J. Microbiol. Biotechnol. (2006), **16**(1), 153–157

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Sequence-Based Screening for Putative Polyketide Synthase Gene-Harboring Clones from a Soil Metagenome Library

JI, SANG CHUN, DOCKYU KIM, JUNG-HOON YOON, TAE-KWANG OH, AND CHOONG-HWAN LEE*

Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-333, Korea

Received: March 9, 2005
Accepted: April 26, 2005

Abstract A soil metagenomic library was constructed using an *E. coli*-fosmid cloning system with environmental DNAs extracted from Kwangreung forest topsoil. We targeted the genes involved in the biosynthesis of bacterial polyketides. Initially, a total of 36 clone pools (10,800 clones) were explored by the PCR-based method using the metagenomic DNAs from each pool and a degenerate primer set, which has been designed based on the highly conserved regions among ketoacyl synthase (KS) domains in actinomycete type I polyketide synthases (PKS Is). Six clone pools were tentatively selected as positive and further examined through a hybridization-based method for selecting a fosmid clone containing PKS I genes. Colony hybridization was performed against fosmid clones from the 6 positive pools, and finally 4 clones were picked out and confirmed to contain the conserved DNA fragment of KS domains. In this study, we present a simple and feasible sorting method for a desired clone from metagenomic libraries.

Key words: Soil metagenome, sequence-based screening, polyketide synthase

Secondary metabolites produced by microbes have been important sources of bioactive molecules such as antibiotics and other pharmaceutical agents over the last 60 years. However, the discovery rate of novel biomolecules using traditional cultivation techniques has significantly decreased during the past couple of years [21]. Recent study in molecular microbiology has shown that more than 99% of the microorganisms present in many natural environments are not readily culturable with currently available techniques [2]. Thus, these uncultured microorganisms are an unexplored reservoir of novel strains that may produce novel natural products [6, 12]. A culture-independent method developed to exploit the genomes of such uncultured majority is the metagenome technology [14], which has provided access to the untapped genomes of unculturable microorganisms.

Microbes produce useful secondary metabolites diverse in both their structure and function, polyketides of which exhibit pharmacologically important activities; for example, antibacterial, antifungal, antiparasitic, antitumor, and agrochemical properties. These polyketides are synthesized through successive condensation of simple building blocks such as acetate, propionate, or butyrate, similar to fatty acid biosynthesis [8–10, 17]. Macrocyclic polyketides are known to be produced principally from *Streptomyces* strains and their related actinomycetes by polyketide synthases (PKSs). Of the three major types of PKSs, type I PKSs (PKS Is) are large multifunctional enzymes that are composed of a succession of modules [3, 5, 16] and catalyze the biosynthesis of most of the non-aromatic polyketides [15] such as erythromycin [13], rifamycin [4, 11], epothilone [20], and myxothiazol [18]. Since ketoacyl synthase (KS) domains are highly conserved among the various domains organizing PKS I type modules, a DNA region of KS has been used as a probe to detect PKS I genes [5, 7].

In the present study, soil samples were collected from Kwangreung forest topsoil (Kyeonggi-do, Korea), and soil DNAs were directly extracted from the samples and further purified through pulsed-field gel electrophoresis, as previously described [14]. A metagenomic library was constructed using an EPIFos™ Fosmid Library Production Kit (Epicentre, U.S.A.), as described by the manufacturer. After calculating the efficiency of *E. coli* EPI100 transformation, chloramphenicol-resistent transformants were divided into a total of 75 metagenomic clone pools in cryotubes, each containing approximately 300 fosmid clones. Through this storage method, we were able to avoid the tedious and time-consuming step to transfer to 96-well microtiter plates. The library was deposited at Microbial Resources Bank, Microbial Genomics and Applications Center at Daejeon, Korea (CAT No.G200104-G200179) by Dr. Seon-Woo

*Corresponding author
Phone: 82-42-860-4294; Fax: 82-42-879-8595;
E-mail: chlee@kribb.re.kr

Lee (Korea Research Institute of Chemical Technology, Daejeon).

Among 75 metagenomic clone pools, 36 clone pools (approximately 10,800 clones) were randomly selected for screening. The fosmid clones grown on LB plate were initially screened for the ability to produce any colored metabolites or to induce a shape change in the *E. coli* host cells. No clone, however, was detected that had the ability to catalyze either reaction. Since the function-based analytic method on plate was not successful in locating the desired clones, the sequence-based approach, such as PCR and colony hybridization, was taken to screen a clone encoding new biosynthetic enzymes with biotechnological potential, in particular PKS Is.

Since PKS Is have the great potential to synthesize useful natural products that are attracting much interest in the biotechnology field, we targeted the genes involved in the biosynthesis of bacterial polyketides, and tried to obtain the gene clusters. PCR amplification was carried out with GeneAmp PCR system 9700 (Perkin-Elmer, Norwalk, CT, U.S.A.). Template DNAs were extracted from 36 clone pools using the Wizard® Plus SV Minipreps DNA Purification System (Promega, Madison, WI, U.S.A.), and custom primers were supplied by Genotec (Daejeon, Korea). The PCR reaction was performed in 20 μl of reaction mixture containing the purified DNA (100 to 300 ng), 10 pmole each of primers, and AccuPower PCR Premix (Bioneer, Daejeon, Korea), under the following conditions: an initial denaturation step at 95°C for 15 min, and 30 cycles of 1 min of denaturation at 95°C, 1 min of hybridization at 65°C, and 1 min of elongation at 72°C, and a final 10 min of extension at 72°C. The primers have previously been designed based on the highly conserved active site of the KS domains that are present in type I PKSs of several actinomycetes [5]: set 1-sense, 5'-CCSCAGSAGCGCST-STTSCTSGA-3' and set 1-antisense, 5'-GTSCCSGTSCC-GTGSGTSTCSA-3'. In our experiments, application of the primers produced ~700 bp fragments in PCR with genomic DNAs from 4 PKS I-producing *Streptomyces* sp. strains, but not from 2 *Bacillus* sp. strains that are known to contain PKS genes nonhomologous to actinomycete PKS Is [5]. Using the primer set confirmed to be specific to actinomycete PKS I target genes, PCR screening was performed with the DNA templates extracted from a total of 36 clone pools. As a result, 6 clone pools (designated #4, 11, 18, 20, 23, and 36) were selected as PCR-positive. The PCR-positive clones produced PCR products of the same size (~700 bp) as in the case of PKS I-producing *Streptomyces* strains, with each band showing different DNA intensity on agarose gel. This result indicates that these clone pools should contain various KS regions derived from actinomycete PKS I genes. The PCR products were purified from an agarose gel with a Geneclean Turbo kit (Q-BIO, Carlsbad, CA, U.S.A.) and then cloned by using the pGEM-T Easy Vector System (Promega). After transformation, 5 recombinant *E. coli* clones were randomly selected from 3 PCR-positive clone pools (# 4, 11, and 20), and the insert DNAs were sequenced. The sequence analysis, using BLAST program [1] against the GenBank database, revealed that the sequences encoded the highly conserved region corresponding to the active site of the KS domains in PKS Is from actinomycetes (*Streptomyces coelicolor*) and cyanobacteria (*Nostoc* sp.) (data not shown). These results confirmed the assumption that the PCR-positive clone pools possess at least one clone containing PKS I genes from diverse origins.

The following next step for selecting a promising clone was colony hybridization against these 6 clone pools. The PKS I-positive clone pools were grown on LB agar plates with chloramphenicol (25 μg/ml), transferred on membranes, and hybridized with PCR products, which were previously confirmed to contain various KS gene fragments and DIG-labeled using the DIG High Prime DNA Labeling Kit (Roche, Indianopolis, IN, U.S.A.). Colony hybridization was performed as recommended by the supplier of the DIG Detection System (Roche). As a result, a total of 4 putative positive clones (designated PKS4 from #4, PKS11 from #11, PKS20 from #20, and PKS36 from #36) were selected. For example, the PKS20 clone made a little deeper violet spot than false spots that were dominantly present on the membrane. However, we could not find any clone in 2 PCR-positive clone pools (#18 and #23) in spite of several tries. Since both #18 and #23 produced a weak DNA band, compared with the rest of the clone pools, this failure was thought to be due to nonspecific binding of the degenerate primers in the PCR step.

Finally, to determine whether the clones have conserved KS regions, PCR amplification was carried out with the

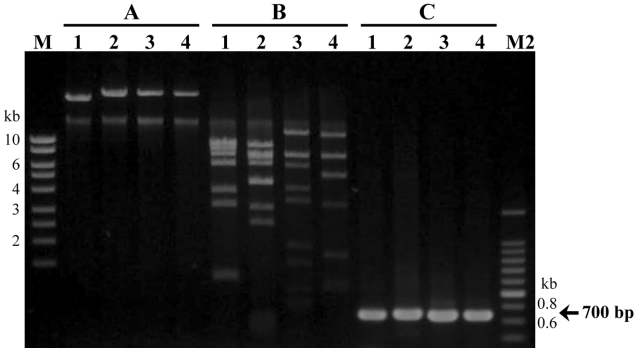

Fig. 1. Nucleotide sequence-based confirmation of the 4 putative positive clones selected by colony hybridization.
(A) Fosmid DNAs extracted from putative positive clones, (B) RFLP profiles resulting from the fosmid DNAs treated with *Bam*HI, and (C) PCR products amplified with the fosmid DNAs and the degenerate primers (set 1-sense and set 1-antisense). M, DNA size marker (1.5–10 kb); M2, DNA size marker (0.4–3 kb); lane 1, PKS4; lane 2, PKS11; lane 3, PKS20; lane 4, PKS36.

Fig. 2. ClustalX alignment of the predicted amino acid sequences of 4 PKS I clones.

The PKS I consensus sequence pfam00109 (http://www.sanger.ac.uk/Software/Pfam) is aligned together. Four PKS I genes and their identities were as follows: PKS4, 62% identity to the *Nostoc punctiforme* PCC73102 polyketide synthase gene (226-amino-acid[aa] alignment; GenBank accession number ZP_00110270.1); PKS11, 60% identity to the *Lyngbya majuscula* CurI gene (225-aa alignment; GenBank accession number AAT70104.1); PKS20, 53% identity to the *Nostoc* sp. strain PCC7120 polyketide synthase gene (224-aa alignment; GenBank accession number AI2140); PKS36, 63% identity to the *Stigmatella aurantiaca* MtaD gene (228-aa alignment; GenBank accession number AAF19812.1). *, Identity; :, strong similarity; ., weak similarity.

fosmid DNAs, showing different *Bam*HI-restriction profiles, extracted from the 4 putative clones (Fig. 1). The PCR products (~700 bp) were cloned into pGEM-T vector and sequenced. The analysis of the predicted protein sequences using the BLAST [1] and SEARCHPKS [22] programs (NCBI and National Institute of Immunology in India, respectively) revealed that all of the obtained sequences encoded the highly conserved KS domain region in PKS I from known microorganisms. Although the primer set used was derived from actinomycetes, high identity values (53–63%) were observed with PKS sequences from cyanobacteria (*Nostoc* and *Lyngbya*) and myxobacteria (*Stigmatella aurantiaca*) (Fig. 2).

Generally, considering transformation efficiency and cloning capacity, metagenomic libraries have frequently been generated using *E. coli* and fosmid or cosmid as the host-vector system, which is preferred in cloning of huge gene clusters required for a specific function, as in PKS genes [7, 16]. For example, the functional PKS system involved in biosynthesis of antifungal pimaricin is composed of 16 ORFs, spanning approximately 85 kb in the *Streptomyces natalensis* genome, which constitute 5 giant multienzyme proteins and 11 additional genes governing modification of the polyketide-derived framework and export [3]. When constructing a metagenomic library, environmental DNAs are apt to be broken into smaller fragments during the DNA isolation, purification, and finally cloning. Thus, the clones with these shortened DNAs would be completely inactive or partially functional at best, producing an intermediate in the course of assembly of a final product. Although PKS gene clusters were cloned, some additional factors would account for a low possibility to pick out an active clone from several hundred thousand clones: a lack of efficient transcription of the metagenome-derived genes in the *E. coli* host, unfolding or incorrect folding of the desired proteins in the host, the absence of factors required

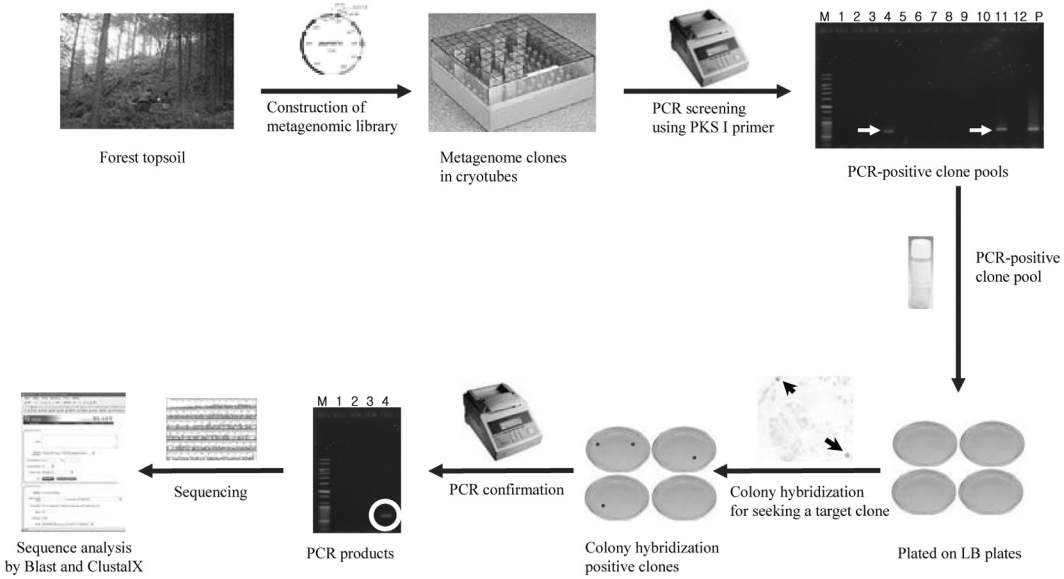

Fig. 3. Experimental setup for screening clones containing PKS genes from a soil metagenomic library (see the text for a detailed description).

for synthesis and folding of the foreign proteins, and so on [19].

It is well known that more than 99% of microbes in nature are uncultivatable with currently available technology. Therefore, the genomes of uncultivated microbes may well represent an unlimited resource for novel genes and natural products. Metagenome technology could lead us to the genomes of such uncultivated majority without the need for cultivation. For accessing the uncovered reservoir of secondary metabolite producers in nature, we constructed a soil metagenomic library that was culture-independent, and targeted the genes encoding enzymes involved in the biosynthesis of polyketides. Since the presently used screening methods demand several laborious and time-consuming steps, we tried to use simple sequence-based screening techniques (PCR and colony hybridization, Fig. 3), and successfully obtained 4 clones that harbor putative PKS I gene fragments, demonstrating that the present work provides a simple and feasible method that is able to sort target clones in a wide variety of metagenome libraries.

Acknowledgment

This research was supported by a grant to Dr. C.-H. Lee from the Microbial Genomics and Applications Center of the 21st Century Frontier Research Program funded by the Ministry of Science and Technology of the Republic of Korea.

References

1. Altschul, S., T. Madden, A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.
2. Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59:** 143–169.
3. Aparicio, J. F., R. Fouces, M. V. Mendes, N. Olivera, and J. F. Martin. 2000. A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. *Chem. Biol.* **11:** 895–905.
4. August, P. R., L. Tang, Y. J. Yoon, S. Ning, R. Müller, T.-W. Yu, M. Taylor, D. Hoffmann, C.-G. Kim, X. Zhang, R. Hutchinson, and H. G. Floss. 1998. Biosynthesis of the ansamycin antibiotic rifamycin: Deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **5:** 69–79.
5. Courtois, S., C. M. Cappellano, M. Ball, F. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, and J. Pernodet. 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* **69:** 49–55.
6. Daniel, R. 2004. The soil metagenome - a rich resource for the discovery of novel natural products. *Curr. Opin. Biotech.* **15:** 199–204.
7. Ginohac, A., C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tuphile, H. Bertrand, T. M. Vogel, G. Perriere, P. Simonet, and R. Nalin. 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* **70:** 5522–5527.
8. Hong, J. S. J., C. Y. Choi, and Y. J. Yoon. 2003. Premature release of polyketide intermediates by hybrid polyketide synthase in *Amycolatopsis mediterrnei* S699. *J. Microbiol. Biotechnol.* **13:** 613–619.
9. Hopwood, D. A. and D. H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **24:** 37–66.
10. Kim, C.-Y., H.-J. Park, and U.-S. Kim. 2003. Heterologous expression of hybrid type II polyketide synthase system in *Streptomyces* species. *J. Microbiol. Biotechnol.* **13:** 819–822.
11. Lee, S. K., C. Y. Choi, J. S. Ahn, K. Y. Cho, C. S. Park, and Y. J. Yoon. 2004. Identification of a cytochrome P450 hydroxylase gene involved in rifamycin biosynthesis by *Amycolatopsis mediterranei* S699. *J. Microbiol. Biotechnol.* **14:** 356–362.
12. Peric-Concha, N. and P. F. Long. 2003. Mining the microbial metabolome: A new frontier for natural product lead discovery. *Drug Discov. Today* **23:** 1078–1084.
13. Reeves, A. R., R. S. English, J. S. Lampel, D. A. Post, and T. J. Vanden Boom. 1999. Transcriptional organization of the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea*. *J. Bacteriol.* **181:** 7098–7106.
14. Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, L. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66:** 2541–2547.
15. Santi, D. V., M. A. Siani, B. Julien, D. Kupfer, and B. Roe. 2000. An approach for obtaining perfect hybridization probes for unknown polyketide synthase genes: A search for the epothilone gene cluster. *Gene* **247:** 97–102.
16. Schwecke, T., J. F. Aparicio, I. Molnar, A. Konig, L. E. Khaw, S. F. Haycock, M. Oliynyk, P. Caffrey, J. Cortes, J. B. Lester, G. A. Bohm, J. Staunton, and P. F. Leadlay. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **92:** 7839–7843.
17. Seok, J. W., E. S. Kim, H. Y. Kang, C. Y. Choi, D. H. Sherman, and Y. J. Yoon. 2003. Site-directed mutagenesis on putative macrolactone ring size determinant

in the hybrid pikromycin-tylosin polyketide synthase. *J. Microbiol. Biotechnol.* **13:** 823–827.
18. Silakowski, B., H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blocker, G. Hofle, S. Beyer, and R. Muller. 1999. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274:** 37391–37399.
19. Streit, W. R. and R. A. Schmitz. 2004. Metagenomics: The key to the uncultured microbes. *Curr. Opin. Microbiol.* **7:** 492–498.
20. Tang, L., S. Shah, L. Chung, J. Carney, L. Katz, C. Khosla, and B. Julien. 2000. Cloning and heterologous expression of the epothilone gene cluster. *Science* **287:** 640–642.
21. Watve, M. G., R. Tickoo, M. M. Jog, and B. D. Bhole. 2001. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176:** 386–390.
22. Yadav, G., R. S. Gokhale, and D. Mohanty. 2003. SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* **31:** 3654–3658.