Theor Appl Genet (1994) 87:934–940

© Springer-Verlag 1994

Detection of genetic diversity and selective gene introgression in coffee using RAPD markers

C. Orozco-Castillo*, K. J. Chalmers, R. Waugh, W. Powell

Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

Recived: 15 June 1993 / Accepted: 23 July 1993

Abstract. RAPD (randomly amplified polymorphic DNA) markers generated by arbitary decamers have been successfully employed to detect genetic polymorphisms between coffee species and between *Coffea arabica* genotypes. The RAPD profiles were used to construct dendrograms and these were consistent with the known history and evolution of *Coffea arabica.* Material originating from Ethiopia and the arabica sub-groups – *C. arabica* var. *typica* and *C. arabica* var. *bourbon* – were clearly distinguished. RAPD analysis therefore reflects morphological differences between the sub-groups and the geographical origin of the coffee material. Species-specific amplification products were also identified, but, more importantly, amplification products specific to *C. canephora* were identified in two *C. arabica* genotypes, Rume Sudan and Catimor 5175. This diagnostic product is therefore indicative of interspecific gene flow in coffee and has biological implications for selective introgressive hybridisation in coffee. Our study demonstrates the power of the polymerase chain reaction technology for the generation of genetic markers for long-lived perennial tree and bush crops.

Key words: Coffee – Diversity – RAPDs – Gene introgression

Communicated by I. Mac Key

* On study leave from: Universidad de San Carlos de Guatemala, Facultad de Agronomia, Ciudad Universitaria, Zona 12, Apartado Postal No. 1545, Guatemala, Central America

Correspondence to: W. Powell

Introduction

Coffee is the world's most valuable agricultural exporting commodity. It contributes approximately £10 billion annually to the economies of more than 50 countries of Latin America, Africa and Asia. Commercial coffee production relies on two species: *Coffea arabica* and *C. canephora.* Higher quality coffee is associated with *C. arabica* and arabica coffee represents 73% of world production and almost all of the production in Latin America. *C. canephora* (robusta coffee) is mainly grown in central and western equatorial Africa and comprises 80% of African production. This is considered to be of low quality and is suitable only for the domestic market.

C. arabica is the only tetraploid ($2n = 4x = 44$) species in the genus *Coffea*, and is indigenous to the highlands of south-western Ethiopia and south-eastern Sudan. It is an inbreeder exhibiting disomic inheritance – and is considered to be a segmental allotetraploid (Charrier and Berthaud 1985). *C. canephora* is an obligate outbreeder with a gametophytic incompatibility system. Disease-resistance genes have been transferred into the *C. arabica* gene pool via natural and artifical interspecific hybridisation between the two species.

The geographical distribution of coffee has been strongly influenced by man. Arabica coffee was introduced to the American continent from the Yemen via Java and the Netherlands in the early 18th century. Historical evidence suggests that the base population descended from only a few trees. The genetic base of the American arabica coffee is therefore considered to be very narrow and represents only a small proportion of the potential genetic variability available within the coffee gene pool. The inbreeding nature of *C. arabica*,

together with the historical perspective on its limited genetic base, has prompted several collecting expeditions. A number of 'living tree' coffee germplasm collections (Kartha et al. 1981) have been established (Carvalho 1952) due to the recalcitrance of coffee seed and tissue culture technology used to establish in-vitro germplasm collections. In both cases there is a need to accurately assess the level of genetic variation in order to minimise duplication and establish core collections. Such core collections should be representative of the genetic variability available within the arabica gene pool and provide a source of germplasm which is accessible to coffee breeders.

Isozymes have been shown to be useful genetic markers for estimating genetic diversity and evaluating population differentiation. However, biochemical assays are limited by the number of polymorphic loci detected. Electrophoretic analysis of *C. arabica* accessions from Ethiopia and Kenya with six enzyme systems failed to reveal polymorphism. These results contrasted to the level of morphological variation detected in the same germplasm (Lowran 1978) and suggested that isozymes may be inappropriate for evaluating diversity in *C. arabica.* The development of molecular biology has resulted in alternative DNA-based procedures for the detection of polymorphism. The most widely used is restriction fragment length polymorphism (Tanksley et al. 1989) but this procedure is time-consuming, labour intensive and costly. Welsh and McClelland (1990) and Williams et al. (1990) reported a novel technique based on the amplification of random DNA sequences by the polymerase chain reaction (PCR) with arbitrary primers. This PCR-based assay has been used for cultivar identification (Hu and Quiros 1991), phylogenetic and pedigree studies (Heun and Helentjaris 1993), genetic mapping (Williams et al. 1990; Welsh et al. 1991), and the estimation of outcrossing rates (Fritsch and Reisberg 1992). The advantages of this technique include its simplicity, speed, and the requirement for only small amounts of relatively-crude genomic DNA (Rafalski et al. 1991; Waugh and Powell 1992). Single-primer DNA amplification is particularly relevant to perennial tree crops and has been used for clone identification in banana and cocoa (Kaemmer et al. 1992; Wilde et al. 1992), population differentiation (Chalmers et al. 1992; Russell et al. 1993), and genetic mapping (Carlson et al. 1991; Roy et al. 1992). Despite its economic and agricultural importance coffee has not benefited extensively from the technological developments which have been applied to other cash crops. In this manuscript we demonstrate that randomly amplified polymorphic DNA (RAPD) can be used for the genetic characterisation of coffee germplasm, the estimation of relatedness, and the detection of interspecific gene introgression.

Materials and methods

Plant material

The coffee genotypes studied are represented by 22 *C. arabica* accessions, one natural interspecific hybrid (Hybrid de Timor), three *C. canephora accessions* and one *C. liberica* accession (Table 1).

Total genomic DNA isolation

DNA was isolated from fresh or freeze-dried leaf material using a modification of the method described by Gawel and Jarret (1991). Leaf material (10 g fresh weight or 2 g freeze-dried) was ground to a fine powder in liquid nitrogen using a pestle and mortar with the addition of 400 mg of Polyclar AT. Extraction buffer [100 ml comprising 2% CTAB, 100 mM Tris HCl (pH 8.0), 1.4 M NaCl, 20 mM EDTA, 0.1% DTT] pre-heated to 65 °C, was added and the extract incubated at 65 °C for 30 min with occasional mixing. The samples were extracted with 75 ml of chloroform for 15 min and centrifuged at 5,000 rpm for 5 min. The aqueous supernatant was filtered through muslin cloth, mixed with an equal amount of ice-cold popan-2-ol and left at room temperature for 15 min to precipitate DNA. Following centrifugation at 10,000 rpm for 10 min, the supernatant was discarded and the DNA pellet drained by inverting the tubes. The DNA was resuspended in 2 ml of TE buffer [10 mM Tris HCl (pH 7.5), 1 mM EDTA] and incubated at 65 °C for 15 min after the addition of 20 μl of RNase A (10 mg/ml). Any insoluble material

Table 1. Coffee genotypes studied together with their country of origin

Species	Accession	Country of origin
Coffea arabica	1. N-39	Tanzania
	2. RS-510	Sudan
	3. Blue Mountain	Jamaica
	4. Typica	Guatemala
	5. Pache	Guatemala
	6. Mundo Novo	Brazil
	7. Caturra	Brazil
	8. Pacas	El Salvador
	9. Anacafe M-87	Guatemala
	10. Caturra	Brazil
Coffea arabica	11. ET6 A2 16695	Ethiopia
	12. ET25 A4 16712	Ethiopia
	13. ET41 A7 16725	Ethiopia
	14. ET11 CA7 16700	Ethiopia
	15. ET19 A3 16708	Ethiopia
	16. ET27 AF 16714	Ethiopia
	17. ET47 A4 16729	Ethiopia
Coffea canephora	18. Robusta 3751	Indonesia
	19. Robusta 3753	Indonesia
	20. Robusta 3580	Congo
Hybrids	21. Hybrido de Timor	Indonesia
	22. Catimor 8660	Brazil
	23. Catimor 5175	Portugal
	24. Catimor 11670	Columbia
	25. Catimor 12870	Brazil
	26. Catimor 8667	Brazil
Coffea liberica	27. Anon	Guatemala

was removed by centrifugation for 5 min at 14,000 rpm and the supernatant removed to clean tubes. The DNA was re-precipitated by the addition of 2 vol of 100% ethanol (ice-cold) and recovered by briefly centrifuging at 14,000 rpm. The DNA samples were vacuum dried, resuspended in 1 ml of TE buffer and stored at 4 °C.

DNA amplification (RAPD)

PCR reactions (50 μl) contained approximately 100 ng of genomic DNA, dATP, dCTP, dGTP and dTTP each at 100 μM final concentration, 200 nM of primer, 1 × *Taq* polymerase buffer and one unit of *Taq* XL polymerase (Northumbria Biologicals Ltd). Each reaction was overlaid with 100 μl of mineral oil of prevent evaporation. The random sequence 10-mer primers used in this study (Tabel 2) were synthesised on an Applied Biosystems 391 PCR-mate oligonucleotide synthesiser. Samples for enzymatic amplification were subjected to 45 repeats of the following thermal cycle: 1 min at 92 °C, 2 min at 35 °C and 2 min at 72 °C. Fragments generated by amplification were separated according to size on 2% agarose gels run in 1 × TBE [89 mM Tris HCl (pH 8.3), 89 mM boric acid, 5 mM EDTA], stained with ethidium bromide and visualised by illumination with ultraviolet light (312 nm).

Hybridisation analysis

Amplified DNA was transferred to Hybond N^+ (Amersham) nylon blotting membrane by the alkaline method of Reed and Mann (1985). Fragments to be used as probes were prepared using a 'Prep a gene' DNA purification kit. Isolated fragments were labelled by random priming (Feinberg and Vogelstein 1984) with ^{32}P-dCTP (3000 Ci mmol^{-1}; ICN Biomedicals), and used to probe the prepared blots with standard procedures (Maniatis et al. 1982). After hybridisation, the blots were washed in several changes of 0.1 × SSC, 0.1% SDS at 65 °C and the hybridising fragments revealed by autoradiography.

Table 2. Primers used for the detection of polymorphism in coffee

Primer	Sequence
SC10-04	5′ TACCGACACC
SC10-15	5′ GCTCGTCAAC
SC10-20	5′ ACTCGTAGCC
SC10-22	5′ CTAGGCGTCG
SC10-25	5′ CGGAGAGTAC
SC10-30	5′ CCGAAGCCCT
SC10-33	5′ TCGCCATAGC
SC10-35	5′ GTGCGGACAG
SC10-36	5′ TCACCGAACG
SC10-37	5′ GCCAATCCTG
SC10-38	5′ GACCCCGGCA
SC10-44	5′ CCAGGAAGCC
SC10-47	5′ ATAGCTCGCC
SC10-49	5′ CCACGAGCAT
SC10-50	5′ ACGCGCTGGT
SC10-55	5′ GGGAGACGTA
SC10-56	5′ CCAGCGTCTA
SC10-57	5′ GCTGGAAGCG
SC10-63	5′ CCTTGCGCTT
SC10-64	5′ CCAGGCGCAA
SC10-66	5′ AGTGGGCGCA
SC10-69	5′ GACGCTCTCC
SC10-70	5′ TTGGCCGCGA
SC10-71	5′ CTGGCGTAGT
SC10-73	5′ TCGGCCCTCG
SC10-74	5′ CGGACTTGGG
SC10-75	5′ ACCCAGCCAC
SC10-77	5′ AGATAGCGGG
SC10-78	5′ TCGGAGCGGT
SC10-84	5′ TGTGGGCATG

Data analysis

Estimates of similarity are based on the number of shared amplification products (Nei and Li 1979). Principal coordinate analysis and single linkage cluster analysis (Kempton and McNicol 1990) were performed with the Genstat 5 Statistical package.

Results and discussion

Initially the level of polymorphism detected with RAPD markers was assayed in five *C. arabica* accessions: N-39, Blue Mountain, RS 510, Catimor, and Hybrido de Timor. Of the 30 RAPD primers used, 25 detected polymorphism with an average of three polymorphic loci per primer. An example of the polymorphism detected with primer SC10-33 is shown in Fig. 1a. In order to extend the analysis a further 22 coffee accessions were evaluated (Table 1). An example of the level of polymorphism detected with primer SC10-15 is shown in Fig. 1b.

The RAPD amplification products generated can be classified into two types: constant (monomorphic) and variable (polymorphic). These differences can be used to examine and establish systematic relationships (Hadrys et al. 1992).

Considering only the variable products, the relationship between species and accessions within species was examined by single linkage cluster analysis (Fig. 2). There is a clear separation of the two diploid species *C. liberica* and *C. canephora* from the arabica genotypes. The arabica genotypes form three distinct groups: Ethiopian-derived germplasm, bourbon and typica types. Historically coffee from Yemen gave rise to two distinct types: *C. arabica* var. *typica* and *C. arabica* var. *bourbon* which was introduced to South America through the island of La Réunion. Morphological differences exist between the two groups, with the bourbon type having a more compact, upright growth habit and being generally higher yielding. RAPD analysis therefore reflects morphological differences between arabica sub-groups. In order to assess whether the clustering of populations based on RAPDs could be further resolved, principle component analysis was used to examine the shared fragment data available for the 27 accessions. In Fig. 3, the first two principle components account for 47% of the total variation observed and reveal a clear separation of the three

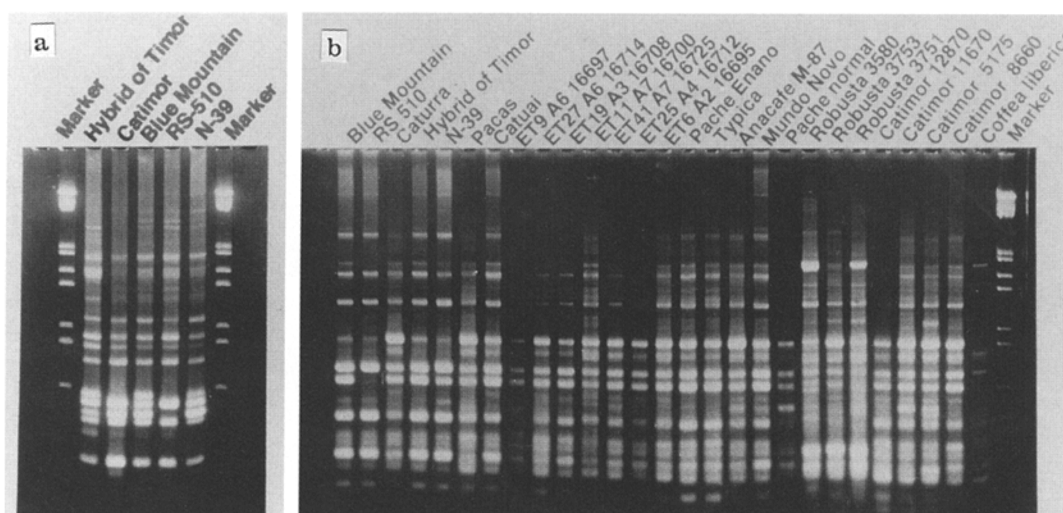

Fig. 1. **a** RAPD profiles of five *Coffea arabica* genotypes with primer SC10-33. **b** RAPD profiles of the 27 *Coffea* genotypes listed in Table 1 after amplification with primer SC10-15

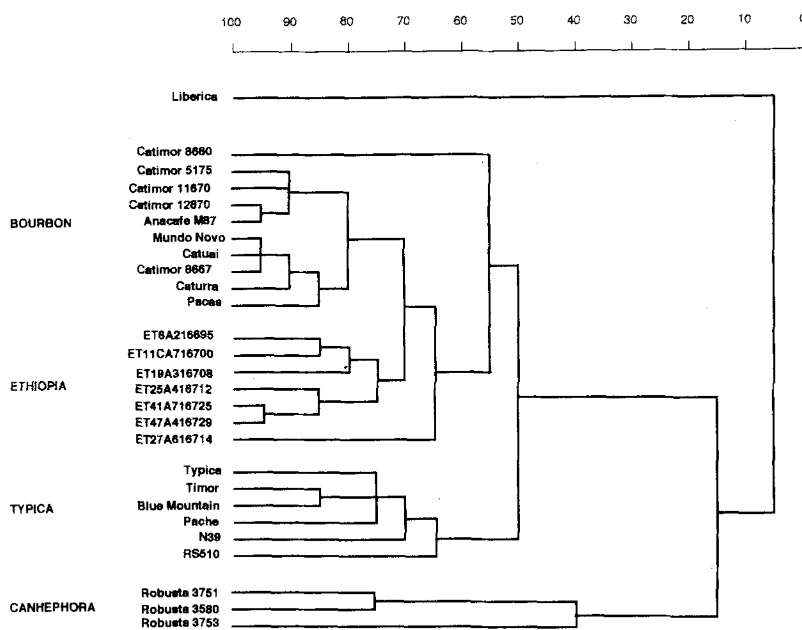

Fig. 2. Dendrogram of the *Coffea* accessions listed in Table 1 based on single linkage cluster analysis. Only 41 the informative polymorphic amplification products were used in the analysis.

species. Furthermore, the arabica accessions are again separated into three distinct groups representing the bourbon and typica types, and Ethiopian-derived germplasm.

In order to facilitate the presentation of both constant and variable data we have used an alternative graphical method termed 'bandmap' (Powell et al. 1991). The 'bandmap' (Fig. 4) examines the relationship between genotypes and RAPD-derived amplification products. Hence, the presence of an amplification product is represented by a filled box and the ordering of genotypes is exactly that generated by the dendrogram of the genotypes based on single linkage cluster analysis (Digby and Kempton 1987). Thus genotypes which share a common amplification product are more likely to be placed close to each other. Amplification product re-ordering is based on relative frequencies. The advantages of this form of data presentation has been described previously (Powell et al. 1991; Wilde et al. 1992). The potential of RAPDs to identify diagnostic markers for strain identification in mice (Welsh et al. 1991) and cultivar characterisation in plants (Hu and Quiros 1991; Klein-Lankhorst et al. 1991) has also been demonstrated. Similar conclusions can be made

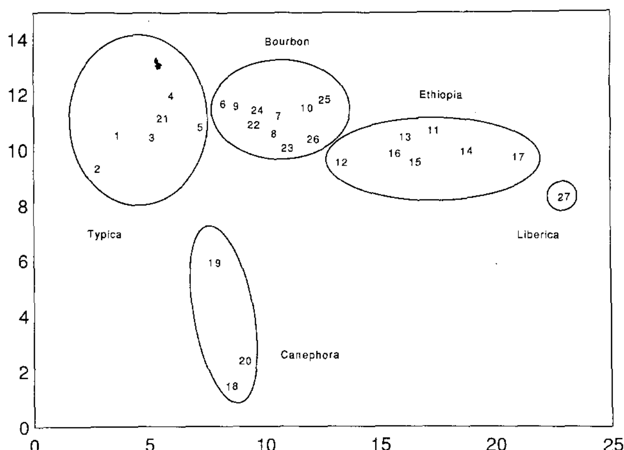

Fig. 3. Principle co-ordinate analysis of the 27 coffee accessions analysed using the data set derived from 41 polymorphic markers

Fig. 4. 'Bandmap' of shared polymorphic amplification products

from the data presented in Fig. 4. For example amplification products numbers 20 and 22 are present in six accessions all of which are representative of the typica group of arabica coffee. These products are therefore diagnostic for the *C. arabica* var. *typica* sub-group. The 'bandmap' provides a convenient method of genome

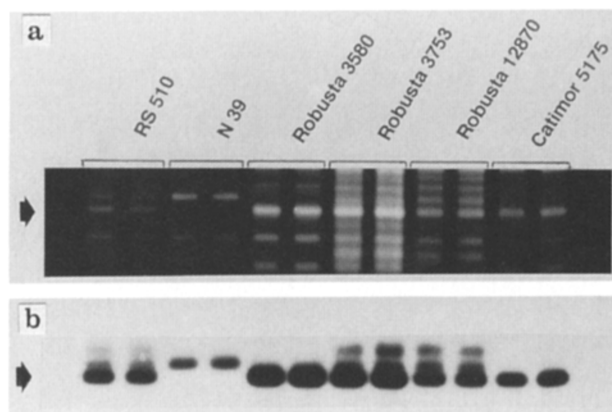

Fig. 5. **a** Amplification products from RS510, N39, Robusta 3753, 3751, 12870 and Catimor 5175 generated with primer SC10-30. The product diagnostic of interspecific gene flow is *arrowed*. **b** The arrowed band was used to probe a Southern-blot of amplified fragments generated by SC10-30. Hybridisation occurs to Canephora, Rume Sudan RS-510 and Catimor 5175 but not N39 (hybridisation to the upper band is the result of probe contamination)

scanning to locate amplification products which are taxonomically useful.

Of particular relevance to coffee is the detection of RAPD amplification products which are species-specific. Both natural and artificial interspecific hybridisation have been a feature of *C. arabica* evolution and improvement. A number of markers are specific to *C. canephora* but more importantly some RAPD primers detected loci which were present in the *C. canephora* accessions but also present in the *C. arabica* genotypes at low frequency. For example product 31 (Fig. 4) is present in the three canephora accession and in two *C. arabica* genotypes (Rume Sudan RS-510 and Catimor 5175). This product is detected with primer SC10-30 and the RAPD profiles are shown in Fig. 5a. In order to provide further molecular proof of the identity of the 0.2-kb product from *C. canephora* in Rume Sudan RS-510 and Catimor 5175 the amplification product arrowed in Fig. 5 was excised, labelled with ^{32}P-dCTP and used to probe Southern blots of the SC10-30 amplification products. Fig. 5b clearly demonstrates that the labelled product hybridises to canephora, Rume Sudan RS-510 and Catimor 5175 products but not to N39. This species-diagnostic product is therefore indicative of interspecific gene flow in coffee.

The *C. arabica* accession Rume Sudan RS-510 was identified in seed collected from wild coffee growing on the Boma Plateau, Rume Valley, south east Sudan (Thomas 1942). As opposed to material collected in the south west highlands of Ethiopia, Rume Sudan RS-510 is derived from truly wild populations of *C. arabica* which have not been subjected to human interference (Charrier and Berthaud 1985). Furthermore, this area

of Sudan is one of the few regions where *C. arabica*, *C. canephora* and *C. liberica* co-exist (Charrier and Berthaud 1985). Rume Sudan RS-510 is also one of the best sources of resistance to coffee berry disease, caused by the fungus *Colletotrichum coffeanum* (van der Vossen and Walyaro 1980). The Catimor accessions were obtained from segregating crosses between Caturra and Hybrid de Timor. Hybrid de Timor is the result of natural hybridisation between *C. arabica* and *C. canephora* (Rodrigues et al. 1975) and is an important donor of resistance genes to coffee berry disease and leaf rust (*Hemileia vastatrix*). The Catimor accessions were produced by backcrossing Caturra with Hybrido de Timor with selection for disease resistance at each stage of crossing (van der Vossen, 1985). Both Rume Sudan RS-510 and Catimor 5175 are characterised by having an 0.2-kb product derived from *C. canephora* and both arabica genotypes are known to possess genes conferring resistance to coffee berry disease.

The results demonstrate the ability of RAPD markers to reliably differentiate between *C. arabica* sub-groups (bourbon and typica types) and provide a molecular tool to examine the distribution of genetic diversity of *Coffea* sp. In addition, RAPDs have been used to detect, natural, interspecific introgression between diploid *C. canephora* and the *C. arabica* accession Rume Sudan RS-510. The same dignostic product also identifies artificial introgression in the case of Catimor 5175. It is therefore likely that this RAPD product provides a marker for selective introgressive hybridisation in coffee. These results, which are in general agreement with the previously-reported phylogenetic studies from a number of sources (Charrier and Berthaud 1985), demonstrate that RAPD markers have the potential to complement both conventional and biotechnological approaches to coffee improvement.

Acknowledgements. The research is supported by the European Community, International Scientific Cooperation Programme. W. Powell and R. Waugh are funded by the Scottish Office Agricultural and Fisheries Department.

References

Carlson JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffelde C, Ruthedge R (1991) Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor Appl Genet 83:194–200

Carvalho A (1952) Taxonima de *Coffea arabica* L. 6. Caracteres morfologicos dos haploides. Bragantia 12:201–212

Chalmers KJ, Sprent JI, Simons AJ, Waugh R, Powell W (1992) Patterns of genetic diversity in a tropical tree legume (*Gliricidia*) revealed by RAPD markers. Heredity 69:465–472

Charrier A, Berthaud J (1985) Botanical classification of coffee. In: Clifford MN, Wilson KC (eds) Coffee: botany, biochemistry and production of beans and beverage. Croom Held Ltd Publishers, Beckenham, UK, pp 13–47

Digby PGN, Kempton RA (1987) Multivariate analysis of ecological communities. Chapman and Hall, London

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity (Addendum). Anal Biochem 137:266–267

Fritsch P, Reisberg LH (1992) High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. Nature 359:633–636

Gawel NJ, Jarret RL (1991) A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol Biol Rep 9:262–266

Hadrys H, Balick M, Schierwater B (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Mol Ecol 1:35–63

Heun M, Helentjaris T (1993) Inheritance of RAPDs in F_1 hybrids of corn. Theor Appl Genet 85:961–968

Hu J, Quiros CF (1991) Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep 10: 505–554

Kaemmer D, Afza R, Weising K, Kahl G, Novak, FJ (1992) Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp.). Biotechnology 10:1030–1035

Kartha KK, Mroginski LA, Pahl K, Leung NL (1981) Germplasm preservation of coffee (*Coffea arabica* L.) by in-vitro culture of shoot apical meristems. Plant Sci Lett 22:301–307

Kempton R, McNicol J (1990) Graphical methods for multivariate data. Scottish Agricultural Statistics Service

Klein-Lankhorst R, Rietreld P, Machiels B, Veukeuk R, Weide R, Gebhardt C, Koorneef M, Zabel P (1991) RFLP markers linked to the root knot nematode resistance gene *Mi* in tomato. Theor Appl Genet 81:661–667

Louarn J (1978) Diversité comparée des descendances de *Coffea arabica* obtennes en autofé condation en fécondation libre an Tonkoui 75–78. Orstom, Man, Ivory Coast

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Nei M, Li W-H (1979) Mathematical model for studying genetical variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 74:5267–5273

Powell W, Phillips MS, McNicol JW, Waugh R (1991) The use of DNA markers to estimate the extent and nature of genetic variability in *Solanum tuberosum* cultivars. Ann Appl Biol 118:423–432

Rafalski JA, Tingey SV, Williams JGK (1991) RAPD markers – a new technology for genetic mapping and plant breeding. Agbiotech New Inf 3:645–648

Reed KC, Mann DA (1985) Rapid transfer DNA from agarose gels to nylon membranes. Nucleic Acids Res 13:7207–7221

Roy A, Frascaria N, Mackay J, Bousquent J (1992) Segregating random amplified polymorphic DNA (RAPDs) in *Betula alleghaniensis*. Theor Appl Genet 85:173–180

Russell JR, Hosein F, Johnson E, Waugh R, Powell W (1993) Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol Ecol 2:64–70

Rodrigues CJ Jr, Bettencourt AJ, Rijol L (1975) Races of the pathogen and resistance to coffee rust. Annu Rev Phytopathol 13:49–70

Tanksley SD, Young ND, Pat AH, Bonierbale MW (1989) RFLP mapping in plant breeding – new tools for an old science. Bio/Technology 7:257–264

Thomas AS (1942) The wild arabica coffee of the Boma Plateau Anglo-Egyptian Sudan. Empire J Exp Agric 10: 207–212

Vossen HAM van der (1985) Coffee selection and breeding. In: Clifford MN, Wilson KC (eds) Coffee: botany, biochemistry and production of beans and beverage. Crom Held, Beckenham, UK, pp 48–96

Vossen HAM van der, Walyaro DJ (1980). Breeding for resistance to coffee berry disease in *Coffea arabica* L. II. Inheritance of the resistance. Euphytica 29:777–791

Waugh R, Powell W (1992) Using RAPD markers for crop improvement. Trends Biotechnol 10:186–191

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Welsh J, Petersen C, McClelland M (1991) Polymorphisms generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping. Nucleic Acids Res 20:303–306

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

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531–6535