12

Family Shuffling with Single-Stranded DNA

Wenjuan Zha, Tongbo Zhu, and Huimin Zhao

1. Introduction

Family shuffling, which generates chimeric progeny genes by recombining a set of naturally occurring homologous genes, is an extremely powerful approach for in vitro protein evolution. In comparison with other in vitro protein evolution methods, family shuffling has the advantage of sampling a larger portion of the sequence space that has been proven functionally rich by nature. So far, the most widely used technique to carry out family shuffling is DNA shuffling ***(1–5)***. However, one significant drawback associated with this method is the low frequency of chimeric genes (recombined gene products) in the shuffled library ***(6)***, which may be largely owing to the annealing of DNA fragments derived from the same parental genes (homo-duplex formation) whose probability is much higher than that of hetero-duplex formation. This is true even when the sequence homologies among parental genes are higher than 80% ***(7)***.

To address this problem, Kikuchi et al. ***(7)*** developed a modified family shuffling method—family shuffling with restriction enzyme-cleaved DNA fragments—based on the DNA shuffling method, which involves the fragmentation of the parental genes using restriction enzymes rather than using DNase I. In addition, the same research laboratory developed a second modified family shuffling method, family shuffling with single-stranded DNA (ssDNA) ***(8)***. The latter method is also based on the DNA shuffling method and uses single-stranded DNA templates rather than double-stranded DNA (dsDNA) templates for DNase I fragmentation. Since use of single-stranded DNA as templates will decrease the probability of homo-duplex formation, the percentage of the parental genes in the shuffled library should be significantly reduced. As a demonstration, this method was used to recombine two catechol 2,3-

From: *Methods in Molecular Biology, vol. 231: Directed Evolution Library Creation: Methods and Protocols*
Edited by: F. H. Arnold and G. Georgiou

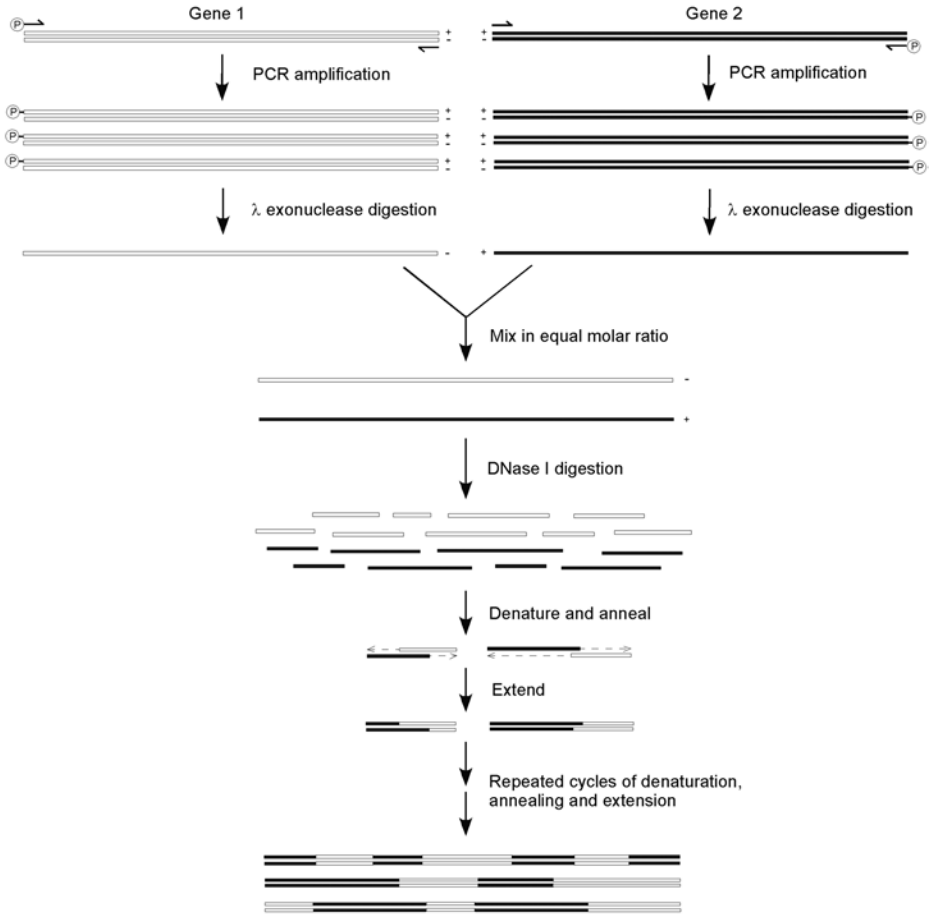

Fig. 1. Schematic representation of ssDNA-based family shuffling. For simplicity, only two DNA templates are shown. Double stranded DNAs with one strand phosphorylated at 5'-terminus are prepared from DNA templates carrying the target sequences using conventional PCR and digested by lambda exonuclease. The two single-stranded DNA templates are mixed at equal molar ratio followed by random fragmentation with DNase I. After the removal of the oligonucleotides and the templates, the homologous fragments are reassembled into full-length chimeric genes in a PCR-like process. The full-length genes may be amplified by a standard PCR and sub-cloned into an appropriate vector.

dioxygenase genes, *nahH* and *xyl*E. It was found that this ssDNA-based DNA shuffling was able to generate chimeric genes at a rate of 14%, much higher than the rate of less than 1% obtained by the original dsDNA-based DNA shuffling method ***(8)***. However, the method Kikuchi et al. used to prepare single-

stranded DNA templates requires the subcloning of the target genes to a phagemid vector and the use of a helper phage. The procedure is time-consuming and inefficient. Moreover, the isolated single-stranded DNA templates all contain the backbone DNA of the phagemid vector, which could become problematic in certain cases since the DNA fragments from the backbone DNA may interfere with the fragment reassembly process of the parental genes.

Here we describe an improved ssDNA-based family shuffling method in which a simple and convenient approach is used to prepare single-stranded DNA templates. The procedure is illustrated in **Fig. 1**. For simplicity, only two parental genes are shown. First, double-stranded DNAs with one strand phosphorylated at 5'-terminus are prepared from DNA templates carrying the target sequences using conventional PCR. For each target sequence, only one of the two PCR primers is 5'-phosphorylated. The primers should be designed such that the 5'-phosphorylated strands from different parental genes are complementary to each other. Second, single-stranded DNA templates are prepared using lambda exonuclease digestion. In the presence of lambda exonuclease, the 5'-phosphorylated strands will be degraded very rapidly. Third, the two single-stranded DNA templates are mixed at equal molar ratio, followed by random fragmentation with DNase I. Fourth, the digested single-stranded DNA fragments are reassembled into full-length chimeric genes.

As a test case, we used this method to shuffle two genes encoding the ligand binding domain of estrogen receptor α and β (ERα-LBD and ERβ-LBD), which share 66% sequence homology at the DNA level. The shuffled gene products were subcloned into an *E. coli* vector pET26b(+) (Novagen), and five transformants were randomly selected for DNA sequencing. It was found that all five genes were chimeric. In comparison, we also shuffled ERα-LBD and ERβ-LBD using the conventional dsDNA based DNA shuffling method. None of the five randomly selected transformants obtained from the shuffled products were chimeric genes. Therefore, it appears that ssDNA based DNA shuffling is indeed a more efficient family shuffling method than dsDNA based DNA shuffling, especially when the parental genes to be shuffled have relatively low sequence homology.

2. Materials

1. DNA templates containing the target sequences to be recombined (*see* **Note 1**).
2. Oligonucleotide primers (*see* **Note 2**).
3. Lamda exonuclease and its 10X reaction buffer (New England BioLabs, Beverly, MA).
4. 10X DNase I digestion buffer: 500 m*M* Tris-HCl, pH 7.4, 100 m*M* $MnCl_2$.
5. 10 m*M* Tris-HCl, pH 7.4.
6. DNase I: 10 U/μL (Roche Diagnostics, Indianapolis, IN).
7. *Taq* DNA polymerase and its 10X reaction buffer (Promega, Madison, WI).

8. *Pfu*Turbo DNA polymerase and its 10X reaction buffer (Stratagene, La Jolla, CA).
9. 10X dNTP mix: 2 m*M* of each dNTP (Roche Diagnostics, Indianapolis, IN).
10. Agarose gel electrophoresis supplies and equipment.
11. MJ PTC-200 thermocycler (MJ Research Inc., Watertown, MA).
12. EZ load precision molecular mass ruler (Bio-Rad, Hercules, CA).
13. QIAquick PCR purification kit (QIAgen, Valencia, CA).
14. QIAEX II gel extraction kit (QIAgen, Valencia, CA).
15. Centri-Sep columns (Princeton Separations, Inc., Adelphia, NJ).

3. Methods

1. For each target sequence, combine 1–10 ng DNA template, 10X *Taq* reaction buffer, 0.5 μ*M* each primer (one of the two primers is 5'-phosphorylated), 10X dNTP mix (2 m*M* each dNTP), and 2.5 U *Pfu*/*Taq* DNA polymerase (1:1) in a total volume of 100 μL (*see* **Note 3**).
2. Run the PCR reaction using the following program: 96°C for 2 min, 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, followed by 72°C for 7 min (*see* **Note 4**).
3. Purify the PCR products according to the manufacturer's protocol in the QIAquick PCR purification kit. Estimate DNA concentrations using gel electrophoresis and EZ load precision molecular mass ruler or using a UV-Vis spectrophotometer.
4. For each target gene, combine 3–5 μg PCR product, 10X lambda exonuclease reaction buffer, and 10 U lambda exonuclease in a total volume of 50 μL. Incubate the reaction mixture at 37°C for 30 min–60 min (*see* **Note 5**).
5. Separate the digestion products by gel electrophoresis. Purify the single-stranded DNA products using a QIAEX II gel purification kit (*see* **Note 6**). Estimate DNA concentrations using gel electrophoresis and EZ load precision molecular mass ruler (*see* **Note 7**).
6. Mix ~0.5 μg of each purified ssDNA product. Dilute the mixture to 45 μL in 10 m*M* Tris-HCl, pH 7.4, and add 5 μL of 10X DNase I digestion buffer.
7. Equilibrate the mixture at 15°C for 5 min before adding 0.6 U of DNase I. Digestion is done on a thermocycler using the following program: 15°C for 2 min and 90°C for 10 min (*see* **Note 8**).
8. Purify the cleaved DNA fragments using Centri-Sep columns according to the manufacturer's protocol.
9. Combine 10 μL of the purified fragments, 2 μL of 10X *Taq* buffer, 2 μL of 10X dNTP mix, 0.5 U *Taq* DNA polymerase and sterile dH_2O in a total volume of 20 μL (*see* **Note 9**). No primer is added.
10. Run the reassembly reaction using the following program: 3 min at 96°C followed by 40 cycles of 30 s at 94°C, 1 min at 55°C, 1 min + 5 s/cycle at 72°C, and finally 7 min at 72°C (*see* **Note 10**).
11. Run a small aliquot of the reaction mixture on an agarose gel. A smear extending through the size of the expected full-length gene should be visible. Otherwise, run an additional 10–20 cycles, or lower the annealing temperature.

12. Combine 1 μL of the assembly reaction, 0.3–1.0 μ*M* each primer, 10 μL of 10X *Taq* buffer, 10 μL of 10X dNTP mix (2 m*M* each dNTP), and 2.5 U *Taq*/*Pfu* (1:1) mixture in a total volume of 100 μL (*see* **Note 11**).
13. Run the PCR reaction using the following program: 96°C for 2 min, 10 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 45 s followed by another 14 cycles of 30 s at 94°C, 30 s at 55 °C and 45 s + 20 s/cycle at 72°C. The final step of elongation is at 72°C for 7 min.
14. Run a small aliquot of the reaction mixture on an agarose gel. In most cases, there is only a single band at the correct size. Purify the full-length PCR product using the QIAquick PCR purification kit, or other reliable method. If there is more than one band, purify the product of correct size using the QIAEX II gel purification kit.
15. Digest the product with the appropriate restriction endonucleases, and ligate to the desired cloning vector.

4. Notes

1. Appropriate templates include plasmids carrying target sequences, cDNA or genomic DNA carrying the target sequences, sequences excised by restriction endonucleases, and PCR amplified sequences.
2. Primer design should follow standard criteria including similar melting temperatures and elimination of self-complementarity or complementarity of primers to each other. Free computer programs, such as Primer3 at Biology Workbench (http://workbench.sdsc.edu), can be used to design primers. Typically, primers should also include unique restriction sites for subsequent directional sub-cloning. For each target sequence, only one of the two primers used to amplify the DNA should be 5'-phosphorylated. In the case of two target sequences, primers should be 5'-phosphorylated in a complementary manner (*see* **Fig. 1**), such that the resultant single-stranded DNAs are complementary. In the case of more than two target sequences, one of the target sequences is selected to generate a single-stranded DNA complementary to those from the rest of target sequences. 5'-Phosphorylated primers can either be ordered directly from various DNA synthesis service providers or prepared from regular primers using T4 polynucleotide kinase.
3. *Pfu*Turbo DNA polymerase is selected because of its high fidelity in DNA amplification and the generation of blunt-ended products. Other DNA polymerases with proofreading activity such as *Vent* DNA polymerase (New England Biolabs, Beverly, MA) or a mixture of *Taq* DNA polymerase and *Pfu* DNA polymerase (1:1) may also be used. *Taq* DNA polymerase alone is not recommended since it cannot generate blunt-ended and 5'-phosphorylated products that are preferred substrates for lambda exonuclease.
4. The elongation time depends on the gene size. For genes of 1 kb or less, an elongation time of 1 min will suffice. For genes larger than 1 kb, the elongation time should be adjusted accordingly. If the yield of the PCR product is low, set up several PCR reactions for each target sequence and combine the products.
5. The extent of the digestion reaction should be checked by running a 2 μL aliquot of the reaction mixture along with an aliquot of undigested DNA as a control on

an agarose gel. Single-stranded DNA will migrate faster than the double-stranded counterpart. Note that the extent of phosphorylation of the PCR primers is critical for lambda exonuclease digestion. It is likely that some double-stranded DNA will still be present in the reaction mixture even after digestion for a long time or with excess enzymes. Those undigested DNAs are presumably the PCR products derived from the non-phosphorylated primers. Furthermore, since lambda exonuclease will also degrade single-stranded DNA substrates, albeit at a greatly reduced rate, digestion of the PCR products for a long time or using excess enzymes may significantly reduce the final yield of the purified single-stranded DNA. Thus, it is highly recommended to follow the progress of the digestion reaction in a test run by taking out an aliquot of the reaction mixture every 10 min after initial 20 min incubation and analyzing them on the gel. Normally, the band of double-stranded DNA becomes lighter as the incubation time increases, indicating more double-stranded DNA has been converted into single-stranded DNA. Inactivation of lambda exonuclase by heating at 75°C for 10 min after digestion is recommended by most manufacturers. However, this step may be omitted if the digestion products are immediately subjected to agarose gel electrophoresis.

6. Lambda exonuclease is a highly processive 5' to 3' exodeoxyribonuclease that selectively digests the phosphorylated strand of double stranded DNA and has significantly reduced activity toward single-stranded DNA ***(9)***. In many cases, lambda exonuclease cannot completely digest the phosphorylated strand of double-stranded DNA even using 10-fold excess of enzymes and longer incubation time. Thus, the single-stranded DNA must be purified from the gel after electrophoresis.
7. Since ethidium bromide binds less efficiently to ssDNA than to dsDNA, the ssDNA concentration estimated using gel electrophoresis and EZ load precision molecular mass ruler will always be lower than its real concentration.
8. A small aliquot (5 μL) of the reaction mixture should be taken and electrophoresed on an agarose gel before the purification step. The size of the cleaved DNA fragments is expected to be around 50 bp. If fragments larger than 100 bp are seen, add more DNase I and repeat the same digestion step.
9. Any commercially available thermostable DNA polymerase can be used to reassemble the DNA fragments. However, the use of DNA polymerases with proofreading activity such as *Pfu* DNA polymerase (Stratagene, La Jolla, CA), *Vent* DNA polymerase (New England Biolabs, Beverly, MA), or *Pfx* DNA polymerase (Invitrogen Life Technolgies, Carlsbad, CA) can minimize the introduction of point mutations in the chimeric progeny genes. When setting up reactions with these polymerases, it is very important to add the polymerase last since, in the absence of dNTPs, the 3'to 5'exonuclease activity of the polymerase can degrade DNAs.
10. The number of cycles depends on the fragment size. Assembly from small fragments may require more cycles than assembly from large fragments. Low annealing temperature may also be required for reassembling small fragments and for reassembling fragments derived from genes with low homology. The increasing elongation time for subsequent cycles correlates to the increasing fragment size as full- length chimeric genes are created.

11. *Pfu* DNA polymerase or other high fidelity DNA polymerases have a much slower processivity than *Taq* DNA polymerase, leading to lower extension rates. Processivity can be significantly improved (without significant loss in fidelity) by using a 1:1 ratio (1.25 Units each in a 100 µL reaction) of *Taq* DNA polymerase and *Pfu* DNA polymerase, rather than *Pfu* DNA polymerase alone ***(10)***. If this is done, an appropriate 10X reaction buffer for *Taq* DNA polymerase should be used in place of the 10X *Pfu* reaction buffer. It is noteworthy that certain commercial polymerase blends such as *Taq*/*Pwo* DNA polymerase (Expand High Fidelity PCR system from Roche Diagnostics) or *Taq*Plus Precision/*Pfu* (*Taq*Plus Precision PCR system of Stratagene) will yield similar results.

References

1. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391,** 288–291.
2. Ness, J. E., Welch, M., Giver, L., et al. (1999) DNA shuffling of subgenomic sequences of subtilisin. *Nat. Biotechnol.* **17,** 893–896.
3. Chang, C. C., Chen, T. T., Cox, B. W., et al. (1999) Evolution of a cytokine using DNA family shuffling. *Nat. Biotechnol.* **17,** 793–797.
4. Christians, F. C., Scapozza, L., Crameri, A., Folkers, G., and Stemmer, W. P. (1999) Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* **17,** 259–264.
5. Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T., and Furukawa, K. (1998) Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat. Biotechnol.* **16,** 663–666.
6. Coco, W. M., Levinson, W. E., Crist, M. J., et al. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* **19,** 354–359.
7. Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) Novel family shuffling methods for the in vitro evolution of enzymes. *Gene* **236,** 159–167.
8. Kikuchi, M., Ohnishi, K., and Harayama, S. (2000) An effective family shuffling method using single-stranded DNA. *Gene* **243,** 133–137.
9. Mitsis, P. G. and Kwagh, J. G. (1999) Characterization of the interaction of lambda exonuclease with the ends of DNA. *Nucleic Acids Res.* **27,** 3057–3063.
10. Zhao, H. and Arnold, F. H. (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25,** 1307–1308.