

Construction of a Microsatellite-enriched Library from Hevea Brasiliensis

SAFIAH ATAN^{*}, F.C. LOW^{*#} AND N.M. SALEH^{**}

A microsatellite library was constructed from Hevea brasiliensis by cloning DNA fragments of between 200 bp – 800 bp in length. These were ligated to pBluescript KS + phagemid as the vector and transformed into Escherichia coli DH5 α . The library appeared to be highly enriched with simple sequence repeats; (GACA)_n – 10%, (GATA)_n – 9%, (GA)_n – 34% and (GC)_n – 9%.

The oldest and most commonly used DNA marker technique is restriction fragment length polymorphisms (RFLPs). RFLP is based on the ability of an endonuclease to recognise a specific DNA sequence (recognition site) and to cleave at this recognition site, thereby producing DNA fragments of various lengths. However, changes in DNA sequence, such as base addition, deletion or substitution, will result in alteration in these recognition sites, resulting in length variation after digestion of the DNA by that same endonuclease. In the RFLP technique, DNA polymorphisms can be detected by the use of an appropriate hybridisation probe consisting of a cloned DNA fragment. Positive hybridisation signals will be obtained when a DNA probe anneals to a DNA sequence because it is either wholly or partly homologous to that DNA sequence^{1, 2}. Genetic maps consisting of RFLP markers have been constructed for a

number of plants in order to assist in breeding programmes³⁻⁵. The usefulness of the map is enhanced when it is used in conjunction with other conventional markers, e.g., morphological and biochemical markers⁶. In recent years, several new classes of molecular markers have gained popularity. One of these is microsatellites or simple sequence repeats (SSRs). Microsatellites consist of stretches of short tandem repeat elements (1–5 bp) which are scattered throughout the genome⁷. As molecular markers, microsatellites have all the characteristics of being very useful because they behave according to Mendelian laws⁸, are able to distinguish between two or more individuals⁹ and are abundant throughout the genome of the organism studied, viz. humans, animals and plants¹⁰⁻¹².

In this study, we attempted to construct a microsatellite-enriched library for *Hevea*

^{*} Rubber Research Institute of Malaysia, P.O. Box 10150, 50908, Kuala Lumpur, Malaysia

^{**} Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[#] Corresponding author

brasiliensis so that it will serve as a source of informative probes for DNA fingerprinting and genetic mapping. At the same time, the nature and frequency of occurrence of microsatellites in the *Hevea* genome might be learned in the course of the construction of this library.

MATERIALS AND METHODS

Plant Materials

The microsatellite library was constructed according to a protocol adapted from Olander *et al.*¹³ Soft, young, light green *H. brasiliensis* clone GL1 leaves were used for DNA extraction. Harvested leaves were frozen immediately in liquid nitrogen and stored at -70°C until ready for use.

DNA Extraction and Digestion

Total genomic DNA was extracted by the method of Low *et al.*¹⁴ A sample of genomic DNA (300 ng) was digested sequentially with *EcoRV*, *HaeIII*, *HinfI*, *HpaII*, *MspI* and *TaqI* with appropriate buffers at 37°C for 16 h according to the manufacturers' instructions. Digestion was terminated by the addition of 1X loading buffer containing 0.006% (w/v) bromophenol blue, 0.6% (w/v) Ficoll 400, 16 mM EDTA. DNA fragments were separated through 1.2% agarose gel in 1X TAE and a 100-bp ladder was used as size markers. That portion of the agarose gel containing DNA which corresponded to 200–800 bp (as indicated by the size markers) was cut out with a sterile scalpel. The DNA was subsequently electroeluted from the gel according to Sambrook *et al.*¹⁵ The concentration of the resultant fragments was estimated by

comparison with known concentrations of λ DNA in a gel.

The ends of the DNA fragments were repaired by adding *Escherichia coli* DNA polymerase (Klenow fragment) in a reaction mix containing 200 μ l of digested DNA, 30 μ l of 10X nick translation buffer [0.5 M Tris-HCl, pH 7.5, 0.1 M magnesium sulphate, 1 mM dithiothreitol (DTT)], 500 μ g/ml bovine serum albumin (BSA) and 10 μ l of Klenow fragment (4 U/ μ l). The mixture was incubated for 10 min at 16°C. Distilled water and 24 μ l of 2'-deoxynucleoside 5'-triphosphates (dNTPs) containing 2.5 mM of each dNTP namely dATP, dCTP, dGTP and dTTP were added to a final volume of 300 μ l. The mixture was incubated at 16°C for a further 30 min. Purification of the DNA was performed by phenol:chloroform (1:1, v/v) extraction. DNA fragments were precipitated with two volumes of ice-cold ethanol in the presence of 0.3 M sodium acetate, pH 5.2, and stored at -20°C overnight. DNA precipitates were collected after centrifugation and washed twice with 70% ethanol. The DNA pellet was redissolved in sterile distilled water.

Preparation of pBluescript KS+

Phagemid vector pBluescript KS+ (Stratagene, USA) was used for cloning. pBluescript KS+ (10 μ g) was digested with 50 U of *SmaI* in the appropriate buffer. The cleaved phagemid was dephosphorylated with 0.5 U of shrimp alkaline phosphatase (United States Biochemical, USA) at 37°C for 1 h. The reaction was terminated by heating the reaction mixture at 65°C for 15 min. After purification with phenol:chloroform followed by

ethanol precipitation, the phagemid DNA was dissolved in a minimal volume of sterile distilled water.

Ligation

DNA fragments (for cloning) and pBlue-script KS+ at 50 and 10 ng/ μ l, respectively, were used at two weight ratios of vector DNA to insert DNA of (1:5) and (1:10). In order to prevent the formation of hairpin loops and to maintain DNA strand separation, these vector and insert DNA samples were initially incubated together at 55°C for 5 min and then plunged into ice. Ligation was then carried out at 15°C overnight. The 35 μ l of ligation mix contained 25 mM Tris-HCl, pH 7.4, 5.0 mM MgCl₂, 5.0 mM DTT, 0.25 mM spermidine, 1.0 mM ATP, 1.25 mM hexamine cobalt chloride, 10 μ g/ml BSA and 2 μ l of T4 DNA ligase (1 U/ μ l) (Boehringer Mannheim, Germany). The control reaction differed from the test reaction in that it contained all the reactants except insert DNA fragments.

Success of the ligation was confirmed by electrophoresis of an aliquot of the ligation reaction through agarose gel (1%) against control ligation reaction. A successful ligation would appear as a DNA smear with a higher molecular weight compared to the control reaction.

Transformation

Transformation was conducted according to the protocol provided by Stratagene (USA). *E. coli* DH5 α was used as the bacterial host. Competent cells (DH5 α) were prepared according to the method of Sambrook *et al.*¹⁶

Plating of Transformants

The transformed cells were pelleted and resuspended in 50 μ l of LB medium. They were then plated over selective LB-agar plates. The selective plates were prepared by spreading 40 μ l of X-gal (20 μ g/ μ l) and 4 μ l of IPTG (23.8 μ g/ μ l) over LB containing ampicillin (50 μ g/ml) on each plate. Plates containing transformed cells were incubated at 37°C for 16 h. White colonies (putative recombinant transformants) were picked with sterile toothpicks. These were inoculated in 3 ml of LB medium and grown overnight for subsequent amplification of DNA inserts.

Insert Amplification by PCR

Insert DNA was amplified by the polymerase chain reaction (PCR). Amplification reaction was carried out in 1X *Taq* DNA polymerase buffer (10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl₂, 50 mM KCl), 100 μ M dNTP, 200 nM SK/KS primers (Stratagene, USA), 1 U *Taq* DNA polymerase and 3 μ l bacterial culture solution, in a final volume of 25 μ l. The reaction was topped with one drop of mineral oil. DNA amplification was carried out through a programme of one cycle of 5 min at 95°C, 5 min at 48°C; 35 cycles of 90 sec at 72°C, 45 sec at 94°C, 45 sec at 48°C and a final 10 min extension step at 72°C. Amplification success was demonstrated by visualisation after agarose gel electrophoresis and staining in ethidium bromide. The amplified products were transferred to nylon filters by Southern blotting. The nylons were subsequently hybridised with oligonucleotide probes.

Generation of Oligonucleotide Probes

Synthetic concatenated oligonucleotides containing repeated sequences such as $(GACA)_n$, $(GATA)_n$, $(GA)_n$ and $(GC)_n$ of a few hundred bps in length were generated by amplification of their respective basic tandem repeat elements. These were $(GACA)_4$, $(GATA)_4$, $(GA)_8$ and $(GC)_8$. Success in amplification of these oligonucleotides was confirmed by visualisation of the PCR products after gel electrophoresis. Since the number of times these SSRs were repeated in each probe was not determined, these concatenated oligonucleotide probes were probably heterogeneous in length.

Screening of Positive Clones

Concatenated oligonucleotides were labelled with γ - P^{32} -ATP by a 5'-end labelling kit from United States Biochemical (USB, USA) and used as hybridisation probes. They were hybridised with filters containing amplified inserts as described above. Positive hybridisation signals were located and respective clones which harboured these inserts were identified.

RESULTS AND DISCUSSION

Library Construction

Digestion of genomic DNA by the six selected REs appeared to result in complete digestion, since an abundance of low molecular weight fragments were obtained.

Success of ligation was confirmed by agarose gel electrophoresis. Successful ligation resulted in an increase in molecular

weight of the vector. In contrast, the control, which was devoid of insert DNA, was unchanged in its molecular weight. Comparison of the resultant ligated vector suggested that the weight ratio of vector DNA to insert DNA of (1:5) was better than (1:10), since a higher proportion of high molecular weight ligated vector was obtained at that ratio (*Figure 1*).

A transformation efficiency of 5×10^6 transformants/ μ g vector DNA was obtained. Though this value appeared low in comparison to commercially available competent cells which are reported to transform at an efficiency of 1×10^8 transformants/ μ g vector DNA, the value obtained was nonetheless considered to be satisfactory, since the competent cells used in the above experiments were prepared in-house and has been stored for some time.

Screening of the Library

The success of cloning was confirmed by PCR amplification. Nearly 400 putative transformants were obtained, but only 281 (70%) were screened and 121 were found to harbour DNA inserts (*Figure 2A*). This indicated that the generated library contained approximately 43% of positive recombinants harbouring insert DNAs of 200 to 800 bp in length.

A few clones, e.g. clones #76 and #83, were shown to contain two bands after PCR-amplification (*Figure 2A*). Sequence homology between one of the primers used for amplification and the insert DNA might have resulted in amplification of two DNA fragments (bands); a major DNA fragment from amplification of

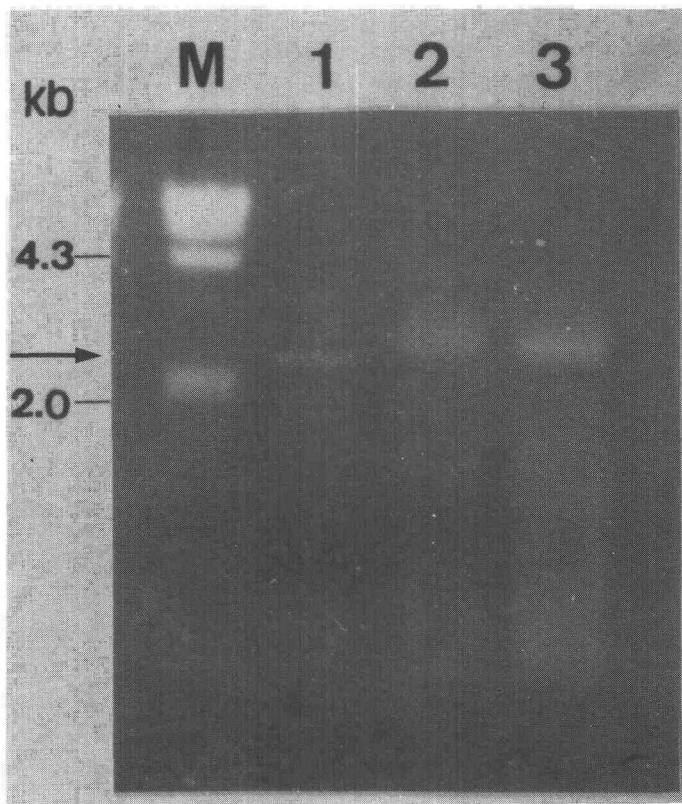


Figure 1. Analysis of success of ligation reaction.

M: Mol. wt. marker, λ /HindIII

1: Control ligation reaction

2: Ligation reaction at a weight ratio of (1:5) vector DNA to insert DNA

3: Ligation reaction at a weight ratio of (1:10) vector DNA to insert DNA

sequences flanking the cloning site and another band from between one of the flanking sequences and an internal region of the insert DNA. However, the disparity between the intensity of these two DNA bands and their relative sizes suggested that it was improbable. Since the higher molecular weight fragment was much brighter than the lower molecular weight fragment, the band of lower intensity was probably a product of contamination from a neighbouring clone, which appeared to

be of similar size as the insert DNA. Although the reason for the above is uncertain at present, analysis of nucleotide sequence of these inserts would definitely shed some light on the question.

Enrichment of Microsatellites in the Library

Hybridisation of these recombinant clones with various SSR sequences suggested that the library was enriched with microsatellite sequences (Figures 2B and 2C).

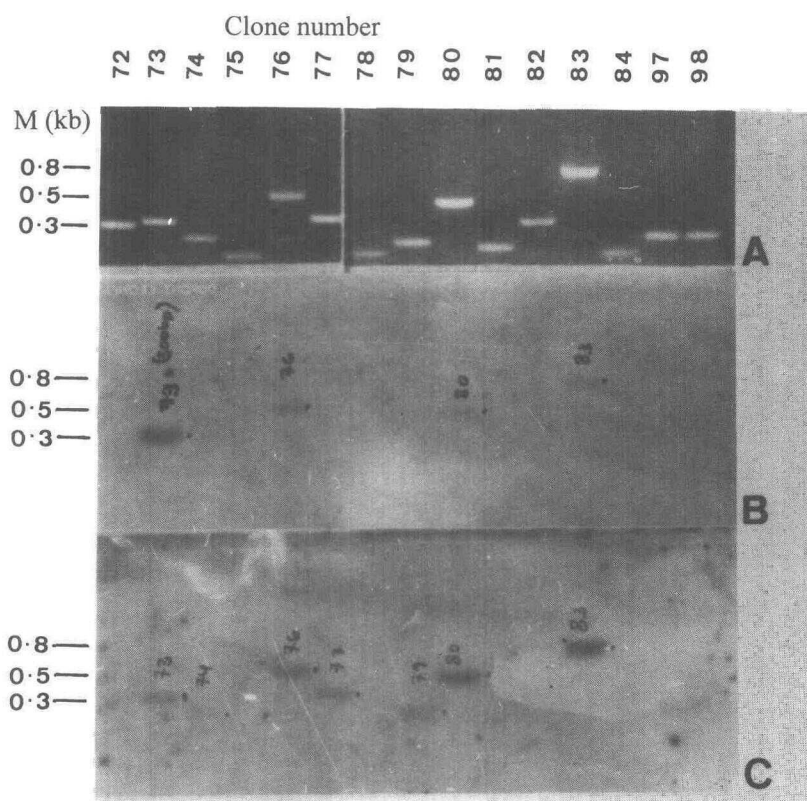


Figure 2. Screening of clones in microsatellite-enriched library.

A: Positive clones after PCR amplification with SK/KS primers

B: After hybridisation with concatenated oligonucleotide probe $(GACA)_n$ of heterogeneous lengths

C: After hybridisation with concatenated oligonucleotide probe $(GA)_n$ of heterogeneous lengths

M: Mol. wt. marker, 100 bp ladder

Four concatenated SSR sequences of heterogeneous lengths (a few hundred bp) were used as probes to test the enrichment of the library with micro-satellites. These were tetranucleotide repeats $(GACA)_n$, $(GATA)_n$ and dinucleotide repeats $(GA)_n$ and $(GC)_n$. The library appeared to be particularly enriched with $(GA)_n$ dinucleotide repeats (Table 1).

The high incidence (34%) of clones which were enriched with $(GA)_n$ dinucleo-

tide repeats in this *Hevea* microsatellite library may reflect the enrichment of this particular dinucleotide microsatellite in the *Hevea* genome. This high proportion of GA repeats in *Hevea* is not uncommon in trees. The enrichment appeared to be similar to that in *Pinus radiata*¹⁸ which was found to have a high content of GA and CA microsatellites. Similarly, Condit and Hubbell¹⁹ reported that AG repeats were 20–40% more abundant than AC repeats in all

TABLE 1 PERCENTAGE OF CLONES POSITIVE TO SSR PROBES OF HETEROGENEOUS LENGTHS

Concatenated SSR probes	% Positive clones
(GACA) _n	10.00
(GATA) _n	9.00
(GA) _n	34.00
(GC) _n	9.00

TABLE 2 PERCENTAGE OF CLONES POSITIVE TO MULTIPLE SSR PROBES OF HETEROGENEOUS LENGTHS

Multiple Concatenated SSR probes	% Positive clone
(GATA) _n , (GA) _n , (GC) _n	1.40
(GATA) _n , (GA) _n	6.40
(GACA) _n , (GA) _n	5.60
(GATA) _n , (GC) _n	0.70
(GA) _n , (GC) _n	6.40

six tropical forest plants which they had examined.

Two separate surveys on plant microsatellite sequences covering 34 and 28 species by Morgante and Olivieri²⁰ and Wang *et al.*²¹, respectively, revealed that (AT)_n was the most abundant, with (AG)_n as the next most abundant dinucleotide repeat sequence in plants²¹. The status of dinucleotide microsatellites in *Hevea* is unclear at present, since only two dinucleotide sequences out of six were used in this study. However, between the two microsatellite (GA)_n and (GC)_n which were examined, (GA)_n appear to predo-

minate. The abundance of microsatellite (GC)_n is probably inaccurate. When used as hybridisation probes, GC repeats will self-hybridise. This will reduce the availability of single-stranded sequences as hybridisation probes thus resulting in decreased hybridisation signals. Similarly, (AT)_n was not used in this study because of its tendency to self-hybridise. On the other hand, the absence of self-hybridisation with concatenated oligonucleotide GA, lends greater confidence in its hybridisation results when it is used as a probe.

Compound Microsatellites

Several recombinant clones in the library were found to hybridise with more than one SSR (Table 2). This would suggest that the stringency of washing of the nylons was not high enough to remove non-specific hybridisation, or intermolecular ligation of two or more DNA fragments had occurred before ligation with a dephosphorylated vector, or these clones contained more than one class of SSR otherwise known as compound microsatellites²². Of these three possibilities, stringency of washing is the least likely. Either or both of the remaining possibilities could have resulted in the observed putative compound microsatellites. Whether these clones were indeed true compound microsatellites would be evident if fewer such clones were obtained in a repeat experiment where ligation was carried out with dephosphorylated DNA fragments (after repair or filling-in of the cohesive ends) and *Sma*I-digested dephosphorylated vector. This repeat experiment would be carried out at a later date, to confirm the above. In the meantime, sequencing results from one of these clones

(#76) indicated that it was an imperfect compound repeat (results not shown). The occurrence of compound microsatellites has not only been documented^{18, 22, 23}, but also appeared to be proportionally higher in plant than in human genome²¹. Notwithstanding that some clones might be true compound microsatellites, a portion of these putative compound microsatellites could have arisen from intermolecular ligation of multiple DNA fragments preceding cloning. This would be confirmed later, after the repeat experiment described above had been carried out as well as when sequencing of all the putative compound microsatellite clones had been completed.

ACKNOWLEDGEMENTS

We are grateful for financial support from the Intensified Research Priority Area (IRPA) Programme of the Malaysian Government. The International Atomic Energy Agency (IAEA) is gratefully acknowledged for a grant awarded under its FAO/IAEA Co-ordinated Research Programme on the Use of Novel Techniques for Detection and Characterisation of Genetic Variation in Vegetatively Propagated Crop Plants.

Date of receipt: August 1996

Date of acceptance: June 1997

REFERENCES

1. BOTSTEIN, D., WHITE, R. L., SKOLNICK, M. AND DAVIES, R. W. (1980) Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphism. *Am J Hum Genet*, **32**, 314.
2. SOLLER, M. AND BECKMANN, J. S. (1985) Restriction Fragment Length Polymorphisms and Animal Genetic Improvements. *Reviews in Rural Sciences*, **6**, 25
3. HELENTJARIS, T. (1987) A Genetic Linkage Map for Maize Based on RFLPs. *Trends in Genetics*, **3**, 217
4. BERNATZKY, R. AND TANKSLEY, S. D. (1986) Majority of Random cDNA Clones Correspond to Single Loci in the Tomato Genome. *Mol Gen Genet*, **203**, 8
5. LANDRY, B. S., KESSELI, R. V., FARRARRA, B. AND MICHELMORE, R. W. (1987) A Genetic Map of Lettuce (*Lettuce sativa* L.) with Restriction Fragment length Polymorphism, Isozyme, Disease resistance and Morphological Markers. *Genetics*, **116**, 331
6. SHIN, J. S., CHAO, S., CORPUZ, L. AND BLAKE, T. (1990) A Partial Map of the Barley Genome Incorporating Restriction Fragment Length Polymorphism, Polymerase Chain Reaction, Isozyme, and Morphological Marker Loci. *Genome*, **33**, 803
7. HEARNE, C. M., GHOSH, S. AND TODD, J. A. (1992) Microsatellite for Linkage Analysis of Genetic Traits. *Trends in Genetics*, **8**, 288
8. WELSH, J., PETERSON, C. AND McCLELLAND, M. (1991) Polymorphisms Generated by Arbitrarily-primed PCR in the Mouse: Application to Strain Identification and Genetic Mapping. *Nucl Acids Res.*, **19**, 303
9. QUELLER, D. C., STRASSMAN, J. E. AND COLIN, R. H. (1993) Microsatellite and Kinship. *Tree*, **8**, 285
10. THOMAS, M. R., MATSUMOTO, S., CAIN, P. AND SCOTT, N. S. (1993) Repetitive DNA of Grapevine: Classes Present and Sequences Suitable for Cultivar Identification. *Theor Appl Genet.*, **86**, 173

11. JEFFREYS, A.J., WILSON, V. AND THEIN, S.L. (1985) Hypervariable "Minisatellite" Region in Human DNA. *Nature*, **314**, 67.
12. HOPKINS, B., O'CONNEL, F.M. AND HOPKINS, J. (1991) Use of DNA Fingerprinting in Paternity Analysis of Closely-related Exmoor Ponies. *Equine Vet J*, **23**, 277
13. OSTRANDER, E.A., JONG, P.M. RINE, J. AND DUYK, G. (1992) Construction of Small-insert Genomic DNA Libraries Highly Enriched for Microsatellite Repeat Sequences. *Proc. Nat Acad Sci USA*, **89**, 3419
14. LOW, F.C., SITI ARIJA MAD ARIF, CHOW, K.S., WAN RAHAMAN WAN YAACOB AND GALE, M.D. (1990) Restriction Fragment Length Polymorphism as Probes to Plant Diversity. *Conservation of Plants Genetics Resources Through In Vitro Methods*. ISBN 967-99915-2-0 **199**, 110
15. SAMBROOK, J., FRITSCH, E.F. AND MANIATIS, T. (1989) *Electroelution into Dialysis Bags Molecular Cloning A Laboratory Manual* 2nd. ed. Cold Spring Harbour Laboratory Press pp 6.28.
16. SAMBROOK, J., FRITSCH, E.F. AND MANIATIS, T. (1989) Preparation of Fresh or Frozen competent *E. coli*. *Molecular Cloning A Laboratory Manual*. 2nd ed. Cold Spring Harbour Laboratory Press pp 1.76
17. SAMBROOK, J., FRITSCH, E.F. AND MANIATIS, T. (1989) Lysis by Boiling. *Molecular Cloning A Laboratory Manual* 2nd ed. Cold Spring Harbour Laboratory Press pp 1.29
18. SMITH, D.N. AND DEVEY, M.E. (1994) Occurrence and Inheritance of Microsatellites in *Pinus radiata*. *Genome*, **37**, 977
19. CONDIT, R. AND HUBBELL, S. (1991) Abundance and DNA Sequence of Two-base Repeat Regions in Tropical Tree Genomes. *Genome*, **34**, 66.
20. MORGANTE, M. AND OLIVIERI, A.M. (1993) PCR-amplified Microsatellites as Markers in Plant Genetics. *Plant J*, **3**, 175.
21. WANG, Z., WEBER, J.L., ZHONG, G. AND TANKSLEY, S.D. (1994) Survey of Plant Short Tandem DNA Repeats. *Theor Appl Genet.*, **88**, 1.
22. WEBER, J.L. (1990) Informativeness of Human (dc-dA)_n (dG-dT)_n Polymorphisms. *Genomics*, **7**, 524.
23. DOW, B.D., ASHLEY, M.V. AND HOWE, H.F. (1995) Characterisation of Highly Variable (GA/CT)_n Microsatellites in the Bur Oak, *Quercus macrocarpa*. *Theor Appl Genet.* **91**, 137.