Randomly amplified polymorphic DNA (RAPDs) markers for the characterization of cocoa germplasm

J. Russell*, E. Johnson**, F. Hosein**, R. Waugh*, and W. Powell*

* Cell and Molecular Genetics Department, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland

** Cocoa Research Unit, The University of West Indies, St. Augustine, Trinidad, West Indies

The genetic improvement of any organism depends upon the existence, nature and the extent of the genetic variability available for manipulation. Thus, current plant breeding and future biotechnological approaches to crop improvement will not only require access to this variability but will also be dependent on the maintenance, conservation and management of germplasm resources. These factors are particularly relevant to tropical tree species such as cocoa. In this paper, we describe our work on the development and exploitation of molecular methods to monitor genetic variability in *Theobroma*. The rationale for the use of Randomly Amplified Polymorphic DNA (RAPD) is described, and the advantages and disadvantages of RAPDs relative to Restriction Fragment Length Polymorphism (RFLP) and morphological markers are discussed. Recent successful examples of our research to genetically fingerprint *Theobroma cacao* clones with RAPD markers will be reported together with the use of PCR based polymorphic assay procedures to quantify the extent of genetic differentiation between and within cocoa populations. Furthermore, the spatial distribution of the genetic variability detected based on RAPD data is shown to be related to the geographical origin of the cocoa accessions studied. Information of this nature will be of value in surveying gene pools and identifying areas of maximum diversity. We demonstrate that the results obtained from the use of RAPD markers will facilitate the sound scientific basis for the collection, conservation and utilization of cocoa genetic resources.

Introduction

The genetic improvement of any organism depends upon the existence, nature and extent of the genetic variability available for manipulation. Current and future plant breeding programmes (and indeed biotechnological approaches) will not only require access to this variability, but will also be dependent upon the conservation and management of biodiversity. This issue is particularly relevant to tropical tree species such as cocoa, (*Theobroma cacao* L.). In Central and South America, widespread deforestation has resulted in the designation of cocoa as a priority crop for conservation. However, the preservation of perennial tree crop germplasm, both *ex situ* and *in situ*, presents a range of problems. The recalcitrance of cocoa seeds together with their heterozygous nature means that cocoa germplasm has to be maintained in living tree collections. Such field gene banks are often difficult and expensive to maintain and emphasizes the importance of a thorough and systematic evaluation of the available genetic variability. In this review, we describe some of our recent research on developing molecular polymorphic assay procedures for cocoa.

Methods of detecting variability

Botstein *et al.* (1980) first suggested that restriction fragment length polymorphism (RFLPs) has the capacity to generate large numbers of genetic markers. Since that time, RFLPs have been used extensively in plants for germplasm characterization, phylogenetic studies and creation of linkage maps (Helentjaris and Burr, 1989; Powell, 1992). It has also been proposed that RFLPs may be used for the characterization of plant genetic resources (Helentjaris and Burr, 1989). However, the cost and labour involved in evaluating large number of samples with RFLPs is prohibitive. An alternative polymorphic assay procedure based on the polymerase chain reaction (PCR) has recently been developed (Williams *et al.*, 1990; Rafalski *et al.*, 1991; Waugh and Powell, 1992). The procedure is based on the use of single random 10 base oligonucleotide primers which are used to amplify random sequences from the target genome. This form of molecular marker has been termed randomly amplified polymorphic DNA (RAPD) and has the advantage of being technically simple, quick to perform, requires only small amounts of DNA and involves no radioactivity. RAPDs are therefore well suited for use in the evaluation and characterization of plant genetic resources. A similar approach but based on shorter (5 bp) oligonucleotide primers combined with the use of polyacrylamide gel electrophoresis and silver staining has also been used for DNA fingerprinting (Caltano-Annolles *et al.*, 1991a, b).

Application of RAPDs to cocoa

Fingerprinting and elimination of duplicates

RAPDs have been successfully used to fingerprint cocoa clones representing the three main cultivated sub-populations: Criollo, Forastero and Trinitario (Wilde *et al.*, 1992). The use of single primers allowed each of the clones evaluated to be uniquely characterized. An example of the molecular profile generated with one primer is given in Figure 1. The level of variation detected with RAPDs was greater than that observed for any other polymorphic assay procedure. In order to evaluate the usefulness of RAPDs further, a set of *Theobroma* clones were evaluated, which are difficult to distinguish using isozymes and morphological traits. Nine primers were used to evaluate the cocoa accessions listed in Table 1. Each primer generated between 5 and 13 amplification products (Table 2). Primers also differ in their capacity to differentiate between individual cocoa genotypes. A minimum of three primers: SC10-22, SC10-45 and SC10-44 are required to uniquely fingerprint each of the cocoa accessions analysed. Examples of the molecular profile generated by SC10-22 is shown in Figure 2. Based on these results it is clear that RAPDs can be used to evaluate and characterize cocoa genetic resources in field gene banks. Furthermore, RAPDs are technically simple and require only small amounts of DNA (100 µg) which can be extracted using mini-prep procedures from juvenile cocoa tissue.

Elimination of duplicates is an important objective for most gene banks. In this situation, the question relates to whether identical molecular profiles obtained from different trees represent a common genotype. If the DNA comes from a different tree then the probability of a correspondence is the probability of a random tree having the same RAPD profile as the tree being evaluated. Assuming that each PCR amplification product is independent, this probability is given by X^i, where X is the proportion of shared products and i is the number of amplification products generated in the tree being studied (Jeffreys *et al.*, 1985). It is therefore possible to calculate the probability of a chance match for each of the cocoa populations studied (Table 3). For both the IMC and LCTEEN populations the probability is very low indicating that two trees exhibiting identical molecular profiles when evaluated with the nine primers used in this study are likely to be duplicates. The chance match probabilities are much higher for the PA populations and this reflects the high degree of similarity exhibited by this group of genotypes. Although the chance match probability is less than 5%, it would be advisable to examine this group of genotypes with a wider range of primers.

Figure 1 Amplification of genomic DNA from 13 cocoa genotypes with primer SC-10-9.

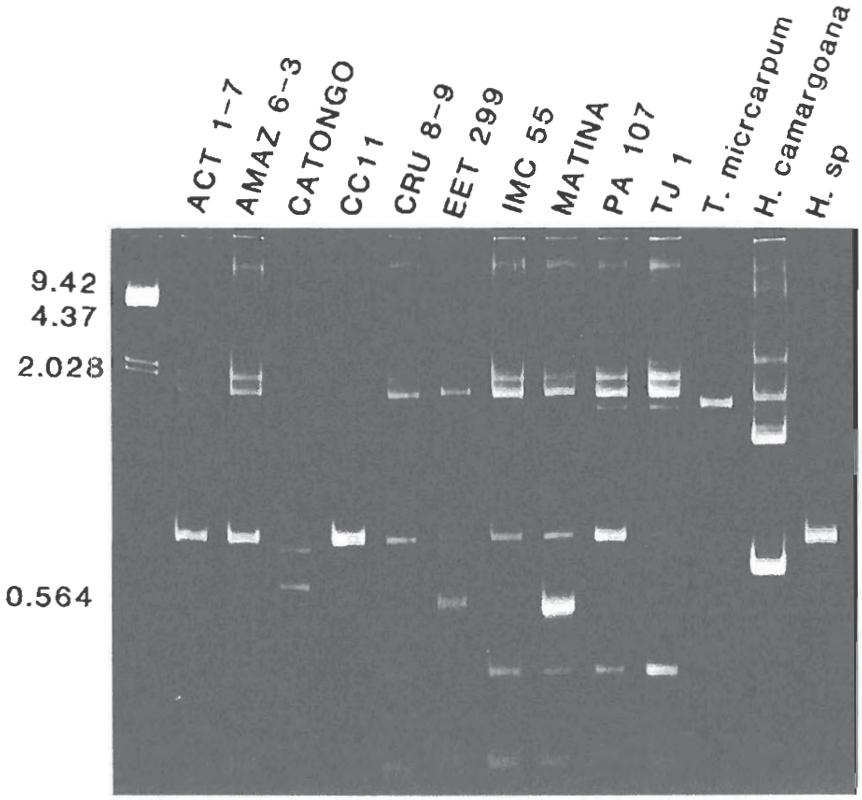


Figure 2 Amplification products generated with SC10-22

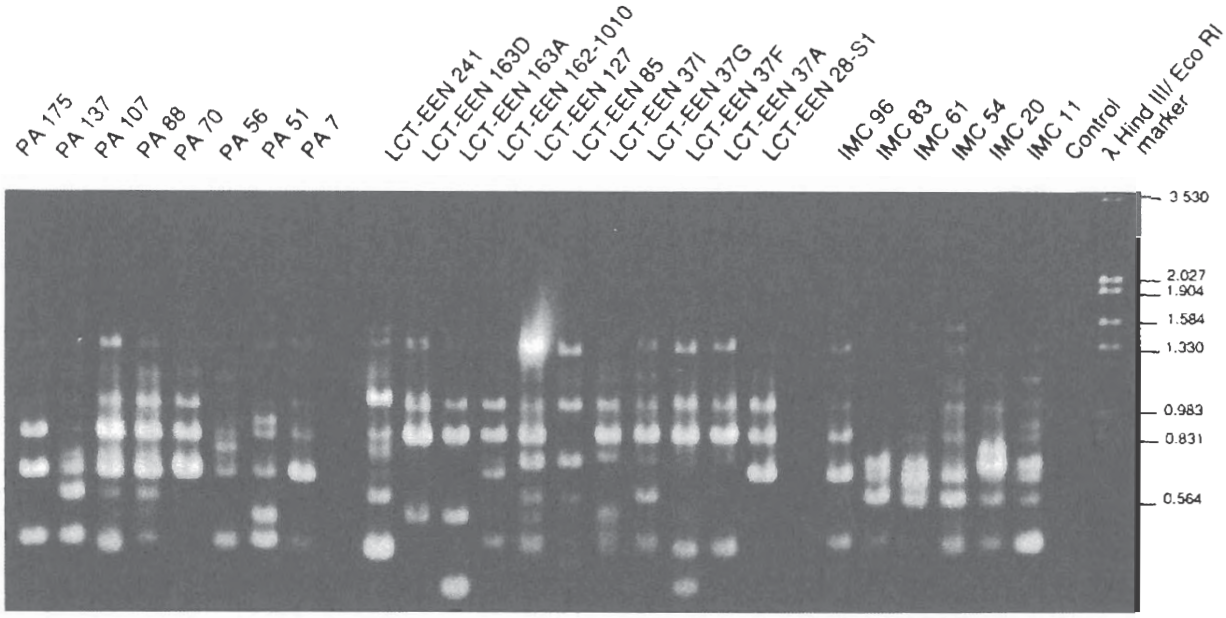


Table 1 The cocoa genotypes studied and information on their geographical origin

Accession Code	Country of Origin	Province	Locality
IMC 11	Peru	Loreto	River Nanay/Iquitos
IMC 20	Peru	Loreto	River Nanay/Iquitos
IMC 54	Peru	Loreto	River Nanay/Iquitos
IMC 61	Peru	Loreto	River Nanay/Iquitos
IMC 83	Peru	Loreto	River Nanay/Iquitos
IMC 96	Peru	Loreto	River Nanay/Iquitos
LCT-EEN 28-S1	Ecuador	Napo	River Napo
LCT-EEN 37A	Ecuador	Napo	River Napo
LCT-EEN 37F	Ecuador	Napo	River Napo
LCT-EEN 37G	Ecuador	Napo	River Napo
LCT-EEN 37I	Ecuador	Napo	River Napo
LCT-EEN 85	Ecuador	Zamora-Chinchipe	River Yacuambi
LCT-EEN 127	Ecuador	Napo	Coca-Lago Agrio
LCT-EEN 162-1010	Ecuador	Napo	Coca-Lago Agrio
LCT-EEN 163A	Ecuador	Napo	Coca-Lago Agrio
LCT-EEN 163D	Ecuador	Napo	Coca-Lago Agrio
LCT-EEN 241	Ecuador	Pastaza	River Villano
PA 7	Peru	Loreto	River Parinaria
PA 51	Peru	Loreto	River Parinaria
PA 56	Peru	Loreto	River Parinaria
PA 70	Peru	Loreto	River Parinaria
PA 88	Peru	Loreto	River Parinaria
PA 107	Peru	Loreto	River Parinaria
PA 137	Peru	Loreto	River Parinaria
PA 175	Peru	Loreto	River Parinaria

Table 2 The number of amplification products generated with nine primers for the IMC, LCTEEN and PA cocoa populations

Primer	No. of polymorphic loci				Number of phenotypes			
	No. of loci	IMC	LCTEEN	PA	IMC	LCTEEN	PA	Total
SC10-7	9	4	5	0	5	6	1	11
SC10-16	5	1	2	1	2	3	2	4
SC10-19	6	0	4	1	1	6	2	8
SC10-22	12	1	11	5	2	11	5	18
					5	10	3	15
SC10-25	13	3	12	2	2	7	4	12
SC10-37	8	1	5	2	3	3	4	8
SC10-44	8	2	2	2	3	3	1	6
SC10-45	6	3	2	0	2	6	3	11
SC10-49	8	1	5	4				
Totals	75	16	48	17				
% loci polymorphic		21.3%	64.0%	22.7%				

Table 3 Chance match probabilities (P) for each cocoa tree analyzed

	PA175	PA137	PA107	PA88	PA70
P	0.0496	0.0391	0.0334	0.0334	0.036
	IMC96	IMC83	IMC61	IMC54	IMC20
P	4.6×10^{-5}	2.1×10^{-5}	4.6×10^{-5}	2.7×10^{-5}	2.7×10^{-5}
	LCTEEN241	LCTEEN163D	LCTEEN163A	LCTEEN162-1010	LCTEEN127
P	5.8×10^{-4}	8.9×10^{-4}	1.9×10^{-4}	5.8×10^{-4}	4.6×10^{-4}
	LCTEEN27G	LCTEEN37F	LCTEEN37A	LCTEEN 28-51	
P	3.0×10^{-4}	3.7×10^{-4}	2.3×10^{-4}	1.1×10^{-3}	
	PA56	PA51			
P	0.0334	0.0391			
	IMC 11				
P	3.5×10^{-5}				
	LCTEEN85	LCTEEN371			
P	3.7×10^{-4}	4.6×10^{-4}			

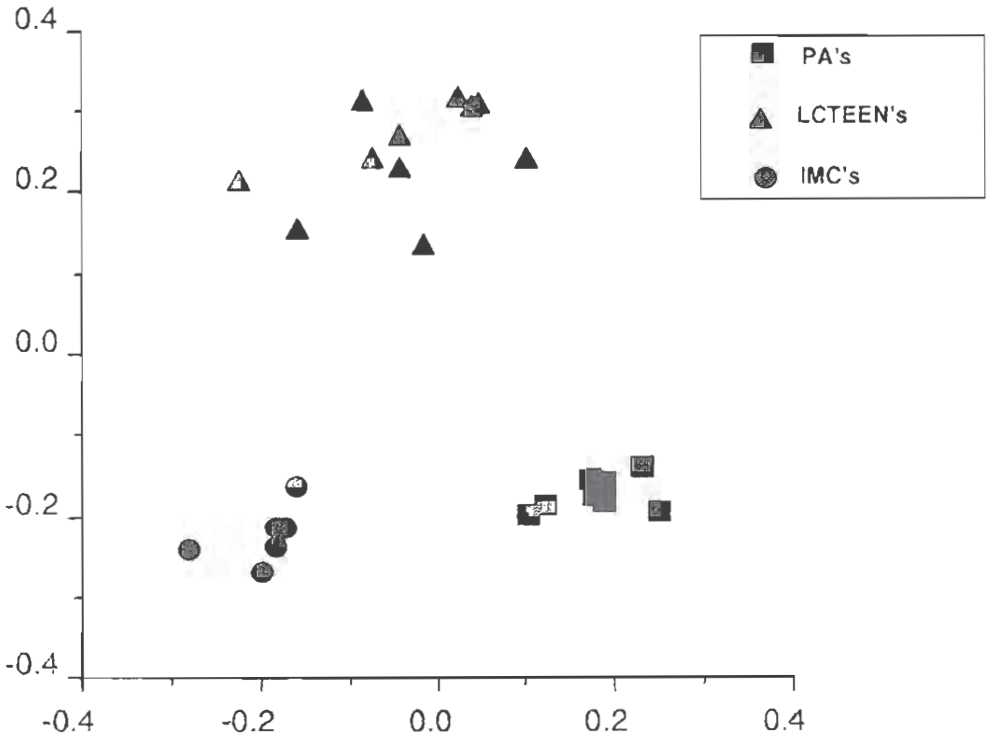
Table 4 Partitioning of genetic diversity revealed by nine primers into between and within population components using Shannon's diversity index

Primer	IMC	H_o PA	LCTEEN	H_{pop}	H_{asp}	H_{pop}/H_{asp}	$(H_{asp}-H_{pop})/H_{asp}$
SC10-7	2.253	0.000	2.051	1.435	2.140	0.670	0.330
SC10-16	0.651	0.811	1.323	0.928	1.391	0.667	0.333
SC10-19	0.000	0.544	2.300	0.948	2.209	0.429	0.571
SC10-22	0.651	2.750	3.461	2.287	3.449	0.663	0.337
SC10-25	2.251	1.406	3.097	2.251	3.508	0.642	0.358
SC10-37	0.651	1.549	2.091	1.430	2.851	0.502	0.498
SC10-44	1.253	1.962	1.540	1.585	2.798	0.566	0.434
SC10-45	2.253	0.000	1.241	1.165	2.272	0.513	0.487
SC10-49	0.651	1.549	2.051	1.417	2.322	0.610	0.390
Mean	1.179	1.175	2.128	1.494	2.549	0.585	0.415

Distribution of variability

Information obtained from the use of RAPDs extend beyond the generation of fingerprints. The molecular data obtained from RAPDs may be used to provide a measure of the distribution of genetic variability over populations and geographical areas. The distribution of variability between and within populations was calculated using the Shannon-Weaver information (Lewontin, 1972). Estimates of diversity was calculated as $H_0 = -\_ pi \log_2 pi$, where pi is the phenotypic frequency. Most diversity was detected in the LCTEEN populations (Table 4). However, two primers SC10-7 and SC10-45 detected most variation in the IMC rather than the LCTEEN population. H_{pop} provides a measure of the average diversity within populations. Primers SC10-22 and SC10-25 detect the most within population variability and SC10-16 the least amount of variability. An examination of the relative proportion of diversity present with-

Figure 3 Principal co-ordinated analysis of the 25 cocoa accessions examined.



in (H_{pop}/H_{ssp}) and between (H_{ssp}-H_{pop})/H_{ssp} populations indicates that on average a higher level of diversity is maintained within rather than between populations. This trend is not evident for primer SC10-19 where most of the variation is detected between populations. The LCTEEN populations (Allen, 1988) exhibited the greatest amount of variability and this reflects the sampling strategy which was designed to broaden the genetic variation available to the breeders. In contrast, the IMC and PA populations were selected for resistance to plant pathogens (Pound, 1938) and do not represent the range of natural diversity. Furthermore, analysis of the RAPD data using principal co-ordinate analysis (Fig. 3) identified population differentiation, which is related to the geographical origin of the cocoa accessions. Information of this nature based on RAPDs may therefore be of value in surveying gene pools and identifying areas of maximum diversity.

Future developments

Technology transfer

In principle, RAPD markers can provide a set of molecular descriptors, which can facilitate the evaluation and rationalization of cocoa gene banks. However, to utilize RAPDs effectively the methodology must be suitable for application in the countries which manage the collections. Recently, we have been involved in transferring RAPD technology from SCRI to the Cocoa Research Unit (CRU), Trinidad. This has involved training Cocoa Research Unit staff, purchasing and testing equipment and supplying detailed protocols and trouble-shooting guides. The successful establishment of RAPD technology at the CRU will provide a focus for large-scale characterization of cocoa germplasm.

Linkage maps and tagging genes

The detection of polymorphism in plants provides an opportunity to develop detailed linkage maps. Easily scored molecular markers provide a more direct and reliable method for identifying desirable genes via their physical association in genetic markers. Marker-based technologies can therefore improve the speed and precision with which new cocoa cultivars are generated. Already RFLPs have been used to create linkage maps in *Picea glauca* (Tulsieram *et al.*, 1992) and genetic linkage maps of cocoa could also be constructed. Targets for such mapping could include incompatibility loci and genes conditioning resistance to Blackpod disease (*Phytophthora* spp.) and Witches Broom disease (*Crinipellis perniciosa*).

The establishment of tight linkage between a molecular marker and gene of interest does, in theory, provide a route to cloning the gene. Map-based gene cloning has three main requirements: close linkage to the gene of interest, a library of large DNA fragments, usually obtained by using years artificial chromosomes (YACs) as vectors and the capacity to transform the plant of interest. Although this technology is not yet available in cocoa it is worth noting that the genome size of *Theobroma cacao* is relatively small. Figueira *et al.* (1992) used flow cytometry to estimate the size of the cocoa genome = 0.85 picograms, which is comparable to rice. Thus, the amount of repetitive DNA present in the cocoa genome is likely to be low.

Alternative polymorphic assay procedures based on PCR

Recently it has been demonstrated that a high level of allelic variation exists in the repeat number for simple sequence tandem repeats known as microsatellites (Weber and May, 1989; Tantz, 1989). The DNA sequences flanking the simple sequence repeats (SSR) are conserved allowing the selection of PCR primers that will amplify the intervening SSR. Although most of the information on SSRs has emerged from research on humans (Weber, 1990) and mice (Dietrich *et al.*, 1992), initial results indicate that SSRs do occur in plants. For example,

Condit and Hubbell (1991) demonstrated that $(AC)_n$ and $(AG)_n$ repeat sequences were abundant in five tropical tree species and *Zea mays*. Markers resulting from SSR polymorphisms are transmitted in a co-dominant manner and can therefore provide important tools for plant genetic resource characterization.

Conclusions

Molecular methods provide greater resolution than morphological or biochemical markers and allow the genetic structure of individual genotypes and populations to be better defined. RAPDs provide a quick and cost-effective method for characterizing large numbers of samples. They are therefore likely to play a pivotal role in fingerprinting clones, identifying duplicate accessions and assembling representative core collections. In this way, RAPDs can help in making germplasm collections more useful to breeders and ultimately assist in the development of improved cocoa cultivars.

References

Allen, J.B. (1988) Geographical variations and population biology in wild *Theobroma cacao*. Ph.D. Thesis, Edinburgh University.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.*, **32** 314-331

Caetano-Anolles, G., Bassam, B.J. and Gresshoff, P.M. (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technology*, **9** 533-557

Condit, R. and Hubbell, S.P. (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. *Genome*, **34** 66-71

Dietrich, W., Katz, H., Lincoln, S.E., Hee-Sup Shin, Friedman, J., Dracopoli, N. and Lander, E.S. (1992) A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics*, **131** 423-447

Figueira, A., Janick, J. and Goldsbrough, P. (1992) Genome size and DNA polymorphism in *Theobroma cacao*. *Amer. Soc. Hort. Sci.*, (in press)

Helentjaris, T. and Burr, B. (eds) (1989) Development and application of molecular markers to problems in plant genetics. *Curr. Commun. Mol. Biol.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Individual specific "fingerprints" of human DNA. *Nature, London*, **316** 76-79

Lewontin, R.C. (1972) The apportionment of human diversity. *Evol. Biol.*, **7** 381-398

Pound, F.J. (1938) Cacao and witchbroom disease of South America. Port-of-Spain, Trinidad

Powell, W. (1992) Plant Genomes, Gene Markers and Linkage Maps. In: Moss, J.P. (ed.) *Biotechnology and Crop Improvement in Asia*, 297-322. ICRISAT publication

Rafalski, J.A., Tingey, S.V. and Williams, J.G.K. (1991) RAPD markers a new technology for genetic mapping and plant breeding. *Agbiotech. News and Information*, **3**(4) 645-648

Tantz, D. (1989) Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucl. Acid. Res.*, **17** 6463-6471

Tulsieram, L.K., Glanbit, J.C., Kiss, G. and Carlson, J.E. (1992) Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. *Biotechnology*, **10** 686-690

Waugh, R. and Powell, W. (1992) Using RAPD markers for crop improvement. *Trends in Biotechnology*, 186-192

Weber, J.L. (1990) Informativeness of human $(dC\text{-}dA)_n(dG\text{-}dT)_n$ polymorphisms. *Genetics*, **7** 524-430

Weber, J.L. and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Amer. J. Hum. Genet.*, **44** 388-396

Wilde, J., Waugh, R. and Powell, W. (1992) Genetic fingerprinting of *Theobroma* clones using Randomly Amplified Polymorphic DNA markers. *Theor. Appl. Genet.*, **83** 871-877

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nuc. Acid. Res.*, **18**(22) 6521-6535