Gene synthesis by circular assembly amplification

Duhee Bang & George M Church

Here we report the development of a gene-synthesis technology, circular assembly amplification. In this approach, we first constructed exonuclease-resistant circular DNA via simultaneous ligation of oligonucleotides. Exonuclease- and subsequent mismatch cleaving endonuclease–mediated degradation of the resulting ligation mixture eliminated error-rich products, thereby substantially improving gene-synthesis quality. We used this method to construct genes encoding a small thermostable DNA polymerase, a highly repetitive DNA sequence and large (>4 kb) constructs.

Gene synthesis has an increasingly important role in biological research. Recent applications have culminated in the development of new antibiotics by *de novo*–designed polyketide synthase gene clusters[1] and in the creation of a highly active synthetic mammalian retrotransposon[2]. Gene synthesis is typically carried out via polymerase cycling assembly (PCA) that involves the assembly of a pool of overlapping complementary oligonucleotides via PCR[3]. Ligation-based assembly followed by PCR also has been used to synthesize the phiX genome[4]. Both of these commonly used methods, however, are highly error-prone (0.2–0.3%) because of errors originating in the synthetic oligonucleotides whether from single-oligonucleotide columns or massively parallel chips[5].

Various methods have been devised to reduce gene synthesis error rates. Of these approaches, a two-cycle assembly process is the most common, refering to the synthetic process up to cloning. In this two-cycle process small DNA fragments (300–500 bp) are first synthesized (typically via PCA)[6,7]. After sequence validation, a second round of assembly is applied to stitch these small fragments together into the desired gene. Other selection approaches include the use of mismatch-sensitive hybridization[5] and enzymatic mismatch cleavage[8,9]. DNA mismatch–binding protein also has been used to remove error-containing DNA[10].

We developed a one-cycle gene synthesis method that substantially improves synthesis quality. This approach uses three different tiers of selection (**Fig. 1**). First, single-stranded oligonucleotides are assembled into circular molecules under a highly stringent annealing condition (70 °C) such that most error-containing oligonucleotides do not anneal (tier one). By subsequently subjecting the ligation mixture to exonuclease treatment we select for circular molecules with the desired sequence (tier two). Finally, by using a mismatch-cleaving endonuclease we convert circular DNA containing residual errors to a linear form that is degraded by exonucleases still present in the solution (tier three). A final PCR is done to amplify the remaining undegraded DNA from the reaction mixture.

To develop our circular assembly amplification method, we focused on the construction of the 1,056-bp *Dpo4* gene[11]. We chose to synthesize *Dpo4* because it is one of the smallest DNA polymerases compatible with thermal cycling PCR (352 codons rather than 834 for *Taq* polymerase) and hence ripe for synthetic improvements. We first designed codon-optimized *Dpo4* DNA sequences[12]. We designed 24 plus strand oligonucleotides and 23 minus strand oligonucleotides, each ~50 base pairs, with a melting temperature of 60 °C using the nearest-neighbor method[13] (see **Supplementary Methods** online for *Dpo4* and oligonucleotide sequence information). Additionally, we designed one more guiding oligonucleotide (24^{th} minus strand) to join 5′ and 3′ ends of the *Dpo4* sequence.

We ligated the 5′-phosphorylated oligonucleotides using a thermostable ligase at 70 °C with or without the guiding

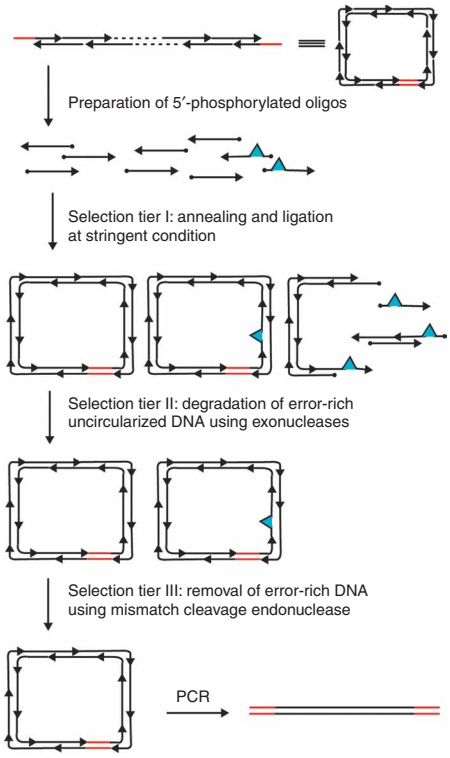

Figure 1 | Schematic representation of circular assembly amplification. Oligonucleotides were designed as described in text. Cyan triangles represent errors.

Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. Correspondence should be addressed to D.B. (dbang@genetics.med.harvard.edu) or G.M.C. (http://arep.med.harvard.edu/gmc/email.html).

RECEIVED 3 OCTOBER; ACCEPTED 25 OCTOBER; PUBLISHED ONLINE 25 NOVEMBER 2007; DOI:10.1038/NMETH1136

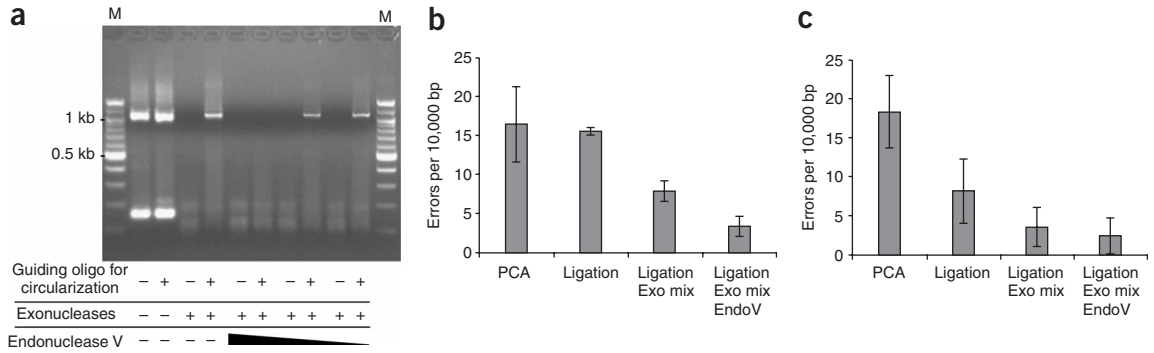

Figure 2 | *Dpo4* gene constructed by circular assembly amplification of 48 oligonucleotides. (**a**) Constructs resulting from amplification of ligation mixtures by PCR with and without a guiding oligonucleotide, subsequently treated with exonucleases and with different concentrations of mismatch-cleaving endonuclease as indicated. M, 100 bp ladder. (**b**,**c**) Number of errors per 10,000 bp resulted from synthetic *Dpo4* made by various methods. Experiments are performed at annealing temperature 65 °C (**b**) and 70 °C (**c**). For each experiment, we sequenced 10–12 clones and repeated each experiment 2–4 times. Error bars, s.d.

oligonucleotide, and then amplified the ligation mixture by PCR. As expected, both ligation mixtures yielded the desired products (**Fig. 2a**). However, when we introduced a mixture of exonuclease I for degradation of ssDNA, and exonuclease III and lambda exonucleases for degradation of dsDNA, only the oligonucleotide pool containing the guide oligonucleotide yielded product (**Fig. 2a**). Use of insufficient exonuclease treatment lead to incomplete exonuclease degradation (**Supplementary Fig. 1a** online), which could be minimized by using an excess of exonucleases to ensure that sequence from error-containing oligonucleotides is eliminated. Exonuclease treatment substantially reduced yield as would be expected if it was selecting for oligonucleotides of full length and correct sequence. In additional experiments, we introduced a mismatch-cleaving endonuclease (endonuclease V from *Escherichia coli* based on the comparison of the mismatch-cleavage efficiencies of different endonucleases[8]). The intensities of the bands resulting from PCR amplification of the endonuclease-treated mixture were highly dependent on the quantity of the mismatch-cleaving enzyme (**Fig. 2a**) owing to a nonspecific activity of the enzyme.

To characterize our assembly method we cloned and sequenced the resulting products. For points of reference, we carried out circular assembly amplification at a less stringent annealing temperature (65 °C). We also carried out conventional PCA to synthesize *Dpo4* (**Supplementary Fig. 1b**,**c**). The sequencing results are summarized in **Figure 2b**,**c**, **Supplementary Figure 2** and **Supplementary Table 1** online. Exonuclease degradation substantially improved the error rate (0.036%). Additional endonuclease treatment further reduced the synthesis error rate to 0.025%. We observed that increased temperatures for annealing and ligation also contributed to error reduction to a moderate degree (**Fig. 2b**,**c**). In contrast, gene synthesis by conventional PCA method resulted in high error rate of 0.183%.

We used our method to generate a 1.6-kb gene construct (**Supplementary Fig. 3** online), suggesting that this method will be applicable to a majority of genes. We further developed our circular assembly amplification method by adopting recently reported USER enzyme strategy[14] to synthesize even larger genes. This strategy takes advantage of the USER Friendly Cloning kit (New England Biolabs; a mixture of uracil DNA glycosidase and DNA glycosylase-lyase endo VIII), in which a deoxyuridine-excision reaction by the enzyme mix components generates 3′ overhangs on PCR-amplified DNA prepared by the use of primers containing deoxyuridines in the place of deoxythymidines.

In our USER-mediated circular assembly amplification, we used large 3′ overhangs (20 bp or more) by incorporating two deoxyuridines into PCR primers. The large overhangs allowed us to confer higher stringency (melting temperature of 70 °C) during a circular ligation of dsDNA (**Fig. 3** and **Supplementary Methods** for experimental procedures). First we prepared 3 dsDNA fragments of a *Pfu* DNA polymerase (*Pfu*(1–811), *Pfu*(812–1,554) and *Pfu*(1,555–2,325)) via circular assembly amplification (**Fig. 3b**). We then generated 3′ overhangs on the *Pfu* polymerase gene fragments using USER enzyme and constructed full-length circular structures by ligation of the 3 gene fragments followed by exonuclease treatment. We then amplified the full-length circular ligation product, *Pfu*(1–2,325), by PCR. Sequencing of a resulting clone revealed that there were no errors, confirming that this USER-mediated circular assembly amplification approach can be used to synthesize large (>2 kbp) genes without errors.

Many genes and genomes are composed of low-complexity sequences. Assembly of these sequences using conventional gene-synthesis methods is highly challenging because promiscuous annealing events among homologous oligonucleotides result in synthesis of heterogeneous products. We reasoned that exonuclease degradation of the circular ligation mixture would eliminate incorrectly annealed DNA sequences, thereby selecting a desired circular DNA molecule for the next amplification step. Thus, we set out to assemble ~300 bp human minisatellite region composed of five tandem repeats of 45 bp each with 97% homology (see **Supplementary Fig. 4a** online for a target sequence). We constructed the desired sequence by circular assembly of 16 oligonucleotides (**Supplementary Fig. 4b**,**c**). In contrast, conventional methods including a PCA resulted in heterogeneous products. However, we found that even circular assembly amplification did not yield a desired product when we tried to assemble 14 tandem repeats of the 45-bp fragment.

We then reasoned that even large tandem repeats can be prepared by effectively masking highly repetitive regions in a dsDNA structure via USER-mediated circular assembly amplification strategy (**Supplementary Methods** and **Supplementary Fig. 5a** online). We amplified *Dpo4* (1,052 bp) by PCR using four primer sets containing two deoxyuridines. We introduced PCR-amplified

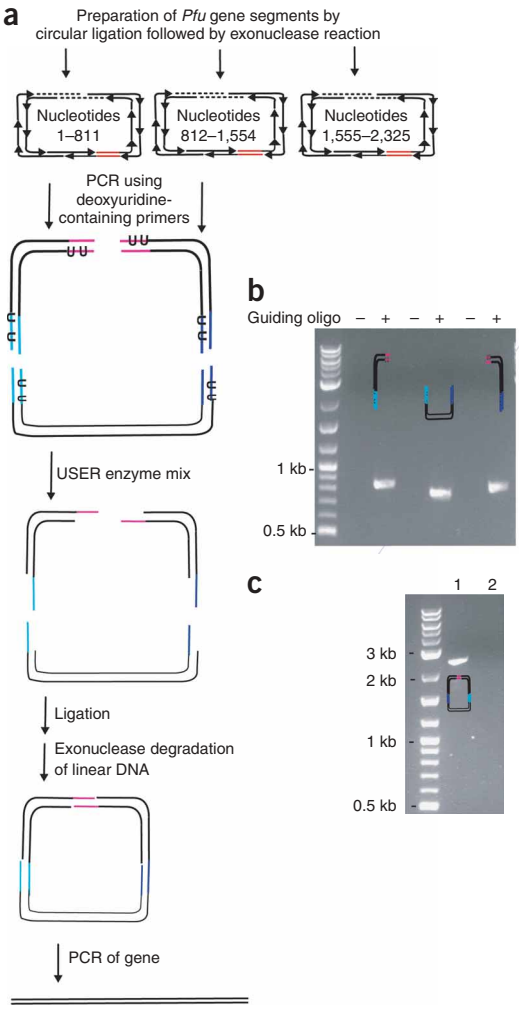

Figure 3 | Circular assembly amplification for the synthesis of *Pfu* DNA polymerase. (**a**) Schematic representation of the synthesis. (**b**) PCR products resulting from the circular assembly amplification of the *Pfu* polymerase gene fragments with or without a guiding oligonucleotide followed by exonuclease treatment. (**c**) PCR products resulting from the USER-mediated circular ligation of the three *Pfu* gene fragments assembled in **b** (circular ligation; lane 1) or after linear ligation (lane 2) followed by exonuclease treatment.

pUC19 DNA as a fifth segment for the circular assembly process. We assembled the five DNA molecules by ligation of the dsDNA containing 3′ overhangs at 70 °C. Transformation followed by characterization of clones showed that four *Dpo4* tandem repeats (~4,400 bp) and a pUC19 fragment (~2,700 bp) were correctly assembled (**Supplementary Fig. 5b**).

Our circular assembly amplification method uses three tiers of selection to reduce gene synthesis error rates by at least a factor of 7 compared to the conventional PCA method. Exonucleases can be used to eliminate a majority of errors occurring during gene synthesis by using an exonuclease-resistant circular structure for the assembly target. This method promises to substantially cut the cost of gene synthesis, as the assembly of an ~1-kb gene (an average length of a gene) can be done in single cycle with less sequencing required to find a perfect construct. Moreover, this method can be used to construct highly repetitive DNA sequences, and our strategy for construction of tandem gene repeats constitutes an effective way of increasing copy number for expression while minimizing the cost overhead of replicating, transcribing and translating the plasmid backbone.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank M. Umbarger, M. Price and J. Aach for critical comments on the manuscript; F. Issacs for comments on experimental design; an anonymous reviewer's suggestion to test repetitive DNA sequence synthesis. We acknowledge funding from US Department of Energy for GTL Center support. D.B is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG#1911-06).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Menzella, H.G. *et al. Nat. Biotechnol.* **23**, 1171–1176 (2005).
2. Han, J.S. & Boeke, J.D. *Nature* **429**, 314–318 (2004).
3. Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. *Gene* **164**, 49–53 (1995).
4. Smith, H.O., Hutchison, C.A., III, Pfannkoch, C. & Venter, J.C. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
5. Tian, J. *et al. Nature* **432**, 1050–1054 (2004).
6. Kodumal, S.J. *et al. Proc. Natl. Acad. Sci. USA* **101**, 15573–15578 (2004).
7. Xiong, A.S. *et al. Nucleic Acids Res.* **32**, e98 (2004).
8. Fuhrmann, M., Oertel, W., Berthold, P. & Hegemann, P. *Nucleic Acids Res.* **33**, e58 (2005).
9. Young, L. & Dong, Q. *Nucleic Acids Res.* **32**, e59 (2004).
10. Carr, P.A. *et al. Nucleic Acids Res.* **32**, e162 (2004).
11. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. *Cell* **107**, 91–102 (2001).
12. Richardson, S.M., Wheelan, S.J., Yarrington, R.M. & Boeke, J.D. *Genome Res.* **16**, 550–556 (2006).
13. SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. USA* **95**, 1460–1465 (1998).
14. Geu-Flores, F., Nour-Eldin, H.H., Nielsen, M.T. & Halkier, B.A. *Nucleic Acids Res.* **35**, e55 (2007).