Effective DNA fragmentation technique for simple sequence repeat detection with a microsatellite-enriched library and high-throughput sequencing

Keisuke Tanaka¹, Rumi Ohtake¹, Saki Yoshida², and Takashi Shinohara²

¹NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo, Japan and ²Junior College of Tokyo University of Agriculture, Tokyo, Japan

BioTechniques 62:180–182 (April 2017) doi 10.2144/000114536

Keywords: DNA fragmentation; high-throughput sequencing; microsatellite enrichment; restriction enzyme digestion; Myrtaceae; simple sequence repeat; sonication

Supplementary material for this article is available at www.BioTechniques.com/article/114536.

Two different techniques for genomic DNA fragmentation before microsatellite-enriched library construction—restriction enzyme (*Nla*III and *Mse*I) digestion and sonication—were compared to examine their effects on simple sequence repeat (SSR) detection using high-throughput sequencing. Tens of thousands of SSR regions from 5 species of the plant family Myrtaceae were detected when the output of individual samples was >1 million paired-end reads. Comparison of the two DNA fragmentation techniques showed that restriction enzyme digestion was superior to sonication for identification of heterozygous genotypes, whereas sonication was superior for detection of various SSR flanking regions with both species-specific and common characteristics. Therefore, choosing the most suitable DNA fragmentation method depends on the type of analysis that is planned.

Molecular markers have been used in various biological fields for over 30 years. The application of molecular markers based on simple sequence repeat (SSR) (i.e., microsatellite) regions enables robust assays with high resolution and reliability. Recently, high-throughput DNA sequencing technologies have been used for SSR detection (1–3). This approach may provide additional potentially useful molecular markers; however, developing such molecular markers is often difficult,

and the data obtained typically include only one sequence with annotation information for the SSR flanking regions. An important factor in selecting appropriate SSRs for the evaluation of genetic diversity in a given population is the identification of heterozygous genotypes among SSR regions to avoid the detection of null alleles.

Here, we propose using restriction enzymes for the DNA fragmentation step before microsatellite-enriched sequencing library construction. Use of restriction

enzymes not only reduces the sequencing cost per sample, but could also help distinguish among alleles. To evaluate this hypothesis, we compared SSR detection rates using two different DNA fragmentation techniques: restriction enzyme digestion and traditional sonication.

Five Myrtaceae plants (*Feijoa sellowiana*, *Myrciaria dubia*, *Psidium guajava*, *Psidium littorale*, and *Syzygium samarangense*) were collected from a greenhouse at the Setagaya campus (Tokyo, Japan), the Miyako Subtropical Farm (Okinawa, Japan), and the Isehara Farm (Kanagawa, Japan) of the Tokyo University of Agriculture. Genomic DNA was extracted from the leaf tissue using the cetyltrimethylammonium bromide (CTAB) method (4).

One microgram of genomic DNA was fragmented by digestion with *Nla*III (New England Biolabs, Ipswich, MA) or *Mse*I (New England Biolabs), or by shearing to an average fragment size of 500 bp using an Adaptive Focused Acoustics sonicator (Covaris, Woburn, MA). DNA libraries were then constructed using an NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs). Size selection was conducted using AMPure XP magnetic beads (Beckman Coulter, Brea, CA) with the approximate insert size set to 400–600 bp. The adaptor-ligated DNAs were amplified by 15 cycles of high-fidelity PCR using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs) and NEBNext Multiplex Oligos for Illumina (New England Biolabs). Subsequently, the PCR products were purified into a 20 μ L volume via a clean-up step using AMPure XP magnetic beads (Beckman Coulter).

The cleaned-up products were mixed with 1 μ L (from a 100 μ M stock in TE buffer) biotinylated SSR probe (GA)₁₀ (Hokkaido System Science Co. Ltd., Hokkaido, Japan) (one of the probes typically used for microsatellite enrichment), incubated at 95°C for 10 min, and placed on ice. The mixtures were hybridized by incubating at 60°C for 1 h. After washing with 20 μ L Dynabeads MyOne streptavidin C1 (Life Technologies, Carlsbad, CA), streptavidin beads were resuspended in 29 μ L 6 $\times$

METHOD SUMMARY

Two genomic DNA fragmentation methods—restriction enzyme digestion and sonication—were compared for use in simple sequence repeat (SSR) detection. After microsatellite-enriched library construction and high-throughput sequencing of DNA fragments generated using each method, SSR regions were detected and compared through integration of paired-end reads and clustering of integrated reads.

saline-sodium citrate (SSC) buffer, added to each hybridized mixture, and incubated at room temperature for 30 min. The mixtures were washed with 2× SSC buffer once and then with 1× SSC buffer twice. Next, 23 µL of the SSR-enriched fragments was amplified by 15 cycles of high-fidelity PCR. These amplified products were purified to a 20-µL volume via a clean-up step using AMPure XP magnetic beads.

Library quality and concentration were assessed using an Agilent Bioanalyzer 2100 system (Agilent Technologies, Waldbronn, Germany) and a KAPA Library Quantification Kit (Kapa Biosystems, Woburn, MA). The library mixture was sequenced by 2 × 300-bp paired-end sequencing using a MiSeq (Illumina, CA). The read data were submitted to the DDBJ Read Archive (http://trace.ddbj.nig.ac.jp/dra/index_e.html) (accession no. DRA004725).

Raw reads containing adaptors were removed using Trimmomatic version 0.32 (5). Additionally, quality filtering was run using FASTX-Toolkit version 0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). The preprocessed paired reads were integrated by the minimum overlapping length parameter of “min 60” using FLASH version 1.2.11 (6). Furthermore, the integrated reads with the same sequence were clustered using CD-HIT version 4.6 (7). The clustered reads including SSR regions were searched using SSRIT version 0.0.1 (8). The listed sequences were identified as either homozygous or heterozygous genotypes by counting the number of identical sequences for the detected SSR

Figure 1. Venn diagram based on annotation information for the simple sequence repeat (SSR) flanking regions in five Myrtaceae plants. The values inside each frame show the numbers of unique or common genes found among five species, and the proportions are indicated in parentheses. The values are determined based on fragmentation by *Nla*III (red), *Mse*I (green), and sonication (black).

flanking region. Moreover, the sequences were annotated with the top-hit description using the local BLAST program with the following settings: (i) execution program: BLASTn; (ii) database: the NCBI nonredundant nucleotide database nt; and (iii) E-value: 1e-4.

The numbers of paired reads ranged from 2 × 1,296,448 to 2 × 1,708,634. SSR detection was performed using these read

data. After searching SSR regions from the clustered reads using the overlapping length parameter of min 60, most motifs (82%–89%) were composed of (CT)_n, (TC)_n, (GA)_n, or (AG)_n (Supplementary Figure S1). This result could be explained by the application of a biotinylated probe, which was utilized to enrich the SSR regions. Various probe conditions required for capturing the SSR regions have been reported previously (9–11). Our protocol also showed that target SSR regions could be captured efficiently. Although differences in the numbers of paired reads were small, the number of detected microsatellite regions differed markedly depending on whether fragmentation was carried out with restriction enzyme digestion or with sonication. This could be due to differences in genomic coverage for detecting SSRs because detection using restriction enzyme-digested genomic DNA requires the restriction site to flank the SSRs, whereas detection using sonicated genomic DNA can cover the entire genome.

Among the SSRs shown in Supplementary Figure S1, probe-related SSR regions were characterized based on genotypic frequency (Table 1). All species tested with both fragmentation conditions showed high rates of homozygous

Table 1. Frequencies of homozygous and heterozygous genotypes in probe-related simple sequence repeat (SSR) regions.

Species	Fragmentation	Homozygous genotype (Mono-allele)	Heterozygous genotype (Di-allele)*	(Other)
<i>Feijoa sellowiana</i>	<i>Nla</i> III	8,997 (81.25%)	1,332 (12.03%) [2–46]	744 (6.72%)
	<i>Mse</i> I	10,160 (84.31%)	1,236 (10.26%) [2–40]	655 (5.44%)
	Sonication	24,266 (99.92%)	20 (0.08%) [2–10]	0 (0.00%)
<i>Myrciaria dubia</i>	<i>Nla</i> III	11,935 (79.13%)	1,768 (11.72%) [2–40]	1,379 (9.14%)
	<i>Mse</i> I	12,092 (81.23%)	1,650 (11.08%) [2–54]	1,144 (7.69%)
	Sonication	32,168 (99.93%)	24 (0.07%) [2–2]	0 (0.00%)
<i>Psidium guajava</i>	<i>Nla</i> III	8,494 (79.24%)	1,308 (12.20%) [2–34]	917 (8.55%)
	<i>Mse</i> I	9,989 (83.49%)	1,192 (9.96%) [2–44]	784 (6.55%)
	Sonication	22,778 (99.94%)	14 (0.06%) [2–4]	0 (0.00%)
<i>Psidium littorale</i>	<i>Nla</i> III	19,235 (79.94%)	3,032 (12.60%) [2–56]	1,794 (7.46%)
	<i>Mse</i> I	20,069 (82.17%)	2,628 (10.76%) [2–62]	1,728 (7.07%)
	Sonication	26,639 (99.84%)	42 (0.16%) [2–4]	0 (0.00%)
<i>Syzygium samarangense</i>	<i>Nla</i> III	13,490 (80.76%)	2,094 (12.54%) [2–52]	1,119 (6.70%)
	<i>Mse</i> I	18,640 (84.34%)	2,254 (10.20%) [2–52]	1,206 (5.46%)
	Sonication	27,348 (99.96%)	12 (0.04%) [2–4]	0 (0.00%)

*The values within square brackets show the range of allele size differences for the same loci in di-alleles. Other alleles in the heterozygous genotype were more than di-allelic (e.g., tri-, tetra-, and penta-allelic).

Superb Sensitivity. No Thermocycler.

Researchers in the UK developed an assay that can detect as little as 14pg/µl of purified RNA from a plant infected with Yam mosaic virus; comparable to the gold standard RT-PCR method, but without the need of a thermocycler.

Want to know more? Read the proof: twistdx.co.uk/yam

genotypes and low rates of heterozygous genotypes. However, the heterozygous genotype rate for all species was substantially higher after fragmentation using restriction enzymes than after fragmentation by sonication (18.75%–20.86% for *Nla*III, 15.66%–18.77% for *Mse*I, and 0.04%–0.16% for sonication). Additionally, the probability of detecting the heterozygous genotype was higher with fragmentation by *Nla*III than by *Mse*I. Thus, our approach using restriction enzyme digestion could effectively identify heterozygous genotypes. Although all five Myrtaceae plants used in this study are diploid, approximately one-third of the identified heterozygous genotypes in all species were more than tri-allelic. This may reflect the occurrence of either multiple homologous copies or PCR errors during library construction.

We next determined the numbers of genes included in the detected SSR flanking regions that were unique or common among the five species (Figure 1). The detected SSR flanking regions had both species-specific and common characteristics. Notably, more of the SSR flanking regions could be annotated with the sonicated DNA targeting the whole genome than with the restriction enzyme-digested DNA.

Our comparison of two different methods for fragmenting genomic DNA revealed that the use of restriction enzymes was more suitable for the identification of heterozygous genotypes than sonication, whereas sonication was superior for the detection of various SSR flanking regions with both species-specific and common characteristics. We conclude that the choice of DNA fragmentation should depend on the type of analysis that DNA will be used for. In particular, the effective detection of a heterozygous genotype through DNA fragmentation by restriction enzymes is expected to contribute to the development of molecular markers for molecular-assisted breeding and population genetics that require clear distinction among alleles.

Author contributions

K.T. contributed to the methodology development, data analysis, and writing of the manuscript. R.O. contributed to the library construction and the sequencing operations. S.Y. contributed to the concept of the study and sample preparation. T.S. contributed to the concept of the study and the sample preparation, and also supervised the study.

Acknowledgments

We thank the Miyako Subtropical Farm and the Isehara Farm of Tokyo University of Agriculture for providing the plant materials. This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology-Supported Program for the Strategic Research Foundation at Private Universities (S1311017).

Competing interests

The authors declare no competing interests.

References

1. Abdelkrim, J., B.C. Robertson, J.A.L. Stanton, and N.J. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques* 46:185-192.
2. Jennings, T.N., B.J. Knaus, T.D. Mullins, S.M. Haig, and R.C. Cronn. 2011. Multiplexed microsatellite recovery using massively parallel sequencing. *Mol. Ecol. Resour.* 11:1060-1067.
3. Malausa, T., A. Gilles, E. Meglecz, H. Blanquart, S. Duthoy, C. Costedoat, V. Dubut, N. Pech, et al. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol. Ecol. Resour.* 11:638-644.
4. Doyle, J. and J.L. Doyle. 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* 19:11-15.
5. Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
6. Magoc, T. and S.L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957-2963.
7. Li, W. and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658-1659.
8. Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11:1441-1452.
9. Acquadro, A., E. Portis, and S. Lanteri. 2003. Isolation of microsatellite loci in artichoke (*Cynara cardunculus* L. var. *scolymus*). *Mol. Ecol. Notes* 3:37-39.
10. He, G., R. Meng, M. Newman, G. Gao, R.N. Pittman, and C.S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 3:3.
11. Stajner, N., J. Jakse, P. Kozjak, and B. Javornik. 2005. The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). *Plant Sci.* 168:213-221.

Received 07 October 2016; accepted 27 February 2017.

Address correspondence to Takashi Shinohara, Junior College of Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo, 156-8502, Japan. E-mail: t3shinoh@nodai.ac.jp

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