5512–5513 Nucleic Acids Research, 1994, Vol. 22, No. 24

Rapid PCR-based characterization of sequences flanking microsatellites in large-insert libraries

J.Rothuizen* and M.van Raak
Utrecht University, Department of Clinical Sciences of Companion Animals, Yalelaan 8, 3508 TD Utrecht, The Netherlands

Received July 18, 1994; Revised and Accepted October 21, 1994

Microsatellite polymorphisms are widely being employed as a major tool in genome analysis. Analysis of length polymorphisms in the PCR depends on the elucidation of the sequence of their flanking regions. Different methods have been developed for the efficient characterization of the flanking sequences. Small insert genomic libraries directly suitable for sequencing of the entire insert and highly enriched for the desired microsatellite sequence may be obtained by a number of strategies (Ostrander *et al.*; Karagyozov *et al.*, 1993; Kandpal *et al.*, 1994). The limitation of small insert libraries is, however, that the inserts are too small for physical mapping by *in situ* hybridization. Microsatellites from large insert genomic libraries suitable for the construction of both a genetic and a physical map are routinely analyzed by subcloning into small insert libraries. This elaborate approach may be circumvented by direct outward sequencing with a set of six primers complementary to the repeat motif having one of the six theoretically possible 3′ flanking bases (Yuille *et al.*, 1991). Disadvantages of the latter procedure are that for both flanks twelve sequence reactions are required, and that ambiguous results can be obtained if the repeat is imperfect, compound or palindromic. We have developed an alternative method that is more generally applicable to analysis of flanking sequences of microsatellites and have compared the results with those of outward sequencing according to Yuille *et al.* (1991).

Large insert recombinants were obtained from a previously characterized (Rothuizen and Wolfswinkel, 1994) genomic dog liver library in λ EMBL-3 SP6/T7 (Clontech, Palo Alto, USA). Plaque lifts were detected by chemiluminescence after hybridization with digoxygenin-labeled $(CA)_{15}$ and ten positive clones were chosen at random. Microsatellite containing clones were grown in liquid culture and DNA was then isolated with Wizard lambda preps (Promega, Leiden, The Netherlands). Insert sizes, as determined by releasing them by digestion with *Xho*I and electrophoresis on agarose, ranged from 11–18 kb (mean 16 kb). PCR-based elucidation of the regions flanking the microsatellites was performed as summarized in Figure 1. Ten μg of DNA of each recombinant were digested with 20 U of restriction endonuclease for 3 hours in 50 μl reactions, to produce blunt-ended fragments. Endonucleases were *Eco*RV, *Hae*III, *Rsa*I, *Ssp*I, or the combination of the four enzymes (20 U each) in fourfold digests. These enzymes do not cleave $(CA)_n$ stretches. Fragment sizes were determined by electrophoreses in 1% agarose gels by EtBr staining. The fragment containing the microsatellite was identified by hybridization with the $(CA)_{15}$ probe after Southern blotting onto positively charged nylon membrane (Boehringer Mannheim, Germany). In all fourfold digests the microsatellites occurred in a small fragment of 200–900 bp (mean size ± 450 bp); the relevant fragment after digestion with the single enzymes varied from 0.8–5.4 kb with *Hae*III and *Rsa*I giving the smaller products. Further procedures were performed with the fourfold digests. Two ng of digested DNA were ligated onto 4 pmol of a blunt-ended adaptor (Uni-amp, Clontech) in a 10 μl reaction. The products were then amplified in the PCR using one primer complementary to the adaptor (5′CCTCTGAAGGTTCCAGAATCGATAG) and another primer having a 5′ biotin label, either $(CA)_{15}$ or $(GT)_{15}$ (primers assembled by Pharmacia/Biotech, Roosendaal, The Netherlands). PCR was performed in 50 μl containing 25 pmol of each primer, 2.5 μl of the ligation mixture, 0.2 mM dNTP, 1.5 mM $MgCl_2$, 1.25 U *Taq* polymerase (Promega), and 1× polymerase buffer. The cycling temperature profile was 1 min 94°C, 1 min 60°C, and 2 min 72°C, for 35 cycles. The amplified biotinylated flanking region of the repeat was isolated and purified by capturing onto 50 μl streptavidin-labeled magnetospheres (10 μg/μl; Dynal, Oslo, Norway) and triplicate washing with buffer consisting of 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, and 2 M NaCl. The isolated flanking region was denatured with 0.15 M NAOH for 5 min, and the non-biotinylated strand was removed by washing once with 0.15 M NaOH, and four times with the above wash buffer. Single stranded magnetosphere-coupled DNA was then dissolved in 10 μl water and sequenced with the primer complementary to the adaptor, using a standard technique (T7 polymerase sequencing kit, Pharmacia/Biotech). Annealing was allowed for 20 min at 37°C. Samples were electrophoresed for 75 and 150 min. Results were compared with outward sequencing using six primers $(CA)_{15}$ having a 3′ anchoring base for each flank according to Yuille *et al.* (1991).

Nine of the ten clones gave a clear sequence of the microsatellite flank and part of the $(CA)_n.(GT)_n$ stretch itself following magnetic capture of the biotinylated $(CA)_n$ or $(GT)_n$ strand of the PCR product. Typically, some 30 bases of the microsatellite could be read; thereafter the sequence reaction lost its specificity due to the fact that annealing-out-of-range (Mariat and Vergnaud, 1992) of the non-anchored biotinylated PCR primer had produced a mixture of products of different length as visualized by agarose gel electrophoresis. The length of the sequence obtained from the end of the fragment to the start of the microsatellite varied from 48 to 220 bases (mean 92 bases),

* To whom correspondence should be addressed

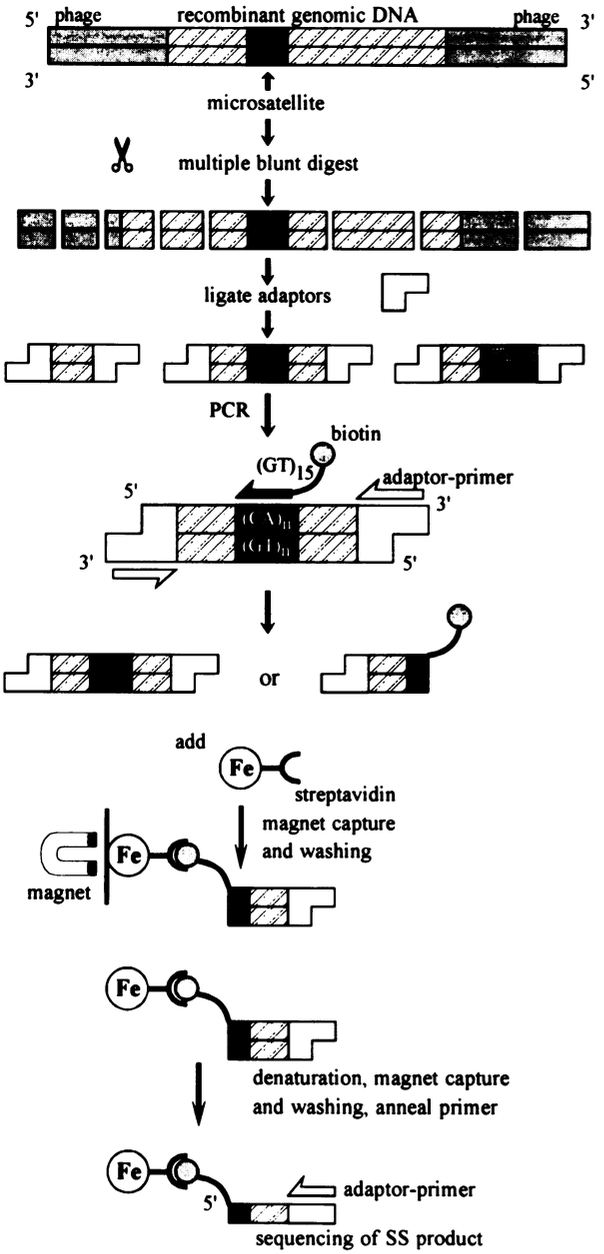

Figure 1. Schematic representation of bacteriophage restriction by four blunt cutting enzymes, ligation to an adaptor, amplification of the fragments and one flanking region with a primer secondary to the adaptor and a 5'-biotinylated $(CA)_{15}$ or $(GT)_{15}$, isolation of the biotinylated flanking region of the microsatellite, and subsequent solid phase sequencing.

which was sufficient for development of a flanking primer in all cases. For one clone no sequence could be obtained from either of the two flanks. In comparison, direct outward sequencing according to Yuille *et al.* (1991) was successful in five of the ten clones. In the other five no specific sequence reaction was obtained, including the one that could not be read by the other method. The four microsatellites that could be sequenced by the present method but not by that of Yuille *et al.* were all interrupted by one or more other sequence stretches. Both methods gave the same sequence of the flanking region of the microsatellite in five clones.

The present technique permits the rapid isolation of microsatellite-containing regions of large insert recombinants and sequencing of the flanking regions. It is less amenable to ambiguities due to compound or imperfect constitution of the repeats because the sequence reaction is primed at the unique adaptor structure. Only microsatellites composed of palindromic repeats are theoretically not suitable for this approach, since this will result in two different biotinylated products in the PCR and a mixture of two templates in sequencing. In contrast, in methods in which the sequence primer anneals within the microsatellite, imperfections in the repeat may give priming at different sites. The present method is well suited to analysis of large numbers of clones, since restriction enzyme mix, adaptors, and primers for PCR and sequencing of the flanking microsatellites flanks may be routinely used for any clone containing a microsatellite with a given basic unit. The entire procedure can be completed within a day. The number of sequence reactions to obtain a flanking region is three times less than in the method of Yuille *et al.* We think that the most efficient use is in the sequence determination of one flank for which a flanking PCR primer towards the repeat is then developed. This primer permits a sequence reaction through the repeat to elucidate the structure of the other flank. The entire microsatellite and its two flanks can be analyzed in two sequence reactions instead of twelve as in outward sequencing. The present method gives the flanking sequence directly adjacent to the repeat, permitting the design of PCR primers very close to the microsatellite, and it is relatively insensitive to irregular or complex sequences within the microsatellite.

REFERENCES

1. Yuille,M.A.R., Goudie,D.R., Affara,N.A. and Ferguson- Smith,M.A. (1991) *Nucleic Acids Res.* **19**, 1950.
2. Ostrander,E.A., Jong,P.M., Rine,J. and Duyk,G. (1992) *Proc. Natl Acad. Sci. USA* **89**, 3419–3423.
3. Karagyozov,L., Kalcheva,I.D. and Chapman,V.M. (1993) *Nucleic Acids Res.* **21**, 3911–3912.
4. Kandpal,R.P, Kandpal,G. and Weismann,S.M. (1994) *Proc. Natl Acad. Sci. USA* **91**, 88–92.
5. Mariat,D. and Vergnaud,G. (1992) *Genomics* **12**, 454–458.
6. Rothuizen,J. and Wolfswinkel,J. (1994) *Theoret. Appl. Genet.*, in press.