PROSPECTING METAGENOMIC ENZYME SUBFAMILY GENES FOR DNA FAMILY SHUFFLING BY A NOVEL PCR-BASED APPROACH
Qiuyan Wang 1, Huili Wu 1, Anming Wang 1, Pengfei Du 1, Xiaolin Pei 1, Haifeng Li 1, Xiaopu Yin 1, Lifeng Huang 1, Xiaolong Xiong 2
Center for Biomedicine and Health 1, College of Life Sciences 2, Hangzhou Normal University; Hangzhou 310012, P. R. China
Running head: TMGS-PCR based DNA family shuffling
Address correspondence to: Qiuyan Wang, Center for Biomedicine and Health, Hangzhou Normal University, No. 222 Wenyi Road, Hangzhou 310012, Zhejiang, China. Fax: (86)571-28865630, E-mail: qywanzhnu@gmail.com.

DNA family shuffling is a powerful method for enzyme engineering, which utilizes recombination of naturally occurring functional diversity to accelerate laboratory-directed evolution. However, the use of this technique has been hindered by the scarcity of family genes with the required level of sequence identity in the genome database. We describe here a strategy for collecting metagenomic homologous genes for DNA shuffling from environmental samples by truncated metagenomic gene specific PCR (TMGS-PCR). Using identified metagenomic gene specific primers, twenty-three 921 bp truncated lipase gene fragments, which shared 64 to 99% identity with each other and formed a distinct subfamily of lipases, were retrieved from 60 metagenomic samples. These lipase genes were shuffled, and selected active clones were characterized. The chimeric clones show extensive functional and genetic diversity, as demonstrated by functional characterization and sequence analysis. Our results indicate that homologous sequences of genes captured by TMGS-PCR can be used as suitable genetic material for DNA family shuffling with broad applications in enzyme engineering.

Directed evolution mimics the processes of Darwinian evolution in a test tube, combining random mutagenesis and/or recombination with screening or selection for enzyme variants that have the desired properties (1). Directed evolution through random mutagenesis of a single starting sequence can result in significant improvement toward the required function. However, this strategy has the disadvantage of having to screen an extremely high number of candidates to find the rare positive mutants that are needed as the genetic material for DNA shuffling (5). A more potent directed evolution strategy is known as DNA family shuffling; it refers to the recombination of equivalent genes from natural homologous families rather than random mutagenesis of a single gene (6). This approach takes advantage of the fact that most of the deleterious mutations have long ago been removed by natural selection, while diverse function has been created by the positive mutation interchange. The reassortment of proven mutations yields a higher frequency of functional progeny sequences, and because multi-gene shuffling starts with more than one parental sequence, it accesses a broader range of progenitor combinations. These attributes make the approach more efficient, reducing loss-of-function mutations dramatically so that fewer progeny molecules need to be screened to discover superior performers (7-9).

Although DNA family shuffling has the above-mentioned advantages over random
mutagenesis, there are only dozens of published reports using it, unlike the thousands published that use random mutagenesis. The modest number of published cases of DNA family shuffling is indicative of its limitation: reliance on natural homologous gene materials. It is always frustrating searching for qualified homologous sequences in the genome database. Recently, several techniques for the combinatorial generation of protein chimeras in regions of low homology have been developed that can circumvent homology dependence for a DNA shuffling process. These techniques include non-homologous random recombination (10), sequence homology-independent protein recombination (11) and incremental truncation for the creation of hybrid enzymes (ITCHY) (12-14). However, random non-homologous recombination results in the creation of libraries containing a large number of out-of-frame or otherwise inactive clones. The theoretical size of such libraries renders them unsuitable for manual screening because a vast number of clones must be interrogated to identify highly active enzymes (14). However, Castle et al. have provided an alternative and ultimately complementary solution to this problem: building synthetic libraries using a subset of diversity from an alignment (15). This method has been proven to be successful for providing sufficient genetic diversity for directed enzyme evolution, even based on information from alignments with few genes of low homology. However, the need to develop more alternatives to collecting natural homologous genes efficiently for DNA family shuffling remains urgent.

In the past decade, advances in the field of 'metagenomics' have dramatically revised our view of biodiversity (16). Considering the estimation that >99% of microorganisms in most environments are not amenable to culturing, very little is known about their genomes. The isolation, archiving and analysis of environmental DNA (or so called ‘metagenomes’) have enabled us to mine microbial diversity, allowing us to access their genomes (17-20). Two strategies are generally used to screen and identify novel genes from metagenomic libraries: function-based and sequence-based screening. In function-based screening, clones expressing desired traits are selected from libraries, and aspects of the molecular biology and biochemistry of the active clones are analyzed. Conversely, sequence-based screening is not dependent on the expression of cloned genes in heterologous hosts. Generally, it is based on the conserved DNA sequences of target genes. Hybridizations or PCRs are performed based on the deduced DNA consensus (21).

We report here the development of a homologous gene collection method termed “truncated metagenomic gene specific PCR” (TMGS-PCR). This approach led to success in tapping into homologous sequences of enzymes from environmental DNA with PCR with a designated metagenomic starting gene and experimentally validated truncated gene specific primers (Fig. 1). Sequences analysis and functional characterization of the resulting shuffled library demonstrated that TMGS-PCR can provide homologous genes with great genetic diversity for DNA family shuffling.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains, plasmids and growth conditions** - E. coli DH5α and pBluescript SK+ (Stratagene, La Jolla, CA, USA) were used as host and vector, respectively, for cloning of the lipase. E. coli strain BL21-CodonPlus (DE3)-RIPL (Stratagene, La Jolla, CA, USA) and pET-24a(+) (Novagen, Inc., Madison, WI) were used as host and vector, respectively, for heterologous expression of the lipase. Plasmid pMD18-T (TaKaRa Ltd., Otsu, Japan) was used for gene cloning and sequencing. E. coli cells were routinely grown in Luria–Bertani (LB)
medium at 37°C, supplemented with 100 µg/ml ampicillin, 50 µg/ml streptomycin sulfate, 10 µg/ml tetracycline HCl and 12.5 µg/ml chloramphenicol, if needed.

**Library construction** - A total of 60 environmental samples were collected from terrestrial microenvironments in China and were named as S01-60. DNA extraction from samples was carried out using the PowerSoil® DNA Isolation Kit (MOBIO, Carlsbad, CA, USA) according to the manufacturer's instructions. Electrophoresis for DNA from environmental samples was performed with a Bio-Rad CHEF DR II system (Bio-Rad, Richmond, CA, USA). A metagenomic library of the S05 sample was constructed using the approximately 40 kb end-repaired DNA fragments and CopyControl™ Fosmid Library Production Kit (Epicentre Technologies, Madison, WI, USA) using protocols provided by the manufacturer.

**Screening and subcloning** - For detection of lipolytic activity, transformed cells were spread immediately after electroporation on LB agar containing 1% (w/v) tributyrin (Merck, Darmstadt, Germany) and incubated overnight at 37°C. After 3–5 days at room temperature or 4°C, colonies were monitored for the presence of clear halos (22). Sau3A I partially digested fosmid DNA from the positive clone was ligated to BamHI-linearized and dephosphorylated pBluescript SK+ vector, and the ligation mixture was used to transform E. coli DH5α. The new clones were examined on the same type of indicator plates for lipase activity. A tributyrin-positive clone was selected and sequenced. The gene encoding LipS05 was amplified by PCR using Lip24aF and Lip 24aR primers, 5′-CGCCATATGATGGAATTTATCCCTG-3′ and 5′-TAGCTCGAGTTGGCACGCAGCG-3′, which resulted in Nde I and Xho I sites at the 5′ and 3′ ends, respectively. The digested PCR fragment was inserted between the Nde I and Xho I sites of pET-24a(+), creating pET-24a-LipS05. The recombinant plasmid was transformed into competent cells of E. coli strain BL21-CodonPlus (DE3)-RIPL using the CaCl₂ method following standard protocols.

**Homologous gene amplification** - Sets of primers, which were assigned designations comprising the first three letters Lip referring to lipase and the sequence length (bp) of the product, were used to amplify related lipase fragments from environmental DNA samples. The product was amplified using a BIO-RAD DNA Engine Peltier Thermal Cycler. The conditions were as follows. Thirty cycles of reactions were performed under the conditions of denaturation at 94°C for 1 min, annealing at 58°C for 1 min, and an extension step at 72°C for 1 min. This sequence was repeated 35 times followed by a 10-min final extension step at 72°C. All PCR products were cloned into the TA vector and confirmed by DNA sequencing.

**Sequences analysis** - The Lipase Engineering Database (http://www.led.uni-stuttgart.de/) was searched using the BLASTP program (23). Multiple sequence alignments were constructed using the CLUSTAL-X program (24). To determine the evolutionary relationship of environmental lipases with other lipases, phylogenetic analysis was performed with the amino acid sequences of environmental lipases and 50 lipases obtained from Lipase Engineering Database using neighbor-joining methods with 1000 bootstrap replications performed by MEGA version 4.0 (25).

**DNA shuffling** - The chimeric library was constructed using a modified DNA shuffling method (26). Twenty-four ~1.3 kb DNA chimeric parent genes containing the LipS01-50 truncated fragments were amplified using Lip24aF/R primers. The reaction mixture (50 µl) contained 1 µl of Pfu polymerase (2.5 units/µl) (Stratagene, La Jolla, CA), 1 µl of 20 µM each primer, 30 ng of each plasmid DNA, 5 µl of
dNTP (2.5 mM each) and 5 µl of 10x Pfu buffer. The resulting parent DNA fragments were then digested with bovine pancreas DNase I (TaKaRa Ltd., Otsu, Japan) as follows. A 25 µl volume of solution containing 0.05 µg of each parent DNA was mixed with 6.25 µl of 0.2 M Tris-HCl (pH 7.5), 0.25 µl of 10 mg/ml BSA and 0.25 µl of 0.1 M MnCl₂. Digestion was initiated with the addition of 2.5 µl of DNase I (0.5 unit/ml) into the mixture at 37°C. Following incubation for 30 min, the digestion was stopped by adding 0.5 M EDTA to a final concentration of 50 mM and heat inactivating at 95°C for 10 min. The digested fragments were separated by gel electrophoresis. The desired 100-200 bp DNA fragments were isolated and purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA).

Reassembly of the DNase I digested fragments was conducted in a 50 µl reaction mixture containing 37 µl of fragment DNA (~3 µg), 5 µl of 10x Pfu buffer, 8 µl of dNTP (2.5 mM each) and 1 µl (2.5 units) of Pfu polymerase. A progressive hybridization PCR cycling program was used for reassembly. The reassembled reaction mixture (1 µl) along with primers Lip²⁴ᵃ⁻¹⁻ᵣ was used to amplify the full-length genes using PCR. Then the NdeI-XhoI restriction fragment was inserted into pET-24a(+) at the corresponding sites, and the resulting plasmid was used to transform E. coli BL21-CodonPlus (DE3)RIPL by electroporation.

Expression, refolding and purification - A single colony of selected mutants was used to inoculate 5 mL LB, containing 50 µg/ml kanamycin, 50 µg/ml streptomycin sulfate, 10 µg/ml tetracycline HCl and 12.5 µg/ml chloramphenicol, and the culture was agitated at 225 rpm overnight at 37°C. The cells were grown in 1 L of LB media at 37°C until OD₆₀₀ nm=0.8, then protein expression was induced using 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the incubation temperature was lowered to 20°C for 20 h.

Inclusion bodies were solubilized using the protein refolding kit (Novagen), according to the manufacturer’s protocol, with slight modifications. E. coli BL21 cells expressing lipases were harvested, resuspended in 0.1 culture volume of wash buffer (20 mM Tris-HCl pH 7.5/10 mM EDTA/1% Triton X-100), and broken by sonication. After centrifugation, the pellet was washed twice with 0.1 culture volumes of wash buffer and then solubilized at room temperature in solubilization buffer (0.1 M glycine, pH 11/0.3% N-lauroylsarcosine) at a concentration of 5 to 10 mg/ml wet weight of the pelleted cell debris, including the inclusion bodies. For refolding, the denatured protein was dialyzed twice (for 3 h and 12 h) against 50 sample volumes of dialysis buffer (20 mM Tris-HCl, pH 8.5) supplemented with 0.1 mM DTT, twice again (3 h each) against dialysis buffer, and then once again (12 h) against 25 sample volumes of dialysis buffer supplemented with 1.0 mM reduced glutathione/0.2 mM oxidized glutathione. All dialysis steps were performed at 4°C.

Soluble protein was incubated with Ni-NTA agarose (Qiagen, Hilden, Germany) for 1 h at 4°C, and the mixture was loaded onto a chromatography column. The column was washed with a buffer containing 100 mM sodium phosphate (pH 8.0), 200 mM sodium chloride, and 10 mM imidazole. The N-terminal His-tagged lipase was eluted from the column with the same buffer containing 500 mM imidazole. The protein was dialyzed overnight against 20 mM sodium phosphate (pH 8.0) and concentrated with polyethylene glycol (PEG) 20000. Then the purified enzyme was stored at 4°C and used within three days for enzyme assays. The purified enzymes were assayed by SDS-PAGE followed by Coomassie brilliant blue G-250 staining and the protein...
concentration was determined by the Bradford method.

**Activity assay** - The time course of the esterase-catalyzed hydrolysis of p-nitrophenyl alkanoate esters was followed by monitoring the production of p-nitrophenyl at 405 nm in 1 cm pathlength cells with UV-Visible spectrophotometer (Shimadzu Co., Kyoto, Japan). In the standard assay, 20µl of 10 mM p-nitrophenyl butyrate (pNPC4), p-nitrophenyl caprylate (pNPC8) or p-nitrophenyl laurate (pNPC12) solution was added to the reaction system for a final concentration of 0.2 mM in 50 mM phosphate buffer (pH 8.0) and incubated at 25°C. Then, the reaction was started with the addition of 20 µl of the enzymatic solution. The background hydrolysis of the substrate was deducted with a reference sample of identical composition to the incubation mixture, except that lipase was omitted. One unit of enzymatic activity was defined as the amount of protein releasing 1 µmol of p-nitrophenyl from p-nitrophenyl alkanoate esters per minute (27).

**Nucleotide sequence accession numbers** - Nucleotide sequences described in this paper have been submitted to GenBank under the following accession numbers: GU942683 to GU942706.

**RESULTS**

**Isolation of starting lipase gene** - To isolate the starting lipase gene from a metagenomic sample, S05 environmental DNA library was constructed using a CopyControl fosmid library production kit (Epicentre) according to the manufacturer’s instructions. This process was followed by the screening of the lipolytic activity on a tributyrin agar plate. A positive clone showing the lipolytic activity among approximately 30,000 colonies was selected (Fig. 2A). Sequence analysis of the short insert DNA obtained by subsequent subcloning experiments revealed the presence of one open reading frame consisting of 1,311 nucleotides, encoding a protein (LipS05) with a molecular mass of 48 kDa. A BLAST search in the Lipase Engineering Database ([http://www.led.uni-stuttgart.de/](http://www.led.uni-stuttgart.de/)) revealed that the LipS05 can be aligned to several lipases around the catalytic site (E value <0.01). Those sequences showing the highest homology were from *Rhodococcus sp.* (gi|111021394|; identity, 25/76 32%; similarity, 39/76 51%), *Coprinopsis cinerea* (gi|169844727|; identity, 27/82 32%; similarity, 45/89 50%), *Streptomyces griseus* (gi|182439566|; identity, 19/58, 32%; similarity, 41/82 50%) and *Cryptococcus neoformans* (gi|134109957|; identity, 25/89 28%; similarity, 48/89 53%). LipS05 and the homologous uncharacterized lipases have a conserved active-site motif consisting of the pentapeptide GXSXG (Fig. 2B) that is characteristic of lipases (28).

The LipS05 gene was subcloned into the pET-24a(+) vector and overexpressed in *E. coli* BL21-CodonPlus (DE3)-RIPL. The supernatant of cell extract from *E. coli* harboring pET-24a-LipS05 showed a significantly elevated activity compared with that of *E. coli* harboring pET-24a without the subcloned lipase sequence. Therefore, high-throughput functional screening can be applied to detect the phenotypic variability of mutants of LipS05 using this expression system (29).

**Recovery of homologous sequences** - Since there are no conserved domains in the LipS05 sequence, only sequences that correspond to tens of amino acids that allow alignment, it is not possible to design degenerate primers to isolate related family genes from other environmental samples as in traditional PCR-based metagenomic screening. Sets of primers specific to arbitrary locations of the LipS05 sequence were designed (Table 1). Initially, overlapping primers targeted terminal to the LipS05 sequence were designed to amplify full length homologous genes from environmental DNA samples. For full length Lip1311 primer sets, no more than 3 specific amplification products of
the correct size (1311 bp) were generated from reactions for 60 templates. For the other two overlapping terminal primers, Lip1275 and Lip1247, none of the products were observed by gel analysis. Despite repeated attempts to optimize PCR conditions, only nontarget amplification products were obtained for the above-mentioned primers. Therefore, sets of primers with a series of progressively longer deletions were designed to collect more homologous genes. For the Lip1095 primer set, four specific amplification products with a size of 1095 bp were found under similar PCR conditions. While for Lip923 primer set, 20 specific amplification products was generated under the same PCR condition. A total of 60 inserts were sequenced from those 20 clone libraries with inserts from at least 3 randomly selected clones sequenced from each library. Twenty-three novel lipase gene fragments, with conserved pentapeptide GXSXG and between 68 to 97% nucleotide identity to LipS05, were retrieved. Corresponding to soil sample name, these lipase genes were named LipS01-1, LipS01-2, LipS03, LipS05-2, LipS07, LipS08-1, LipS08-2, LipS09, LipS10, LipS11, LipS12, LipS13, LipS16, LipS17, LipS19, LipS30, LipS33, LipS48-1, LipS48-2, LipS49, LipS50-1, LipS50-2 and LipS51, respectively. To confirm the correlation between the number of specific PCR products and the length of the truncated fragments, the Lip481 primer set, which had further truncation of sequence length, was applied in PCR screening of those environmental DNA samples, which resulted in 16 specific amplification products under similar reaction conditions. Optimized amplification reaction conditions and an increase in the number of sequenced clones may result in more novel sequences retrieved. As the isolated sequences were of the appropriate number and length for following DNA family shuffling, no attempts were made to isolate more sequences. Twenty-three 923 bp retrieved lipase genes and the starting gene LipS05 were chosen for use as genetic materials for DNA family shuffling.

Sequence analysis of LipS05 and subfamily sequences - A multiple alignment was constructed to illustrate conserved and variable sites within the retrieved lipase subfamily. The classical signature of lipases, the catalytic site (GXSXG), was also conserved in the subfamily. The pairwise identities of the protein sequences ranged from 64% to 99% (68–99% DNA sequence identity). This high sequence identity showed the close relationship between these proteins and indicated they can be considered to belong to the same subfamily. In spite of this high pairwise identity between sequences, 141 variable sites with different homology levels were also observed in multiple alignments (Fig. 4). Therefore, a combination of 141 variable sites with a minimum of two amino acids change in one site would comprise at least \(2^{141}\), or \(2.8 \times 10^{42}\), different sequences. Although the vast majority of the sequence space that can be explored will be contingent on the recombining capacity of DNA family shuffling, it can be assumed that the DNA family shuffling of those homologous genes will result in extensive genetic diversity.

The phylogenetic tree constructed by the NJ method for the lipase family proteins is shown in Fig. 5. All of the lipases from environmental samples cluster closely together on the left, while all of the known lipases cluster loosely together on the right. The lipases map together on a branch of the tree separate from those known lipases and form a clearly distinct group. The phylogenetic tree clearly indicates that novel lipases form a well-supported clade (NJ bootstrap, 100%) that is loosely related to the other known lipases (bootstrap, <50%).

Analysis of shuffled clones - To explore whether those homologous genes can be used as suitable genetic materials for DNA family shuffling, we created a shuffled library with 24 homologues and characterized a small number of
chimeric clones. The shuffled library of chimeras was transformed into *E. coli* and screened for lipolytic activity on a tributyrin agar plate (Fig. 6A). About 10% of the transformants formed clearing halos, indicating expression of active enzyme. The size of the resulting clear halos in active clones was uniform, indicating that phenotypic diversity was generated from the chimeric library. To gain a deep insight into the functional diversity of the chimeric mutants, the three active clones with the largest halos were overexpressed and purified to apparent homogeneity. The clones’ specific activity on pNPC4, pNPC8, and pNPC12 were measured to demonstrate varied activity on short-, medium-, or long-chain p-nitrophenyl alkanoate esters substrates of the mutant enzymes. The results demonstrated that the mutants exhibited a distinct substrate specificity, with significantly varied specific activity on different substrates (Fig. 6B). The mutants C3 and E5 showed a preference for the short-chain pNPC4, while mutant D5 had its maximal activity on the medium-chain pNPC8. Although both C3 and E5 exhibited maximal activity on pNPC4, an obvious difference in activity on pNPC4 and pNPC8 was observed between them: for mutant C3, the activity on pNPC8 was 54% relative to that of pNPC4, while mutant E5 demonstrated a comparable activity on pNPC4 and pNPC8. All of the mutants indicated similar substrate specificity for the long-chain substrate pNPC12, with no more than 20% of maximal activity on their preferred substrates. In addition to this distinct substrate specificity, the mutants also had a definite difference in their specific activity on those substrates. For the short-chain substrate pNPC4, the mutant C3 showed the highest specific activity (87.4 U mg⁻¹) among the mutants, which was 1.5 and 3.5 fold higher than that of D5 and E5, respectively. For the medium-chain substrate pNPC8, although the mutant D5 showed the maximal activity relative to its activity on the short- and long-chain substrates, the value of its specific activity (56.2 U mg⁻¹) was slightly higher (about 1.3 fold) than that of the other two mutants. For the long-chain substrate pNPC12, the mutant C3 showed the highest specific activity with 17.4 U mg⁻¹, which was more than 2 fold more than that of the other two mutants. To investigate the correlation between structure and observed activity, these three clones were sequenced. The sequence of C3 is a chimera of LipS33, LipS11, LipS30, LipS11, LipS01-2, LipS08-2 and LipS11. The E5 gene encodes a chimeric enzyme with sequence elements originating from LipS01-2, LipS08-2, LipS30, LipS33, LipS11, LipS30, LipS08-2, LipS33, LipS13, LipS12, LipS13, LipS30 and LipS16. D5 contained sequence elements from LipS09, LipS48-2, LipS30, LipS13 and LipS30 (Fig. 6C). As such strikingly high crossover frequency was observed in the selected clones, another 7 unselected chimeric clones in small library were sequenced to determine the average crossover frequency in the library. A per-gene average of 8 crossovers and 3 spontaneous mutations were observed in sequenced genes.

**DISCUSSION**

In the present study, a novel metagenomic gene-specific strategy was chosen to retrieve related genes, in contrast to traditional PCR-based metagenomic screening, which uses known gene information to design degenerate primers corresponding to conserved regions of an enzyme family (21). Total DNA extracted directly from environmental samples does not typically contain an even representation of the population’s genomes within the sample. Therefore, the targeted homologous genes of known sequences might be overshadowed by sequences from unknown dominant microbial populations in environmental DNA, which would severely decrease the efficiency and specificity of gene amplification and affect the product yield (18). In fact, LeCleir G et al. failed to retrieve...
any chitinase sequences from environmental samples using known sequences based PCR approach, despite evidence suggesting high levels of chitinolytic activity in the sample (31). A previous study retrieved only one lipase sequence from the environmental DNA of olive oil percolation using known sequences based PCR approach, even with extensive analysis of conserved regions and careful primer design (32). In contrast, metagenomic gene-specific PCR uses an gene identified by functional screening of soil DNA library, instead of a known gene in database as the startpoint for retrieving family genes, which has a inherent property of biasing for genomes of dominant microorganisms of soil DNA. This distinguishing feature make metagenomic gene a good start point for retrieving metagenomic homologous family genes from soil DNA samples by PCR. It is worth noting that 20 bands of correct size and without any non-specific bands was observed in the gel analysis of the products amplified from DNA of 60 soil samples with metagenomic gene specific Lip923 primer set (Fig. 3). That was achieved by only one experiment without any attempt at optimizing reaction conditions. Nevertheless, neither our method nor traditional metagenomic sequence-based PCR guarantees acquisition of full-length family genes because identical sequences are conserved in the interior regions of subfamily genes rather than at terminal regions (6).

Phylogenetic analysis indicated that the metagenomic gene specific amplification strategy offers a chance to access to a novel subfamily that was previously unknown and completely uncharacterized. Traditional degenerate primer-based PCR for metagenomic screening is based on the conserved DNA sequences of known family genes. Generally, the amino acid identities of newly discovered sequences from a metagenome to those in the protein databases have been above 50% and usually around 80–98%, which is rarely possible to shed light on a novel gene family (33-37).

The functional diversity of chimeric phenotypes generated by DNA-family-shuffling was evident in experiments assessing both activity on indicative plates and substrate specificity of purified mutants. The varied size of the clear halos around the mutant clones indicated an apparently varied lipolytic activity. In addition, characterization of the specific activity on substrates with various chain lengths provides definite evidence for the functional diversity of chimeric clones. Such a property of diversity is expected and is indeed a hallmark of successful libraries created by DNA family shuffling (6,7). The crossover frequency of chimeric clones in this study is higher than that reported for normal enzyme shuffling and lower that that of a shuffled library created with methods for generating highly recombined genes, such as RICHT (38). Given that we used the same shuffling procedure as in normal enzyme shuffling reports to create our chimeric library, the high crossover frequency can be attributed to the diversity of the genetic material we used. In fact, there are usually no more than 4 homologous genes of enzymes from microorganisms used to shuffle in reports using DNA family shuffling (38). Additionally, in the 13,310 bases sequenced, 3 spontaneous mutations were found. This rate of 1 mutation per 4,436 bases is well within the number of spontaneous mutations expected from the rounds of PCR used to generate the fragment and template DNA and to amplify the chimeric products. Further, no unexpected insertions, deletions, or rearrangements were detected by DNA sequencing. These data suggest that the DNA shuffling procedure using genetic material collected by TMGS-PCR can creates a fair qualified chimerical library.

CONCLUSION

In conclusion, we have presented a novel TMGS-PCR approach with unique features that
distinguish it from the traditional PCR-based metagenomic screening method. Functional analysis of the shuffled library, coupled with the sequence information from the selected chimeric progeny, indicated that this method can provide high-quality genetic materials for DNA family shuffling. We expect that TMGS-PCR could provide a new alternative to collecting metagenomic homologous genes of microbial enzymes for DNA family shuffling, which is an excellent tool for the irrational design of biocatalysts with tailored properties. Meanwhile, one can also use TMSG-PCR to generate breeding stock and then use “synthetic shuffling” to breed divergent sequences for efficient directed enzyme evolution.

REFERENCES


**FOOTNOTES**

Acknowledgement: The work was supported by National Natural Science Foundation of China (30900253, 20906016), Natural Science Foundation of Zhejiang Province (Y4080317).

The abbreviations used are: TMGS-PCR, truncated metagenomic gene specific polymerase chain reaction; p-nitrophenyl butyrate (pNPC4); p-nitrophenyl caprylate (pNPC8); p-nitrophenyl laurate (pNPC12).
FIGURE LEGENDS

Fig. 1. Schematic overview of homologous genes collected by truncated metagenomic gene specific amplification. Initially, the metagenomic starting gene was isolated from environmental samples by functional screening. Based on the sequence of this identified gene, truncated gene specific primers were used to amplify homologous genes from different environmental samples. The retrieved homologous family genes were subjected to shuffling to generate the chimeric libraries for isolated clones with desired properties.

Fig. 2. Isolation of starting lipase from metagenomic library. (A) Detection of lipase on the plates containing tributyrin. (B) Multiple alignment of the partial amino acid sequences containing the conserved GxSxG motifs of LipS05 and other lipases. GI number and organisms: gi|111021394|, Rhodococcus sp; gi|169844727|, Coprinopsis cinerea; gi|182439565|, Streptomyces griseus; gi|134109957|, Cryptococcus neoformans. The putative catalytic triad residues are shaded and denoted by arrowheads.

Fig. 3. Agarose gel electrophoresis of TMGS-PCR products amplified from DNA of 60 soil samples using Lip923F/R primer set. M indicates the 5.0 kb ladder lane.

Fig. 4. Amino acid sequence alignment of lipases. The alignment quality curve is displayed below the alignment. Asterisks, colons and dots indicate completely, strongly and weakly conserved positions, respectively.

Fig. 5. Phylogenetic analysis of lipase sequences (partial sequence, 307 aa). The data are illustrated as a neighbor-joining tree based on 24 new lipases and 50 previously known sequences from the Lipase Engineering Database. The name at the nodes is assigned according to those in the Lipase Engineering Database and is shown for the major clades only. Each geographic origin is represented by a grey dot.

Figure 6. Characterization of chimerical clones. (A) Lipase activity screens. The chimeric lipase library was plated to rich medium containing 1% tributyrin, grown one day, and incubated for 5-7 additional days at 4°C. Transformants exhibiting diverse activity produce clear variably-sized halos. (B) Specific activity on varied substrates for selected chimeric mutants obtained by family shuffling. Data reported are the mean of at least two independent experiments. (C) Sequence of selected chimaeric mutants obtained by family shuffling. The segments derived from different environmental lipases are shown as white bars and noted as their name above the bars. Because of sequence identity in crossover, shown as black bars, the exact derivation of the sequence cannot be determined exactly. The aminoacid point mutations is indicated by carets (v) above the bars.
TABLE LEGENDS

Table 1 Primers used in environmental sample amplification.
Figure 2

A

Halo

LipS05

B

gi111021394  CACCTCMTTSCGTTQAN  155

gi11182439565  GAPAADYGSQSMMPN  150

LipS05  NNFRANLGHGSGLIIT  151

gi1169844727  PGRLNLHDEEGGELDCR  305

gi1134109957  PGREINLHCHEGGDDCR  215
Figure 5
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’→3)</th>
<th>$Tm$ (°C)</th>
<th>Location within gene</th>
<th>Size of Product (bp)</th>
<th>No. of amplification products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip1311F</td>
<td>ATGGAATTTATTCGCCCTTCATCTTTTGG</td>
<td>62.03</td>
<td>1-30</td>
<td>1311</td>
<td>0</td>
</tr>
<tr>
<td>Lip1311R</td>
<td>TTATTTGGGACGCAGCGGCTTGGAC</td>
<td>63.69</td>
<td>1287-1311</td>
<td>1311</td>
<td>0</td>
</tr>
<tr>
<td>Lip1275F</td>
<td>TCATCTTTGTGCCGGCATCACC</td>
<td>63.73</td>
<td>20-42</td>
<td>1275</td>
<td>0</td>
</tr>
<tr>
<td>Lip1275R</td>
<td>GCTTGACCAGCGGATCCCCATTT</td>
<td>63.8</td>
<td>1273-1294</td>
<td>1275</td>
<td>0</td>
</tr>
<tr>
<td>Lip1247F</td>
<td>GCATCACCAGGACGTATTTGCAGA</td>
<td>63.73</td>
<td>35-57</td>
<td>1247</td>
<td>0</td>
</tr>
<tr>
<td>Lip1247R</td>
<td>ATCCCATTTTCAGGCAGCTGCAC</td>
<td>61.96</td>
<td>1259-1280</td>
<td>1247</td>
<td>0</td>
</tr>
<tr>
<td>Lip1095F</td>
<td>CGTGTGCGCTTCCACCAGATG</td>
<td>65.66</td>
<td>115-136</td>
<td>1095</td>
<td>4</td>
</tr>
<tr>
<td>Lip1095R</td>
<td>GAAGTGACGGACAATGATCGATGC</td>
<td>63.62</td>
<td>1185-1209</td>
<td>1095</td>
<td>4</td>
</tr>
<tr>
<td>Lip923F</td>
<td>CCCATCAAGGAATCATTCAGGAACTGC</td>
<td>63.48</td>
<td>196-223</td>
<td>923</td>
<td>20</td>
</tr>
<tr>
<td>Lip923R</td>
<td>TCCAGATATCCATAATCCCTCAGGGTCAC</td>
<td>64.85</td>
<td>1090-1118</td>
<td>923</td>
<td>20</td>
</tr>
<tr>
<td>Lip481F</td>
<td>CATTGCTTGGGCGGATGATCATCACC</td>
<td>65.04</td>
<td>415-441</td>
<td>481</td>
<td>16</td>
</tr>
<tr>
<td>Lip481R</td>
<td>TAACGCCTGCACACGCGGCAGCTTCTCT</td>
<td>67.68</td>
<td>867-896</td>
<td>481</td>
<td>16</td>
</tr>
</tbody>
</table>
Prospecting metagenomic enzyme subfamily genes for dna family shuffling by a novel pcr-based approach
Qiuyan Wang, Huili Wu, Anming Wang, Pengfei Du, Xiaolin Pei, Haifeng Li, Xiaopu Yin, Lifeng Huang and Xiaolong Xiong

J. Biol. Chem. published online October 20, 2010

Access the most updated version of this article at doi: 10.1074/jbc.M110.139659

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2010/10/20/jbc.M110.139659.full.html#ref-list-1