JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Mäki, Anita; Rissanen, Antti; Tiirola, Marja

Title: A practical method for barcoding and size-trimming PCR templates for amplicon sequencing

Year: 2016

Version: Published version

Copyright: © The Author(s) 2016.

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Mäki, A., Rissanen, A., & Tiirola, M. (2016). A practical method for barcoding and size-trimming PCR templates for amplicon sequencing. BioTechniques, 60(2), 88-90. https://doi.org/10.2144/000114380

A practical method for barcoding and size-trimming PCR templates for amplicon sequencing

Anita Mäki, Antti J. Rissanen, and Marja Tiirola
Department of Biological and Environmental Science, University of Jyväskylä, Finland

BioTechniques 60:88-90 (February 2016) doi 10.2144/000114380
Keywords: next-generation sequencing; PCR; ligation; primer; polyclonality

Supplementary material for this article is available at www.BioTechniques.com/article/114380.

Sample barcoding facilitates the analysis of tens or even hundreds of samples in a single next-generation sequencing (NGS) run, but more efficient methods are needed for high-throughput barcoding and size-trimming of long PCR products. Here we present a two-step PCR approach for barcoding followed by pool shearing, adapter ligation, and 5´ end selection for trimming sets of DNA templates of any size. Our new trimming method offers clear benefits for phylogenetic studies, since targeting exactly the same region maximizes the alignment and enables the use of operational taxonomic unit (OTU)-based algorithms.

The efficiency of next-generation sequencing (NGS) of PCR amplicons has increased via sample barcoding (1), which facilitates multiplex sequencing of numerous samples and genes (such as ribosomal RNA or protein-coding genes) in the same run. Barcodes can be added to the PCR amplicons either by ligation or by performing the PCR amplification with fusion primers, which include both the barcodes and sequencing adapters. When analyzing the sequence diversity of 10 genes in 100 samples, barcoding with fusion primers would require 1000 barcoded primers (10 genes × 100 samples), making the task both laborious and expensive without dual indexing, which is only available on the Illumina MiSeq platform. Another challenge when amplifying target genes with well-established universal primer pairs is related to fragment size optimization. Two commonly used NGS platforms, Illumina and Ion Torrent, recommend maximum fragment lengths of about 300 and 400 nucleotides, respectively (2). Many previously established PCR primer pairs produce much longer amplicons, which must be cut for optimal NGS sequencing. Here we show how library preparation can be simplified with a two-step PCR protocol with M13-tagged primers and how the sample pool can be cut to a certain length all at once instead of performing shearing, adapter ligation, and size selection for each sample separately. Our protocol was validated by sequencing archaeal *16S rRNA* genes from environmental samples using the Ion Torrent (Life Technologies Corporation, Carlsbad, CA) chemistry (with sequencing adapters IonA and P1), but the same template preparation principles are also valid for the other NGS chemistries.

Our cost-efficient, labor-reducing method begins with amplification of each gene library by two-step PCR using barcoded primers, followed by pooling the libraries together (Figure 1). Shearing, ligation with a sequencing 3´ end adapter (P1 on the Ion Torrent platform), and size selection of the amplicons takes place in a single tube, and, very importantly, the process produces sequencing templates with full 5´ ends. Amplicons must be phosphorylated and blunt-ended to effectively ligate adapter P1. In this reaction, ends that are not sheared enzymatically (e.g., with Life Technologies' Ion Shear Plus reagent kit) are not phosphorylated, which prevents ligation of P1 to the 5´ end of the IonA adapter. Two overhanging deoxythymidine nucleotides in the P1 adapter (Supplementary Table S1) prevent the adapter from ligating in a false orientation, and phosphorothioate backbone modification protects the two overhanging nucleotides from exonuclease activity. Fragments that are also sheared from the 5´ side (the IonA side) having P1 on both ends are not efficiently amplified in the subsequent PCR and are not selected during the bead enrichment step, which selects IonA-positive beads. Thus, this method facilitates complete selection of sequences with full 5´ ends.

An M13 linker has been used in nested PCR to reduce the need to invest in fluorescent primers for microsatellite genotyping (3) and for the sequencing of amplicons from different exons of the human epidermal growth factor receptor (EGFR) gene using 454 chemistry (Roche) (4). Recently, barcoding with a similar two-step PCR approach with a 16-bp head-sequence has been designed for the Illumina platform (5) using templates of different sizes. Here, pooling barcoded libraries before shearing and final adapter ligation allows size optimization all

METHOD SUMMARY

Here we present a new protocol combining PCR and adapter ligation for next-generation sequencing (NGS) template preparation that greatly improves sample multiplexing. During library construction, the 5´ sequencing adapter is incorporated in a two-step PCR with universal barcoded M13-tailed primers, and the 3´ adapter is ligated to the pooled and sheared PCR fragments in a single tube, steps that assist in the focused sequencing of the 5´ ends of long PCR fragments. When using long fusion primers for template preparation, selection of the area where polyclonality is detected is required.

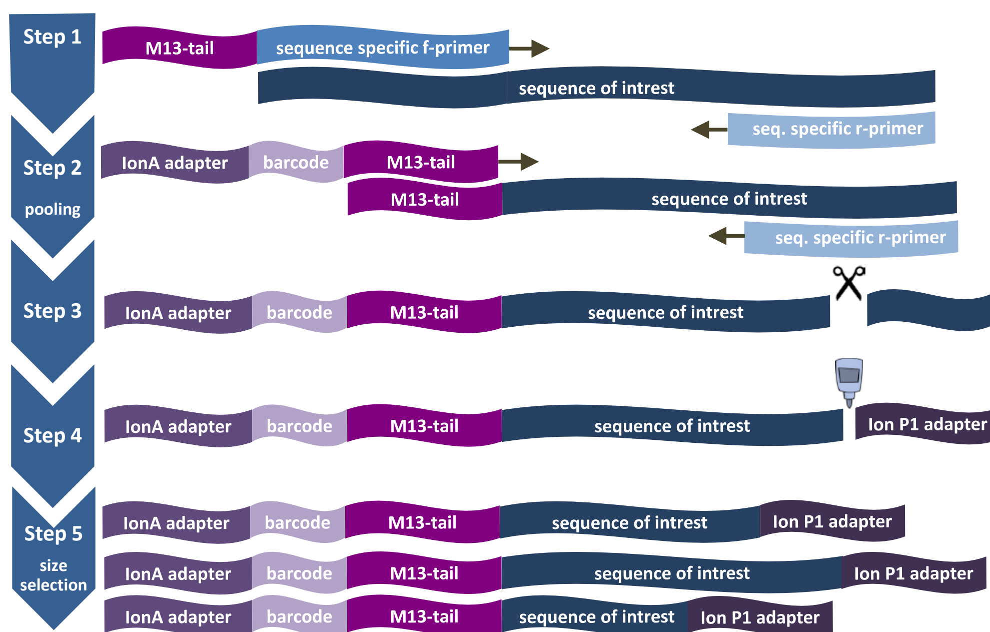

Figure 1. Combining PCR and ligation techniques for barcoding and trimming of long PCR fragments for next-generation sequencing (NGS) library preparation. In Step 1, an M13-tail is incorporated into the PCR products. In Step 2, the 5´ sequencing primer (IonA) and barcodes are incorporated by exploiting the M13 tail. Barcoded samples are then pooled together, and shearing, ligation of the P1 adapter (3´ sequencing primer), and size selection (Steps 3–5) are performed in a single tube.

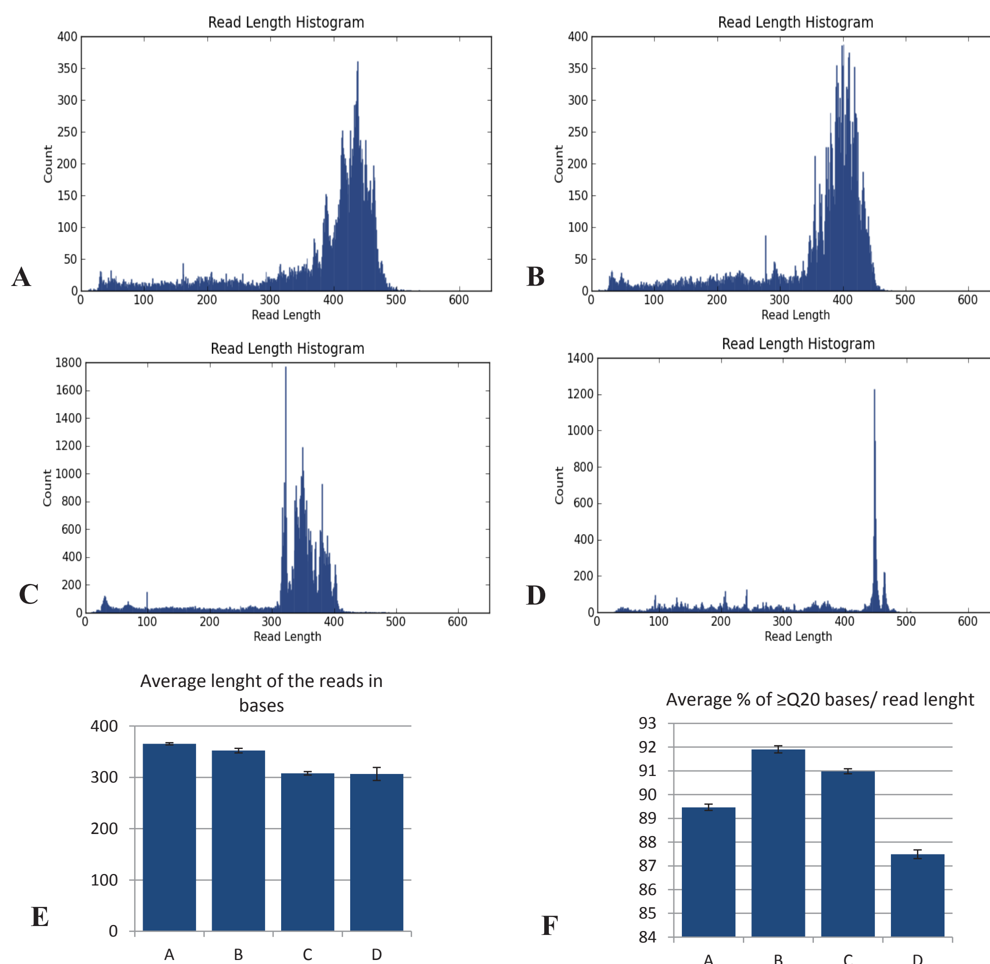

Figure 2. Comparison of read-length histograms and quality bar graphs for the same sequencing run on the Ion Torrent platform containing amplicons from various template preparation methods. The libraries for sequencing were prepared by (A) our method using DNA fragments of the archaeal *16S rRNA* gene; (B) our method using DNA fragments of the gene encoding the α subunit of the archaeal methyl-coenzyme M reductase (*mcrA*); (C) the fusion method (no fragmentation, M13-usage, or ligation) using DNA fragments of the bacterial *16S rRNA gene; or (D) our method with M13-tail usage at the 5´ end but without shearing of the DNA fragments (oversized fragments >500 bp with 3´ P1 adapter incorporated by PCR) of the archaeal mcrA* gene. Accordingly, the bar graphs show the average lengths (±SEM, n = 10) of the reads (in bases) (E) and the average read-specific percentage (±SEM, n = 10) of the bases with quality scores ≥Q20 (F) in these treatments. The average read lengths and the quality of reads were significantly higher for our method for the *mcrA* gene (B) than in the method where DNA fragments were left oversized (D) (read length: t = 9.18, P < 0.001; quality: t = 18.8, P < 0.001). All libraries were added in equimolar concentrations to the emulsion PCR before the main sequencing.

at once, as the Ion Torrent platform is sensitive for long template sizes.

To demonstrate our method using Ion Torrent chemistry, we sequenced fragments of the archaeal *16S rRNA* gene from environmental samples (see protocol in the Supplementary Material). Lyophilized slurries of lake sediment samples were extracted with a Power Soil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA). Partial archaeal *16S rRNA* genes were amplified with forward primer M13–340F and reverse primer 1000R (Supplementary Table S2). Barcodes were added to each amplified sample with another six cycles of PCR where M13-tailed forward primer IonA_bc_M13 was annealed to the M13 sequence of the first PCR products. Amplicon size and yield were checked via agarose gel electrophoresis, and the purification of the PCR products was performed with the Agencourt AMPure XP purification system (Beckman Coulter, Brea, CA). DNA yield was determined with a Qubit 2.0 Fluorometer and a dsDNA HS Assay Kit (Thermo Fisher, Cambridge, UK), and the samples were pooled together. The pooled sample was further purified with AMPureXP, fragmented all at once using an Ion Shear Plus reagent kit (Life Technologies), and, with the same all at once principle, the P1 adapter (Supplementary Table S1) was ligated into fragmented DNA products using the Ion Plus Fragment Library kit (Life Technologies). DNA fragments were size-selected with the Pippin Prep system (Sage Science, Beverly, MA). Amplification of the size-selected fragments was performed using the Platinum PCR SuperMix High Fidelity kit (Life Technologies). Quantitation and size control were performed with the Ion Library TaqMan Quantitation kit (Life Technologies) and with the Agilent Bioanalyzer 2100 (Agilent Technologies, Stockport, UK) using the Agilent High-Sensitivity dsDNA kit. Emulsion PCR with the Ion OneTouch system and Ion OT2 400 kit (Life Technologies) (quality control included), templated bead enrichment, and sequencing with the Ion Personal Genome Machine (PGM) with an Ion PGM Sequencing 400 Kit and Ion 314 chip (Life Technologies) were performed in accordance with the manufacturer's instructions.

A comparative sequencing test was performed on the Ion Torrent platform using equimolar concentrations of libraries representing 4 template preparations (Figure 2): (*A*) our proposed method using DNA fragments of the archaeal *16S rRNA* gene; (*B*) our method using DNA fragments of the gene that encodes the α subunit of the archaeal methyl-coenzyme M reductase (*mcrA*); (*C*) the fusion method (no fragmentation, M13-usage, or ligation) using DNA fragments of the bacterial *16S rRNA* gene; and (*D*) our method with M13-tail usage at the 5´ end but without shearing of the DNA fragments (oversized fragments >500 bp with 3´ P1 adapter incorporated by PCR) of the archaeal *mcrA* gene. Comparatively good average read lengths were achieved using our library preparation method (Figure 2E). The average percentage of the bases per reads whose quality scores were ≥Q20 dropped to the lowest number using oversized *mcrA* libraries (Figure 2F).

We also compared our method with the standard Ion Torrent adapter ligation protocol (Supplementary Figure S1) to study *18S rRNA* genes from phytoplankton samples. The data were analyzed using Mothur (6). The standard method did not yield intact 5´ ends, as forward adapter ligation needs sheared ends. This severely reduced the length of the overlapping area of DNA fragments in subsequent sequence alignments. In contrast, the proposed method, which retains the 5´ ends, maximized the alignment length (data not shown). Thus, the information content utilized in the operational taxonomic unit (OTU) (e.g., at the standard $OTU_{0.97}$ level) clustering and taxonomic classification of OTUs was higher with our method. Our approach, therefore, leads to more accurate identification and more efficient taxonomic classification of OTUs of marker genes (e.g., rRNA or functional genes) than the standard method.

After sequencing the amplicons with long fusion primers, filtering of the polyclonal sequences required adjustment changes in the check region. Using Ion Torrent software (Torrent Suite 4.2.1), polyclonality is, by default, checked during flows 12–70. With extended primer lengths, filtering has to be based on the later region: flows 120–160 (see protocol in the Supplementary Material).

Although NGS is increasingly becoming automated, preparation of multi-sample templates still requires many steps and much manual work. Here, we have shown that exploiting a universal head-sequence (such as M13) and rearranging the order of the steps in the template preparation offers a practical alternative to standard barcoding methods.

Author contributions

A.M. performed the molecular biology study and wrote the paper. M.T. and A.J.R edited the manuscript and offered valuable advice and intellectual contributions. A.M., A.J.R., and M.T. all contributed to the study design.

Acknowledgments

We thank Elina Virtanen for her dedicated technical assistance with PGM, and Anu Mikkonen and Veli-Mikko Puupponen for helping with the bioinformatics. A.M. was supported by Academy of Finland Project 260797 and ERC Micro-RIP 615146 awarded to M.T., and A.J.R was supported by Academy of Finland Project 140964.

Competing interests

The authors declare no competing interests.

References

1. **Parameswaran, P., R. Jalili, L. Tao, S. Shokralla, B. Gharizadeh, M. Ronaghi, and A.Z. Fire.** 2007. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res. *35*:e130.
2. **van Dijk, E.L., H. Auger, Y. Jaszczyszyn, and C. Thermes.** 2014. Ten years of next-generation sequencing technology. Trends Genet. *30*:418-426.
3. **Schuelke, M.** 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. *18*:233-234.
4. **Daigle, D., B.B. Simen, and P. Pochart.** 2011. High-throughput sequencing of PCR products tagged with universal primers using 454 life sciences systems. *Curr Protoc Mol Biol. Chapter 7*:Unit7.5.
5. **Herbold, C.W., C. Pelikan, O. Kuzyk, B. Hausmann, R. Angel, D. Berry, and A. Loy.** 2015. A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. Front Microbiol. *6*:731.
6. **Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, et al.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*:7537-7541.

Received 18 May 2015; accepted 02 November 2015.

Address correspondence to Anita Mäki, Department of Biological and Environmental Science, PO Box 35, FI-40014 University of Jyväskylä, Finland. E-mail: anita.maki@jyu.fi

To purchase reprints of this article, contact: biotechniques@fosterprinting.com