Nucleic Acids Research, 1996, Vol. 24, No. 21 **4369–4371**

Single locus microsatellites isolated using 5′ anchored PCR

P. J. Fisher[1,2,*], R. C. Gardner[2] and T. E. Richardson[1]

[1]New Zealand Forest Research Institute Ltd, Private Bag 3020, Rotorua, New Zealand and [2]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

Received May 14, 1996; Revised and Accepted September 10, 1996

ABSTRACT

Microsatellites are widely used as genetic markers because they are co-dominant, multiallelic, easily scored and highly polymorphic. A major drawback of microsatellite markers is the time and cost required to characterise them. We have developed a novel technique to reduce this cost by producing a microsatellite-rich PCR profile from genomic DNA which was cloned to yield a genomic library enriched for microsatellites. Sequence data and subsequent allele scoring within pedigrees revealed that these microsatellites retained their original repeat length and segregated normally. This technique permits genomic amplification with only one specific primer. Together with enrichment, the savings in primer costs reduces the cost of microsatellite characterisation considerably.

Microsatellite loci are widely used for DNA fingerprinting (1), linkage map construction (2,3), paternity analysis (4), and various population genetic studies (5) because they are co-dominant, easily scored and highly polymorphic (6). A major drawback of single locus microsatellite markers is the time and cost required to isolate and characterise each locus, a process typically involving library construction and screening, DNA sequencing, PCR primer design and PCR optimisation. To capture some of the polymorphism associated with microsatellite loci without the expense of single locus isolation, two multi-locus DNA profiling approaches have been developed which target microsatellite regions in the genome. One approach is to probe genomic DNA with oligonucleotides complementary to microsatellites (7,8). A second approach is PCR using primers containing microsatellite repeats (9,11). While both approaches have proven useful for multi-locus DNA profiling, the repeat length polymorphism of individual microsatellite loci is typically not assayed.

Some PCR-based approaches, however, have employed PCR primers containing microsatellite motifs at the 3′ end and three or more non-repetitive bases at the 5′ end (9,10,12). The 5′ bases are expected to ‘anchor’ the primer to the 5′ end of a microsatellite. In practice this approach has proven difficult to optimise for complex genomes and there has been little evidence that variation in microsatellite repeat length at the primer binding sites has been preserved (9). These difficulties arise because the PCR primers fail to anchor at the 5′ end of microsatellites, instead slipping to the 3′ end of the microsatellites during PCR. Each amplified fragment therefore contains exactly the number of repeats found in the primers and any repeat length polymorphism is lost.

This paper describes the development of a 5′ anchoring procedure that consistently anchors PCR primers at the 5′ ends of microsatellites, amplifying two close and inverted simple sequence repeats and the region between them. This technique offers a number of advantages compared with previously published protocols. Firstly, amplification of genomic DNA yields PCR products containing at least two microsatellites (one on each end of the amplicon) which retain their original repeat length. The multi-locus profiles produced from 5′ anchored primers should be more polymorphic than those from non-anchored primers. Secondly, this technique can be used to produce libraries which are highly enriched for single locus microsatellites. Lastly many of these individual loci can be amplified with only one locus-specific primer which further reduces the cost of microsatellite discovery.

Primer design and PCR testing

A degenerate primer (PCT4) was designed with the sequence KKVRVRV$(CT)_6$, where K = G/T, V = G/C/A and R = G/A. The $(CT)_6$ component of the primer was designed to anneal to $(GA)_n$, where $n \geq 6$, and the seven 5′ nucleotides in the primer form the ‘anchor’. The two most 5′ nucleotides in the anchor were designed to anneal to any nucleotide, with G used to pair with C or T, and T used to pair with A or G. The next five nucleotides, designated ‘blocking’ bases, were designed so that they would not pair with GA repeats, but were otherwise as redundant as possible. Thus V (G, A or C) will pair with any nucleotide except A, while R (G or A) will not pair with G. Also, R will not pair with A because T was excluded from the base to avoid T-G binding. These redundancies mean that this primer should be complementary to one in six of all possible random sequences adjacent to GA repeats.

Determination of the PCR conditions that produced consistently anchored amplicons was crucial. Excessive stringency should produce no amplicons, whereas low stringency will permit slippage of the primer to the 3′ end of the targeted microsatellite loci resulting in loss of repeat length variation. To prevent this slippage, the annealing stringency had to be high enough so that amplification occurred only when one or more of the five

*To whom correspondence should be addressed at: New Zealand Forest Research Institute Ltd, Private Bag 3020, Rotorua, New Zealand. Tel: +64 7 347 5699; Fax: +64 7 347 9380; Email: fisherp@tawa.fri.cri.nz

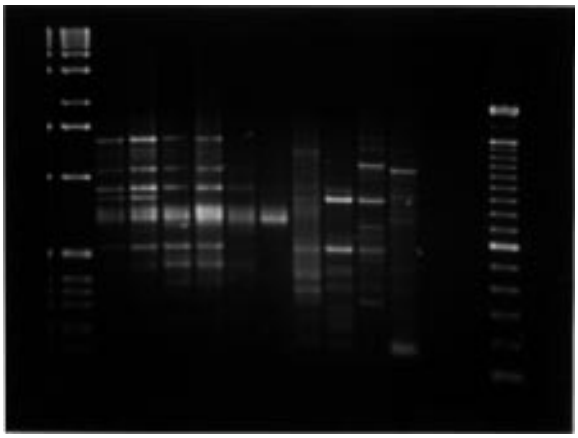

Figure 1. PCR profiles produced from genomic DNA. Lane 1, 1 kb DNA ladder; lanes 2–5, *P.radiata* using primer PCT4 with 1.5, 1.75, 2.0 and 2.5 mM $MgCl_2$ respectively. The PCR profiles shown in lanes 6–13 were performed at 1.5 mM $MgCl_2$. Lane 6, *P.taeda*; lane 7, *P.strobus*; lane 8, *Ps.menziesii*; lane 9, *C.macrocarpa*; lane 10, *L.esculentum*; lane 11, *A.chinensis*; lane 12, *E.coli*; lane 13, no DNA control; lane 14, 100 bp DNA ladder. The annealing temperature for PCR was 59°C for the first five cycles, followed by 57°C for 35 cycles.

blocking bases paired with the template. For the initial optimisation experiments, plasmids containing *Pinus radiata* microsatellites with known 5′ flanking sequence were used. With these plasmids we could manipulate the degree of primer–plasmid complementarity and then detect successful 5′ anchoring based on the size of the amplified product over a range of annealing temperatures and $MgCl_2$ concentrations. These experiments revealed that the 'anchoring window' was approximately ±4°C for a sequence which was perfectly complementary to the five blocking bases. (That is, with annealing temperatures of 53–61°C, primer PCT4 amplified fragments of the size expected for faithful 5′ anchoring.) However, when the sequence matched at only two of the five blocking bases, this window was reduced to ±1°C, and a high proportion of the amplicons resulted from unanchored amplification (data not shown).

The effect of $MgCl_2$ concentration on the number of fragments produced by PCT4 directed amplification of *P.radiata* genomic DNA is shown in Figure 1 (lanes 2–5). A discrete banding pattern was observed above a background smear. As the $MgCl_2$ levels increased, a higher proportion of background smear was observed, and there were changes in the visible banding pattern, indicative of a decrease in annealing stringency.

Amplification of multi- and single-locus profiles

Following these results, genomic DNA from *P.radiata*, *P.taeda*, *P.strobus*, *Pseudotsuga menziesii*, *Cupressus macrocarpa*, *Actinidia chinensis*, *Lycopersicum esculentum* and *Escherichia coli* was amplified using primer PCT4 at 1.5 mM $MgCl_2$ (Fig. 1). Several PCR products were observed in the profiles of each species except *E.coli*, which is believed to lack GA repeats (11). The profiles show clear inter-specific polymorphisms. The profiles of *P.taeda* (lane 6) and *P.radiata* (lane 2) are very similar, suggesting that many of the binding sites may be shared by these two pine species.

PCR products produced by PCT4 amplification of *P.radiata* genomic DNA were cloned into pGEM-T (Promega). Sequence analysis of eight randomly selected clones revealed that each was unique, each contained a $(CT)_n$ microsatellite at both ends of the insert, and the 16 microsatellites ranged from 6 to 12 repeats. This pattern of repeat lengths is entirely consistent with successful 5′ anchoring of the primers during PCR amplification. To identify the longest microsatellites within this library we probed ~1000 colonies at high stringency (final wash at 60°C, 0.1× SSC) with a $(GA)_{16}$ oligonucleotide. Fifteen positive clones were identified and sequenced; 13 of these were unique. Each clone contained two terminal $(CT)_n$ microsatellites, and in each case, at least one microsatellite had $n \geq 14$. The presence of terminal microsatellites containing more than six repeats confirmed that anchoring had been successful. Five of the clones had one or more additional microsatellites in the internal sequence, including one clone (number 7 in Table 1) that had seven tandem repeats of a 95 bp region, each containing a microsatellite. In total, the 13 clones contained 19 $(CT)_n$ microsatellite loci with $n \geq 14$.

To test for polymorphism and Mendelian transmission, specific primers were designed to 10 of the longest microsatellite loci and used to amplify *P.radiata* genomic DNA. An important feature of this approach is that only one additional primer was required to amplify the terminal microsatellites. The internal microsatellites were amplified by two specific primers, as normal (13). Figure 2 shows one of the polymorphic loci (locus 5.16) which was amplified using primer PCT4 and one locus-specific oligonucleotide. The size of one of the alleles amplified from one individual tree (lane 5) corresponded exactly with that obtained from the sequenced plasmid (derived from the same tree). Seven other microsatellite loci were amplified with locus specific primers in conjunction with PCT4 and in every case, the size of one allele from this tree corresponded with that obtained from the cloned plasmid. This data confirmed that PCT4 was reliably anchoring at the 5′ end of the microsatellite. Mendelian inheritance of polymorphic loci was confirmed by genotyping two parental trees, haploid megagametophytes from each tree and several progeny (Fig. 2). The parents (lanes 5 and 16) and their progeny (lanes 6–15) each have two alleles of the expected sizes, and the haploid megagametophytes (lanes 1–4 and 17–20) have one allele each (consistent with the maternal parent). Radioactive labelling of amplicons was obtained by incorporating [α-^{32}P]dCTP in the PCR reaction. Surprisingly, few low molecular weight bands were observed after amplification of genomic DNA using PCT4 and the specific primer. However, if background bands were obtained from PCT4 amplification, they could be avoided by end-labelling the specific primer.

This 5′ anchored approach should be applicable to different types of di-, tri- and tetranucleotide repeats and to any genome containing abundant microsatellites. This technique has two major advantages over other methods for characterising single locus microsatellites. Firstly, construction of an enriched library is technically simple and fast. Secondly, once microsatellite loci are identified, most loci require the design and synthesis of only one additional PCR primer, thereby reducing the cost per locus. The presence of additional internal microsatellites was an unexpected bonus, implying that there may be clustering of microsatellites in some genomic regions.

Table 1. Microsatellites isolated with the 5′ anchor technique

Clone	5′ terminal repeat	Internal repeat(s)	3′ terminal repeat
1	$(CT)_{14}$	$(CT)_6$	$(AG)_6$
2	$(CT)_{15}.(AT)_{16}$		$(AG)_6$
3	$(CT)_{18}$		$(AG)_6$
4	$(CT)_{18}$		$(AG)_6$
5	$(CT)_{19}$		$(AG)_6$
6	$(CT)_{21}$	$(CT)_6$	$(AG)_6$
7	$(CT)_6$	$(TC)_{21},(CT)_{19},(CT)_{19},(TC)_{15},(CT)_{14},(CT)_{13},(CT)_{10}$	$(AG)_6$
8	$(CT)_{25}$	$(AG)_{20}$	$(AG)_8$
9	$(CT)_8$	$(CT)_{25},(CT)_6,(CT)_6$	$(AG)_6$
10	$(CT)_{27}.(AT)_{22}$		$(AG)_6$
11	$(CT)_{17}$	$(CT)_{29},(CT)_6$	$(AG)_6$
12	$(CT)_7$	$(TC)_{33},(CT)_6,(CT)_6$	$(AG)_6$
13	$(CT)_{53}$		$(AG)_6$

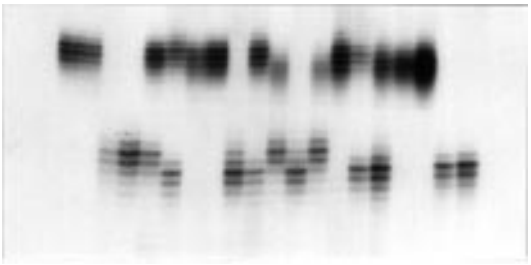

Figure 2. Segregation of *P.radiata* locus 5.16. PCR was performed using primer PCT4 and a specific primer (5′-GAAGCCTCAACATTGATT-CATG-3′). Lanes 1–4, megagametophytes from maternal parent; lane 5, maternal parent; lanes 6–15, progeny of the cross between the parents; lane 16, paternal parent; lanes 17–20, megagametophytes from paternal parent. The annealing temperature for PCR was 61°C for 10 cycles, followed by 57°C for 35 cycles.

Procedures for extraction of plasmid DNA (14) and plant genomic DNA (15) have been described. Multi-locus PCRs were performed in a Techne PHC-3 thermocycler using 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 3 U *Taq* DNA polymerase, 50 pmol PCT4, 30 ng genomic DNA, 0.2 mM each of dNTPs and deionised water (to 25 µl) with a 50 µl oil overlay. The amplification profile consisted of 94°C (3 min), 5 cycles of 93°C/59°C*/72°C (30 s each), 35 cycles of 93°C/57°C*/72°C (30 s each), and 1 cycle at 72°C (2 min). Agarose gel electrophoresis and subcloning were as described (14). Radioactively labelled PCRs were performed using 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1 U *Taq* DNA polymerase, 20 pmol PCT4, or 4 pmol of each specific primer, 10 ng genomic DNA, or 250 fg plasmid DNA, 0.2 mM each of dATP, dGTP and dTTP, 0.02 mM dCTP, 0.3 pmol [α-^{32}P]dCTP and deionised water (to 10 µl), with a 50 µl oil overlay. Radioactive labelled PCR products were electrophoresed in 6% polyacrylamide/urea gels (14) using Life Technologies S2 apparatus. Ligations were carried out according to the manufacturers' instructions with Promega T4 DNA ligase using 50 ng pGEM-T vector (Promega). *Escherichia coli* DH10-β cells were transformed with ligation mix using a Bio-Rad Gene Pulser, according to the manufacturers' instructions in cuvettes with a 0.1 cm electrode gap. The transformed colonies were lifted onto Du Pont hybridisation transfer nylon membranes according to the manufacturers' instructions. The colonies were probed with a $(GA)_{16}$ oligonucleotide (14), which was end-labelled with [γ-^{32}P]ATP according to the enzyme manufacturers' instructions (Boehringer). Positive clones were sequenced using an ABI 373 automated sequencer. Specific primers were designed using PRIMER software (16).

*The annealing temperature of PCRs which use two primers other than PCT4 varied depending on the T_m of the primers.

ACKNOWLEDGEMENTS

Research of the authors is supported by a New Zealand PGSF-FRST grant (CO4310) and the FRI-GEENZ joint venture.

REFERENCES

1 Thomas, N.R. and Scott, N.S. (1993) *Theor. Appl. Genet.* **86**, 985–990.
2 Bowcock, A.M., Ruiz-Linares, A., Tomforhde, J., Minch, E., Kidd, J.R. and Cavalli-Sforza, L.L. (1994) *Nature* **368**, 455–457.
3 Hazan, J., Dubay, C., Pankowiak, M., Becuwe, N. and Weissenbach, J. (1992) *Genomics* **12**, 183–189.
4 Amos, B., Schlötterer, C. and Tautz, D. (1993) *Science* **260**, 670–671.
5 Bell, C.J. and Ecker, J.R. (1994) *Genomics* **19**, 137–144.
6 Queller, D.C., Strassmann, J.E. and Hughes, C.R. (1993) *Trends Ecol. Evol.* **8**, 285–288.
7 Bierworth, S., Kahl, G., Weigand, F. and Weising, K. (1992) *Electrophoresis* **13**, 115–122.
8 Richardson, T., Cato, S., Ramser, J., Kahl, G. and Weising, K. (1995) *Nucleic Acids Res.* **23**, 3796–3799.
9 Zietkeiwicz, E., Rafalski, A. and Labuda, D. (1994) *Genomics* **20**, 176–183.
10 Wu, K., Jones, R., Danneberger, L. and Scolnik, P.A. (1994) *Nucleic Acids Res.* **22**, 3257–3258.
11 Weising, K., Atkinson, R.C. and Gardner, R.C. (1995) *PCR Methods Applic.* **4**, 249–255.
12 Wolff, K., Zietkiewicz, E. and Hofstra, H. (1995) *Theor. Appl. Genet.* **91**, 439–447.
13 Litt, M. and Luty, J.A. (1989) *Am. J. Hum. Genet.* **44**, 397–401.
14 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
15 Janssen, B.-J. and Gardner, R. (1993) *Plant Cell Rep.* **13**, 28–31.
16 Lincoln, S.E., Daly, M.J. and Lander, E.S. (1991) PRIMER Version 0.5: A computer program for automatically selecting PCR primers.