,5-diketo-D-gluconic acid reductases. *Appl Environ Microbiol* **67:** 4206–4214.

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.

Gonzalez, J.M., Portillo, M.C., and Saiz-Jimenez, C. (2005) Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. *Environ Microbiol* **7:** 1024–1028.

Hayashi, H., Abe, T., Sakamoto, M., Ohara, H., Ikemura, T., Sakka, K., and Benno, Y. (2005) Direct cloning of genes encoding novel xylanases from the human gut. *Can J Microbiol* **51:** 251–259.

Hayashi, K., Nakazawa, M., Ishizaki, Y., Hiraoka, N., and Obayashi, A. (1986) Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucleic Acids Res* **14:** 7617–7631.

Hummelshoj, L., Ryder, L.P., Madsen, H.O., and Poulsen, L.K. (2005) Locked nucleic acid inhibits amplification of contaminating DNA in real-time PCR. *BioTechniques* **38:** 605–610.

Jiang, Z., Zhang, X., Deka, R., and Jin, L. (2005) Genome amplification of single sperm using multiple displacement amplification. *Nucleic Acids Res* **33:** e91.

Kilstrup, M., and Kristiansen, K.N. (2000) Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res* **28:** E55.

Laghi, L., Randolph, A.E., Malesci, A., and Boland, C.R. (2004) Constraints imposed by supercoiling on *in vitro* amplification of polyomavirus DNA. *J Gen Virol* **85:** 3383–3388.

Latorra, D., Arar, K., and Hurley, J.M. (2003) Design considerations and effects of LNA in PCR primers. *Mol Cell Probes* **17:** 253–259.

Levin, J.D., Fiala, D., Samala, M.F., Kahn, J.D., and Peterson, R.J. (2006) Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers. *Nucleic Acids Res* **34:** e142.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., and Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* **19:** 225–232.

Lorenz, P., and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* **3:** 510–516.

Monstein, H.J., Olsson, C., Nilsson, I., Grahn, N., Benoni, C., and Ahrne, S. (2005) Multiple displacement amplification of DNA from human colon and rectum biopsies: bacterial profiling and identification of *Helicobacter pylori*-DNA by means of 16S rDNA-based TTGE and pyrosequencing analysis. *J Microbiol Methods* **63:** 239–247.

Mullis, K.B., and Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155:** 335–350.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120:** 621–623.

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

Raghunathan, A., Ferguson, H.R., Jr, Bornarth, C.J., Song, W., Driscoll, M., and Lasken, R.S. (2005) Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* **71:** 3342–3347.

Rector, A., Tachezy, R., and Van Ranst, M. (2004) A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *J Virol* **78:** 4993–4998.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66:** 2541–2547.

Sanjay, K.S., Nielsen, P., Koshkin, A.A., and Wengel, J. (1998) LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem Commun* **4:** 455–456.

Silahtaroglu, A.N., Tommerup, N., and Vissing, H. (2003) FISHing with locked nucleic acids (LNA): evaluation of different LNA/DNA mixmers. *Mol Cell Probes* **17:** 165–169.

Stokes, H.W., Holmes, A.J., Nield, B.S., Holley, M.P., Nevalainen, K.M., Mabbutt, B.C., and Gillings, M.R. (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl Environ Microbiol* **67:** 5240–5246.

Sunna, A., and Bergquist, P.L. (2003) A gene encoding a novel extremely thermostable 1,4-beta-xylanase isolated directly from an environmental DNA sample. *Extremophiles* **7:** 63–70.

Tang, K., Utairungsee, T., Kanokratana, P., Sriprang, R., Champreda, V., Eurwilaichitr, L., and Tanapongpipat, S. (2006) Characterization of a novel cyclomaltodextrinase expressed from environmental DNA isolated from Bor Khleung hot spring in Thailand. *FEMS Microbiol Lett* **260:** 91–99.

Triglia, T., Peterson, M.G., and Kemp, D.J. (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res* **16:** 8186.

Uchiyama, T., and Watanabe, K. (2006) Improved inverse PCR scheme for metagenome walking. *BioTechniques* **41:** 183–188.

Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005)

Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* **23:** 88–93.

Vester, B., and Wengel, J. (2004) LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* **43:** 13233–13241.

Wahlestedt, C., Salmi, P., Good, L., Kela, J., Johnsson, T., Hokfelt, T., *et al.* (2000) Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proc Natl Acad Sci USA* **97:** 5633–5638.

Yokouchi, H., Fukuoka, Y., Mukoyama, D., Calugay, R., Takeyama, H., and Matsunaga, T. (2006) Whole-metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using phi29 polymerase. *Environ Microbiol* **8:** 1155–1163.

