MOLECULAR DIVERSITY AND EVOLUTIONARY STUDIES IN THE GENUS *COFFEA*

CARLOS ALFONSO OROZCO CASTILLO

Ph D
UNIVERSITY OF EDINBURGH
NOVEMBER, 1995.

Coffea arabica var. Caturra

Coffea eugenioides

ABSTRACT

Genetic variation, phylogenetic relationships and evolution of *Coffea* species are important considerations in the evaluation and conservation of coffee genetic resources as well as in coffee breeding. PCR-based molecular markers have been used to assess genetic diversity, to detect gene introgression in *Coffea arabica* and *Coffea* species, and sequence analysis has been used to examine phylogenetic relationships within the genus *Coffea*. RAPD markers detected genetic polymorphism both between coffee species and between *C. arabica* genotypes. The RAPD data were consistent with the known history and distribution of C. *arabica*. Material originating from Ethiopia and the arabica sub-groups - *C. arabica* var. typica and *C. arabica* var. bourbon - could be clearly distinguished. RAPD analysis reflected morphological differences between the sub-groups and the geographical origin of the *C. arabica* material. Species-specific amplification products were identified. Amplification products which were present in *Coffea canephora* were also identified in two *C. arabica* genotypes. This product is possibly indicative of interspecific gene flow between these two species and could potentially have biological implications for selective introgressive hybridisation in coffee.

The phenetic relationships between *Coffea* accessions

representing 11 of the most important *Coffea* species employed in current breeding programmes were also examined using RAPD markers. Chloroplast and mitochondrial genome specific sequence tagged sites were used to complement the RAPD data. Estimates of variability based on the number of shared amplification products placed the species into three distinct groups which were consistent with derived chloroplast DNA phenotypes, the geographical origin of the species and previous studies based on morphological characteristics and RFLPs. *Coffea eugenioides* (2n=2x=22) exhibited the greatest similarity to the cultivated *C. arabica* (2n=4x=44) and it suggested that this species may represent its maternal progenitor. Sequence analysis of chloroplast DNA genomic regions also reflected the origin and geographical distribution of the coffee species. In addition, the data indicated that *C. eugenioides* is the maternal progenitor of the tetraploid *C. arabica*. Examination of sequences from a specific nuclear DNA PstI genomic clone revealed two distinct sequences (alleles) in *C. arabica*. The sequence of one of these alleles was identical to the sequence for *C. eugenioides*, and the other sequence was identical to a sequence for *C. canephora* and *Coffea brevipes*. While these results confirmed *C. eugenioides* as the maternal ancestor of *C. arabica*, they also suggested that *C. canephora* or *C. brevipes* were the paternal parent. The information generated by sequence analysis was consistent with data based on RAPD markers

and specific sequence tagged sites (STS). The data generated in this study were broadly consistent with classical taxonomical comparisons, but provide more specific and valuable information on the phylogenetic relationships of the genus *Coffea* and the evolution of *C. arabica*. The results presented here provide important information for the identification of suitable gene pools for inclusion in hybridization programmes to incorporate favourable genes into commercial *Coffea* species. In addition it provides a quantitative measure of the amount of genetic diversity present in coffee germplasm available to coffee breeders.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. R. Waugh, Dr. W. Powell and Dr. W. Spoor for their help, guidance and encouragement throughout the period of this study.

I would like to thank the staff of CMG for their technical support, when it was needed, particularly to Dr. Ken Chalmers for his technical advice in the first year of my studies.

I would like to thank Dr. Frank Wright for his help and statistical expertise in the generation of the phylogenetic trees presented in this study. I also would like to thank the staff of Data Processing, Visual Aids, and Library at SCRI for their cooperation in the different aspects involved in the writing and preparation of this thesis. Special thanks to Mr. P. Smith for help in the use of some computer programmes and proof reading of the thesis, and to Mr. G. Menzies for preparation of some of the figures included in this study. I am grateful to the staff of the Glass House for growth and care of the plants.

I wish to express my gratitude to the European Community, International Scientific Cooperation Programme, for the financial support of the research and to the Scottish Crop Research Institute for the use of all its excellent facilities over the duration of this research.

DEDICATION

I would like to dedicate this thesis to:

My wife Irma de Orozco
For her love, patience, continued support and encouragement during the time of my studies.

My sons Juan Carlos,
Pablo Ivan,
Ian Josue,
my daughter Alba Mariela
for their love and understanding of the limited time I had for them in the last period of my studies.

My mother
Alba T. Castillo de Orozco
For her love and encouragement to achieve this degree.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figures in Appendix

ABBREVIATIONS

Adh	Alcohol dehydrogenase
AMPPD	3-(2'-spiroadamantane)-4(3''phosphoryloxy)-phenyl-1,2-dioxetane
AS-PCR	Allele-specific PCR
AP-PCR	Arbitrarily primed PCR
APS	Ammonium persulphate
AT	Adenine, thymine
bp	Base pair
CAPS	Cleaved amplified polymorphic sequence
Ci	Curie
cM	centimorgan
cp	chloroplast
CATB	Cetyltrimethylammonium bromide
cv.	Cultivar
dATP	2'Deoxyadenosine 5'-triphosphate
dCTP	2'Deoxycytosine 5'-triphosphate
dGTP	2'Deoxyguanosine 5'-triphosphate
dUTP	2'Deoxyuracil 5'-triphosphate
dig-dUTP	Digoxigenin-2'Deoxyuracil 5'-triphosphate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic Acid
dNTP	2'Deoxyribonucleotide triphosphate
dTTP	2'Deoxythymidine 5'-triphosphate
$EDTA(Na)_2$	Ethylenediaminetetra-acetic acid, disodium salt

EtBr	Ethidium bromide
g	Relative Centrifuge Force (RCF)
GC	Guanine, cytosine
HCl	Hydrochloric acid
IAA	Isoamyl alcohol
kb	Kilobase
KCl	Potassium chloride
LB	1 % bacto-tryptone, 0.5 % yeast extract, 1% NaCl.
MDE	Mutation Detection Enhancement
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
mt	Mitochondria
IPTG	Isopropyl thiogalactoside
NaCl	Sodium chloride
$NaClO_4$	Sodium perchlorate
nc	nuclear
ng	nanograms
NaOH	Sodium hydroxide
NJ	Neighbor-joining
O.D.	Optical density
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PHYLIP	Phylogeny Inference Package
QT(L)	Quantitative trait (loci)
RAPD	Random Amplified Polymorphic DNA
rDNA	Ribosomal DNA

RCF	$(1.12x10^{-5})(rpm)^2r$ r = radius in cm measured from center of spindle to bottom of rotor bucket
Repel silane	0.2 % v/v dimethyldichlorosilane in 1,1,1, trichloroethane
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
RNase A	Ribonuclease A
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SCRI	Scottish Crop Research Institute
SDS	Sodium dodecyl sulphate
SSCP	Single strand conformational polymorphism
SSC	150 mM NaCl, 15mM tri-sodium citrate (pH 7.0)
SSPE	180 mM NaCl, 10 mM $NaPO_4$, 1 mM EDTA (pH 7.4)
Taq.	*Thermophilus aquaticus*
TBE	0.089M Tris, 0.089M boric acid, 0.002M EDTA
TE	10mM Tris HCl, 1mM EDTA
TEMED	N,N,N,N'-tetramethylethylenediamine
Tm	DNA melting temperature
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase

CHAPTER 1. INTRODUCTION

1.1 IMPORTANCE OF COFFEE PRODUCTION

Coffee is one of the most important crops in many Latin American, African and Asian countries, where it contributes approximately £ 10 billion annually to their economy (Anon., 1995a). The economic value of coffee in the tropical and subtropical countries is reflected in the high percentage of the population employed in its cultivation, harvesting, processing, shipping and marketing.

In 1994, the total world area of coffee planted was estimated to be 10,927,000 ha. In the crop-year 1993-94, world coffee bean production was 93,538,000 bags of 60 kg each, and the estimated production for the crop year 1994-95 is 94,306,062 bags (Anon., 1995b). Coffee represents one of the major export earning revenues in Latin American countries, e.g. Guatemala (38.9%), Colombia (58%), El Salvador (59%), Brazil (11%) and Costa Rica (26.4%). For most coffee producing countries in Africa, it also represents the main export earner; for example, Ethiopia (62%), Tanzania (35%), Madagascar (37%), Uganda (93%), Ivory Coast (37%), Burundi (87%) and Kenya (26%) (Anon., 1989). Brazil and Colombia dominate the world coffee market, accounting for 34 percent of world coffee exports, and 40 percent of total world production (Carvalho, 1985).

Even so, coffee exports are vitally important to many countries throughout Africa, Asia and Latin America. In many regions of the world, coffee is cultivated by small farmers, who depend on the income from the crop for subsistence. In Latin America, Colombia is a typical example, with 300,000 coffee farmers growing an average of 3.4 hectares. In Indonesia, 650,000 farmers grow coffee on plots averaging one hectare in size (Anon., 1989). Given its importance, particularly to producers in developing countries, it is vitally important that the latest advances in biotechnology be applied to the improvement of coffee production and hence to the economies of these countries.

Commercially, two coffee species are used; *C. arabica* and *C. canephora* Pierre (robusta coffee). *C. arabica* is cultivated at medium and high altitudes, and accounts for an estimated 70 % of the commercial world coffee and for practically all of that produced in Latin America (Carvalho, 1985). This species is also grown in some African countries such as Ethiopia, Tanzania and Kenya. Robusta coffees are mainly grown in the other African coffee-growing countries at low altitudes.

1.2 HISTORY OF COFFEE

C. arabica is considered to have originated in the highlands of south-western Ethiopia. From here it was taken

to Yemen, where it was grown commercially in the 17th century. Later, coffee was taken to the Dutch colony of Java. A few seeds were taken to the Botanical Garden of Amsterdam, and from here, finally introduced into America, via the Dutch colony of Surinam in the first decade of the 18th century (Berthaud and Charrier, 1988). These plants gave rise to a variety called Typica. In the 19th century other coffee plants were introduced into South America through the Island of La Réunion, formerly called Bourbon. Seeds from these plants were the basis of the Bourbon coffee variety (Carvalho et al., 1969). The history of dissemination of *C. arabica* is summarized in Figure 1.1 and illustrates the separate introductions of *C. arabica* into Latin America.

At the end of the 19th century there was an increasing interest in other coffee species. *C. canephora* was discovered in Tanzania in 1861 and cultivated later in the African Atlantic Coast. *C. liberica* and *C. stenophylla* were discovered in West Africa at the end of 17th century; and *C. congensis* was found in Central Africa. Even though cultivation attempts were made with several species, only three of them were commercially successful: *C. arabica*, *C. canephora*, and *C. liberica*. *C. arabica* is well adapted to the highlands whereas *C. canephora* and *C. liberica* grow well in lowland tropical areas (Van der Vossen, 1985). A tracheomycosis epidemic, caused by the fungus *Fusarium*

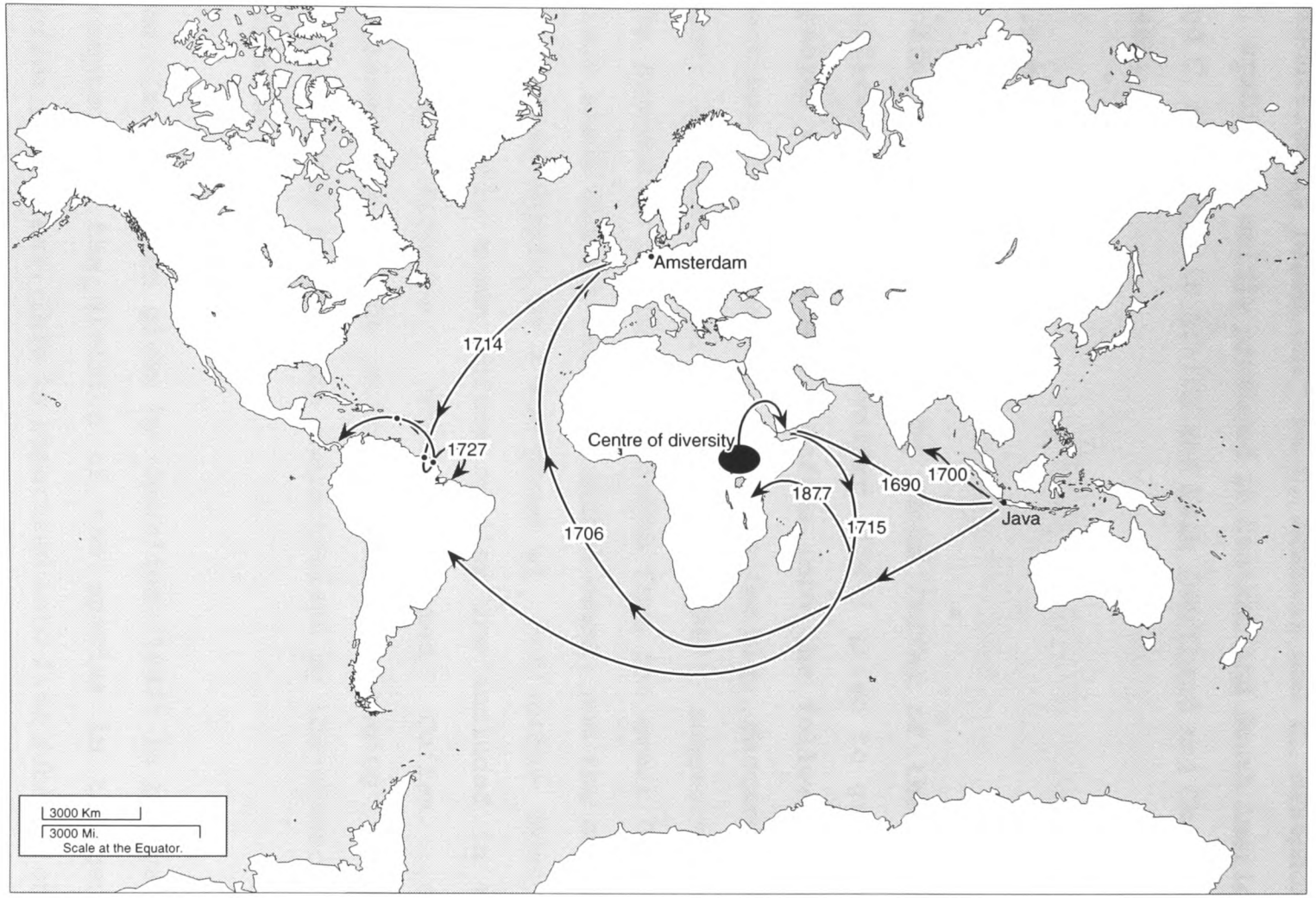


Figure 1.1. History of dissemination and early cultivation of *C. arabica*.
Taken from *Berthaud and Charrier*, 1988.

xylarioides, eliminated *C. liberica* species in the field in Africa and Asia between 1940 and 1950. Now the entire production is based only on C. arabica and *C. canephora*. *C. arabica* is mainly produced in Central and South America, and *C. canephora* in Africa and Asia (Berthaud and Charrier, 1988)

1.3 BOTANY

Coffee trees belong to the genus *Coffea* of the family *Rubiaceae*. Initially Chevalier (1947) tried to group the species within the genus *Coffea* into the following four sections: *Argocoffea* Pierre, *Paracoffea* Miq., *Mascarocoffea* Chev., and *Eucoffea* K. Schum. Leroy (1967), suggested that the *Argocoffea* should be excluded from the genus *Coffea*, since seeds did not resemble coffee beans; and the section *Paracoffea* should be a sub-genus of *Psilanthus*. Most of the presently known coffee species are included in the sections *Eucoffea*, actually named *Coffea*, and *Mascarocoffea*. The coffee species belonging to the *Mascarocoffea* section are characterised by the absence of caffeine.

The classification given by Chevalier (1947) is generally accepted for the division of the species in the genus *Coffea* (*Eucoffea*). This is separated into five subsections, *Nanocoffea*, *Pachycoffea*, *Erythrocoffea*, *Melanocoffea* and

Mozambicoffea based on the following criteria: tree height, leaf thickness, fruit colour, and geographical distribution. The species included in each subsection are shown in table 1.1.

Table 1.1. Grouping of Species in the Subsection *Eucoffea* according to Chevalier (1947)

Subsections	Species
Erythrocoffea	*C. canephora* *C. arabica* *C. congensis*
Pachycoffea	*C. abeokutae* *C. liberica* *C. klainii* *C. oyemensis* *C. dewevrei*
Melanocoffea	*C. stenophylla* *C. carissoi* *C. mayombensis*
Nanocoffea	*C. humilis* *C. brevipes* *C. togoensis*
Mozambicoffea	*C. schumanniana* *C. eugenioides* *C. kivuensis* *C. munifindiensis* *C. zanguebariae* *C. racemosa* *C. ligustroides* *C. salvatrix*

C. zanguebariae has been separated from *C. pseudozanguebariae* described by Bridson (1982), based on morphological characteristics, one of them being the absence or presence of fruit stalks. Bridson (1987)

described a new species under the complete name of *C. sessiliflora*, initially named *Coffea sp.* and equivalent to *C. zanguebariae* Form A. There is some divergence between the species *C. liberica* and *C. dewevrei*, however, the current opinion is that they are the same species, with two distinct forms; *C. liberica* var. *liberica* and *C. liberica* var. *dewevrei* (usually called Excelsa coffee) (Anthony et al., 1985).

1.4 REPRODUCTIVE SYSTEMS

C. arabica is an autogamous and selfcompatible species. Coffee flowers first appear when the young plant attains an age of two to three years. Buds are formed on the leaf axil of the plagiotropic branches or less frequently in the leaf axil of the orthotropic young branches. Two to five individual compound inflorescences usually develop in each leaf axil. The inflorescences have a short axis, two pairs of bracts and bracteoles and one to five flowers. The outcrossing frequency in *C. arabica* has a range from 7 to 15 % (Carvalho and Krug, 1949; Van der Vossen, 1974).

Diploid species in the genus *Coffea* are allogamous with a gametophytic system of self-incompatibility (Berthaud, 1980). Since diploid species are self-incompatible, there is a high level of heterozygosity in populations.

1.5 CYTOLOGY

The basic chromosome number of the genus *Coffea* is n=11 chromosomes, which is also characteristic of most of the genera of the family *Rubiaceae*. In the genus *Coffea* all species are diploid (2n=2x=22) with the exception of *C. arabica* which is tetraploid (2n=4x=44). *Mascarocoffea* species are diploid (2n=2x=22). Haploid plants (2n=22) have however been found in *C. arabica*, and are known as the variety Monosperma. In the same species, polyploids have also been found, such as triploids (3n=33) and pentaploids (5n=55) (Sybenga, 1960). Colchicine treatment has been used for the induction of autotetraploids in *C. canephora* (Berthou, 1975), and these have been used in coffee breeding programmes.

1.6 GENETIC DIVERSITY

There is considerable genetic variability in the genus *Coffea*, especially around the centre of diversity. Various methods have been used to describe this variation including morphological descriptors and numerical taxonomy (Chevalier, 1947; Leroy 1967, 1980); studies of the frequency distribution of incompatibility alleles within and between populations (Berthaud, 1980); genetic analysis of progenies from controlled crosses (Louarn, 1982); isozyme variation (Berthou et al., 1980); and recently

RFLP analysis (Cros et al., 1993). In spite of these studies, the extent and distribution of this apparent variability has not been totally and accurately described.

C. arabica has apparently high variability in the centre of diversity (Charrier and Berthaud, 1985)). Morphological characteristics seem to be clearly different among populations from different origins, and within plants of the same family. This was shown by Reyner et al. (1978) and Louarn (1978) using hierarchical variance analysis. In contrast, no differences in populations from Ethiopia and Kenya were found using electrophoretic analysis of isozymes (Berthou and Trouslot, 1977). *C. canephora* showed considerable variation, on the basis of morphological characteristics and floral biology; and this has also been verified by isozyme studies (Berthou and Trouslot, 1977). In *C. stenophylla*, the geographical isolation of the species seems to have produced considerable genetic diversity between populations, as intrapopulation crosses gave distinct hybrid vigour (Charrier and Berthaud, 1985).

1.7 GENETIC RESOURCES

Commercial coffee production is based on two species: *C. arabica* and *C. canephora*. However, there are other species in the genus *Coffea*, which together with the former constitute the gene pool which is useful for breeding

purposes. The collection, conservation and use of these genetic resources is therefore of vital importance, providing the widest possible gene pool for exploitation by coffee breeders.

The main strategy for collection and conservation of *Coffea* genetic resources is the establishment of core collections, which theoretically represent the genetic diversity of a crop species and its relatives, with a minimum of repetitiveness (Brown, 1989). Collection is directed towards the centre of diversity located mainly in the African continent. The current methods for genetic conservation are generally identified as *ex situ* and *in situ*. The first refers to conservation in an artificial habitat or in a habitat different from the original, such as seed banks, field collections and *in vitro* culture. *In situ* conservation refers to the maintenance of genetic resources in natural or original habitats, for example conservation of large tracts of land to protect the indigenous species (Chapman, 1989). The longevity of coffee seed seldom goes beyond one year and seed conservation is therefore inappropriate for long term storage (Kartha et al., 1981). Field collections allow medium-range conservation, of about 50 years, and the need to replace and maintain accessions is lower than that of seed conservation (Singh and Williams, 1984). On a world basis, collections are maintained in various coffee growing

countries. These are important for regional conservation. Collections of *C. arabica* with material from the Ethiopian centre of diversity are present at Jimma (Ethiopia), Turrialba (Costa Rica), Campinas (Brazil), Chinchina (Colombia), Lyamungu (Tanzania), Ruiru (Kenya), Foumbot (Cameroon) Man (Ivory Coast) and Ilaka-Est (Madagascar). There is a unique collection of species of the section *Mascarocoffea* at Kianjavato (Madagascar). The main African coffee species are kept in the living collections located at Divo and Man in Ivory Coast, but some replicates are also maintained in other countries, such as Turrialba (Costa Rica) and Lyamungu (Tanzania).

In vitro techniques provide another option for genetic conservation. This approach is being currently evaluated in some coffee research institutes such as ORSTOM, France (Engelmann et al., 1993) and CATIE, Costa Rica (Berthouly, 1991). The methods still have to be thoroughly tested, and factors such as the ability to propagate any genotype and procedures to limit growth rate remain problematic. In addition, in these very artificial conditions, long-term genetic stability needs to be maintained. *In situ* conservation has the advantage that a large number of coffee trees can be preserved in their original habitat.

Irrespective of the conservation approach, the evaluation of diversity and avoidance of duplicates is of fundamental

importance. Providing access to representative germplasm for breeders and reducing overall cost of maintenance are also important issues. Various approaches have been used to assess diversity in populations and collection. The latest and perhaps most feasible option is the use of molecular markers. One of their main advantages is that they are not affected by environmental factors. The techniques can also be applied regardless of the method of conservation.

1.8 EVOLUTION

The origin of coffee is considered to be in Africa, where wild species can still be found. Its distribution is mainly in the West, Central and East Africa. Different species differ in their distribution area (Charrier and Berthaud, 1985). Some, such as *C. humilis* and *C. stenophylla,* have a very limited distribution in West Africa. Others have a wider distribution, without covering extensive regions. For example, *C. arabica* is found only in south-eastern Ethiopia and on the Boma Plateau in Sudan (south-eastern). In contrast, *C. canephora* and *C. liberica* are widely distributed. *C. canephora* is indigenous to the African equatorial lowland forest zone from Guinea to Uganda. *C. liberica* is found in the south of Guinea, Ivory Coast, Liberia, Ghana, Gabon, Congo and in the North of Angola. The natural distribution of some coffee species is indicated in Figure 1.2

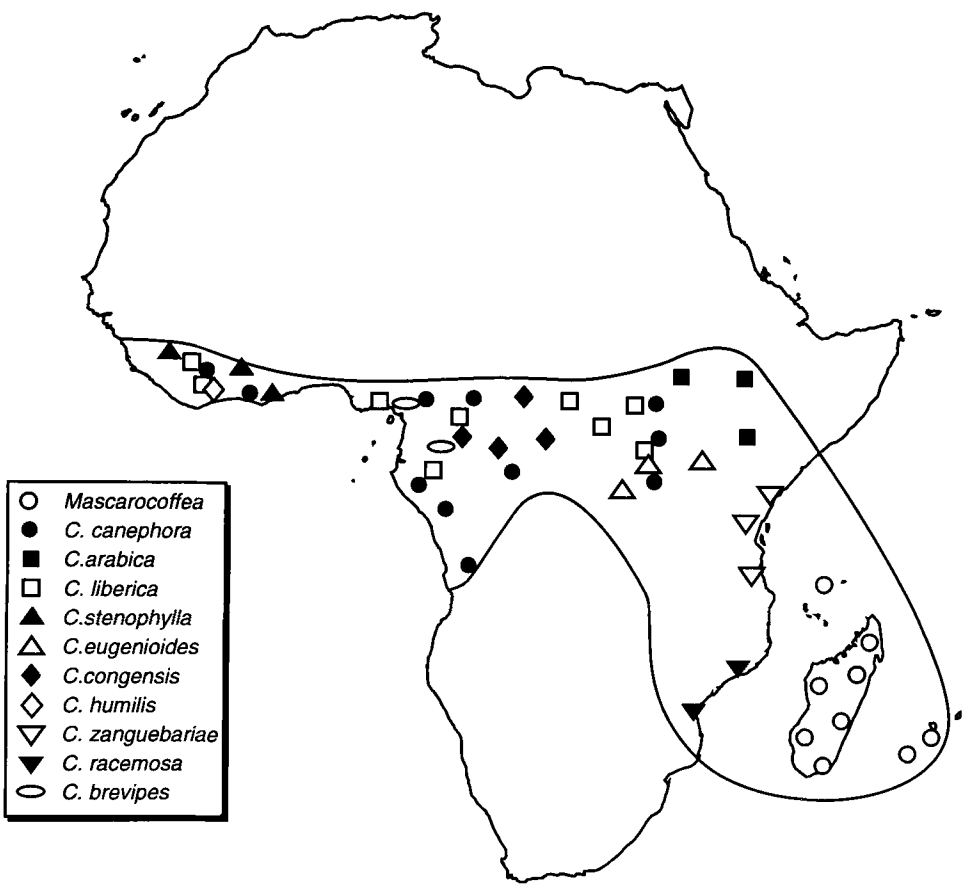


Figure 1.2. Natural distribution of *Coffea* species in Africa, and Madagascar (from Charrier A. and Berthaud J., 1985).

Attempts to determine the origin of coffee have mainly utilised crossability and chromosome homology among species (Carvalho and Monaco, 1968; Vishveshwara, 1963; Charrier, 1978,; Louarn, 1982). However, considerable variation is observed in the level of successful hybridization between species, with the genotype, crossing technique and environment affecting the rate of success. Thus, accurate quantitative information on the genetic relationships of the species is difficult to obtain. As expected, crosses between species within one taxa are less difficult (e.g. *C. canephora* and *C. congensis* in the subsection *Erythrocoffea* are relatively easy to hybridize). In more distant taxa, success has however eventually been obtained (e.g. *C. liberica* and *C. eugenioides*).

Interspecific hybridization conducted by Monaco and Carvalho (1964) showed that *C. arabica* was more closely related to *C. eugenioides* than to *C. canephora* or to *C. congensis*. Hybrids between *C. arabica* and *C. canephora* are highly sterile, not only due to genetic differences, but also to their triploid nature. Mendez (1958), found that triploid hybrids showed very low chromosome pairing, while hexaploids or tetraploids (*C. arabica* x 4n *C. canephora*) were more fertile. Cramer (1957), in his crossability studies, concluded that *C. canephora* and *C. congensis* seem to be closely related species, as a fairly high seed set is obtained in some combinations. Crossability studies by

Louarn (1993) indicated that *C. racemosa*, *C. sessiliflora* and *C. pseudozanguebariae* are interfertile but genetically isolated from the other species by a strong fertility barrier. Cros et al. (1993) found that these species native to east Africa have the lowest DNA content per nucleus (about 1 pg in comparison to 1.6 pg in other species), suggesting that one constitutive element to this fertility barrier could be related to the genomic size difference.

There is no chromosomal differentiation between diploid species of the genus *Coffea* (Bouharmont, 1963; Charrier, 1977; Charrier, 1978), which seems to indicate that these species have maintained the identity of their consensus origin during their evolutionary history, notwithstanding their geographical isolation. This suggests that the genetic diversification of these species would have resulted essentially from genetic differentiation. One important factor in this process would seem to be the geographical isolation of species, resulting in genetic variation through genetic drift and natural selection pressure (Berthaud and Guillaumet, 1978; Leroy, 1982).

C. arabica is the only tetraploid species in the genus *Coffea*, and is indigenous to the highlands of south-western Ethiopia and south-eastern Sudan. Both its diploid meiotic behaviour and the fact that its centre of genetic diversity is situated outside the area of distribution of the diploid

coffee species would suggest an allotetraploid origin (Carvalho, 1952). Grassias and Kammacher (1975), suggest that *C. arabica* should be considered a segmental allotetraploid, where chromosome pairing in meiosis is characterized usually by bi- and multivalent formation, and the parental chromosome sets combined in the diploid hybrid are partially homologous (homoeologous) and thus correspond in segments which permit "heterogenetic pairing" (Rieger et al., 1976).

The ancestors of *C. arabica* have often been assumed to be *C. eugenioides* and *C. canephora, C. liberica* or *C. congensis*, (Carvalho, 1952; Cramer, 1957; Narasimhaswamy, 1962). Monaco and Carvalho (1964) found that hybrids between *C. arabica* and *C. eugenioides* are usually sterile, but at the same time observed several loci which are dominant for the corresponding alleles of *C. arabica*. On the basis of the behaviour of these hybrids and the geographical distribution of *C. eugenioides*, Monaco and Carvalho (1964) suggested that this species might have participated in the origin of *C. arabica*. Narasimhaswamy (1962) suggested that *C. eugenioides* and *C. liberica* may be the ancestors of arabica coffee. On the other hand it was considered (Cramer, 1957) that *C. congensis* possesses many characteristics which indicate that it may also have contributed to the formation of the polyploid species. According to Louarn (1976), meiotic pairing of chromosomes

of the genomes of *C. eugenioides* and *C. canephora* showed greater bivalent pairing than in dihaploids plants of *C. arabica*, suggesting that the former species could be the ancestors of *C. arabica*. The triploids derived from crosses between *C. arabica* and certain diploid species show vigorous growth, but they are usually sterile. The number of bivalents plus trivalents formed during meiosis in interspecific crosses is in general close to 11, suggesting that one genome of *C. arabica* is related to that of the diploid species (Mendez and Biacchi, 1940; Berthaud, 1976; Berthaud, 1977). This has been suggested to indicate that in the genus *Coffea* all species share the same basic genome and have a monophyletic origin. Charrier and Berthaud (1985) suggest that *C. arabica* could have arisen from natural hybridization between two ancestral diploid coffee species followed by unreduced gamete formation. They also suggest that the degree of homology of the two genomes could have been high as a consequence of the monophyletic origin of the participating species.

Palynological observations indicated that *C. arabica* produces two types of pollen, one type related to *C. canephora* and one closely related to the pollen from *C. rhamnifolia*, a xerophytic species indigenous to the coastal regions of Somalia and Kenya (Lobreau-Callen and Leroy, 1980). The important point is that a consensus has not yet been reached as to the origin of *C. arabica*, and further

studies need to be done.

Phylogenetic relationships among *Coffea* species have also been addressed using biochemical and serological affinities. Isozyme studies (Berthou and Trouslot, 1977; Berthou et al., 1980) showed the following results: *C. canephora* and *C. congensis* share the same allozyme variants, but frequency differences observed between the species are larger than between populations of the same species; the genetic distance between *C. liberica* from Ivory Coast and *C. dewevrei* from the Central African Republic is of the same order as that between *C. liberica* and *C. humilis* from the Ivory Coast. This evidence has been interpreted as supporting the distinction between *C. liberica* and *C. dewevrei*, both of which belong to the subsection *Pachycoffea*, even though these species are actually considered to be the same (Anthony et al., 1985). In addition, there is considerable divergence between *C. canephora*, *C. liberica* and *C. eugenioides*; and the extent of enzymic affinities of *C. arabica* and different diploid coffee species are more or less similar. According to Berthou and Trouslot (1977) *C. arabica* could have arisen from *C. eugenioides* and *C. canephora* or *C. congensis* based on complementary electrophoretic bands of acid phosphatases and esterases.

RFLP analysis of chloroplast DNA conducted by Berthou et

al. (1980, 1983) suggests that *C. arabica* and *C. eugenioides* have a similar origin. The same pattern seems also to apply to *C. canephora* and *C. congensis*. With respect to mitochondrial DNA, there is great similarity between *C. arabica*, *C. eugenioides* and *C. congensis*, and considerable divergence between *C. canephora* and *C. arabica* or *C. eugenioides*. The mitochondrial analysis also showed a wide genetic divergence between *C. dewevrei* and *C. liberica* (Berthou et al., 1983)

Cros et al. (1993) also used analysis of chloroplast DNA to study genetic relationships among *Coffea* species. Even though the interspecific variation was very low, it was enough to provide phylogenetic information. Their results suggested a main separation in the group constituted by *C. canephora*, *C. congensis* and *C. brevipes* from the other species. Serological studies (Hofling and Oliveira, 1981) showed that *C. arabica* has more affinity with *C. congensis* and *C. eugenioides* than with *C. canephora*.

The most recent attempt to classify the relationships between coffee species was given by Berthaud and Charrier (1988). They separated the *Coffea* spp. in accordance with biogeographic criteria, supported by interspecific crossability and cytogenetic data. They defined the West and Central African group as *Erythrocoffea*, and the East Africa group as *Mozambicoffea*. Two morphological characters

that were used for the separation of the *Mozambicoffea* group, were the shorter ripening period and lower caffeine content in relation to the other species. The *Erythrocoffea* group defined by Berthaud and Charrier (1988) includes the subsections *Erythrocoffea* and *Nanocoffea* characterized by Chevalier (1947).

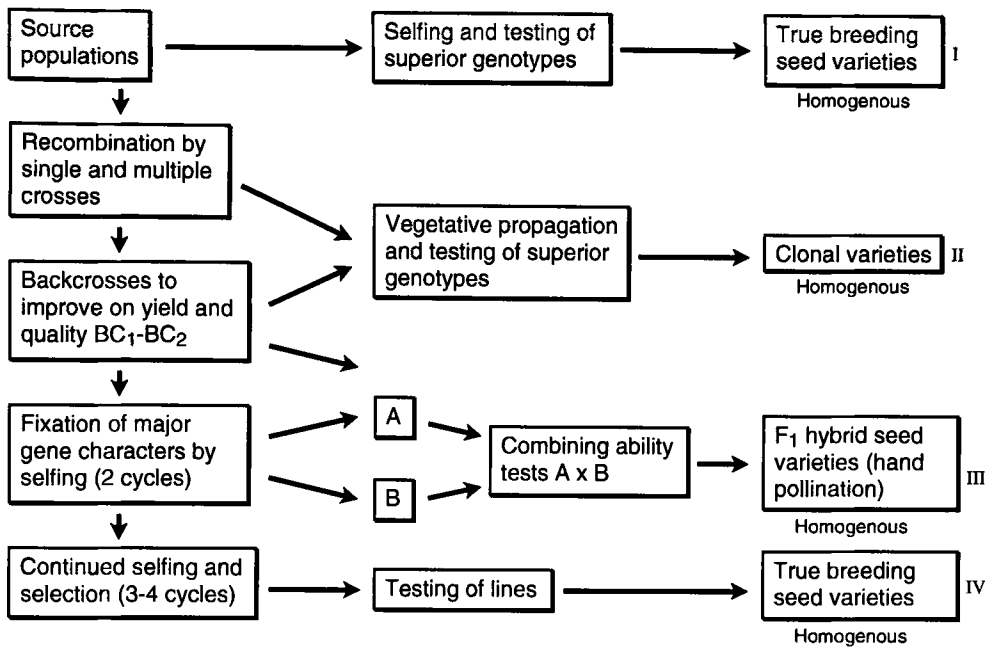
Thus, different studies have examined the genetic relationships between *Coffea* species and the evolution of *C. arabica*. However, further studies are still required, to elucidate clearly the genetic relationships among species and particularly the origin of the tetraploid *C. arabica*.

1.9. COFFEE BREEDING

Coffee breeding programmes are mainly conducted within *C. arabica* and *C. canephora*. However, crosses between these two species are also important, aiming to introgress vigour and disease resistance from *C. canephora* into *C. arabica*. In addition, other wild species have been used to incorporate useful characters into the gene pools of *C. arabica* and *C. canephora* (Carvalho, 1985). The main breeding objectives in coffee are as follows: increasing the genetic yield potential, yield stability, growth characters and yield components, improvement of quality, reduction in caffeine content and resistance to diseases and pests. Additional objectives include drought resistance

and suitability for mechanical harvesting. The yield stability refers to the achievement of high productivity under a wide range of environmental conditions and the ability to overcome biennial bearing (Van der Vossen, 1985). Caffeine content is an important selection criterion in *C. canephora* cultivars, in which the average caffeine content (2 to 3 %) is almost double that of *C. arabica* (1 to 1.3 %). In the latter, importance has been given to the generation of caffeine free varieties (Smith, 1988). In this respect, a low caffeine arabica cultivar named Laurina has been used as a donor (Mazzafera and Carvalho, 1992). Breeding for disease resistance focusses on two major diseases in *C. arabica*: *leaf rust (Hemileia vastatrix)*, present in all coffee producing countries, and *coffee berry disease (Colletotrichum coffeanum)*, present in East and Central Africa. Nematode resistance is an important breeding objective in several countries, especially in Central America, where *Meloidogyne* spp. and *Pratylenchus* spp. are frequently found in soil where coffee is grown. Different levels of tolerance/resistance have been found in *C. arabica* and *C. canephora* populations; however, this character has not been incorporated into commercial *C. arabica* cultivars. An outline of breeding schemes employed in *Coffea* improvement is shown in Table 1.3.

a.



b.

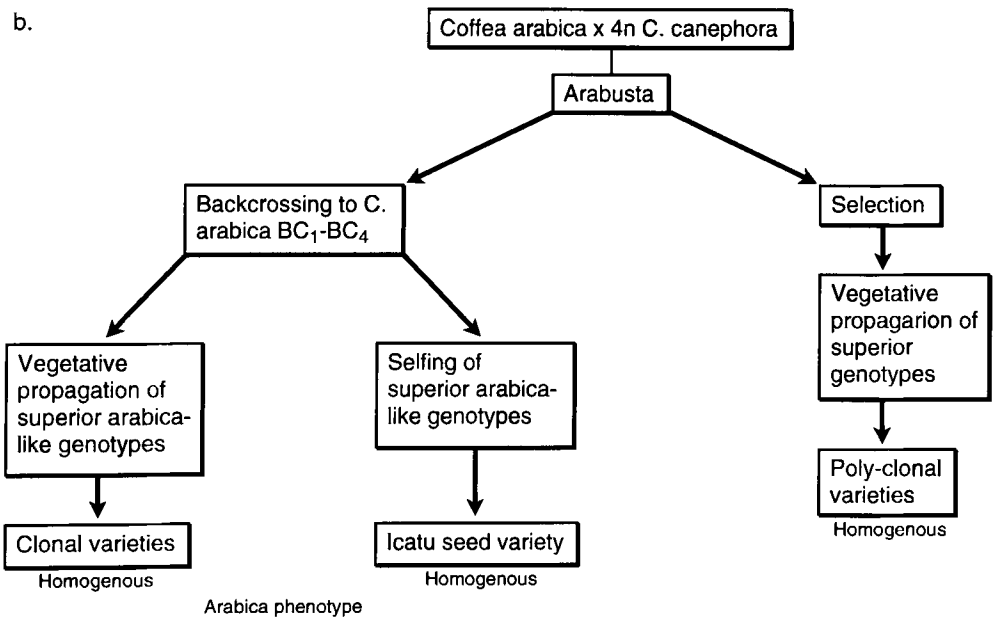


Figure 1.3. Outlines of breeding schemes:
a. The self pollinating arabica
b. Starting from arabica x robusta interspecific hybridization
Taken from Van der Vossen, H. A. M., 1985.

1.10 MOLECULAR MARKERS AND APPLICATIONS

Over the last 20 years a wide range of molecular marker technologies have been developed. These include: protein (isozyme), restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), and more recently a range of markers based on the polymerase chain reaction (PCR). Initially, isozymes were used successfully for certain aspects of plant breeding and genetics (Tanksley and Orton, 1983), but the number of isozyme polymorphisms available is generally limiting. To overcome this difficulty, markers such as RFLPs and RAPDs have been developed which allow analysis of genetic differences at the DNA level. Furthermore, in recent years, additional assays such as Microsatellites or Simple Sequence Repeat (SSRs), Single Strand Conformational Polymorphism (SSCP), Direct Sequencing of PCR products; and Amplified Fragment Length Polymorphism (AFLP) have emerged. These assays are discussed below.

1.10.1 ISOZYMES

Traditionally, plant genetic variation has been measured using morphological data, with the assumption that phenotypic differences have a genetical and environmental basis. Many of the characters that can be used for detection of genetic variation are determined by a number

of genes with alleles at each locus contributing to the phenotype. In addition, there is frequently a strong environmental influence on these phenotypic characters, such as plant height and vigour.

Isozyme markers were initially applied to help overcome the problem faced by measures of genetic differentiation based on morphological characters. Isozymes are functionally similar forms of enzymes that share a common substrate, their only distinction being their electrophoretic mobility or isoelectric point (Markert and Moller, 1959). Isozymes are generally visualized as coloured bands in a gel matrix and are indicative of localized areas of enzyme activity where catalysis of a specific chromogenic substrate takes place. Differences in electrophoretic mobility of enzymes are the result of changes in structural genes coding for polypeptides; therefore differences are the direct result of genetic variation (Crawford, 1983). Several kinds of gels are used to detect polymorphism, including starch, polyacrylamide, agarose and cellulose acetate gels (Wendel and Weeden, 1990). Isozyme assays are relatively simple and easy, however, their great limitation is the small number of systems available. Nevertheless, isozymes have been widely used in plant breeding and genetic studies (Tanksley and Orton, 1983) and in phylogenetic and population studies (Gotlieb, 1981; Hart, 1979; Rick and Tanksley, 1981). They have also been applied to characterize plant accessions,

such as maize (Cardy and Kannenberg, 1982); onion (Pefley and Orozco-Castillo, 1987); cocoa (Johnson et al., 1990), and sorghum (Morden et al., 1990); as well as to discriminate between cultivars in apple (Weeden and Lamb, 1985); and grape (Walters et al., 1989).

1.10.2 RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)

The principle of RFLPs, centres around digestion of DNA samples by restriction endonucleases, which cleave at specific sites in the nucleic acid sequences of the DNA (Botstein et al., 1980). Among a number of individuals, the DNA fragments generated will consist of both similar and different lengths. The proportion of different size fragments between two individuals is an indication of genetic diversity. Identical size restriction fragments from different genotypes are interpreted as representing genetic similarities, whereas different size fragments are interpreted as genetic differences. Visualization of the specific fragments is achieved by hybridizing labelled DNA probes to Southern blots of total genomic DNAs digested with restriction enzymes. The polymorphisms revealed by this approach are due to changes in the DNA such as deletions, insertions, base substitutions and structural rearrangements (Helentjaris, 1987).

RFLPs have a number of advantages as molecular markers.

Some of these include: the absence of environmental or developmental influences on their expression; their codominant inheritance pattern; there is essentially no limitation to the number of markers that can be utilized; the DNA samples and filters prepared from them are useful for a long time and can be repeatedly examined. The limitations of RFLPs are the large amount of DNA required (5-10 µg) and the requirement for radiolabelled probes. However, an alternative to overcome the use of radioactivity is the recent availability of non-radioactive probing procedures. In general, their major disadvantage is the high cost and extensive labour involved in the process. Although for repeated probing of large sample sizes it has been argued that in comparison to PCR based methods, RFLPs compare favourably (Ragot and Hoisington, 1993).

This aside, the application of RFLPs to plant breeding and genetics are numerous, and have been reviewed by different authors (Beckman and Soller, 1986a; Tanksley et al., 1989). Generally, the first step in evaluating the actual feasibility of using RFLPs in plant breeding and genetics is to examine the degree of genetic variability detectable by RFLPs. Burr et al. (1988) and Helentjaris et al. (1985) have reported significant variability among different maize inbred lines when either repetitive or single-copy DNA sequences were used as probes, however, intraspecific RFLPs were very low in tomato (Helentjaris et al., 1986).

Nevertheless, variability has also been detected in a wide variety of other crops (Helentjaris et al., 1985; Tanksley et al., 1989; Menancio et al., 1990).

Perhaps the major use of RFLPs has been in genetic linkage mapping. Constructing a linkage map generally involves the following steps: 1. development of probes; 2. identification of parental lines and polymorphic probes; 3. genotype determination in the segregating populations; and 4. analysis of data to generate a genetic map. The segregation analysis can be performed with a range of populations. F_2 populations provide the most genetical information per individual, but backcross populations and inbred line populations can also be used (Burr et al., 1988). The construction of a linkage map is based on the estimation of recombination frequencies between genetic loci representing different alleles and the determination of the linear order of loci in linkage groups. Several computer software programmes have been developed to assist in the construction of genetic linkage maps (Lander et al., 1987; Ritter et al., 1990), and such maps have been now developed for a high number of crop plants (O'Brien, 1993).

Linkage maps can be used in plant breeding for indirect selection of closely linked agronomic traits. With indirect selection, one does not directly select for the gene of interest, but RFLP markers. If the RFLP markers are indeed

closely linked, they will remain associated during segregation. This allows one to select for the RFLP marker with confidence that the conventional gene will also be present, since only relatively rare recombination events would separate the two. RFLPs have been utilized to detect markers linked to a number of genes in a wide range of crops (e.g. Sarfatti et al., 1989). RFLPs are also being utilized to identify markers linked to genes affecting quantitative traits. In two separate programmes in tomato, selection for specific quantitative trait loci (QTLs) in segregating progenies has led to the development of insect resistance lines (Nienhuis et al., 1987) and tomato with increased soluble solids content (Osborn et al., 1987). The detection of RFLPs linked to traits of a quantitative nature promises to be a very important application to crop breeding and improvement programmes.

RFLP technology has been used to investigate aspects of plant evolution (Song et al., 1988; Debener et al., 1990; Miller and Tanksley, 1990) and research based on RFLP mapping has provided useful information on comparative genome organization in pepper, potato and tomato (Bonierbale et al., 1988; Gebhardt et al., 1989; Tanksley et al., 1988). Phylogenetic relationships among species have been based primarily on polymorphism derived from chloroplast and nuclear DNA (Song et al., 1988, Smith and Sytsma, 1990; Pradhan et al., 1992; Barret et al., 1992;

Palmer et al., 1983). Phylogenies obtained with RFLPs in most species are in general agreement with those determined by conventional methods (Debener et al., 1990; Menancio et al., 1990) even though the validity of deriving phylogenetic (as opposed to phenetic) trees by different statistical approaches is disputed (Debener et al., 1990).

Other applications of RFLPs are: as a diagnostic tool for cultivar, line or individual identification (Soller and Beckmann, 1983); for the analysis of plant diseases, such as strain identification (Hulbert et al., 1988); to speed the introgression of a character controlled by a gene or genes from a wild species or a cultivar to another cultivar (Tanksley et al., 1989); and to monitor somaclonal variation occurring in *in vitro* propagated material (Beckman and Soller, 1986b).

1.10.3 POLYMERASE CHAIN REACTION (PCR)

The *in vitro* amplification of DNA by the polymerase chain reaction (PCR) is a recently developed and alternative method which has helped overcome some of the limitations of conventional RFLPs (Erlich et al., 1991). PCR is technically simple, rapid and requires only small amounts of DNA. It is based on the synthesis of a complementary strand of DNA in the 5' to 3' direction, using a single-stranded template, but starting from a double-stranded

region (White et al., 1991). It involves the use of a thermostable DNA polymerase (Taq) from the thermophilic bacterium *Thermus aquaticus* (Vosberg, 1989). In general, two primers are needed for a PCR reaction, each complementary to opposite strands of the region of DNA to be amplified. After denaturation by heating, the primers are allowed to anneal to the single strands of DNA and direct the synthesis of a new strand facilitated by activating the polymerase by increasing the temperature. This process is then repeated around 30 times, allowing a doubling of the target DNA every cycle. PCR is routinely carried out in a specialized heating block called a thermal cycler. For each cycle the following steps are generally employed: 1. denaturation of DNA into single strands (usually at 92-96 °C for one minute); 2. annealing of the primers to single DNA strands (55 °C or less); and 3. the synthesis of DNA (72 °C for five minutes). After 30 or more cycles, the amplification products are separated by gel electrophoresis, stained with ethidium bromide and visualized by illumination with ultraviolet light.

PCR is now used routinely in molecular biology and in plant research. D'Ovidio et al. (1990) reported that PCR can be used to detect genetic polymorphism in wheat with primer sequences derived from the sequence of a γ-gliadin gene. After gel electrophoresis, the derived amplification products detected genetic polymorphism between the studied

genotypes. Weining and Langridge (1991) showed that PCR revealed genetic polymorphism in barley and wheat, with specific primers derived from the sequence of a α-amylase. Tragoonrung et al. (1992), using PCR and known sequence primers, distinguished two types of polymorphism in barley; one due to insertion/deletions events and the other due to point mutations.

1.10.4 RANDOM AMPLIFIED POLYMORPHIC DNA MARKERS (RAPDs)

A relatively new PCR-based procedure using single primers of arbitrary sequence was developed independently by Williams et al. (1990) and Welsh and McClelland, (1990). Variation detected by this approach has been termed Random amplified polymorphic DNA (RAPD) markers or Arbitrarily Primed-PCR (AP-PCR) markers. Arbitrary sequence ten-mer oligonucleotide sequences containing at least 50 % G/C content and no palindromic sequences are routinely used in this process. The generation of specific products depends on the probability that a given DNA sequence complementary to that of the primer will occur in the genome, on opposite DNA strands, in opposite orientation and within a distance which is amplifiable by PCR. Polymorphism results from changes in the primer binding site, such as point mutations or from changes which modify the size or inhibit the amplification of the target DNA (such as insertions, deletions and inversions). Some advantages of RAPDs over

RFLPs are: 1. a single set of primers can be used for detection of polymorphism in a wide variety of species; 2. they do not require preliminary work, such as that required for RFLP analysis; and 3. they require a small amount of DNA (Waugh and Powell, 1992). One limitation of RAPDs is their dominant nature (only the dominant allele is present in the profile), so mapping approaches have to take this into account. Another disadvantage is their reliability; however, this can be overcome by optimizing the DNA concentration, using equal amounts for every DNA sample, and ensuring the same reaction conditions and thermal profile are maintained during amplification.

RAPD markers have found wide application. Most significantly, RAPDs have been used in the analysis of species not previously explored with molecular markers. For example, in cocoa, RAPDs were used to characterize clones from three sub-populations (Criollo, Forastero and Trinitario) where they were able to discriminate between the geographical origin of these populations (Wilde et al., 1992; Russell et al., 1993). Chalmers et al. (1992), used RAPD markers to examine population structure and genetic variation in *Gliricidia sepium* and *G. maculata*, and Yang and Quiros (1993) used RAPD markers to identify 23 celery cultivars. In a germplasm collection of *Brassica oleracea* L. and *B. rapa* L. RAPDs discriminated specific differences among accessions (Kresowich et al., 1992). Results from

Demeke et al. (1992) indicated that RAPDs can be used for classification at various taxonomic levels (among individuals, cultivars, and species in *Brassica)*. Wachira et al. (1995) used RAPD markers to estimate genetic diversity and taxonomic relationships in varieties of different tea species.

RAPDs have also been used in assays for gene introgression. For example, interspecific nuclear gene flow and presumed hybrid origins of *Iris* spp. were observed using RAPDs by Arnold et al. (1991) and Dawson et al. (1995) showed gene introgression between populations of *Gliricidia sepium* and *G. maculata*.

As with RFLPs, RAPDs have been exploited for the generation of genetic maps (Roy et al., 1992; Reiter et al., 1992; Chaparro et al., 1992). In a population of *Arabidopsis thaliana*, Reiter et al. (1992), quickly established 250 new genetic markers and demonstrated the uselfulness of RAPD markers for the production of saturated genetic maps. Despite the difficulties in producing segregating populations in conifers (Carlson et al., 1991), RAPDs have been used in Loblolly pine to generate a 191 marker RAPD map (Chaparro et al., 1992).

One of the important practical aims of generating genetic maps is the identification of markers linked to favourable

monogenic or polygenic characters. Barua et al. (1993), using near isogenic lines in barley, were able to identify RAPD markers linked to a resistance gene for *Rhynchosporium secalis*. Michelmore et al. (1991), using bulk segregant analysis, targeted a RAPD marker linked to the Dm 5/8 locus conferring resistance to downy mildew in lettuce. Similar approaches have been also taken to identify RAPD markers linked to quantitative characters (QTLs). For example, Chalmers et al. (1993), using double haploid lines in combination with RAPDs and bulk segregant analysis, identified RAPD markers linked to genes that determine a quantitative character termed milling energy in barley.

1.10.5 UNIVERSAL PRIMERS

Universal or consensus primers allow the amplification of DNA from a wide range of species, using only one pair of primers (Kocher, 1992). Such primers are generally designed from a comparison of the nucleotide sequences from a range of species, and their use generally extends to the amplification of homologous regions in distantly related species, using PCR. Kocher et al. (1989) used consensus primers derived from conserved regions of the mitochondrial genome in combination with PCR to study genetic relationships in more than 100 animal species including mammals, birds, amphibians, fishes, and some invertebrates. Using the same principle, Xiong and Kocher (1991) studied

the analogous region in mitochondrial DNA sequences of Drosophila and Black fly. Taberlet et al. (1991) designed primers to amplify noncoding regions of the chloroplast genome and demonstrated that these primers were useful over a wide taxonomic range and therefore of general use in studies of population biology and plant systematics. Fennell (1994) used those primers to determine systematic relationships in *Vicia,* and Russell (1994) designed and used primers which amplified a variable region of the plant mitochondrial small subunit rRNA gene to establish relationships between different *Theobroma* species. These primers have also been used to amplify homologous fragments from different species, including oilseed rape, flax and tobacco. However, the mtDNA sequences showed less variation than the cpDNA.

1.10.6 MICROSATELLITES

Microsatellites (Simple Sequence Repeats, SSRs), consist of short units of tandemly repeated nucleotide motifs. They are found at high frequency in eukaryotic genomes and are highly polymorphic (Weber and May, 1989), especially the dinucleotide repeats (AC)n, (AG)n and (AT)n. SSRs are inherited in a codominant manner, and both alleles are amplified by PCR, using primers specific to the short genomic regions containing the repeated sequence. Polymorphism is observed usually after separation on high

resolution acrylamide gels and SSRs have been used as markers in humans, rat and mouse genomes to generate high-density genetic maps (Todd, 1992; Dietrich, 1992; Serikawa et al., 1992; Hazan et al., 1992). Recently, SSRs have been used to detect polymorphism in plant genomes such as soybean (Morgante and Olivieri, 1993; Powell et al., 1995), *Brassica sp.* (Bell and Ecker, 1994), and rice (Zhao and Kochert, 1992).

The original studies on plants used SSRs that were identified in DNA databases, but their number was relatively low. Consequently, much effort has been invested in identifying sufficient SSR markers, to study the genetics of populations and construct genetic maps.

1.10.7 CLEAVED AMPLIFIED POLYMORPHIC SEQUENCE (CAPS)

This assay is a combination of RFLP and PCR. The contribution of PCR is the amplification of DNA sequences from different individuals. The primers used are derived from either random low copy number genomic clones or DNA sequences from public databases which give information about the locus targeted. The amplified products are digested with different restriction enzymes to identify RFLPs. Polymorphic CAPS are usually codominant. The advantage of this technique is the availability of a wide range of restriction enzymes which can be employed.

However, the main limitation is that the polymorphism detected is generally lower than that generated by RFLPs. This is almost certainly due to the limited size of the amplified fragments (less than 2Kb).

1.10.8 SINGLE-STRAND CONFORMATIONAL POLYMORPHISM (SSCP)

The principle of the technique is that single stranded DNA molecules take on specific sequence-based secondary structures (conformers). In SSCP the genomic region of interest is first amplified by PCR, the resulting double stranded DNA (dsDNA) is then denatured to single-stranded DNA (ssDNA) and separated electrophoretically in a nondenaturing polyacrylamide gel, potentially revealing any polymorphisms present in the sequences assayed. SSCP analysis has been shown to be an effective technique for the detection of single base substitutions (Sheffield et al., 1993) and has been used specifically to detect point mutations in the chloroplast genome of *Nicotiana plumbaginofolia* (King-Ying et al., 1993).

1.10.9 ADDITIONAL POLYMORPHIC ASSAYS

Apart from the polymorphic assays already mentioned, several modifications of these have been developed. These include DAF (DNA amplification fingerprinting), SCARs (Sequence-characterized amplified regions) and AS-PCR

(Allele-specific PCR). DAF, like RAPDs, uses one short primer (5-8 nt in length) of arbitrary sequence to amplify short segments of genomic DNA, generating different DNA extension products. The amplification products are generally analyzed on acrylamide gels and detected by silver staining. SCARs uses specific primers derived from DNA sequences of RAPD bands to observe polymorphism on agarose or polyacrylamide gels (Paran and Michelmore, 1993) and AS-PCR uses primers designed to amplify one specific allele. Specificity in the later is achieved by designing one or both primers so that they partially overlap a site of sequence difference between the amplified alleles (Nickerson et al., 1990).

1.10.10 SEQUENCING OF PCR PRODUCTS

Determining the DNA sequence of homologous DNA fragments is the highest resolution assay available. As automated DNA sequencing is now available, nucleic acid sequence information can be obtained both quickly and efficiently. This is an extremely powerful approach, especially for phylogenetic studies when combined with PCR technology. PCR can be used both to amplify the genomic regions of interest and for subsequent sequencing. In order to streamline the entire process it is usually convenient to screen amplified fragments for preliminary detection of sequence polymorphism by using SSCP (Hayashi, 1992). Importantly,

the level of nucleotide sequence divergence between individuals varies according to the species under study. For example, the maize genome reveals 1% divergence (Shattuck-Eidens et al., 1990), while in melons this is reduced to 0.1 %.

1.10.11 AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP)

This method combines the detection of restriction site changes directly with PCR. In this assay (Zabeau and Voss, 1993), genomic DNA is digested with two restriction enzymes (a rare and a frequent cutter) followed by ligation of double stranded adaptors to the digested restriction sites. From the total set of ligation products, certain restriction fragments are selected and displayed by using a pair of specific primers with up to three 3'nucleotide extensions; the primer specific to the rare cutting enzyme adaptor is labelled with either ^{32}P or ^{33}P. The resulting radioactive amplification products are then separated in denaturing polyacrylamide gels and visualised after autoradiography. The polymorphism revealed by AFLP can be either dominant or codominant, the first due the presence or absence of restriction sites between genomes and the second because of fragment length differences (Powell et al., 1995).

The AFLP assay produces a high multiplex ratio and this

makes the method suitable for relatively quick mapping studies and DNA fingerprinting. However, AFLP's require skilled staff and extensive facilities for the use of radioactivity and a well equipped lab.

1.11 CHLOROPLAST AND MITOCHONDRIAL GENOMIC DNA AND EVOLUTIONARY STUDIES

Plants possess three DNA-containing organelles, the nucleus, the mitochondria and the chloroplast. The amount and size of the DNA in the cytoplasmic organelles is small relative to the nucleus. The knowledge that the chloroplast and mitochondria contain their own DNA has led to intensive studies of the structure and organization of organellar genomes, and of the identity, sequence, and expression of their constituent genes (Palmer, 1985a).

The chloroplast genome is a circular molecule ranging in size from 135 to 160 Kb. In higher plants, it is characterized by a large, ca. 25 kilobase (kb) inverted repeat, which divides the remainder of the genome into large and small single copy regions (Palmer, 1985a; Sugiura, 1989). Recent studies of chloroplast genome evolution have revealed a high degree of conservation in size, conformation, structure, gene content and linear order among major lineages of land plants (Palmer, 1985b; Palmer, 1992; Palmer and Stein, 1986). This conservative

mode of cpDNA evolution suggests that any change in structure, arrangement, or content of the chloroplast genome may have significant phylogenetic implications (Downie and Palmer, 1992).

To date, complete chloroplast sequences have been obtained for five chloroplast genomes; tobacco (*Nicotiana tabacum*) (Shinosaki et al., 1986), liverwort (*Marchantia polymorpha*) (Ohyama et al., 1986), black pine (*Pinus thumbergii*) (Wakasugi et al., 1994), rice (*Oryza sativa*) (Hiratsuka et al., 1989) and *Epifagus virginiana* (Wolfe et al., 1992). This complete sequence information has contributed towards an increasing understanding of the structure, gene content and evolution of the chloroplast genome in higher plants.

Chloroplast genes identified include the complete set of rRNA and tRNA genes and some 25 protein encoding genes. Another 20 polypeptides, (primarily ribosomal proteins) are known to be synthesized within the chloroplast and are primarily encoded by cpDNA (Curtis and Clegg, 1984). The most prominent structural differences found among individual chloroplast genes within species involves introns, for example some tRNA genes in angiosperms contain large introns of 451-949 bp.

Less information is available on mitochondrial genomes; the only completely sequenced genome is from the bryophyte

Marchantia polymorpha (Oda et al., 1992). It contains about 90 different genes and these include three rRNA genes, 27 tRNA genes, and roughly 60 protein coding sequences. Partial sequencing of angiosperm mitochondrial DNAs (mtDNAs) suggest that they contain a broadly similar set of genes (Olmstead and Palmer, 1994). However, they vary widely in size, structure, and gene order (Newton, 1988), making studies exploiting whole-genome restriction site distances difficult to interpret. Plant mitochondrial DNA evolves rapidly in structure and organization, but slowly in sequence, thus, much of the variation in mitochondrial genomes is due to rearrangements (Palmer and Shields, 1984).

In general, molecular investigations of organelle genomes have made it possible to infer features of their evolution, and conversely to exploit organelle DNA variation in order to assess species relationships.

1.12 OBJECTIVES

The general aim of my study was to develop and use molecular markers to study genetic diversity, phylogenetic relationships and evolution of *Coffea* species, and to consider their potential application in the characterization and conservation of coffee genetic resources and in plant improvement.

The specific objectives were:

1. To assess genetic variability and level of gene introgression in *Coffea arabica* germplasm.

2. To estimate the level of genetic diversity in *Coffea* species.

3. To examine phylogenetic relationships within the genus *Coffea* in an attempt to trace the evolution of *C. arabica*.

CHAPTER 2. MATERIALS AND METHODS

2.1. PLANT MATERIAL

The germplasm used in the study on genetic diversity and detection of gene introgression in *C. arabica* cultivars was composed of 24 *Coffea arabica* accessions, from which 10 were cultivated varieties, seven were wild accessions from Ethiopia and six were hybrid accessions. Three *C. canephora* and one *C. liberica* accession were also used. They are listed in Table 2.1.

Eighteen accessions comprising 11 species were used in the study on genetic diversity in the genus *Coffea*. These accessions together with their geographical origin are listed in Table 2.2. The germplasm used in the study on phylogenetic relationships and evolution of coffee was the same as that described in Table 2.2, except *C. canephora* (Robusta 3580) and *C. arabica* (ETA216733), which were replaced by the *C. arabica* accessions RS-510 and Typica.

The *C. arabica* varieties N39 and Hibrido de Timor 1343, and a segregating population derived from a cross between these lines were examined for preliminary genetic mapping studies.

More detailed descriptions of the *C. arabica* accessions

used in this study and their relationships are indicated in the Appendix (Figure A and Table A).

Table 2.1. Coffee genotypes studies 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	Salvador
	9. Anacafe M-87	Guatemala
	10. Catuai	Brazil
Coffea arabica	11. ET6 A2 16695	Ethiopia
	12. ET25 A4 16712	Ethiopia
	13. ET41 A7 16725	Ethiopia
	14 ET11CA7 16700	Ethiopia
	15. ET19 A3 16708	Ethiopia
	16. ET27 A8 16714	Ethiopia
	17. ET47 A4 16729	Ethiopia
Coffea canephora	18. Robusta 3751	Indonesia
	19. Robusta 3753	Indonesia
	20. Robusta 3580	Congo
Hybrids	21. Hybrid de Timor	Indonesia
	22. Catimor 8660	Brazil
	23. Catimor 5175	Portugal
	24. Catimor 11670	Colombia
	25. Catimor 12870	Brazil
	26. Catimor 8667	Brazil
Coffea liberica	27. Anon	Guatemala

Table 2.2. Species and genotypes used to estimate genetic variation in the genus *Coffea*

Species	Accession	Origin	Source
C. stenophylla	Don G.D8A	Ivory Coast	ORSTOM
C. racemosa	Anon	East Africa	ORSTOM
C. pseudozangue-bariae Brids.	Diani	Kenya	ORSTOM
C. humilis	Anon	Ivory Coast	ORSTOM
C. congensis Froehner	D7	Central Africa	ORSTOM
C. sessiliflora	Anon	Central Africa	ORSTOM
C. brevipes	Anon	Ivory Coast	ORSTOM
C. liberica	Anon	Guatemala	ANACAFE
*C. liberica (Excelsa)**	Anon	Guatemala	ANACAFE
C. canephora	Robusta 3751	Indonesia	ANACAFE
C. canephora	Robusta 3753	Indonesia	ANACAFE
C. canephora	Robusta 3580	Congo	ANACAFE
C. eugenioides Moore	Anon	Kenya	ORSTOM
C. eugenioides Moore	Anon	Kenya	ORSTOM
C. arabica	ET57A216737	Ethiopia	CATIE
C. arabica	ET52A216733	Ethiopia	CATIE
C. arabica	N39	Tanzania	LARI
C. arabica	Caturra	Brazil	ANACAFE

* *C. liberica* (Excelsa) is also known as *C. dewevrei*

ORSTOM Institut Francais de Recherche Scientifique pour le Development en Cooperation, Montpellier, France.

ANACAFE Asociacion Nacional del Cafe, Guatemala, Centro America.

CATIE Centro Agronomico Tropical para la Investigacion y Ensenanza, Costa Rica, Centro America.

LARI Lyamungu Agricultural Research Institute, Moshi, Tanzania

2.2 METHODS

2.2.1. DNA EXTRACTION PROCEDURES

2.2.1.1. LEAF-DISC METHOD

DNA was isolated from fresh leaf material using a modification of the method described by Edwards et al. (1991). Samples were collected using the lid of a sterile Eppendorf tube to punch out a disc of leaf material. This ensured a uniform sample size and reduced the possibility of contamination arising from handling the tissue. The tissue was macerated in the original tube at room temperature using disposable Eppendorf grinders, with the addition of 10-20 mg of Polyclar AT (polyvinylpyrrolidone, insoluble). 400 µl of extraction buffer (200 mM Tris HCl, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS, 10 mM mercaptoethanol) were added and the sample vortexed for 5 seconds. The extracts were centrifuged at 13,800*g* for 1 minute and the supernatant transferred to a clean tube. The supernatant was extracted with phenol/chloroform, then chloroform and the resulting aqueous fraction mixed with 300 µl isopropanol and left at room temperature for 2 minutes to precipitate DNA. Following centrifugation at 13,800*g* for 5 minutes, the DNA pellet was vacuum dried and dissolved in 100 µl TE buffer (100 mM Tris HCl, pH 7.5, 1 mM EDTA). Of this sample 2 µl was sufficient for a

standard PCR reaction.

2.2.1.2. LARGE SCALE DNA ISOLATION

DNA was isolated from fresh or freeze-dried leaf material using a modification of the method described by Gawel & Jarret (1991). Ten grammes fresh weight or 5 g freeze-dried leaf material 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) (2% CTAB, 100 mM Tris HCl, (pH 8), 1.4 M NaCl, 20 mM EDTA, 0.1% DTT), preheated to 65°C was added to the tissue and incubated at 65°C for 30 minutes with occasional mixing. The samples were then extracted with 75 ml chloroform for 15 minutes and centrifuged at 4,066*g* for 5 minutes. The aqueous supernatant was filtered through muslin cloth, mixed with an equal amount of ice cold isopropanol and left at room temperature for 15 minutes to precipitate DNA. Following centrifugation at 16,266*g* for 10 minutes, the supernatant was discarded and the DNA pellet drained by inverting tubes for 5 minutes. The DNA was resuspended in 2 ml of TE Buffer (10 mM Tris HCl, (pH 7.5), 1 mM EDTA). Samples were incubated at 65°C after adding 20 µl RNase A (10 mg/ml) for 15 minutes and stored at 4°C. Any insoluble material was removed by centrifugation for 5 minutes at 16,000*g* and the supernatant removed to clean tubes. The DNA was reprecipitated by the addition of two volumes of 100%

ethanol (ice cold) and recovered by briefly centrifuging at 16,000*g*. The DNA samples were dried for 5 minutes,resuspended in 2 ml of TE buffer and stored at 4°C.

2.2.2. POST ISOLATION DNA PURIFICATION

2.2.2.1. CAESIUM CHLORIDE DNA PURIFICATION

This method was used to obtain DNA suitable for RFLP and sequencing analysis. The procedure was as follows: 1.5 g of caesium chloride was added to a volume of 1.5 ml of DNA solution and mixed in a 2.2 ml centrifuge tube. Ethidium bromide (40 µl of 50 µg/ml stock solution) was added to this solution and it was mixed again. Centrifuge tubes were balanced and then centrifuged at 338,240*g* for 4 hours at 25°C. A DNA band, located in the centre of the gradient in the tube, was visible after centrifugation. The DNA band was collected by inserting a needle attached to a 5 ml syringe; first to the top of the tube to allow the air to enter and then in the middle of the tube to collect the DNA, which was placed in a 2.2 ml tube. The ethidium bromide was removed from the DNA solution by repeated extraction with water saturated butanol. Caesium chloride was removed by dialysis against TE buffer. After dialysis, the DNA solution was removed, precipitated with Ethanol/3M Sodium acetate and resuspended in 500 µl of distilled water. The concentration of the DNA was determined either

spectrophotometrically or by running test gels alongside Lambda DNA standards.

2.2.3. ISOLATION OF PLASMID DNA FROM BACTERIAL CELLS

Plasmid DNA was isolated by the alkaline sodium dodecyl sulphate (SDS) protocol described by Maniatis et al. (1982). Five ml overnight cultures of the appropriate bacterial strains were centrifuged at 1513*g* for 5 minutes at room temperature to harvest the cells, and the pellet resuspended in 100 µl of Solution I (50 mM glucose, 25mM Tris-HCl, (pH 8), 10mM EDTA and 5 mg/ml lysozyme) . After incubating for 5 minutes on ice, 200 µl of freshly prepared Solution II (0.2M NaOH, 1% SDS) was added, mixed and incubated on ice for 5 more minutes. One hundred fifty µl potassium acetate stock, 3M with respect to potassium and 5M with respect to acetate (60 µl 5M potassium acetate, 28 µl glacial acetic acid, 11.5 µl water) was added and the solution mixed by inversion. Genomic DNA and bacterial debris were pelleted by centrifugation at 16,000*g* for 5 minutes and the supernatant was extracted with an equal volume of phenol:chloroform. After centrifugation for 5 minutes, the supernatant was transferred to a clean tube and DNA precipitated by adding 2 volumes of 100 % ethanol. DNA was recovered by centrifugation at 16,000*g* for 5 minutes, washed with 70% ethanol, dried and dissolved in 50 µl TE containing 20 µg/ml RNase A.

2.2.4. ISOLATION OF PLASMIDS FOR USE WITH ABI AUTOMATED SEQUENCER

This method was described by Promega as a Modified Wizard Miniprep procedure, to be used in automated sequencing. The steps were as follows: 10 ml overnight cultures were centrifuged at 1513*g* for 10 minutes to collect cells, which were resuspended in 300 µl resuspension buffer and 2 µl of RNase A (20 mg/ml stock). Resuspended cells were transferred to a 2 ml microcentrifuge tube. Three hundred µl cell lysis buffer were added and mixed followed by the addition of 300 µl of neutralising buffer and mixing again. The solution was spun at 16,000*g* for 3 minutes, and the supernatant transferred to a fresh 1.5 ml microcentrifuge tube and spun again for 3 minutes. The clear supernatant was transferred into two microcentrifuge tubes (approximately 400 µl/tube) and 500 µl of Wizard miniprep DNA Purification Resin were added to each tube, mixed and left at room temperature for 5 minutes with occasional mixing. The resin/DNA mix was pipetted into a five ml syringe barrel, and pushed gently into a minicolumn with the syringe plunger. Three mls of column wash solution were pushed through the minicolumn. The minicolumn was transferred to a new tube and spun at 16,000*g* for 1 minute. Again the minicolumn was transferred to a new tube and 100 µl of sterile water heated to 70 °C were added and the columns incubated at room temperature for 1 minute. To

elute the DNA, the minicolumn was spun at 16,000*g* for 1 minute and the eluent re-applied to the top of the minicolumn and spun through again for 1 minute. DNA was analyzed for purity and concentration against known standards, as indicated in section 2.2.5.1.

2.2.5. ESTIMATION OF DNA CONCENTRATION

2.2.5.1. LAMBDA DNA STANDARD

In this procedure comparisons were made between specific volumes of DNA and standard concentrations of lambda DNA on agarose gels stained with ethidium bromide and visualised by illumination with ultraviolet light.

2..2.5.2. O.D. 260/280 METHOD

DNA concentration was calculated spectrophotometrically at 260 nm where an optical density (OD) of 1 was assumed to correspond to approximately 50 µg/ml double stranded DNA (Maniatis et al., 1982). The ratio of the readings at 260 nm and 280 nm provided an estimation of the purity of the nucleic acids. Pure samples of DNA have a ratio of approximately 1.8. Samples contaminated with phenolic compounds or proteins show significantly lower ratios.

2.2.5.3. FLUOROMETER METHOD

DNA concentration was also estimated using a TKO Mini-Fluorometer, exploiting the specific binding of bis benzidimazole (Hoechst 33258) to double stranded DNA. When Hoechst 33258 binds to DNA, its peak excitation spectrum shifts from 356 to 365 nm and its emission spectrum from 492 nm to 458 nm. The emission and detection peaks of the TKO 100 are optimized for detecting the bound form of the dye.

2.2.6 ELECTROPHORESIS OF DNA

2.2.6.1. AGAROSE GEL ELECTROPHORESIS

To resolve digested and undigested DNA and PCR amplified DNA products agarose gels were run in 1x TBE (89 mM Tris HCl (pH 8.3), 89 mM boric acid, 5 mM EDTA) as running buffer. For minigels 30 ml 1 x TBE and 1% agarose (IBI) were heated by microwave until the agarose was fully dissolved. Once cooled to less than 60 °C, molten agarose was poured into preformed minigel casting trays and the required comb(s) put in position. For larger gel systems the TBE/gel mix was increased proportionally. Once gels had completely set, gel formers and comb(s) were carefully removed and the gels covered with 1 X TBE buffer solution. 5 µl of loading dye (0.25% bromophenol blue, 0.25% xylene

cyanol, 40 % glycerol) were added to DNA samples prior to running. To estimate size, a standard size marker was run alongside the samples (Lambda DNA digested with Eco RI and Hind III). After electrophoresis DNA was visualised by staining the gel in ethidium bromide solution (0.5 µg/ml) and visualized on a UV transilluminator (312 nm).

2.2.6.2. POLYACRYLAMIDE GEL ELECTROPHORESIS

Polyacrylamide gel electrophoresis was used for Single Stranded Conformational Polymorphism gels. For a standard 40 cm x 20 cm gel, 50 ml of the following acrylamide mix was used: 12.5 ml of MDE (6% acrylamide, 0.3% bis acrylamide, 1x TBE), 3 ml 10 x TBE, 200 µl of 10 % APS, and 34.5 ml of distilled water. The mixture was degassed in a vacuum desiccator for 10 minutes to release dissolved oxygen. Finally 80 µl of TEMED (N,N,N'N'-tetramethylethylene diamine) was added prior to pouring between two glass plates, one (front plate) was coated with repel silane (0.2 % dimethyldichlorosilane in 1,1,1,-trichloroethane, Pharmacia) and the other (back plate) was coated with bind-silane (A-174, Pharmacia) solution (0.2 % in chloroform). Gels were placed in a horizontal position, combs positioned to form wells approximately 10 mm deep and the gel left to polymerise (approximately 60 minutes). Once polymerised, the gel (still between the glass plates) was clamped in a vertical position within the

gel tank (Gibco-BRL, model V16). Buffer tanks were filled with 1 x TBE buffer solution and the comb removed. Wells were flushed with buffer, using a Pasteur pipette, to remove any unpolymerised acrylamide. Prior to loading, loading dye (95 % formamide, 0.05 % bromophenol blue, 0.05 % xylene cyanol, 20 mM EDTA) was added to DNA samples. These were heated to 95 °C for 5 minutes and snap cooled on ice. Gels were then run for up to 48 hours with 5 Watts constant power. After electrophoresis, gels were silver stained using the procedure outlined in 2.2.6.3.2

2.2.6.3 VISUALISATION OF DNA FRAGMENTS IN GELS

2.2.6.3.1 ETHIDIUM BROMIDE PROCEDURE

The fluorescent dye ethidium bromide was used to visualise DNA following agarose gel electrophoresis. The compound intercalates between DNA molecules which consequently fluoresce under ultraviolet radiation. Once electrophoresis was complete, gels were immersed in ethidium bromide solution for one hour and destained for 15 minutes. DNA was visualised under ultraviolet radiation by placing the stained gel over a UV transilluminator (312 nm). Photography of gels was carried out using a UVP ImageStore 5000 ultraviolet product computer system.

2.2.6.3.2. SILVER STAINING

The BIORAD protocol included in the silver stain kit (catalog No. 161-0443) was used immediately after electrophoresis was complete. Briefly, the gel was removed from the tank and placed in a container with 500 ml of fixative solution (40 % methanol, 10 % acetic acid) for 30 minutes. The first fixative solution was removed, and 500 ml fresh (10 % ethanol, 5 % acetic acid) were added for 5 minutes, and this step repeated. This solution was removed and 200 ml of oxidizer (100 ml/l of kit stock solution containing potassium dichromate and nitric acid) were added to the container for 5 minutes. After removing the oxidizer, 500 ml of distilled deionized water were applied twice. Two hundred ml of the silver reagent (100 ml/l of kit stock solution containing silver nitrate) were then placed in the container for 20 minutes, and washed with 500 ml of distilled deionized water. Two hundred ml of developer (32 g/l of dry chemical blend containing sodium carbonate and paraformaldehyde) were then added to the container until the solution turned yellow or until a brown "smokey" precipitate appeared. The developer was then replaced with fresh developer for 5 minutes and this step repeated. Gel development was then stopped by adding 500 ml of a 5 % acetic acid solution for five minutes. The gel was then washed three times in tap water to remove the stop solution from the gel and photographed.

2.2.7. POLYMERASE CHAIN REACTION (PCR)

2.2.7.1. RANDOM AMPLIFIED POLYMORPHIC DNA (RAPDs)

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

2.2.7.2. ORGANELLE AND NUCLEAR SPECIFIC PRIMERS

PCR was carried out as for random primers, except that the annealing temperature was increased to 50 °C for organelle specific primers and 53 °C for nuclear specific primers.

2.2.8. RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLPs)

2.2.8.1. RESTRICTION DIGESTS, ELECTROPHORESIS AND SOUTHERN TRANSFER

DNA was digested with the following restriction enzymes: EcoRI, XbaI, BamHI and EcoRV (Boeringer). Digestion was carried out in a total volume of 500 µl using 10 µg of DNA, 50 units of enzyme, 10 % 1 x standard digestion buffer (33 mM Tris-acetate pH 7.8, 65 mM potassium acetate, 10 mM magnesium acetate, 4 mM spermidine, 0.5 mM dithiothreitol) and distilled water at 37 °C overnight. 5 µl of loading dye were added to the digested DNA, and the DNA fragments separated according to length on 1 % agarose gels with 1x TBE as the running buffer. DNA was transferred to Hybond N+ (Amersham, UK) using the alkaline blotting procedure of Reed and Mann (1985), using 0.4 M NaOH as the transfer buffer. After blotting, the membranes were washed in 2x SSC (300 mM sodium chloride, 30 mM tri-sodium citrate) and the DNA bound to the membrane by exposing to UV light (312 nm) for 2 minutes followed by baking at 80 °C for 2 hours. Membranes were stored between Whatmann 3MM paper at room temperature until use.

2.2.8.2. PRODUCTION OF THE PROBE

The probes were generated from *C. arabica* DNA as follow: DNA was digested with Pst I and a fraction of DNA between 0.1 kb and 2 Kb ligated into pUC18 (PBI, Cambridge). A small ligation mixture was used to transform maximum efficiency DH5α bacterial cells (Gibco, BRL), and recombinant plasmids identified and isolated as indicated in sections 2.2.3, 2.2.9.3. and 2.2.9.4. To separate inserts from vector, plasmid DNA was digested with Pst I and run on 1 % agarose gels with 1x TBE as the running buffer. Inserts were excised and isolated from agarose gels using Prep-A-Gene (BioRad, UK), according to the procedure recommended by the supplier.

Labelling of the probes was achieved by two methods either by random priming (Feinberg and Vogelstein 1984), with α[^{32}P]-dCTP (3,000 Ci mmol^{-1}; ICN Biomedicals), or by PCR with digoxigenin-dUTP as described below.

2.2.8.3. NON RADIOACTIVE SOUTHERN BLOTTING

Probes were labelled by PCR as follow: a premix was prepared containing the following components: 1 µl containing 1-10 ng of plasmid DNA; 10 µl 10x PCR buffer, 5 µl dATP, dCTP, dGTP nucleotide stock (4mM each), 2 µl dTTP (2mM stock) 1 µl dig-dUTP (0.4 mM stock), 8 µl M13

reverse and forward sequencing primers (12.5 mM stock); 72.8 µl distilled water and 0.2 µl (1 unit) of Taq polymerase. Labelling was achieved by PCR with the following program: 97.5 °C for 20 seconds; 48 °C for 20 seconds; and 73 °C for 3 minutes for 25 cycles. The amplified products were run on a gel to confirm the presence of the labelled product. Two tests were done to see if the plasmid had actually been labelled. First the product should be visible on a gel and second, by running PCR reactions side by side with reaction that contain no dig-dUTP, the unlabelled reaction product runs in front of the labelled product. The slower migration of the labelled product is caused by the 11 atom side arm and digoxigenin attached to the PCR product. The DNA was then precipitated with Ethanol, and washed with 70 % ethanol. The probe was then dried and resuspended in 100 µl of TE.

2.2.8.4. HYBRIDIZATION AND AUTORADIOGRAPHY

When radiolabelled probes were used the procedure was as follows: the prehybridization and hybridization buffers contained 5x SSPE (1x 180 mM NaCL, 10mM NaPO4 pH 7.4, 1mM EDTA), 0.5 % SDS (Sodium Dodecyl Sulphate), 5x Denhardts solution (1x Denhardts solution is 0.02 % bovine serum albumin, 0.02 % Ficoll, 0.02 % polyvinylpyrrolidine) and 100 µg/ml sonicated salmon sperm DNA). Membranes were pre-washed in 2xSSC and prehybridised in 100 ml of buffer in

glass bottles for 4-16 hours at 65 °C in a rotating oven (Hybaid, UK). Labelled probes were denatured either by heating to 100 °C for 5 minutes or by adding 20 µl of 3 M NaOH and leaving at room temperature for 5 minutes. Denatured DNA probes were added to 10 ml of prehybridization buffer at 65 °C. Hybridizations were performed for at least 12 hours at 65 °C in a rotating oven. Membranes were washed at 65 °C in 2xSSC, 0.5 % SDS for 15 minutes, 2xSSC, 0.1 % SDS for 15 minutes and 0.1xSSC and 0.1 % SDS for 15 minutes.

When non radiolabelled probes were used, the method was the following: 50 ml of prehybridization mix was prepared containing the following: 33.5 ml distilled water; 2.5 ml 10% SDS, 1 ml 5 % lauryl sarcosine, 12.5 ml 20x SSC, 0.5 ml denatured herring sperm DNA (10 mg/ml stock) and 0.5 g blocking reagent (BCL). A prehybridization solution (30 ml from 50 ml prepared) was added to the filter in a plastic bag, and held at 68 °C for 5 hours. For hybridization the probe was boiled for 10 minutes, quenched on iced water and then added to the remaining prehybridization solution at 68 °C and mixed thoroughly. The prehybridization solution in the tube was tipped out and the probe solution added and hybridization allowed to proceed for 16 hours at 100 rpm at 68 °C.

After hybridization, filters were washed in 2 x SSC, 0.1% SDS solution at room temperature for 5 minutes. Filters were then washed for two 15 minute periods in 0.1 x SSC, 0.1 % SDS at 68 °C with shaking at 45 rpm followed by a rinse in sterile buffer 1 (100 mM Tris, 150mM NaCl, pH 7.5) then buffer 2 (same composition as buffer 1 plus 1% non-fat dried milk powder) for 30 minutes. Filters were transferred into a plastic bag and antibody conjugate (1/10,000 dilution) in buffer 2 added and the filters shaken at 45 rpm for 30 minutes. Filters were transferred to a sandwich box and washed as follows: twice in Buffer 1 for 15 minutes; once for 30 minutes and three times for 5 minutes in buffer 2; four times in buffer 1 and one in buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM $MgCl_2$, pH 9.5). In all washes, filters were shaken at approximately 45 rpm. Diluted AMPPD (10.6 µl per ml of 10 mg/ml stock solution in buffer 3) was added to the filters in a plastic bag and shaken at 100 rpm for 5 minutes. The AMPPD was then poured off and the filters wrapped in Saran wrap. The membranes were exposed at room temperature against Fuji X-ray film (Fuji, Japan) for 2 hours to 7 days. After autoradiography the membranes were recycled by washing twice with 1000 ml of prewarmed sterile 0.24 M NaOH, 0.1 % SDS at 45 °C for 15 minutes, with shaking, and then 2x SSC solution for 5 minutes.

2.2.9. CLONING OF PCR PRODUCTS

2.2.9.1. PURIFICATION OF PCR PRODUCTS

Amplified PCR products were purified by using the Magic DNA Clean Up System (Promega). The method is based on the binding of large DNA fragments to an insoluble resin, which allows subsequent washing and purification. First, the mineral oil was removed from the surface of the PCR amplification mixture. To 40 µl of PCR amplification product, 100 µl of buffer was added then 1 ml of resin, and the mixture vortexed briefly. The resin-DNA solution was pipetted into a 3 ml disposable syringe, passed over a mini-column, where the resin was retained, and washed with 2 ml 80% isopropanol. The minicolumn was then transferred to a 1.5 ml Eppendorf tube and centrifuged at 16,000*g* for 20 seconds to remove isopropanol. The minicolumn was transferred to a new 1.5 ml Eppendorf tube. 50 µl of water were added to the upper surface of the resin in the minicolumn and incubated for 1 minute. The column was centrifuged at 16,000*g* for 20 sec to elute the bound DNA in distilled water.

2.2.9.2 LIGATION OF PCR PRODUCTS INTO PLASMID VECTORS

Purified PCR-products were ligated into pGEM^{R}-T by standard procedures. The vector is prepared by cutting pGEM-5ZF(+) with EcoR V and adding a 3' terminal thymidine to both ends. Ligation was as follows; 1 µl of T4 DNA Ligase 10X buffer was placed in a 0.5 ml tube, followed by 1 µl of pGEM-T (50 ng). 7 µl of purified DNA was then added, plus 1 µl of T4 DNA ligase. The total ligation reaction containing 10 µl, was incubated for 3 hours at 15 °C, and then heated at 70 °C for 10 minutes. The reaction was then left at room temperature to allow it to cool, used immediately or stored at -20 °C.

2.2.9.3. TRANSFORMATION

Maximum efficiency DH5α cells (Gibco-BRL), were used for transformation. 4 µl of the ligated PCR product: pGEM^{R}-T were placed in 2.2 ml sterile eppendorf tube. Cells were removed from -70 °C storage and placed on ice bath until just thawed. 50 µl of cells were added to the ligated PCR product, placed on ice for 20 minutes, heated for 45 seconds at 42 °C and returned to ice for 2 minutes. 1 ml of LB liquid medium (1% bacto-tryptone, 0.5 % yeast extract and 1 % NaCl) was added and the mixture incubated at 37 °C, for 60 minutes. 100 µl of the transformation culture were spread on LB plates containing 50 µg/ml

ampicillin, 100 µl of 100mM IPTG (isopropylthiogalactoside) and 20 µl of 50 mg/ml X-gal solution and then incubated overnight (24 hours).

2.2.9.4. IDENTIFICATION OF POSITIVE CLONES

The identification of clones was based on blue/white screening. White colonies generally contained the insert (recombinants) while blue colonies did not (non-recombinant). Individual recombinant colonies were streaked on fresh LB plates containing ampicillin, IPTG and X-gal, and incubated overnight again at 37 °C. Individual white colonies from this new plate were placed in tubes containing 2 ml of LB broth and ampicillin (50 mg/ml), incubated overnight at 37 °C, with shaking at 250 rpm, and then used for plasmid preparation. Recombinant plasmid DNA was extracted by the method indicated in section 2.2.3, digested with 5 units of Sph I and Pst I restriction enzymes, at 37 °C for two hours, and then visualized on minigels to determine the approximate size. Those clones containing the right size insert were grown again, this time in 10 ml of LB liquid medium and used for plasmid preparation for automated sequencing. Permanent stocks of the selected clones were made (Maniatis et al., 1982) by mixing 1 ml of an overnight culture with 70 µl DMSO in a 1.8 ml glass vial, flash frozen and stored at -70 °C.

2.2.9.5. SEQUENCING OF CLONED DNA FRAGMENTS

An Applied Biosystems 373 automated DNA sequencer was used to sequence cloned DNA fragments. The general steps were as follows: A reaction mixture was prepared using the Applied Biosystems (ABI) 'PRISM' DNA premix in a 0.2 ml Microamp tube (9.5 µl of ABI 'PRISM' reaction premix, 5 µl DNA template (200 ng/µl), 1 µl of primer (M13 forward and reverse sequencing primers), and 4.5 µl of water) to a total reaction volume of 20 µl. The Microamp tubes containing the reaction mixture were then placed in a Perkin Elmer 9600 thermal cycler, with cycling as follows: 96 °C for 30 seconds, 50 °C for 15 seconds, 60 °C for 4 minutes, with a total of 25 cycles then 4 °C hold. The DNA extension products were extracted twice with phenol/chloroform after adding 80 µl of H_20 to the reaction mixture. The extension products were precipitated by adding 15 µl of 2 M sodium acetate, pH 4.5, and 300 µl of 100 % ethanol. The precipitated DNA was pelleted by centrifugation at 16,000*g* for 15 minutes and vacuum dried. The samples were prepared for running on a 4.75 % denaturing acrylamide gel, (19:1 acrylamide:bisacrylamide, 7 M urea, 10 % APS and 34 µl of TEMED) by adding 4 µl of loading buffer (5:1 Deionised Formamide:50mM EDTA) and denatured. 2 µl were loaded onto the gel. Gels were run for 12 hours at 30 W. Digital Sequence Data was edited using the 'Sequence Navigator' software supplied with the

ABI 373 Stretch instrument.

2.2.10 RAPD AND STS DATA ANALYSIS

To analyze genetic diversity revealed by random and specific primers, different statistical procedures were employed. The principles and definitions of these methods are as follows.

2.2.10.1 NEI AND LI COEFFICIENT

Also referred to as 'shared fragments'. This coefficient is defined as the number of bands shared by two samples divided by the total number of bands in the two samples (Nei and Li, 1979). The similarity between samples can be used to generate a simple matrix, according to the following formula:

$S = 2n_{ij}/(n_i + n_j)$

where:,

n_{ij} = number of shared product presences,

n_i = number of products in i,

n_j = number of products in j.

2.2.10.2 PRINCIPAL COORDINATE ANALYSIS (PCO)

PCO is based on the matrix of similarities derived from Nei and Li coefficients. This approach allows the samples to be plotted on a hypothetical multi-dimensional grid. The dominant patterns in the data are usually reflected along the first few axes of the analysis. These coordinates can then be plotted in a scatter diagram to provide an indication of relationships among the individuals studied. (see Kempton & McNicol, 1990, for details).

In this work, estimates of similarity and PCO were performed with the GENSTAT 5 (1987) statistical package.

2.2.10.3. SINGLE AND AVERAGE LINKAGE ANALYSIS

In single linkage analysis, the similarity between two groups is defined as the similarity between their two nearest (most similar) members. Average linkage analysis defines the similarity between two groups as the average of all the similarity between the members of the first and members of the second group (see Kempton and McNicol, 1990, for details). Linkage analysis was based on the Nei and Li's similarity matrix, and performed with the GENSTAT 5 statistical package.

2.2.11. SEQUENCE DATA ANALYSIS

Five sets of sequences (chloroplast intron, chloroplast intergenic spacer and chloroplast intron plus intergenic, mitochondrial and nuclear) were multiple aligned using the automated method of Higgins & Sharp (1988). The clustalW package (Thompson et al., 1994), was used with the default parameters for this purpose. The resulting alignment was then checked and edited manually.

The PHYLIP package (Felsenstein, 1993) was used to carry out the phylogenetic tree construction for all sequence sets, except mitochondrial. The procedure was as follows:

A phylogenetic tree was constructed using a distance-matrix method. Pairwise distances were calculated between each pair of aligned sequences using the DNADIST program in the PHYLIP package. The distances were corrected for multiple substitutions using Kimura's two parameters method (Kimura, 1983).

A Neighbor-joining tree (Saitou & Nei, 1987) was then constructed from the matrix of paired distances using the NEIGHBOR program in the PHYLIP package. The Neighbor-joining method (NJ) does not assume a constant substitution rate in all lineages, and has been shown by simulation studies (Saitou & Imanishi, 1989) to perform well in

relation to other methods (Kimura 1983).

The tree (topology plus branch lengths) produced by PHYLIP DNADIST and PHYLIP NEIGHBOR was produced under the assumption that the transition/transversion ratio (Ts/Tv) was 2. The PHYLIP DNAML program was used to provide an estimate and a support interval for Ts/Tv. The default values for Ts/Tv used by the DNADIST and DNAML programs (i.e. Ts/Tv =2) were very close to the actual estimated values from the four data sets. All estimates of Ts/Tv were not significantly different from 2. Thus, the phylogenetic trees constructed using Ts/Tv equal to 2 did not require to be re-estimated.

In order to statistically evaluate the tree produced by the NJ method, a technique known as "bootstrapping" was used. This is a general computer-intensive statistical method which involves repeatedly resampling the original data. The PHYLIP programs SEQBOOT, DNADIST, NEIGHBOR, and CONSENSE were used to carry out the bootstrap analysis. A large number (2,000; see below) of shuffled data sets were generated from the original multiple alignment. Distance matrices and phylogenetic trees were calculated for each of these shuffled data sets, and finally a single consensus tree obtained. This consensus tree contains the number of bootstrap trials that support particular groups (clades, clusters) of species in the tree. Thus a group with over 95

% (1,900 out of 2,000 trials) bootstrap support is likely to be statistically significant. As recommended by Hedges (1992), 2,000 resamples were carried out. The resultant percentage P value quoted at each node has an accuracy of +/- one percent. Thus a bootstrap P value of 95 % at a tree node states that 94 % to 96 % of resamples supported that node. In this work, only bootstrap values in excess of 50 % are shown on the trees. Phylogenetic trees were plotted using the DRAWGRAM program in the PHYLIP package. Phylogenetic trees were produced with the help of Dr. Frank Wright at BioSS (Biomathematics & Statistics Scotland, University of Edinburgh).

Chloroplast (intergenic spacer and intron) and nuclear substitution rates were obtained using the fastDNAml program (Olsen et al., 1994), which uses the maximum likelihood algorithm (Felsenstein, 1981). All calculations were based on the phylogenetic tree constructed using the PHYLIP DNADIST and PHYLIP NEIGHBOR programs. Principal coordinate analysis were conducted using the statistical package of Higgins (1992). The percentage divergence among sequences was estimated with the formula employed by O'Donell (1992):

Percentage divergence= {(Ts+Tv+I/D)/sequence length} x 100, where Ts = transition, Tv = transversion and I/D = insertion/deletion. (Multi-base length differences were scored as 1).

CHAPTER 3. USE OF RAPD AND STS MARKERS TO ESTIMATE GENETIC DIVERSITY AND DETECT GENE INTROGRESSION IN COFFEE (*COFFEA ARABICA*)

3.1. INTRODUCTION

Commercial coffee production relies on two species: *C. arabica* and *C. canephora*. Superior quality coffee (flavour) is associated with *C. arabica* and arabica coffee represents 73% of world production and almost all of the production in Latin America. *C. arabica* is grown at medium and high altitudes. *C. canephora* (robusta coffee) is considered to be of low quality, however, it is widely grown in central and western equatorial Africa and comprises 80% of African production. *C. canephora* is also grown in some Asian countries, such as Indonesia, due to its adaptation to low land altitudes and as a substitute for *C. arabica*, which was esentially eliminated from low altitude regions by the attack of leaf rust (*Hemileia vastatrix*). The robusta coffees have a high caffeine content (2 %) and higher content of soluble extracts, which makes it more economical in the manufacture of instant coffee. *C. canephora* is regarded as an important source of disease resistance genes and these have been transferred into the *C. arabica* genepool via natural and artificial interspecific hybridization between the two species.

The genetic base of the arabica coffee is 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. Several coffee germplasm field collections have been established (Carvalho, 1985) due to the recalcitrance of coffee seed, and tissue culture technology have been used to establish in vitro germplasm collections (Kartha et al., 1981). In both cases there is a need to assess accurately the level of genetic variation in order to minimise duplications 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, such biochemical assays are limited by the number of polymorphic loci detected. For example, electrophoretic analysis of *C. arabica* accessions from Ethiopia and Kenya with six enzyme systems did not show polymorphism (Moreno, 1989). These results contrasted with the level of morphological variation detected in the same germplasm (Louarn, 1978) and suggested that this

approach may be inappropriate for evaluating diversity in *C. arabica*.

In this chapter it is demonstrated that an alternative approach, Random Amplified Polymorphic DNA (RAPDs) (Williams et al., 1990; Welsh and McClelland, 1990) can be used for the genetic characterization of coffee germplasm. In addition, RAPD data have also been used to estimate the relationships between individuals and examples are provided where RAPDs detect interspecific gene introgression from *C. canephora* into the arabica gene pool.

3.2 MATERIALS AND METHODS

3.2.1. 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 2.1).

3.2.2 DNA PROCEDURES

DNA isolation and RAPD analysis were performed as described in Chapter 2. The 10-mer primers used in this study are indicated in Table 3.1. A pair of specific primers which amplify the intergenic region between the nuclear genes

encoding U2snRNAs were also used and those are indicated in Table 3.2. Southern analysis was carried out as described in Chapter 2.

Table 3.1. Primers and their base sequence used for the detection of polymorphism in coffee.

Primer	Sequence
SC10-04	5' TACCGACACC 3'
SC10-15	5' GCTCGTCAAC 3'
SC10-20	5' ACTCGTAGCC 3'
SC10-22	5' CTAGGCGTCG 3'
SC10-25	5' CGGAGAGTAC 3'
SC10-30	5' CCGAAGCCCT 3'
SC10-33	5' TCGCCATAGC 3'
SC10-35	5' GTGCGGACAG 3'
SC10-36	5' TCACCGAACG 3'
SC10-37	5' GCCAATCCTG 3'
SC10-38	5' GACCCCGGCA 3'
SC10-44	5' CCAGGAAGCC 3'
SC10-47	5' ATAGCTCGCC 3'
SC10-49	5' CCACGAGCAT 3'
SC10-50	5' ACGCGCTGGT 3'
SC10-55	5' GGGAGACGTA 3'
SC10-56	5' CCAGCGTCTA 3'
SC10-57	5' GCTGGAAGCG 3'
SC10-63	5' CCTTGCGCTT 3'
SC10-64	5' CCAGGCGCAA 3'
SC10-66	5' AGTGGGCGCA 3'
SC10-69	5' GACGCTCTCC 3'
SC10-70	5' TTGGCCGCGA 3'
SC10-71	5' CTGGCGTAGT 3'
SC10-73	5' TCGGCCCTCG 3'
SC10-74	5' CGGACTTGGG 3'
SC10-75	5' ACCCAGCCAC 3'
SC10-77	5' AGATAGCGGG 3'
SC10-78	5' TCGGAGCGGT 3'
SC10-84	5' TGTGGGCATG 3'

3.2.3 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.

Table 3.2. Sequence of nuclear specific primers used for the detection of genetic diversity in coffee.

Name	Sequence	Amplified Region
POTU222-1	5'GCCAAAAGGCGAGAAAGGTATG3'	Intergenic region between genes
POTU2 23-44	5'TAAGATCAAGTGTAGTATCTGTT3'	encodingU2snRNAs

3.3 RESULTS

Initially the level of polymorphism detected with RAPD markers was assayed in four *C. arabica* accessions, N-39, Blue Mountain, RS 510, Catimor, and in the Hybrid de Timor clone 2252/28. Of the 30 RAPD primers used, 25 detected polymorphism with an average of three polymorphic RAPD loci per primer. An example of the polymorphism detected with primer SC10-33 is shown in Figure 3.1. In order to extend the analysis a further 22 coffee accessions were evaluated (Table 2.1). An example of the level of polymorphism detected with primer SC10-15 is shown in Figure 3.2.

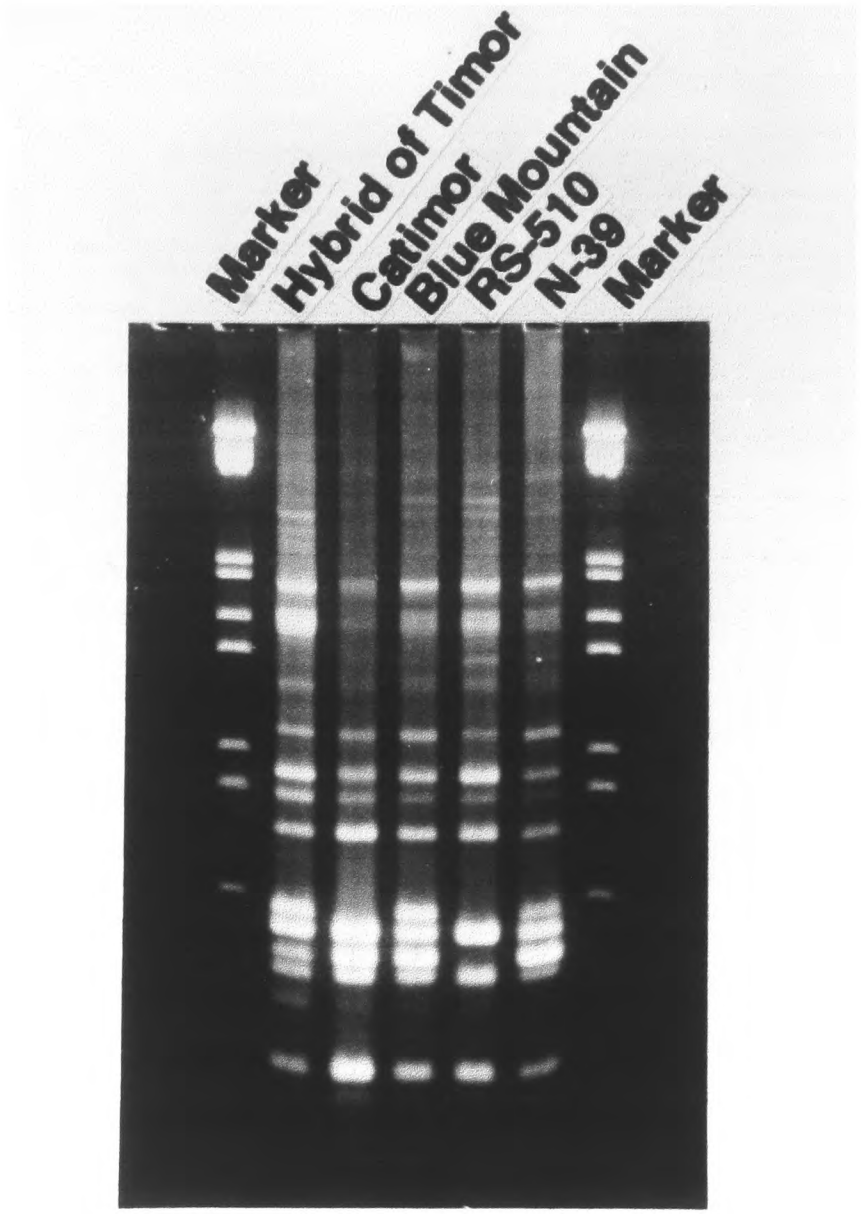


Figure 3.1. RAPD profiles of four accessions of *C. arabica* and Hybrid of Timor with primer SC10-33.

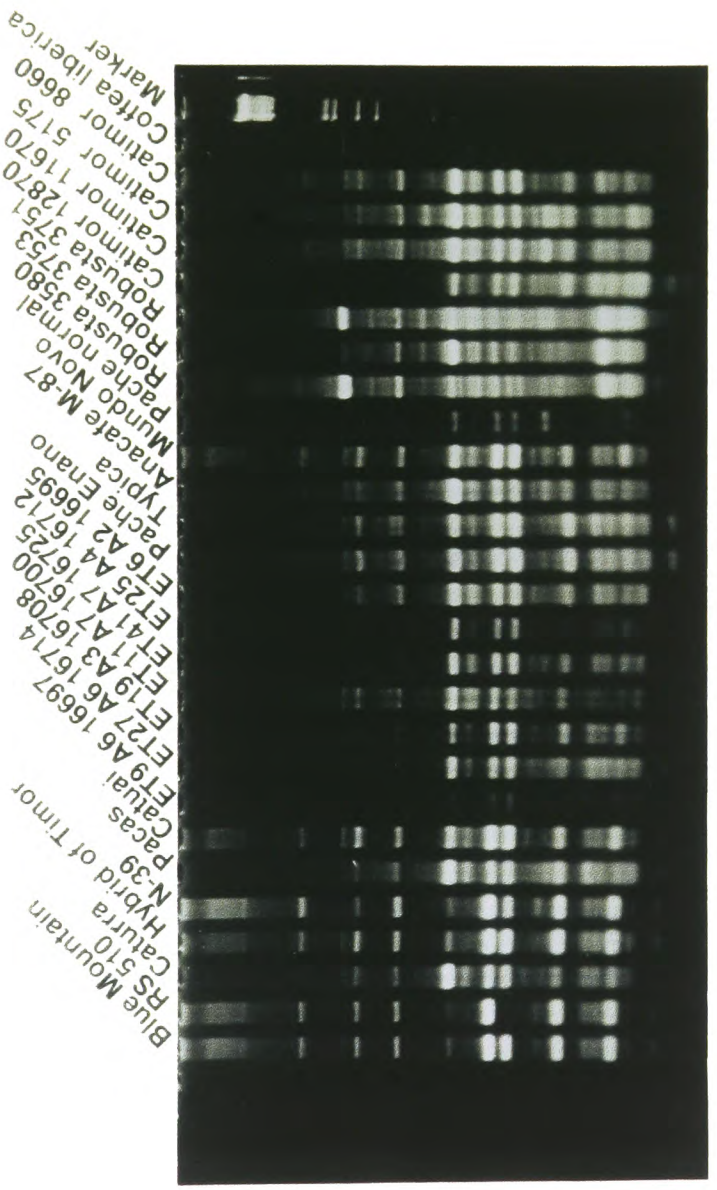


Figure 3.2. RAPD profiles of the 27 *Coffea* genotypes listed in Table 2.1 after amplification with primer SC10-15.

Single linkage cluster analysis was used to examine the relationship between accessions within a species and also between species groups. Only the variable products were used in the analysis. A representative dendrogram shown in Figure 3.3, reflects a clear separation of the two diploid species *C. liberica* and *C. canephora* from the arabica genotypes. There is also a clear separation of the arabica genotypes into three different groups, the first composed of the lines originated from Ethiopia, the second composed of the Bourbon type, and the last composed of the Typica type.

Principal coordinate analysis was conducted to analyze the distribution of variation in the coffee accessions (Figure 3.4). The first and the second component account for 47 % of the total variation. The results supported the separation of coffee species and *C. arabica* genotypes displayed in the dendrogram (Figure 3.3) again forming groups comprising the Ethiopian derived lines, the Bourbon types and the Typica types.

In order to facilitate the presentation of both constant and variable data an alternative graphical method termed "bandmap" has been used (Powell et al., 1991). To display the data the bandmap is shown in Figure 3.5 and it examines the relationship between genotypes and RAPD-derived amplification products. Hence the presence of an

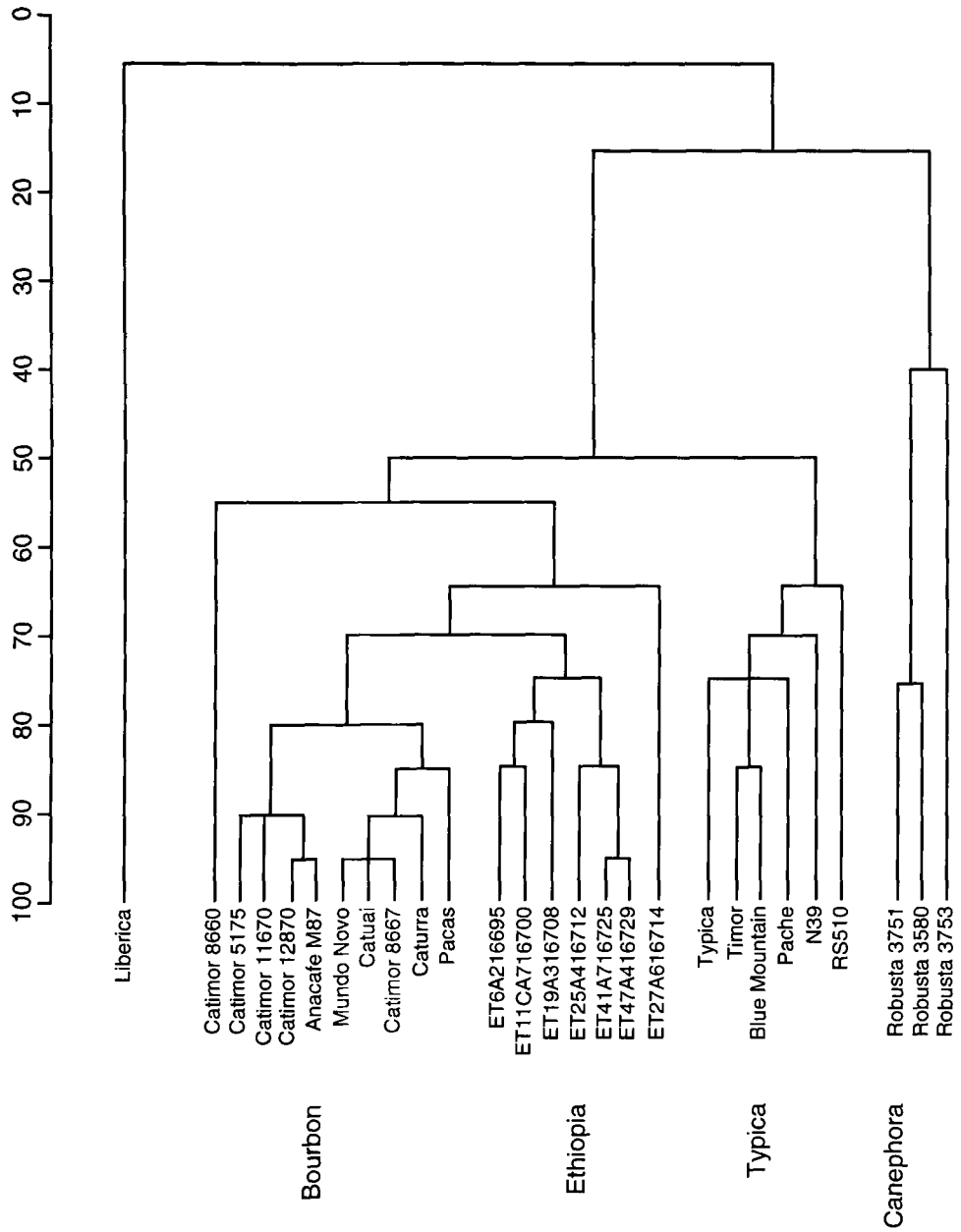


Figure 3.3. Dendrogram of the Coffea accessions listed in Table 2.1 based on single linkage cluster analysis. Only 41 of the informative polymorphic amplification products were used in the analysis.

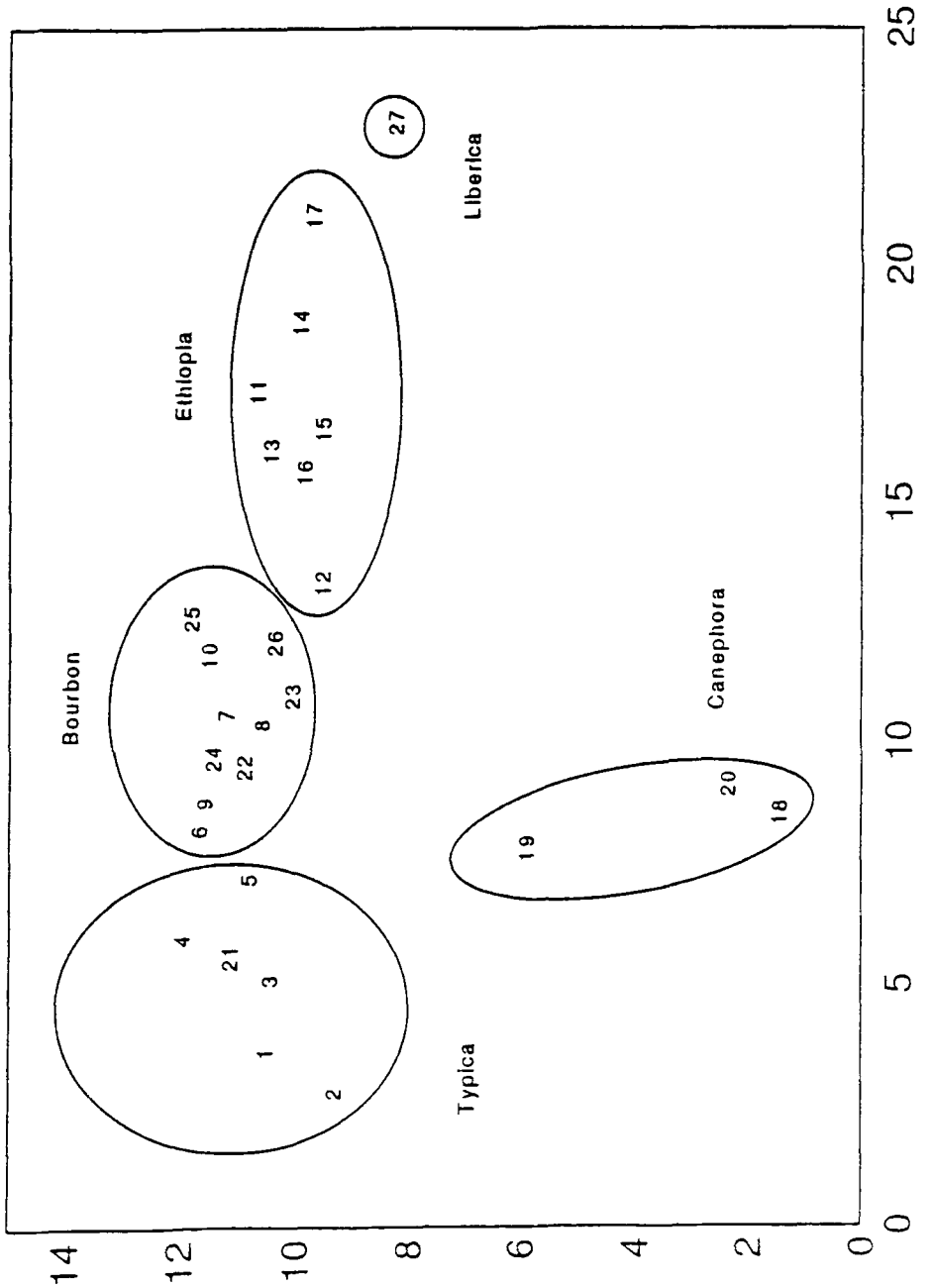


Figure 3.4. Principal co-ordinate analysis of the 27 coffee accessions analyzed using the data set derived from 41 polymorphic markers.

Figure 3.5. 'Bandmap' of shared polymorphic amplification products. Forty one amplification products were generated and these are indicated individually in the *left margin*. The dark box represents the presence of an amplification product. The number in the *right margin* indicates the number of genotypes sharing a given amplification product. The number at the *top* of each column represents the individual coffee genotypes studied as listed in Table 2.1.

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 common amplification products 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 are: the original RAPD information is displayed; similarities between genotypes are highlighted by placing genotypes that share the same amplification products adjacent to one another; attention is drawn to the most and least commonly occurring products; it is immediately obvious whether or not two similar genotypes are in fact identical with respect to their amplification products (Powell et al., 1991; Wilde et al., 1992). Some products such as products 28 and 29 are specific to certain species, in this case for *C. canephora*. It can also be observed that some products are specific for *C. arabica* groups (e.g. products 20 and 22 are present in six accessions that belong to the Typica group). A few specific products which were present in *C. canephora* accessions were also present in some *C. arabica* accessions, such as product 31 in the *C. canephora* genotypes and in the *C. arabica* accessions Rume Sudan RS510 and Catimor 5175.

This is shown clearly in Figure 3.6 where RAPD products

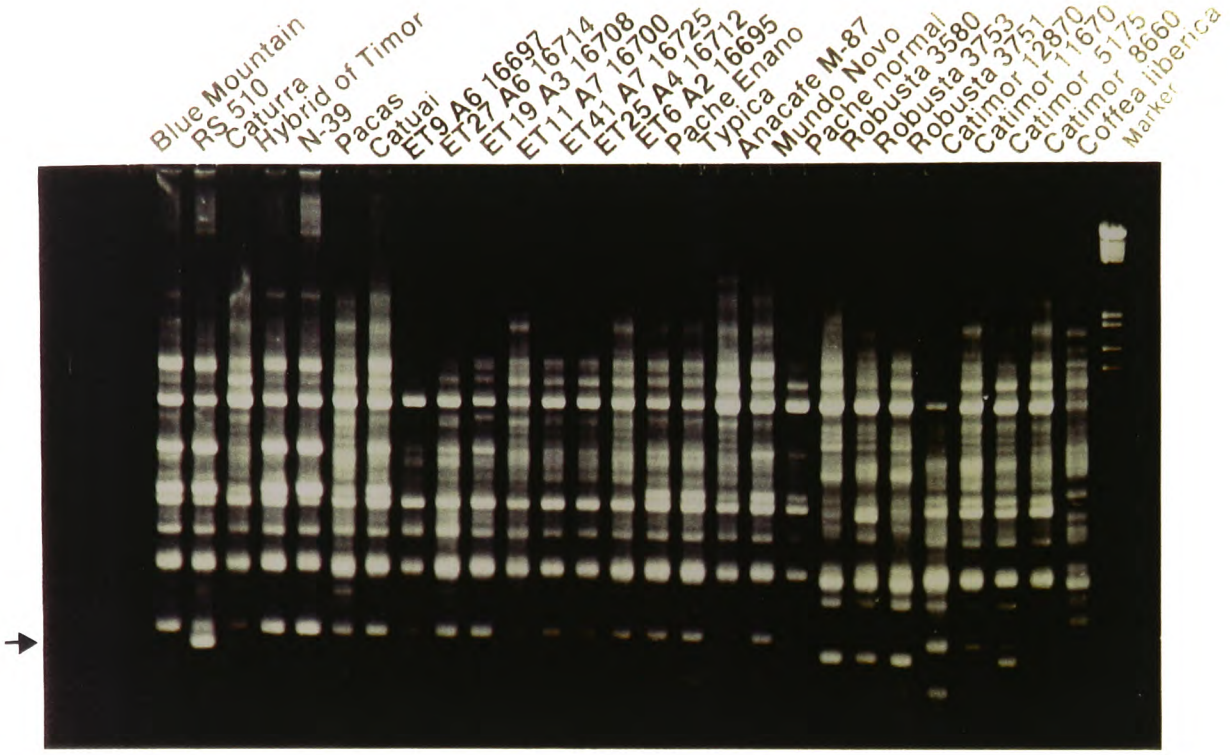


Figure 3.6. Amplification products generated from coffee genotypes, using primer SC10-30. The arrow indicates the amplification product present in Rume Sudan 510, *C. canephora* (Robusta) accessions and Catimor 5175. This unique band pattern reveals the presence of gene introgression in Rume Sudan 510 and Catimor 5175 from *C. canephora*.

derived from the amplification of DNA from the 27 coffee accessions using primers SC10-30 reveal a 0.2 kb product (indicated with an arrow) present in the *C. arabica* varieties RS510 and Catimor 5175, and in the *C. canephora* genotypes.

This unique band pattern appears to reveal the presence of gene introgression in RS510 and Catimor 5175 from *C. canephora*. To exclude the possiblity that the 0.2 kb band was not simply a similar sized non homologous product, the amplification product from *C. canephora* was exised, labelled, and used to probe Southern blots of the SC10-30 amplification products. The results are shown in Figure 3.7. The labelled probe hybridizes to the homologous product in RS510, Catimor 5175, and to the *C. canephora* genotypes, but not to N39.

In addition to random markers, specific primers that amplify noncoding regions of nuclear DNA were used to reveal differences between species and accessions within a species. Primers POTU2.22-1 and POTU2.23-44, which amplify the intergenic region between genes encoding U2snRNAs, revealed clear polymorphism which separate *C. canephora* from *C. arabica*. Furthermore, this pair of primers detected polymorphism among *C. arabica* genotypes (Figure 3.8). The same pair of primers also detected an amplification product that was present in two *C. canephora*

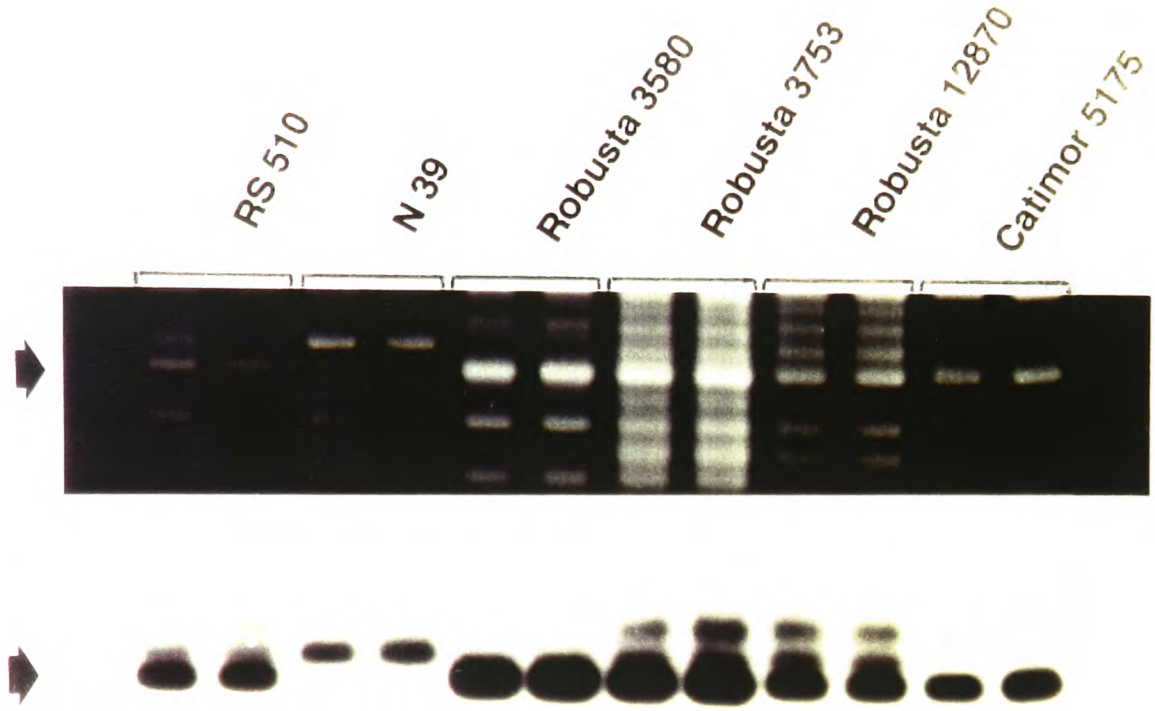


Figure 3.7. The unique band (Figure 3.6) was used to probe a Southern blot of amplified fragments generated by SC10-30. Hybridization occurs to canephora, Rume Sudan 510 and Catimor 5175 but not N39. The product diagnostic of interspecific gene flow is arrowed.

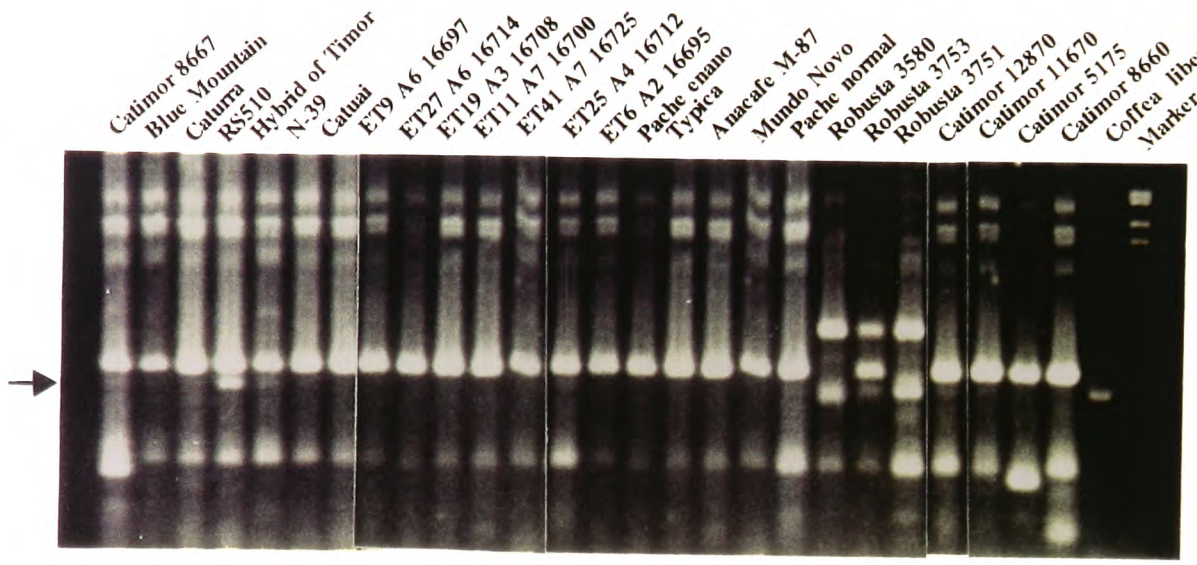


Figure 3.8. Amplification products derived from *C. canephora* and *C. arabica* genotypes, using specific primers POTU2-222 and POTU-2344. The arrow indicates the amplification product present in Rume Sudan 510 and *C. canephora* accessions which demonstrates again the occurrence of gene introgression.

accessions and present in the *C. arabica* genotype Rume Sudan RS510, again possibly indicating the occurrence of natural introgression of genes from *C. canephora* into *C. arabica.*

3.4 DISCUSSION

RAPD markers are being widely used to detect polymorphism and estimate levels of genetic diversity in plant species (Wilde et al., 1992; Chalmers et al., 1993; Wachira et al., 1995) as well as in assays to demonstrate interspecific nuclear gene flow (Arnold et al., 1991; Dawson et al., 1995). Nuclear genome specific sequence tagged sites (STS) markers have also been used to detect polymorphism in plants (Weining and Langridge, 1991; Tragoonrung et al., 1992). In this study, RAPD and nuclear specific STS reflected clear polymorphism in coffee, at the inter and intraspecific level.

The RAPD amplification products generated can be classified into two types: constant (monomorphic) and variable (polymorphic) and these 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. There is a clear separation of the two diploid species *C. liberica* and *C.*

canephora from the arabica genotypes. The *C. 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, principal component analysis was used to examine the shared fragment data available for the 27 accessions. In Figure 3.4, the first two principal components account for 47% of the total variation observed and reveal a clear separation of the three species. Furthermore, the arabica accessions are again separated into three distinct groups representing the Bourbon and Typica types, and Ethiopian-derived germplasm. Graphical representation of the raw data in the Bandmap clearly highlights RAPD products which are diagnostic of a particular species or groups (e.g. products 20 and 22 are only present in the Typica accessions) The bandmap provides a convenient method of genome scanning to locate amplification products which are taxonomically useful. The

potential of RAPDs to identify diagnostic markers has been demonstrated previously for strain identification in mice (Welsh et al., 1991) and cultivar characterisation in plants (Hu and Quiros, 1991; Klein-Lankhorst et al., 1991).

Of particular relevance to coffee is the detection of amplification products which are species-specific and those indicative of interspecific gene flow. Molecular evidence of these events have already been demonstrated in other plants (Dawson et al., 1995; Rieseberg et al., 1988). As in other crops (Rieseberg and Seiler, 1990), both natural and artificial interspecific hybridization have been a feature of evolution and improvement in *C. arabica*. A number of markers are specific to *C. canephora, such as products 28 and 29,* but more importantly some RAPD primers detected loci which were present in the *C. canephora* accessions and also present in the *C. arabica* genotypes at low frequency. For example product 31 (Figure 3.5) is present in the three *C. canephora* accessions and in two *C. arabica* genotypes (Rume Sudan and Catimor 5175). This product was detected with primer SC10-30 and its identity confirmed by Southern analysis (Figure 3.7). Thus, this product is possibly indicative of interspecific gene flow in coffee, which is inferred from the presence of an amplification product which is found in *C. canephora* and in only two *C. arabica* accessions. These could have been

subjected to either natural or artificial hybridization with the former species as discussed below. The potential occurrence of natural gene introgression was also detected by a pair of nuclear STS primers which amplified a product present in *C. canephora* and in the *C. arabica* accession Rume Sudan RS510.

Rume Sudan 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 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 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 a product of natural hybridization 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*). Caturra is a dwarf mutant of the *C. arabica* 'Bourbon' type. The Catimor accessions were produced by backcrossing Caturra with Hybrid de Timor with

selection for disease resistance at each stage of crossing. Both Rume Sudan and Catimor 5175 are characterised by having a 0.2 kb product (SC10-30) derived from *C. canephora* and both arabica genotypes are known to possess genes conferring resistance to coffee berry disease (dominant R- and recessive K- in RS510 and probably dominant T- in Catimor (Van der Vossen and Walyaro, 1980)). This diagnostic product which identifies natural gene introgression in the case of Rume Sudan 510, also identifies artificial introgression in the case of Catimor 5175. It is therefore likely that this RAPD product provides a marker for selective introgressive hybridization in coffee.

In conclusion, the results demonstrate that RAPD markers have the potential to complement both conventional and biotechnological approaches to coffee improvement. More specifically, 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 in *Coffea* spp. In addition, interspecific gene introgression from *C. canephora* into *C. arabica* gene pool was detected which may be associated with the selective introgression of adaptive gene(s) into *C. arabica*. To confirm this possibility, further studies need to be carried out, such as the establishment of a genetic linkage

map, using RAPDs or other molecular markers, where the potential parents would be the two genotypes that showed the diagnostic amplification product for gene introgression, such as RS510 and one of the *C. canephora* accessions (Robusta).

In this chapter the usefulness of RAPDs to estimate genetic diversity between and within *C. arabica* and *C. canephora* germplasm has been demonstrated. Therefore it was considered appropriate to use this molecular assay to analyze a wide range of *Coffea* species, particularly those of relevance in coffee improvement which are maintained in some of the international coffee germplasm collections centres. Thus, a study on genetic variability was initiated, applying RAPDs, chloroplast and mitochondrial specific sequence tagged sites (STS) to a group of species currently used in breeding programmes. The results from this molecular approach are presented and discussed in the following chapter.

CHAPTER 4. EXAMINATION OF INTERSPECIFIC VARIATION AND GENETIC RELATIONSHIPS IN *COFFEA* SPECIES USING PCR-BASED ASSAYS

4.1 INTRODUCTION

While more than 100 distinct taxa have been identified, commercial coffee production relies mainly on two species: *C. arabica* and C. *canephora*. All other coffee species are cultivated to a rather limited extent, or are found only in the wild or in coffee collections maintained in a few coffee research centres. Many of these species, such as *C. eugenioides* and *C. stenophylla* have favourable genes that could be transferred into the commercial species (Carvalho, 1985). As interspecific hybridization can occur frequently in the genus, the effective gene pool which is useful for *C. arabica* breeding practically consists of all *Coffea* species. Of primary importance in the utilization of these resources is the understanding of the phylogenetic relationships of the members of this taxa as the ability to create hybrids is related to the genetic distance between the individuals being crossed.

Conventional methods of taxonomic classification based on morphological characters have been conducted in *Coffea* species (Chevalier, 1947) and relatedness among species using geographical distribution and ecological adaptation

have also been established (Leroy 1980). In addition, crossability and cytological analysis (Carvalho and Monaco, 1968; Charrier, 1977; Berthaud, 1977; Louarn, 1982; Louarn, 1993), biochemical markers (Berthou et al., 1980; Berthaud, 1986), and variation of the cytoplasmic DNAs (Berthou et al., 1983) have been utilized for studying phylogenetic relationships among *Coffea* species. However, these studies have had only limited success in determining relationships, and many inconsistencies still exist. For example, one of the basic unanswered questions in the phylogeny of coffee still concerns the origin of *C. arabica*. It is an inbreeder exhibiting disomic inheritance and is considered to be a segmental allotetraploid (Carvalho 1952; Grassias and Kammacher, 1975) suggesting that it has evolved from a cross between two diploid species. One nuclear genome has been considered to have originated from a member of the genus *Coffea*, the other genome from among the wild species of the subgenus *Eucoffea* or in the neighbouring genera, *Paracoffea*, J.F. Ler or *Psilanthus* Hook F. (Charrier, 1978).

In Chapter 3, RAPDs were used to estimate genetic variation within and between mainly *C. arabica* and *C. canephora* accessions. Given the importance of the *Coffea* gene pool for the improvement of the commercial species, in this chapter a combination of RAPDs, mitochondrial and chloroplast specific sequence tagged sites (STS) have been

used to estimate the level of genetic variation between *Coffea* accessions from 11 different species.

4.2 MATERIALS AND METHODS

4.2.1 PLANT MATERIAL

Eighteen *Coffea* accessions comprising 11 species were used in this study. These accessions together with their geographical origin are listed in Table 2.2.

4.2.2 DNA PROCEDURES

DNA isolation and RAPD procedures were carried out exactly as described in Chapter 2 at least twice on different occasions. Only primers which generated robust amplification profiles and then only major, easily scoreable products were included in the analysis. Thus, of 50 arbitrary sequence 10-mer primers initially screened, a complete data set was obtained from only the 15 listed in Table 4.1.

Table 4.1. Primers and their base sequence used for the detection of polymorphism in *Coffea* species.

Primer	Sequence
SC10-03	5' CAGTTCGAGG 3'
SC10-04	5' TACCGACACC 3'
SC10-05	5' ACTCAGGAGC 3'
SC10-15	5' GCTCGTCAAC 3'
SC10-32	5' CCCCTCGGCT 3'
SC10-33	5' TCGCCATAGC 3'
SC10-37	5' GCCAATCCTG 3'
SC10-44	5' CCAGGAAGCC 3'
SC10-53	5' CAGGGGACGA 3'
SC10-57	5' GCTGGAAGCG 3'
SC10-63	5' CCTTGCGCTT 3'
SC10-73	5' TCGGCCCTCG 3'
SC10-77	5' AGATAGCGGG 3'
SC10-94	5' GGGGTCGATT 3'
SC10-97	5' TCCGGCTTTC 3'

4.2.3 CHLOROPLAST AND MITOCHONDRIAL DNA ANALYSIS

Specific primers that amplify non-coding regions of chloroplast DNA and the V7 region of the small ribosomal RNA gene of the mitochondrial DNA were utilized for detecting genetic diversity among species. The chloroplast and mitochondrial specific primers used are listed in Table 4.2 and were designed according to Taberlet et al., (1991) and Russell (1994). PCR was carried out as for random primers, except that annealing temperature was increased to 50 °C. Amplification products were digested with MboI, RsaI, HaeIII, HindIII, HinfI, DraI, PstI, EcoRI and TaqI

restriction endonucleases. Restriction enzyme digests were carried out according to the manufacturer's instructions (Northumbria Biologicals Limited). Data analysis were performed as outlined previously.

Table 4.2. Sequence of specific primers (chloroplast and mitochondrial regions) used for the detection of genetic diversity in *Coffea* species.

Name	Sequence	Amplified region
CH-A	5'CATTACAAATGCGATGCTCT 3'	Chloroplast DNA trnT(UGU to
CH-B	5'TCTACCGATTTCGCCATATC 3'	trnL(UUA) intergenic region
CH-C	5' CGAAATCGGTAGACGCTACG 3'	Chloroplast DNA
CH-D	5' GGGGATAGAGGGACTTGAAC 3'	trnL(UUA) intron
CH-E	5'GGTTCAAGTCCCTCTATCCC3'	Chloroplast DNA trnL(UAA) to
CH-F	5'ATTTGAACTGGTGACACGAG3'	trnF(GAA) intergenic region
MT-P1V7	5' CTTTCATACAGGTGCTGC 3'	V7 region of plant
MT-P2V7	5' TCACTGGCAGTCCCTCGT 3'	mitochondrial rRNA

CH = chloroplast MT= Mitochondria

4.3 RESULTS

Fifteen RAPD primers were used to screen the 18 accessions listed in Table 2.2. Each primer detected extensive polymorphism, with an average of 4.3 polymorphic RAPD loci being scored per primer from the chosen primers. Products which appeared monomorphic across all 18 accessions were rare. A total of 65 polymorphic products were scored for analysis. Examples of the RAPDs detected with SC10-32 are shown in Figure 4.1.

PCR amplification of the chloroplast DNA sequences revealed no length polymorphism. However, polymorphism was revealed in two of the chloroplast products after restriction endonuclease digestion with RsaI and MboI (Figure 4.2a and 4.2b). *C. arabica*, *C. eugenioides*, *C. humilis* and *C. stenophylla* were found to share a common chloroplast haplotype, *C. canephora*, *C. liberica*, *C. brevipes* and *C. congensis* a second and *C. pseudozanguebariae*, *C. sessiliflora* and *C. racemosa* a third. The mitochondrial amplification products were monomorphic in all species examined.

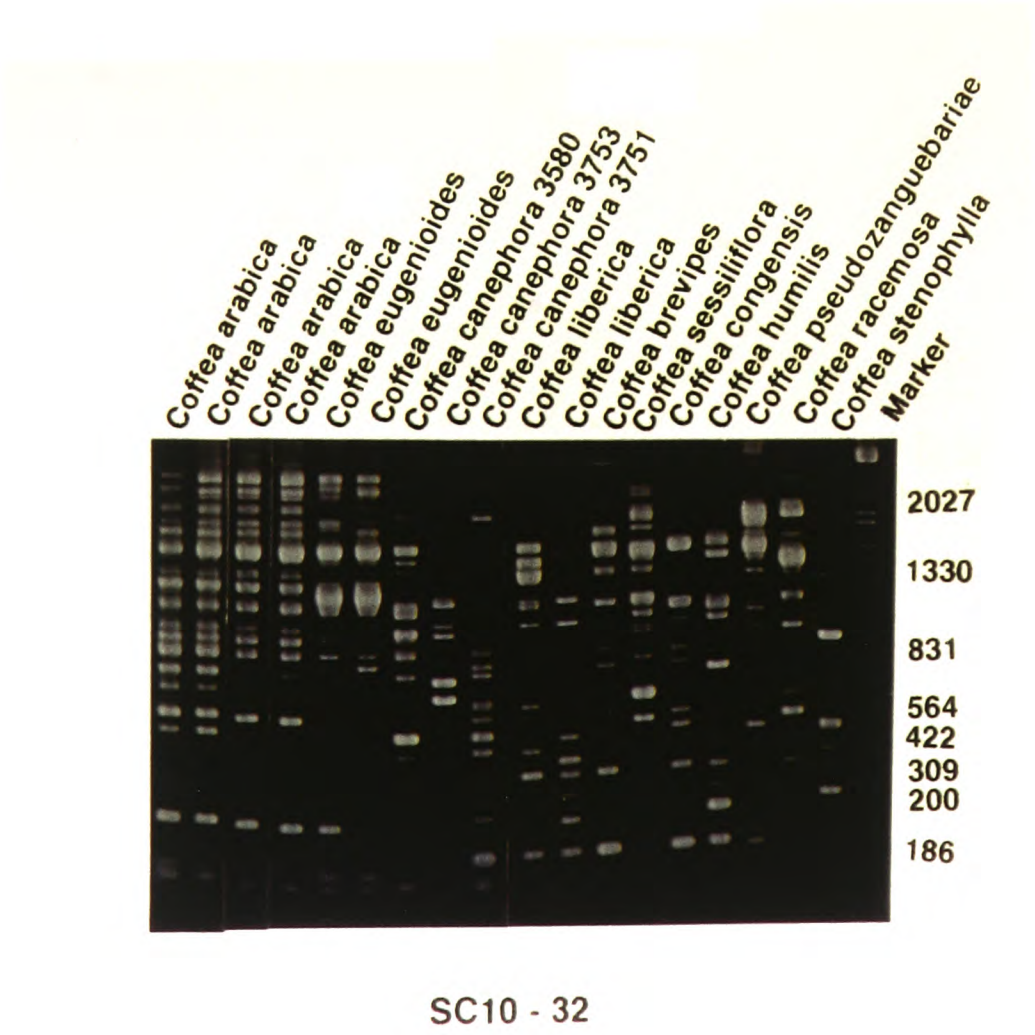


Figure 4.1. RAPD profiles of 18 genotypes from 11 coffee species with primer SC10-32.

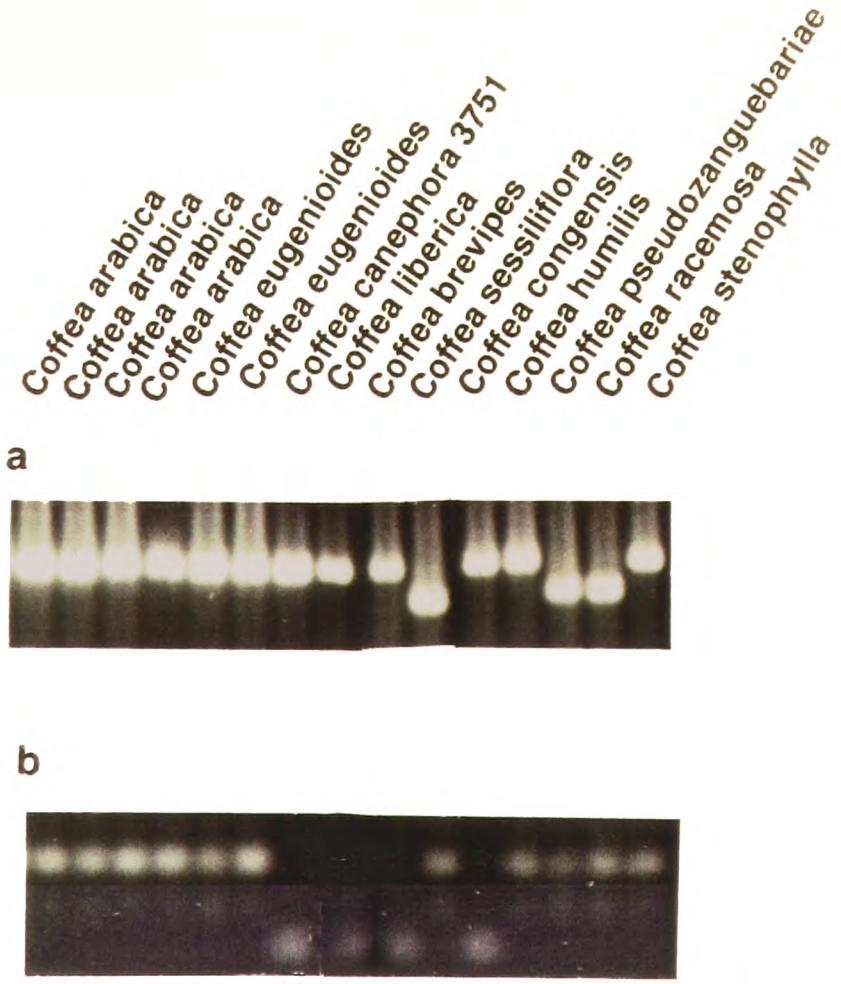


Figure 4.2. Amplification products generated from 18 coffee accessions (11 species), using specific chloroplast primers a) C and D and restriction enzyme Rsa I b) E and F and restriction enzyme Mbo I.

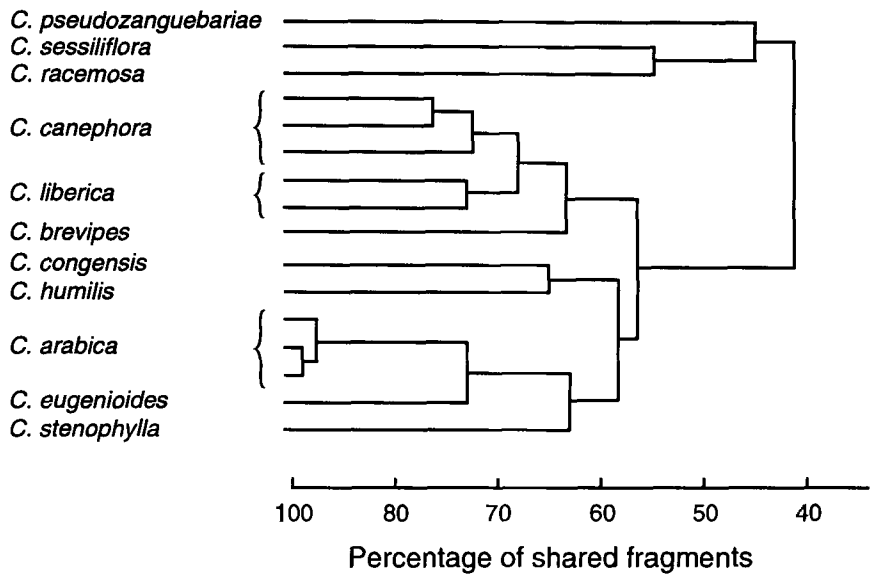
A similarity matrix based on the frequency of shared amplification products (Nei and Li, 1979) using only the variable RAPD products and polymorphic chloroplast loci is given in Table 4.3. For species represented by more than one accession the calculations were based on the frequency of shared products. A number of RAPD products were unique and could be considered diagnostic for a given accession. For the four species represented by more than one accession, the leading diagonal provides an estimate of the within species similarity. The proportion of shared fragments ranges from 0.812 to 1.000.

A dendrogram displaying hierarchical relationships between accessions is given in Figure 4.3a. Even though there are limitations to having only one accession for most of the wild species, giving no measure of intraspecies variation, there is a clear separation of *C. racemosa*, *C. sessiliflora* and *C. pseudozanguebariae* from the other species with 55% of the products differing. Furthermore, while they form an out-group from the other species they are also distantly related to each other and can also be clearly differentiated. Of the remaining species *C. canephora*, *C. liberica* and *C. arabica* show intraspecific variation, which is most pronounced in the diploid species. The exception to this is *C. eugenioides* where the two accessions would appear to be identical. Of particular interest is the similarity between *C. arabica* and *C. eugenioides* which

Table 4.3. Similarity matrix based on the number of shared products. Average similarity within species is represented by the leading diagonal. The remaining values are based on the mean of all accessions within a species.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1. C. stenophylla	n/a										
2. C. racemosa	0.528	n/a									
3. C. pseudozanguebariae	0.449	0.524	n/a								
4. C. humilis	0.667	0.511	0.465	n/a							
5. C. congensis	0.525	0.296	0.360	0.618	n/a						
6. C. sessiliflora	0.509	0.625	0.545	0.531	0.464	n/a					
7. C. brevipes	0.593	0.490	0.383	0.694	0.571	0.500	n/a				
8. C. liberica	0.605	0.366	0.337	0.632	0.687	0.504	0.756	**0.812**			
9. C. canephora	0.592	0.380	0.315	0.572	0.737	0.488	0.699	0.764	**0.820**		
10. C. eugenioides	0.710	0.571	0.462	0.643	0.581	0.483	0.678	0.617	0.644	**1.000**	
11. C. arabica	0.714	0.496	0.387	0.628	0.622	0.577	0.664	0.701	0.723	0.819	**0.986**

a.



b.

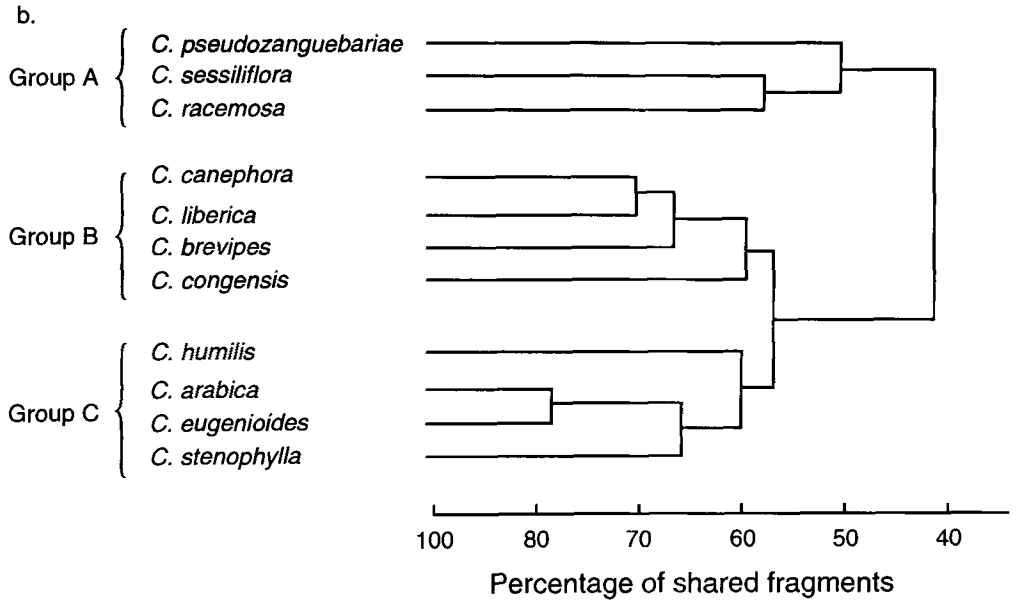


Figure 4.3. Dendrogram of 18 coffee accessions from 11 *Coffea* species analyzed, using data derived from RAPD and STS markers.
a. Based on single cluster analysis.
b. Based on group average clustering analysis.

share 80% of the products scored.

To simplify the representation of the relationships between individual accessions a second dendrogram displaying hierarchical associations between species is given in Figure 4.3b. The dendrogram is generated by group-average clustering where the similarity between two groups is defined as the average similarity of all loci scored in each group. The same general associations were observed allowing the species to be split into three groups (A, B and C in the figure).

In order to assess whether the clustering of populations based on these markers could be further resolved, principal co-ordinate analysis was used to analyze the shared fragment data (Figure 4.4). The first two principal components of this analysis account for 43 % of the total variation. The same general associations were again observed. The principal coordinate analysis also showed that the *C. arabica* accessions cluster most closely to *C. eugenioides*, *C. stenophylla* and *C. brevipes*.

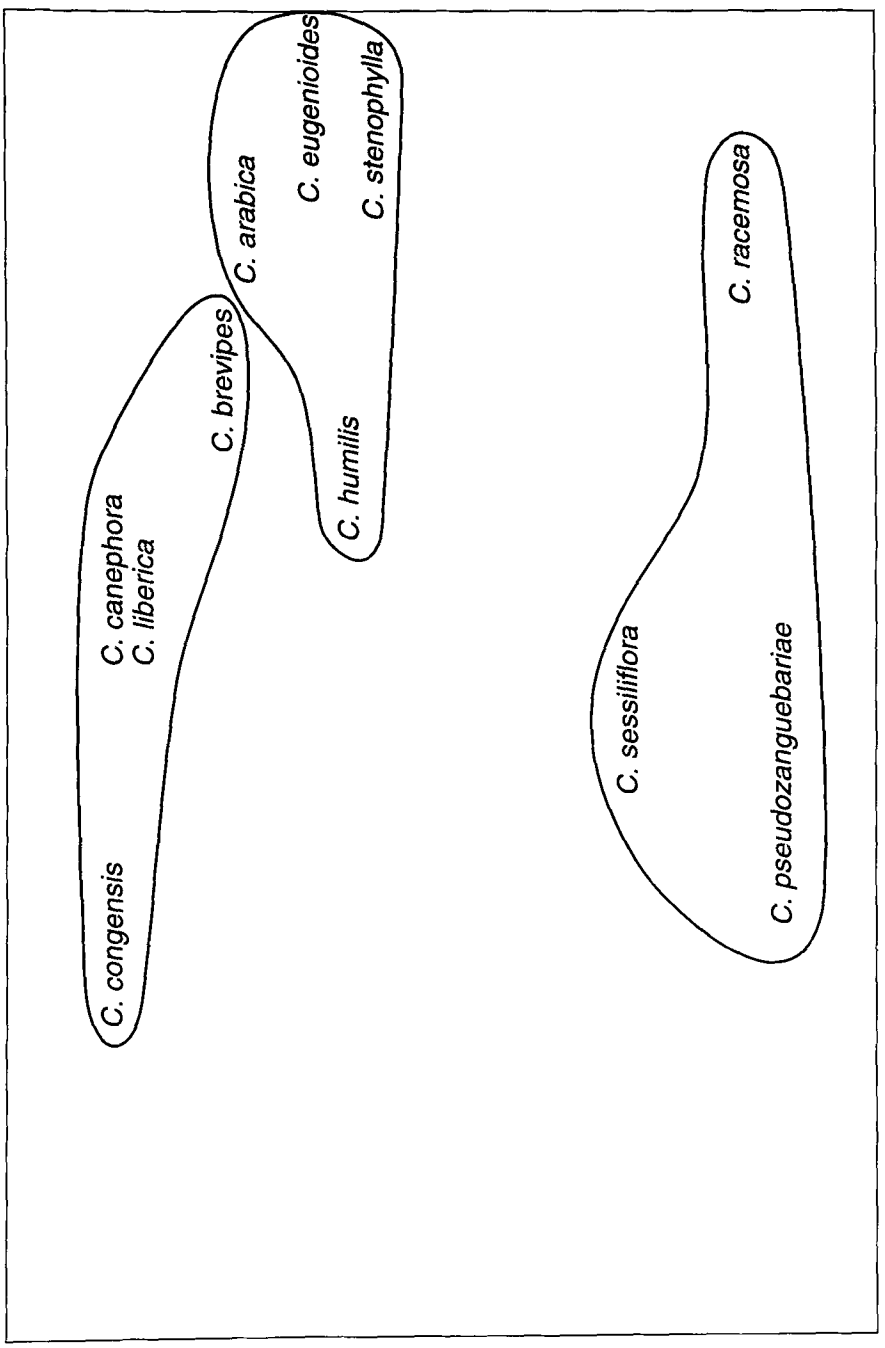


Figure 4.4. Principal coordinate analysis of 11 Coffea species analyzed, using the data set derived from RAPD and STS products.

4.4 DISCUSSION

The genetic relationships in 11 *Coffea* species which are currently important in coffee breeding programmes have been assessed on the basis of RAPD and specific chloroplast and mitochondrial markers. Intra- and interspecific polymorphism was readily detected with RAPD markers (Figure 4.1). The chloroplast markers detected only interspecific polymorphism (Figures 4.2a and 4.2b). *C. eugenioides* was exceptional in that the two accessions of this diploid species appeared to be identical. This was surprising, since being outbreeding and self-incompatible, polymorphism would be expected between accessions, However, The *C. eugenioides* material used in this study was provided as freeze-dried leaf material with a simple identifier attached. It is possible that the two samples were either clones of the same original accession or that a mix-up of the material occurred subsequent to its arrival at SCRI. The low level of intraspecific variation observed for tetraploid *C. arabica* compared to that found in diploid *C. canephora* is in agreement with that found previously using both isozymes and RAPDs (Moreno, 1989; Orozco-Castillo et al., 1994). The restricted genetic variation in *C. arabica* may reflect the tendency of inbreeding crop species towards homozygosity when compared to an obligate outbreeder such as diploid coffee species. Alternatively it may indicate a narrow genetic base or a loss of genetic

diversity during its origin (Hamilton 1974).

There is a close similarity between the dendrograms derived here based on nuclear and organellar DNA markers and classical taxonomic relationships based on morphological characteristics (Charrier and Berthaud, 1985). Of particular interest are the three distinct groupings which broadly correspond to the geographical distribution of these species. Group A, comprised of *C. racemosa*, *C. sessiliflora* and *C. pseudozanguebariae*, originate from South-Eastern Africa and form a distinct out-group with 40% shared fragments (Figure 4.3b). Those forming Group B, *C. congensis*, C. *brevipes*, *C. liberica* and *C. canephora*, are all from Central Africa and have 60 % of shared fragments. The remaining species (Group C) originate from the west and atlantic side of Central Africa. The same results are also observed in the principal coordinate analysis. The chloroplast DNA phenotypes are also different between these groups. This correspondence therefore suggests that geographical factors are important in the speciation process. Removing the chloroplast data from the analysis does not significantly alter the topography of the dendrogram presented (data not shown). Although it may be inappropriate to infer a phylogeny from the type of data presented here, the combination of RAPD and chloroplast DNA markers does appear to reflect a common origin of these species.

It is interesting to notice in the dendrograms and principal coordinate analysis that there is a clear separation of *C. racemosa*, *C. pesudozanguebariae* and *C. sessiliflora* from the other species. The separation of these species in these results is similar to that of Louarn (1993), who found that these species were interfertile between them, but incompatible in interspecific crosses with other species, reflecting a strong fertility barrier. Cros et al. (1993) suggested that the incompatibility of the former species in crosses with other species could be related to genomic size difference, since they have the lowest DNA content per nucleus (about 1 pg). The same broad separation of these species was proposed by Berthaud and Charrier (1988), based on morphological, cytological and geographical distribution of the species.

Of particular interest is the level of similarity between *C. arabica* and *C. eugenioides* (80 % of products shared, Figure 4.3b). Berthou et al. (1983) proposed that *C. arabica* was derived from a close relative of *C. eugenioides*. The results presented here are not in conflict with this theory and both nuclear and chloroplast data indicate that *C. eugenioides* is the most closely related of the diploid species to *C. arabica*. Along with *C. humilis* and *C. stenophylla* these species share the same chloroplast haplotypes suggesting that one of them could be the maternal progenitor of *C. arabica*.

In practical terms, these results provide a better understanding of species relationships within the genus *Coffea* and may facilitate the development of improved coffee germplasm based on alien gene transfer. Various routes to the creation of synthetic amphidiploids have been described for *Coffea* and the data presented here will help in the identification of suitable gene pools for inclusion in such programmes.

In order to extend the analysis of genetic relationships in the genus *Coffea* alternative approaches based on sequence analysis of nuclear and organellar genomic regions may also be pursued. In Chapter 5 this approach is used to examine phylogenetic relationships of *Coffea* species and the evolution of tetraploid *C. arabica*.

CHAPTER 5. SEQUENCE ANALYSIS TO STUDY PHYLOGENETIC RELATIONSHIPS AND EVOLUTION OF THE GENUS *COFFEA*.

5.1 INTRODUCTION

As outlined previously, various studies have been conducted to investigate the genetic affinities between the species of the genus *Coffea*. Despite these, both the relationships between species and the origin of the commercial tetraploid species *C. arabica* are still unclear. On the basis of chromosome pairing observed in haploids it has been suggested that *C. arabica* could be an autotetraploid (Narasimhaswamy, 1962). However, the meiotic behaviour of some interspecific hybrids and the mode of inheritance of duplicated genes indicate that *C. arabica* is an allopolyploid (Carvalho, 1952), or perhaps a segmental tetraploid (Grassias and Kammacher, 1975). Its geographical distribution is characteristic of polyploids, as this falls almost completely outside the range of distribution of the diploid species (Carvalho et al., 1969). Charrier and Berthaud (1985), suggest that *C. arabica* could have arisen from natural hybridization between two ancestral diploid coffee species followed by unreduced gamete formation.

As outlined in the previous two chapters, molecular techniques have been useful for assessing genetic

relationships and tracing the evolution of species with the mainly uni-parental inheritance of organellar genomes allowing the directionality of hybridization events to be determined. The chloroplast genome is especially suitable for constructing molecular phylogenies among plant taxa. Sequence comparisons reveal a low rate of nucleotide substitutions in chloroplast DNA (cpDNA) as a whole, although direct sequence studies reveal rate differences among specific chloroplast genes (Curtis and Clegg, 1984). The noncoding regions of cpDNA display the highest frequency of mutation (Palmer et al., 1988; Clegg and Zurawski, 1992a) and the amplification and sequencing of these regions is especially useful for evolutionary studies. Analysis of noncoding regions of cpDNA could also potentially extend the utility of the molecule at lower taxonomic levels (Curtis and Clegg, 1984; Clegg and Zurawski, 1992b). Mutations in cpDNA are fundamentally of two types: point mutations (single nucleotide pair substitutions) and rearrangements, with several kinds of rearrangements recognized. The most frequent are point mutations and deletions/insertions in noncoding regions. The former are useful for phylogenetic studies at all taxonomic levels (Downie and Palmer, 1992). The rate of change, particularly point-mutational change, is a critical determinant of the level at which a particular genome or DNA sequence is most useful for taxonomic studies (Palmer, 1987).

Sequence differences in the chloroplast genome have been used in molecular systematics (Fennell, 1994), in the determination of hybrids, and in the identification of the maternal progenitor in allopolyploids (Palmer et al., 1983; Erikson et al., 1983; Palmer, 1985a; Smith and Sytsma, 1990). Chloroplast sequences have also been used in the description of new species (Doyle et al., 1990). Gielly and Taberlet (1994) evaluated the use of chloroplast DNA, specifically the trnL-trnF intergenic spacer and the trnL intron, in resolving plant phylogenies, concluding that it reflected clearly the phylogeny of species in the genus *Gentiana L.* as well as providing enough information to separate the genera *Hordeum*, *Triticum* and *Aegilops*.

In contrast, plant mitochondrial DNA (mtDNA) has a lower rate of nucleotide substitution than cpDNA, though it undergoes frequent rearrangements (Palmer and Shields, 1984). Mitochondrial DNA has been used widely in animals to determine phylogeny, since it evolves relatively fast (5 times more than nuclear sequence) and is maternally inherited (Xiong and Kocher, 1991; Hillis et al., 1990; Janczewski et al., 1992). In plants, mtDNA evolves more slowly than nuclear and chloroplast sequences but has nevertheless been used in studying evolutionary relationships (Ishii et al., 1993; Barret et al., 1992). Since the silent substitution rate is significantly lower in plant mtDNA than in any other plant or animal genome

(Sederoff, 1987; Wolfe et al., 1987), comparative sequencing of plant mitochondrial genes may play a more important role in unravelling deep branches of plant phylogeny.

The nuclear genome contrasts strikingly with the chloroplast and animal mitochondrial genomes. Its immense size and complexity, as well as the presence of orthologous (resulting from speciation) and paralogous (resulting from gene duplication) copies of most nuclear genes complicates the use of nuclear sequences in phylogenetic studies (Olmstead and Palmer, 1994). Nevertheless, in analysing phylogenetic relationships, nuclear DNA has been used widely (Sytsma and Schaal, 1985; Gepts and Clegg, 1989; Delseny et al., 1990; Furnier et al., 1990). In plants, due mainly to high substitution rates, nuclear sequences tend to be the most variable. However, nuclear DNA is inherited from two different progenitors and undergoes meiotic recombination which potentially complicates the sequence comparisons. Barbier et al. (1991) suggest that sequences with less probability of crossing over, such as those within or near single copy genes, are appropriate nuclear targets. Using this approach Barbier et al. (1991) were able to differentiate between perennial and annual rice. Nuclear DNA has also been used to study *Populus* species (Barret et al., 1992) and to determine differences between cultivated and wild species of pearl millet (Gepts and

Clegg, 1989).

Comparison of specific DNA sequences provides the highest resolution for assessing genetic and evolutionary relationships. DNA sequence data has the advantage of being independent of other biological characters, in the sense that no assumptions about relationships are necessary to infer phylogenies from sequence data. This means that molecular phylogenies can be contrasted to conventionally derived phylogenies and patterns of character evolution can be examined within the context of molecular trees (Clegg and Zurawski, 1992b). Despite its powerful resolution, this approach has only recently been widely adopted, due to the development of PCR. This technique allows the amplification of sequences from the individuals of interest, using consensus primers, followed by cloning or direct sequencing of double stranded amplified products.

In this chapter two chloroplast non-coding regions selected by Taberlet et al. (1991) have been used to examine molecular phylogeny in *Coffea* spp. Variation in mitochondrial DNA in the V7 region of srRNA gene has also been examined (used previously by Russell (1994)). To provide complementary information, a nuclear sequence derived from an anonymous PstI clone has also been used. The primary objective was to investigate phylogenetic relationships within the genus *Coffea* and evolution of *C.*

arabica.

5.2 MATERIALS AND METHODS

5.2.1 PLANT MATERIAL

The plant material was the same as that used for RAPD analysis, which is described in Table 2.2, with the exceptions indicated in Chapter 2 (2.1). In the nuclear sequence analysis, N39 was not included, on the assumption that most of the *C. arabica* accessions would have the same sequence pattern. Four accessions representing four different genera in the *Rubiaceae* family were used as outgroup species. These were *Vangueria edulis*, *Gardenia thumburgia*, *Mitrostigma axilare* and *Ixora findlaysoniana*.

5.2.2 MOLECULAR PROCEDURES AND SEQUENCE DATA ANALYSIS

The specific chloroplast primers used were identical to those used by Taberlet et al. (1991) to amplify the trnL exon-trnF intergenic spacer region and the trnL(UUA) intron. The mitochondrial primers used were designed by Russell (1994) to amplify the V7 region of plant mitochondrial ribosomal DNA. Primers for nuclear DNA amplification were designed from one of six sequenced clones from a coffee PstI Library. All primers are listed

in Tables 4.4 and 5.1. The SSCP assay, cloning and sequencing of amplified products and analysis of sequence data were as described in Chapter 2.

Table 5.1. Sequences of primers used to amplify and sequence specific region of the coffee nuclear genome.

Name	Sequence	Amplified region
COC1 (forward)	5'TGGGAGAAGGAGACTGAA3'	anonymous nuclear
COC2 (reverse)	5'CGAAAGCCAAATAGAAAA3'	anonymous nuclear

5.3 RESULTS

5.3.1 DETECTION OF POLYMORPHISM USING SINGLE STRANDED CONFORMATIONAL POLYMORPHISM (SSCP)

After amplification of the chloroplast trnL-trnF intergenic spacer and trnL intron, and the mitochondrial V7 region, the products were examined by separation on SSCP gels. The intergenic spacer and intron were polymorphic in different species (Figures 5.1 and 5.2 respectively), while the mitochondrial products were monomorphic (Figure 5.3). The presence of SSCP polymorphism indicated that these regions should reveal differences at the sequence level, which could be phylogenetically informative.

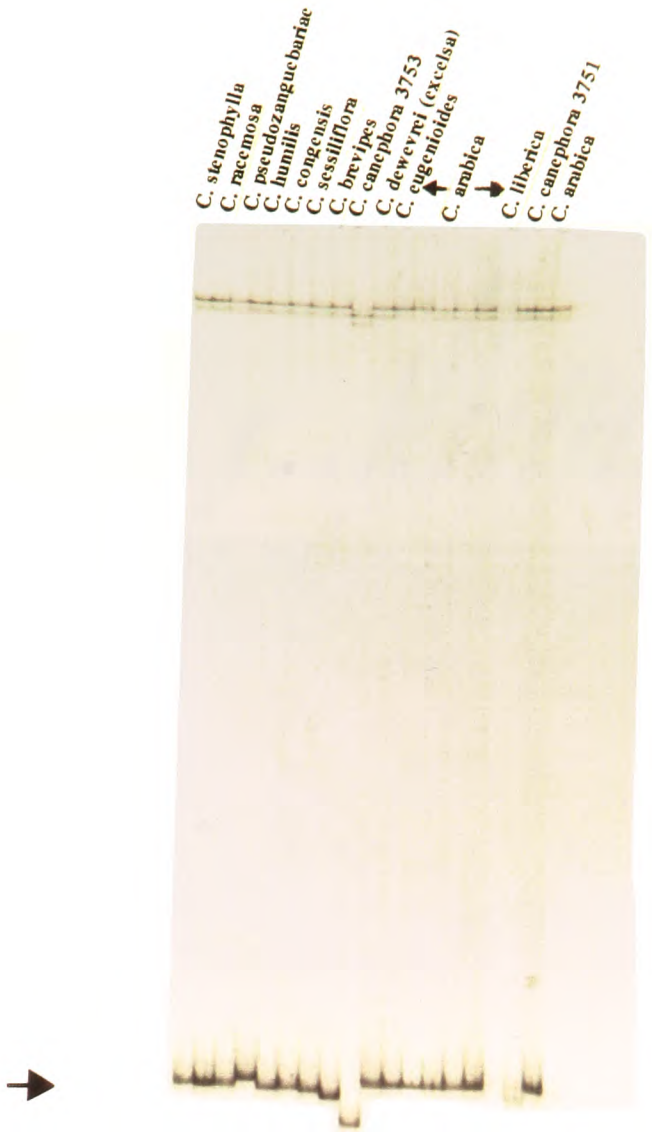


Figure 5.1. Single strand conformational polymorphism (SSCP) of *Coffea* species for the chloroplast trnL-trnF intergenic spacer.

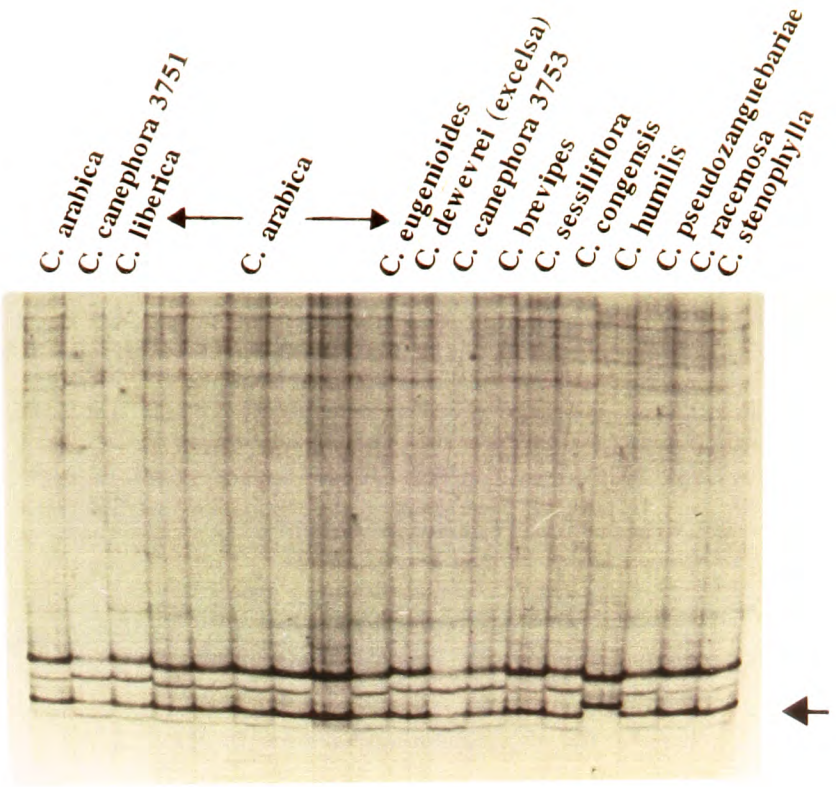


Figure 5.2. Single strand conformational polymorphism (SSCP) of *Coffea* species for the chloroplast trnL-trnF intron.

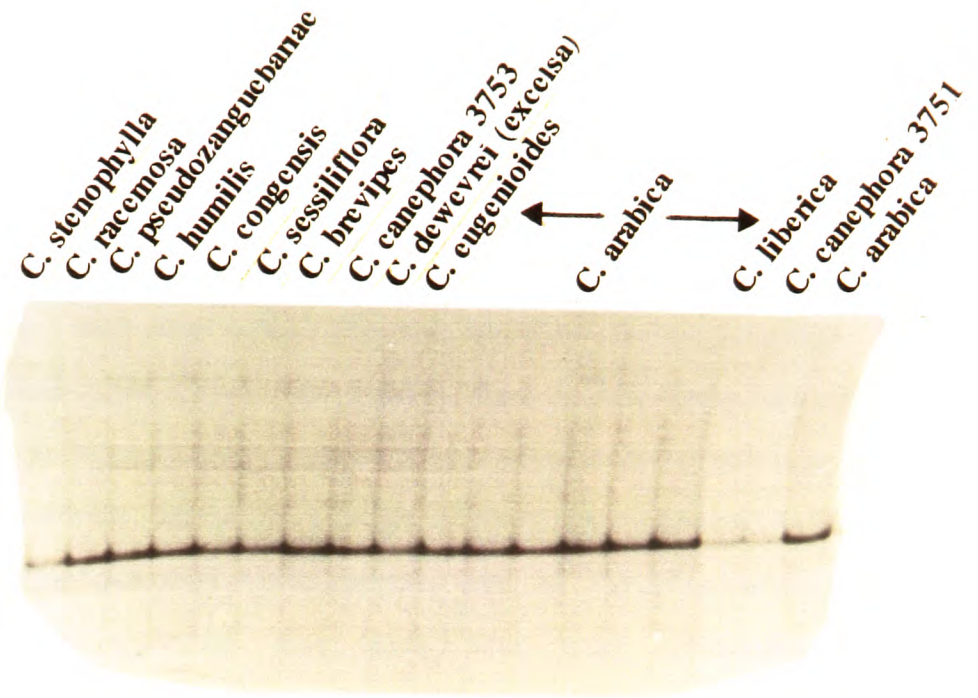


Figure 5.3. Single strand conformational polymorphism (SSCP) of *Coffea* species for the V7 small ribosomal RNA mitochondrial region.

5.3.2 SEQUENCE ANALYSIS OF THE CHLOROPLAST trnL-trnF INTERGENIC SPACER AND trnL INTRON REGIONS

DNA fragments from 18 accessions representing 12 species were cloned and sequenced. For the trnL-trnF intergenic spacer four additional species classified in the family *Rubiaceae* were also sequenced to compare the genetic distances within the genus *Coffea* to those in a different genus. For the trnL intron only *I. findlaysoniana* was sequenced and used as an outgroup. Tables 5.2 and 5.3 give the size of the amplified chloroplast trnL-trnF intergenic spacer and the trnL intron regions respectively for all the genotypes studied. There were no differences between *C. arabica* accessions, therefore, only one accession from this species was included in the subsequent analysis. Figure 5.4 show the complete alignment for 13 coffee accessions and outgroup species for the trnL-trnF intergenic spacer region. The overall rate of nucleotide substitutions was higher than the rate of insertion deletions (16 nucleotide substitutions against 1 insertion-deletion). The transition/transversion rate was of the order of 1.2 (see Table 5.4). Figure 5.5 shows the complete alignment for coffee accessions and the outgroup species for the trnL intron. As in the trnL-trnF intergenic spacer, the percent of nucleotide substitution rate in the trnL intron was higher than that of insertion/deletions (13 substitutions

against 1 insertion/deletion). The transition/transversion rate was 2.2 (Table 5.4).

Table 5.2 Fragment size (bp) for the chloroplast trnL-trnF intergenic spacer in coffee and outgroup species.

GENOTYPES	FRAGMENT SIZE (bp)
1. *C. brevipes*	394
2. *C. canephora* 3753	394
3. *C. canephora* 3751	394
4. *C. arabica* (five accessions)	394
5. *C. congensis*	394
6. *C. eugenioides*	394
7. *C. dewevrei* (exelsa)	383
8. *G. thumburgia*	385
9. *C. humilis*	394
10. *I. findlaysoniana*	390
11. *C. liberica*	394
12. *M. axillare*	391
13. *C. racemosa*	394
14. *C. pseudozanguebariae*	394
15. *C. sessiliflora*	394
16. *C. stenophylla*	394
17. *V. edulis*	371

Table 5.3 Fragment size (bp) for the chloroplast trnL intron in *Coffea* and outgroup species.

GENOTYPES	FRAGMENT SIZE (bp)
1. *C. brevipes*	531
2. *C. canephora* 3753	531
3. *C. canephora* 3751	531
4. *C. arabica* (5 accessions)	530
5. *C. congensis*	531
6. *C. eugenioides*	530
7. *C. dewevrei* (exelsa)	531
8. *C. humilis*	530
9. *I. findlaysoniana*	525
10. *C. liberica*	531
11. *C. pseudozanguebariae*	530
12. *C. racemosa*	530
13. *C. sessiliflora*	530
14. *C. stenophylla*	530

Fig. 5.4. Multiple alignment of *Coffea* and outgroup species for the trnL-trnF intergenic spacer sequences. Shading shows sequence non-identity.

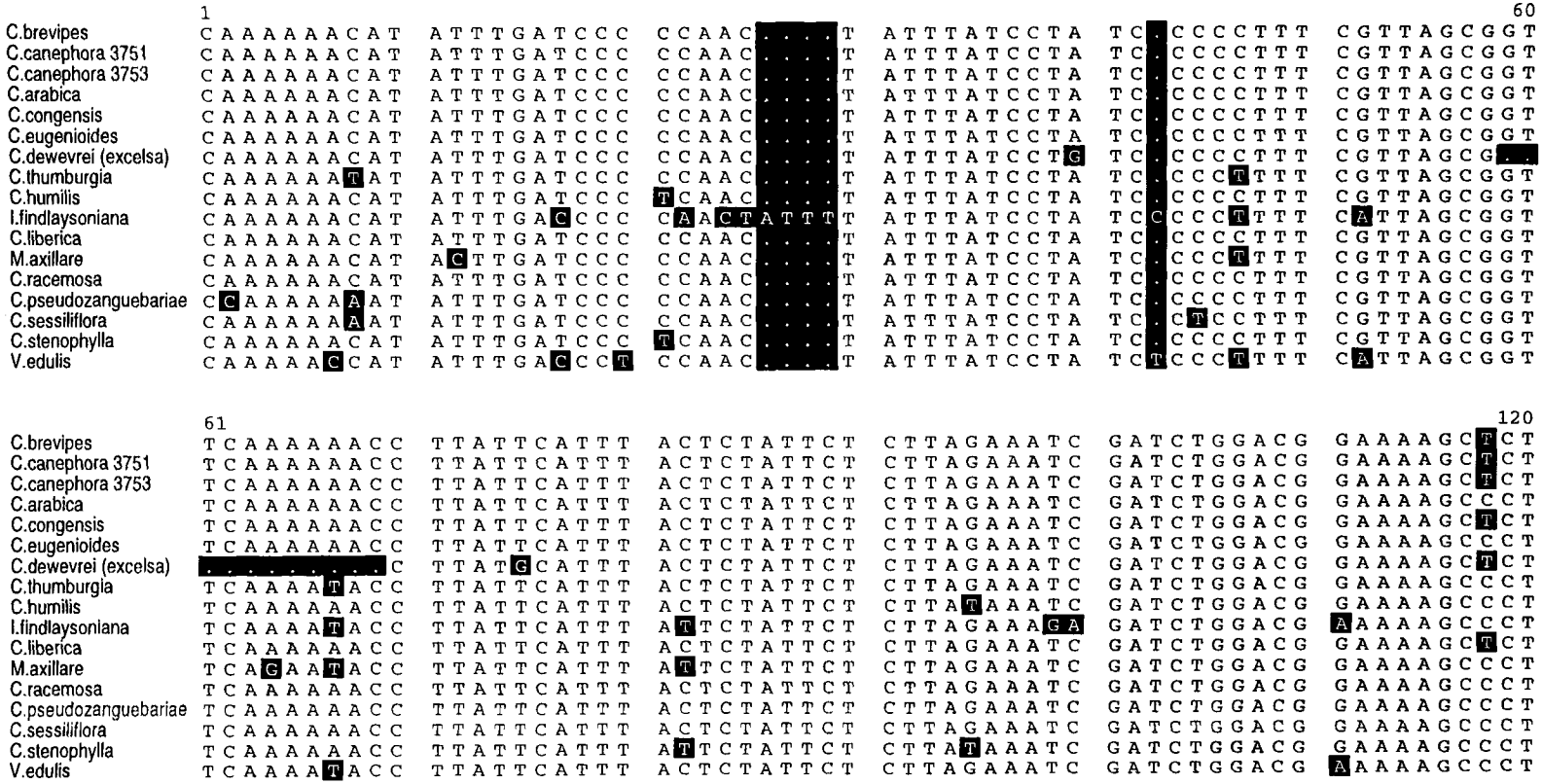


Fig. 5.4. Multiple alignment of *Coffea* and outgroup species for the trnL-trnF intergenic spacer sequences. Shading shows sequence non-identity.

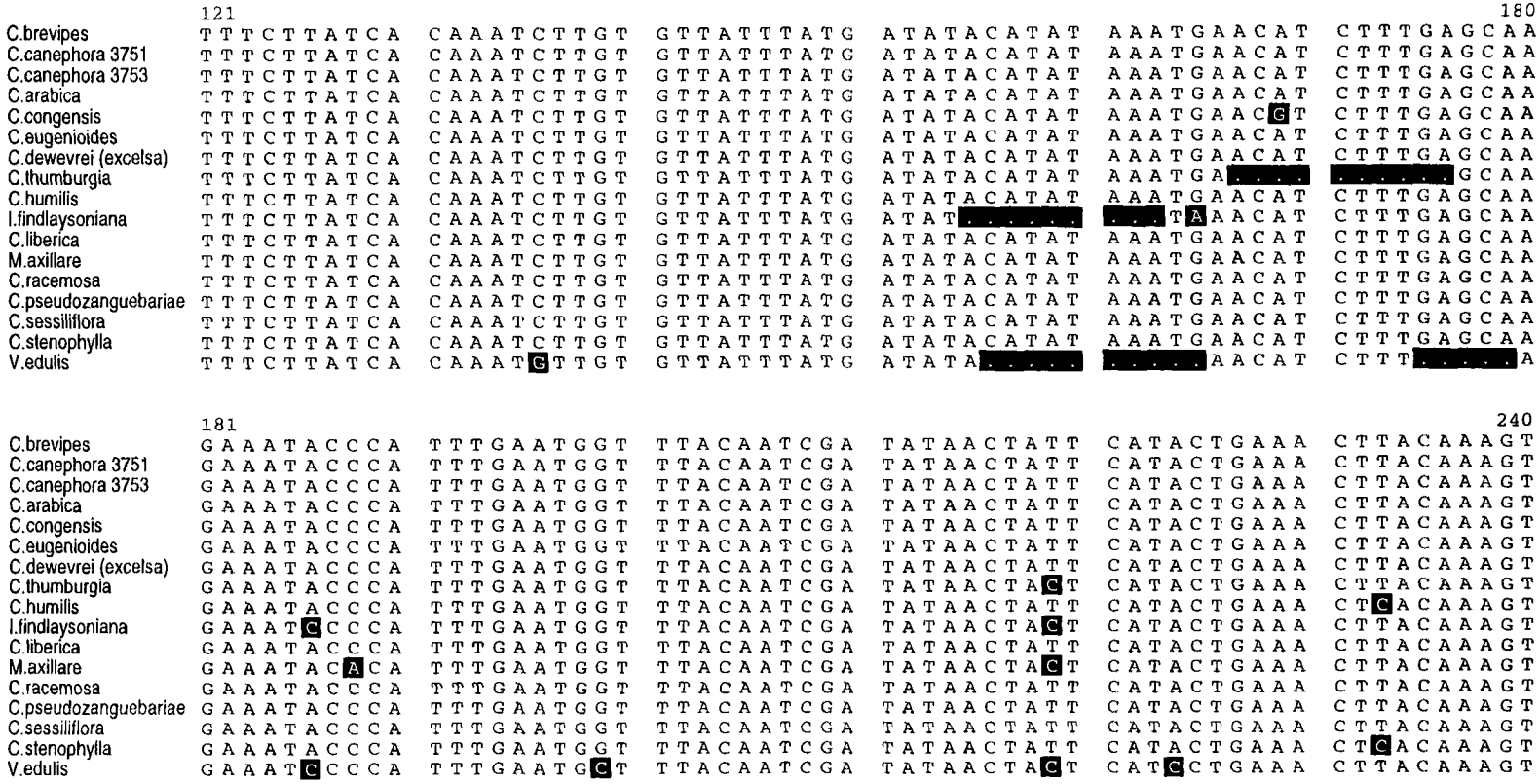


Fig. 5.4. Multiple alignment of *Coffea* and outgroup species for the trnL-trnF intergenic spacer sequences. Shading shows sequence non-identity.

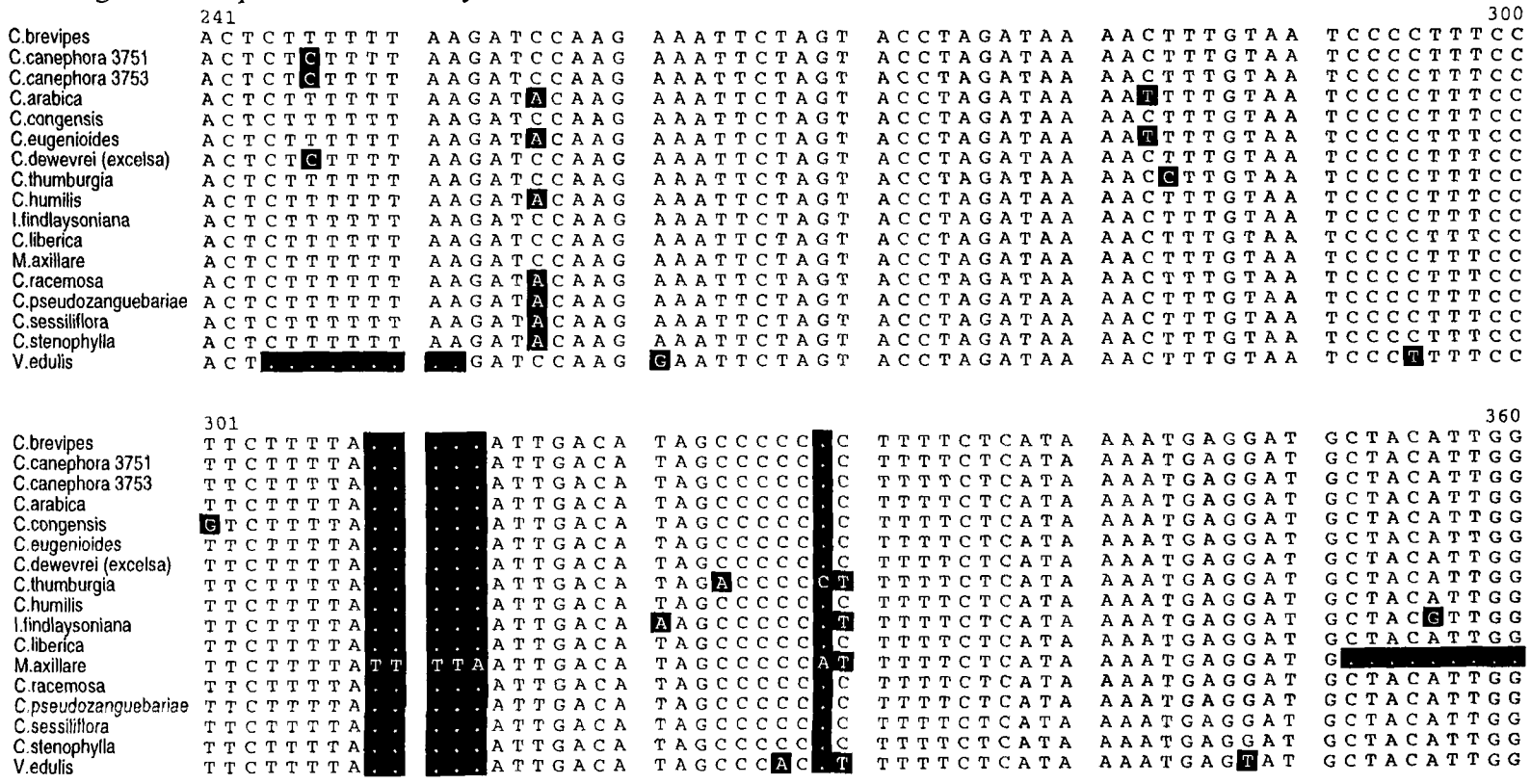


Fig. 5.4. Multiple alignment of *Coffea* and outgroup species for the trnL-trnF intergenic spacer sequences. Shading shows sequence non-identity.

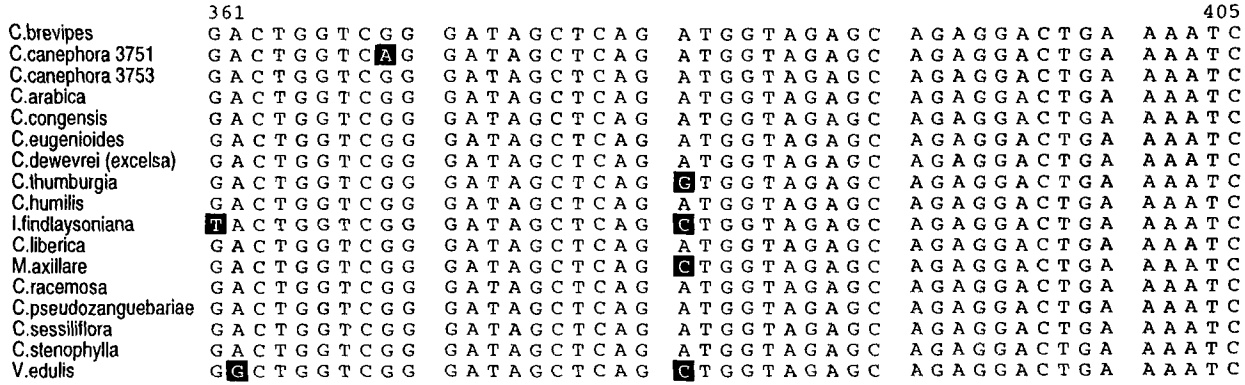


Fig. 5.5. Multiple alignment of *Coffea* spp. and *I. findlaysoniana* for the trnL intron sequences. Shading shows sequence non-identity.

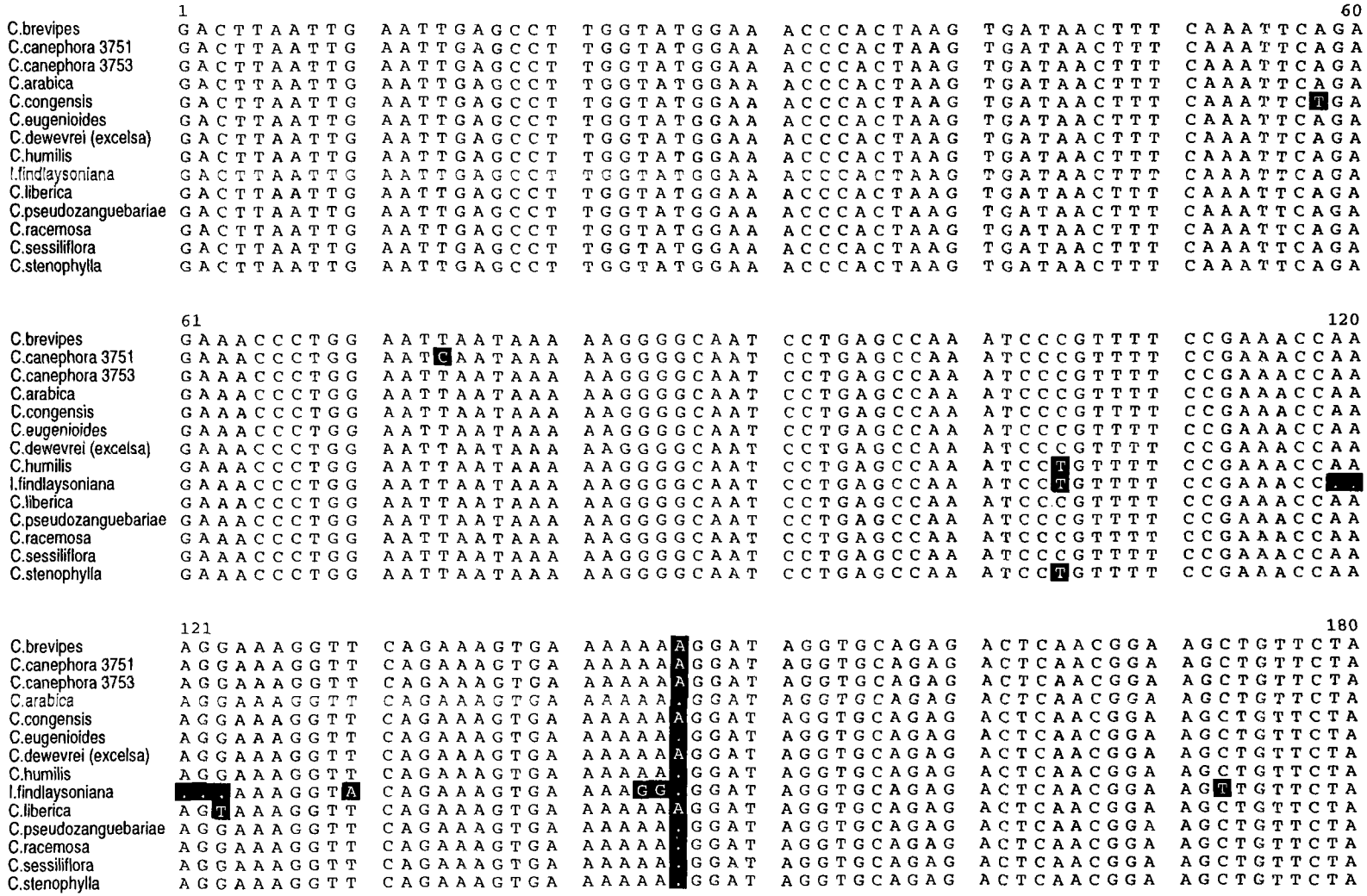


Fig. 5.5. Multiple alignment of *Coffea* spp. and *I. findlaysoniana* for the trnL intron sequences. Shading shows sequence non-identity.

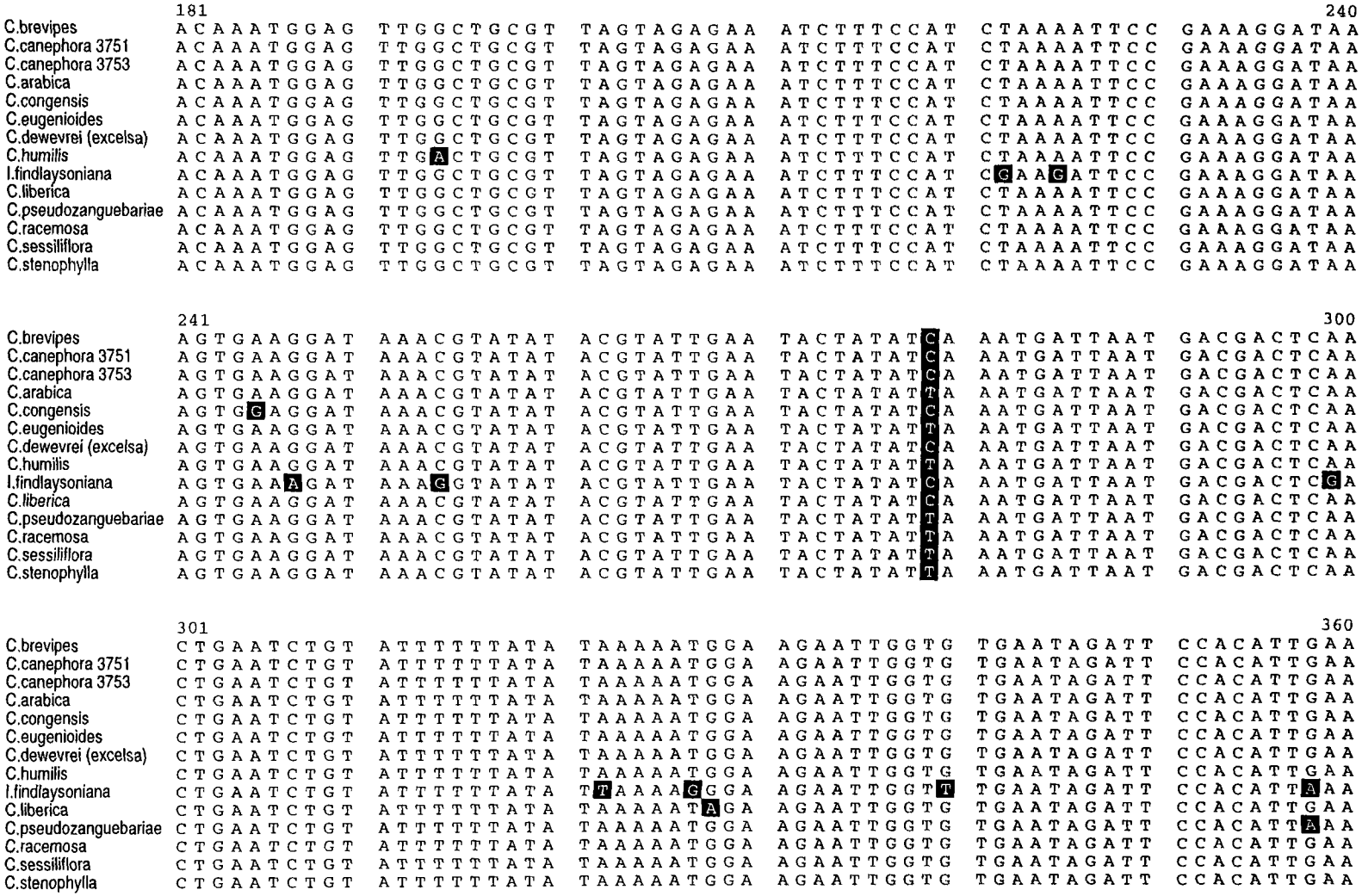


Fig. 5.5. Multiple alignment of *Coffea* spp. and *I. findlaysoniana* for the trnL intron sequences. Shading shows sequence non-identity.

```
                     361                                                                     420
C.brevipes           GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.canephora 3751     GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.canephora 3753     GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.arabica            GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.congensis          GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.eugenioides        GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.dewevrei (excelsa) GAAAGAATCG AATATTCATT GATCAAACGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.humilis            GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
I.findlaysoniana     GAAAGAATCG AATATTCATT GATTAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.liberica           GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGGTAG ATCTTTTCAA
C.pseudozanguebariae GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.racemosa           GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTTTGATAG ATCTTTTCAA
C.sessiliflora       GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA
C.stenophylla        GAAAGAATCG AATATTCATT GATCAAATGA TTCACTCCAT AGTCTGATAG ATCTTTTCAA

                     421                                                                     480
C.brevipes           GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.canephora 3751     GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.canephora 3753     GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.arabica            GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.congensis          GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.eugenioides        GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.dewevrei (excelsa) GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.humilis            GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
I.findlaysoniana     GAATTGATTA ATCGGACGAG AATAAAGATA GAGCCCGGTT CTACATGTCA ATGTCGGCAA
C.liberica           GAATTGATTA ATCAGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA
C.pseudozanguebariae GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT GTACATGTCA ATGTCGGCAA
C.racemosa           GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT GTACATGTCA ATGTCGGCAA
C.sessiliflora       GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT GTACATGTCA ATGTCGGCAA
C.stenophylla        GAATTGATTA ATCGGACGAG AATAAAGATA GAGTCCCGTT CTACATGTCA ATGTCGGCAA

                     481                                                         531
C.brevipes           CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.canephora 3751     CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.canephora 3753     CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.arabica            CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.congensis          CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.eugenioides        CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.dewevrei (excelsa) CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.humilis            CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
I.findlaysoniana     CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.liberica           CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.pseudozanguebariae CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.racemosa           CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.sessiliflora       CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
C.stenophylla        CAATGAAATT TATAGTAAGA GGAAAATCCG TCGACTTTAA AAATCGTGAG G
```

Table 5.4. Transition-transversion rate derived from nucleotide substitutions in chloroplast and nuclear DNA in the genus *Coffea*.

Type of substitution	Intergenic	Intron	Total Chloroplast	Nuclear
Transitions (Ts)				
A--G	3	6	9	7
T--C	7	5	12	9
Total Transition	10	11	21	16
Transversion (Tv) I				
A--T	0	1	1	3
C--G	0	1	1	6
Transversion (Tv) II				
A--C	3	0	3	0
T--G	3	0	3	1
Total transversions	6	2	8	10
Ts/Tv rate	1.2	2.2	1.8	1.6
Ts/Tv interval	(0.6-2.1)	(0.9-4.9)	(0.9-2.8)	(0.7-3.5)

5.3.3 GENETIC DISTANCES AND PHYLOGENETIC TREE DERIVED FROM THE COMBINED SEQUENCE ANALYSIS OF THE CHLOROPLAST trnL-trnF INTERGENIC SPACER AND THE trnL INTRON

Sequence analysis of both chloroplast regions generated broadly similar results (data in appendix), thus, their sequences were placed together to estimate a combined genetic distance and phylogenetic tree (*I. findlaysoniana* was used as an outgroup species). A total average sequence divergence of 3.7 % was found among *Coffea* species. Nucleotide substitutions and addition/deletions appeared to be randomly distributed (see alignments Figures 5.4 and 5.5) and the transition/transversion occurrence and rate are summarised in Table 5.4.

Pairwise genetic distances based on the combined data are shown in the form of a matrix in Table 5.5. Low genetic distance values are indicative of the high similarity between the compared accessions. There was no differences found between *C. arabica and C. eugenioides*. The highest overall distance is between the outgroup species *I. findlaysoniana* and *C. congensis* (0.0456). *C. liberica* and *C. dewevrei* (excelsa), which have been previously considered a single species, exhibit large genetic distances values (0.0088). Two *C. canephora* accessions can also be distinguished (0.0022).

Table 5.5 Pairwise distance matrix for the all chloroplast combined data (trnL-trnF intergenic spacer and trnL intron) sequences. (Using Kimura's 2-parameter distance).

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. C. brevipes													
2. C. canephora 3751	0.0032												
3. C. canephora 3753	0.0011	0.0022											
4. C. arabica	0.0043	0.0076	0.0054										
5. C. congensis	0.0043	0.0076	0.0054	0.0087									
6. C. eugenioides	0.0043	0.0076	0.0054	0.0000	0.0087								
7. C. dewevrei	0.0044	0.0055	0.0033	0.0088	0.0088	0.0088							
8. C.humilis	0.0087	0.0120	0.0098	0.0065	0.0131	0.0065	0.0133						
9. I. findlaysoniana	0.0409	0.0444	0.0420	0.0432	0.0456	0.0432	0.0449	0.0456					
10. C. liberica	0.0043	0.0076	0.0054	0.0087	0.0087	0.0087	0.0088	0.0131	0.0444				
11. C. pseudozanguebariae	0.0076	0.0109	0.0087	0.0054	0.0120	0.0054	0.0122	0.0098	0.0445	0.0120			
12. C. racemosa	0.0054	0.0087	0.0065	0.0033	0.0098	0.0033	0.0099	0.0076	0.0445	0.0098	0.0043		
13. C. sessiliflora	0.0065	0.0098	0.0076	0.0043	0.0109	0.0043	0.0111	0.0087	0.0457	0.0109	0.0033	0.0033	
14. C. stenophylla	0.0087	0.0120	0.0098	0.0065	0.0131	0.0065	0.0133	0.0022	0.0433	0.0131	0.0098	0.0076	0.0087

A representative tree based on the combined data is given in Figure 5.6. The tree topology separated the *Coffea* species into the four main clades: Clade A, comprising *C. arabica* and *C. eugenioides*; Clade B, comprising *C. racemosa*, *C. pseudozanguebariae* and *C. sessiliflora*; Clade C, comprising *C. stenophylla* and *C. humilis*; and Clade D, comprising *C. canephora*, *C. liberica*, *C. dewevrei* (excelsa), *C. congensis* and *C. brevipes*. Clade C (*C. stenophylla* and *C. humillis*) is clearly different from the others (93 % bootstrap). The separation of the other clades are also well supported (> 58 % bootstrap). The outgroup species *Ixora* was significantly different from the *Coffea* species (100 % bootstrap). Figure 5.7 shows the association between the clades in the phylogenetic tree and the distribution of the species in the African continent.

As an alternative approach to illustrate graphically the differences between these species, principal coordinate analysis was used (Figure 5.8). The second and the third components which reflected the best separation of the species account for 32 % of the variation. The results support the separation of species displayed in the tree topology (Figure 5.6) in separating the coffee accessions in four main groups. The second axis discriminates between the groups A, B, and D, and the third between the group C and the remainder.

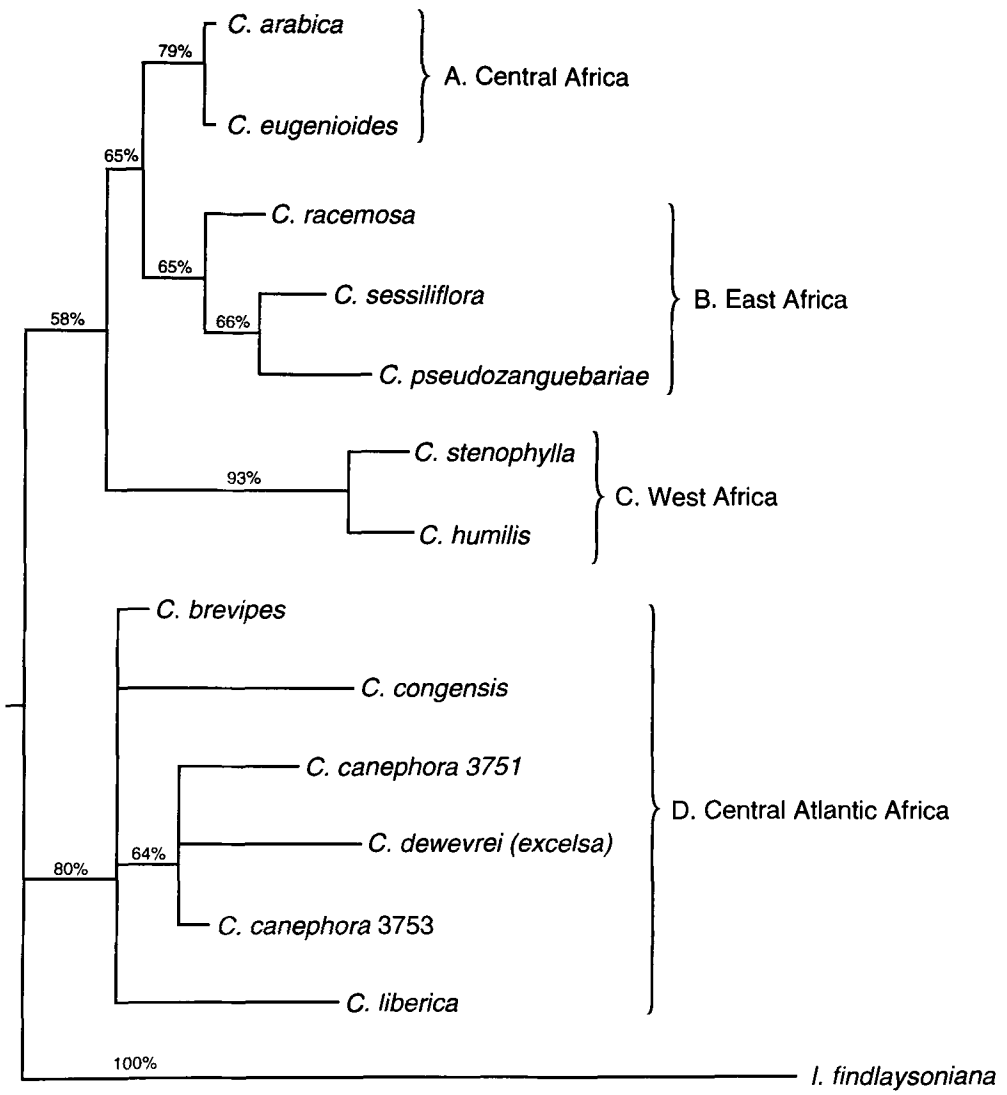


Figure 5.6. Phylogenetic tree (drawgram) of all *Coffea* species and *I. findlaysoniana* for the combined data (chloroplast trnL-trnF intergenic spacer and trnL intron) sequences. Origin of species is also indicated.

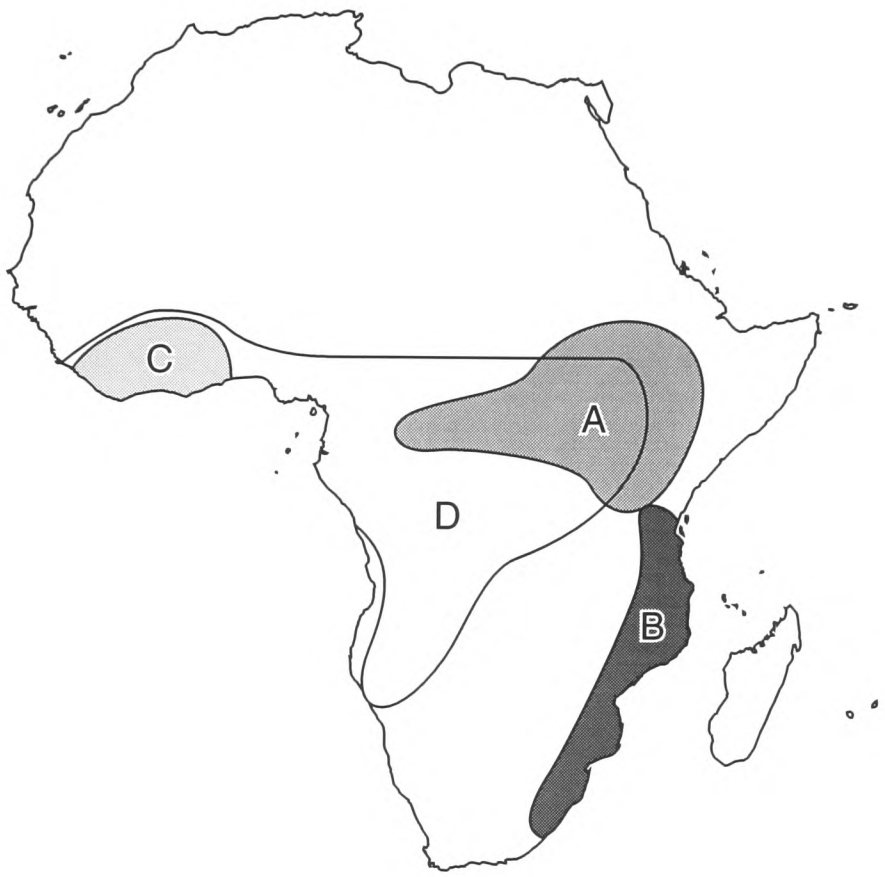


Figure 5.7. Distribution of *Coffea* species in Africa, belonging to the four clades displayed in the phylogenetic tree derived from chloroplast genomic sequences.

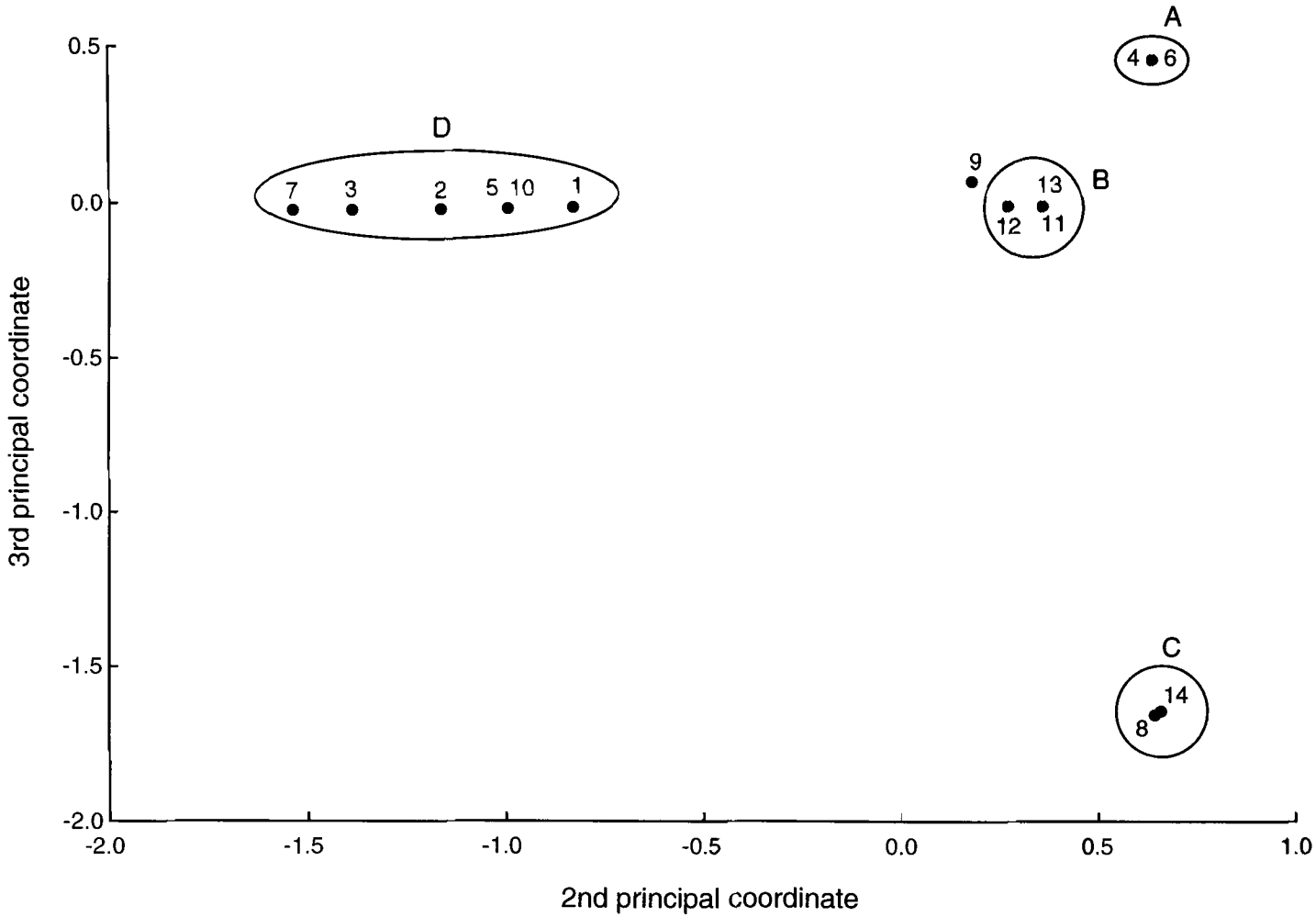


Figure 5.8. Principal coordinate analysis of 14 coffee accessions representing 12 species and one outgroup species for the combined data sequences (trnL-trnF intergenic spacer and trnL intron). The numbers in the figure correspond to the order number in Table 5.3.

5.3.4 SEQUENCE ANALYSIS OF THE MITOCHONDRIAL srRNA V7 REGION

Even though the SSCP analysis revealed no differences in the mt srRNA V7 region, the size of the fragments analyzed were too large to effectively reveal point mutations or minor differences. In an attempt to reveal deeper branches in coffee phylogeny it was therefore decided to carry out sequence analysis on this region as well. Figure 5.9 shows the sequence of the mitochondrial srRNA V7 region for *C. arabica*, var. Typica. Unfortunately, no differences were found among species in this region.

CTTTCATACA GGTGCTGCAT GGCTGTCGTC AGCTCGTGTC GTGAGATGTT
TGGTCAAGTC CTATAACGAG CGAAACCCTC GTTTTGTGTT GCTGAGACAT
GCGCCTAAGG ATAAAGTCTT TGCAACCGAA GTGAGCCGAG GAGCCGAGTG
ACGTGCCAGC GCTACTAATT GAGTGCCAGC ACGTAGCTGT GCTGTCAGTA
AGAAGGTAGC CGGCGCCTTT CGAAGCACTT TCTAGTTAGC GCTTTAGTTT
GATTGCAGCT AGCGCGCTTG ACTAATAAGA AGACTCGGCA TTCAGGCGAG
CCGCCCGGTG GTGTGGTATG TAGTGGGTTT AGTACGCCCC GCCAAAAGGG
CTCCGAAACA AAGAAAAAGG TGCGTGCCGC ACTCACGAGG GACTGCCAGT
GA 402

Figure 5.9. Complete sequence of the V7 small ribosomal RNA mitochondrial region of *C. arabica*, variety Typica. Primers are underlined.

5.3.5 DESIGN OF PRIMERS FOR AMPLIFICATION OF SPECIFIC NUCLEAR SEQUENCES

Six clones were taken from a coffee DNA Pst I library and their DNA sequence was determined. One of the sequences was used to design a pair of primers (Table 5.1), to amplify the corresponding nuclear DNA region in the various accessions and species.

5.3.6 SEQUENCE ANALYSIS OF THE NUCLEAR DNA REGION

DNA fragments from 17 accessions representing 12 species were amplified, cloned and sequenced. Because of potential heterozygosity, four individual clones from each accession were sequenced. Table 5.6 gives the size of the nuclear specific DNA fragment for all the genotypes studied. These ranged from 419 to 424 bases. *C. racemosa* contained a deletion of 61 bases, with a fragment size of 359.

Figure 5.10 shows the complete alignment of the 17 sequences. The rate of nucleotide substitution was higher than the rate of insertion/deletion (23 nucleotide substitutions against eight insertion/deletion) and the transition/transversion rate was 1.6 (see Table 5.4).

Table 5.6 Fragment size (bp) for the nuclear region in *Coffea* species.

GENOTYPES	FRAGMENT SIZE (bp)
1. *C. brevipes*	421
2. *C. canephora* I	421
3. *C. canephora* II	421
4. *C. arabica I*	421
5. *C. arabica* II	420
6. *C. congensis I*	419
7. *C. congensis* II	420
8. *C. eugenioides*	420
9. *C. dewevrei* (exelsa) I	419
10. *C. dewevrei* (exelsa) II	419
11. *C. humilis*	421
12. *C. liberica*	419
13. *C. pseudozanguebariae*	424
14. *C. racemosa*	359
15. *C. sessiliflora*	419
16. *C. stenophylla* I	421
17. *C. stenophylla* II	424

Fig. 5.10. Multiple alignment of *Coffea* species for the nuclear sequences, two different sequences for the same species means two distinct alleles. Shading shows sequence non-identity.

	1					60
C.brevipes	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.canephora I	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.canephora II	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.arabica I	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.arabica II	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.congensis I	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGACTGGG	GAAAGGCTCG	GAGAAGGTTC
C.congensis II	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.eugenioides	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.dewevrei(exc) I	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.dewevrei(exc) II	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.humilis	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.liberica	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.pseudozanguebа	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.racemosa	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.sessiliflora	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAGGGCTCG	GAGAAGGTTC
C.stenophylla I	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC
C.stenophylla II	TGGGAGAAGG	AGACTGAAGC	GGACCGATTT	CTTGATTGGG	GAAAGGCTCG	GAGAAGGTTC

	61					120
C.brevipes	CTTCGGTGTT	GTTTACTGTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.canephora I	CTTCGGTGTT	GTTTACTGTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.canephora II	CTTCGGTGTT	GTTTACTCTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.arabica I	CTTCGGTGTT	GTTTACTGTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.arabica II	CTTCGGTGTT	GTTTACTCTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.congensis I	CTTCGGTGTT	GTTTACTGTG	GTCTCGTTGT	TCCTAAGAAC	GTGCCAGTGG	ATGCTGACAG
C.congensis II	CTTCGGTGTT	GTTTACTGTG	GTCTCGTTGT	TCCTAAGAAC	GTGCCAGTGG	ATGCTGACAG
C.eugenioides	CTTCGGTGTT	GTTTACTCTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.dewevrei(exc) I	CTTTGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.dewevrei(exc) II	CTTTGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.humilis	CTTCGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.liberica	CTTTGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.pseudozanguebа	CTTCGGTGTT	GTTTACTCTG	GTCTCGTTGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.racemosa	CTTCGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.sessiliflora	CTTCGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.stenophylla I	CTTCGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG
C.stenophylla II	CTTCGGTGTT	GTTTACTCTG	GTCTCGTCGT	TCCTAAGAAC	GTGTCAGTGG	ATGCTGACAG

Fig. 5.10. Multiple alignment of *Coffea* species for the nuclear sequences, two different sequences for the same species means two distinct alleles. Shading shows sequence non-identity.

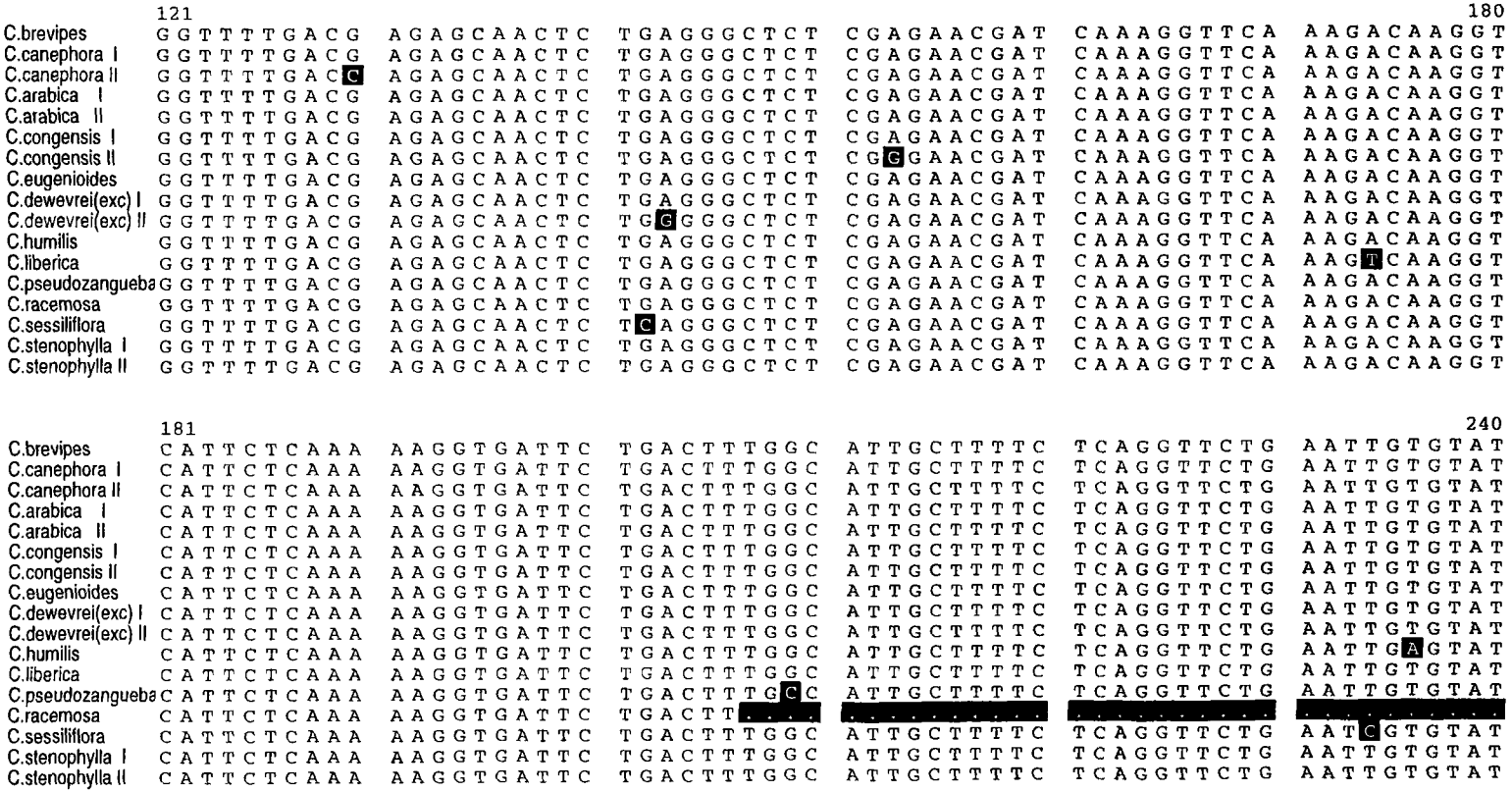


Fig. 5.10. Multiple alignment of *Coffea* species for the nuclear sequences, two different sequences for the same species means two distinct alleles. Shading shows sequence non-identity.

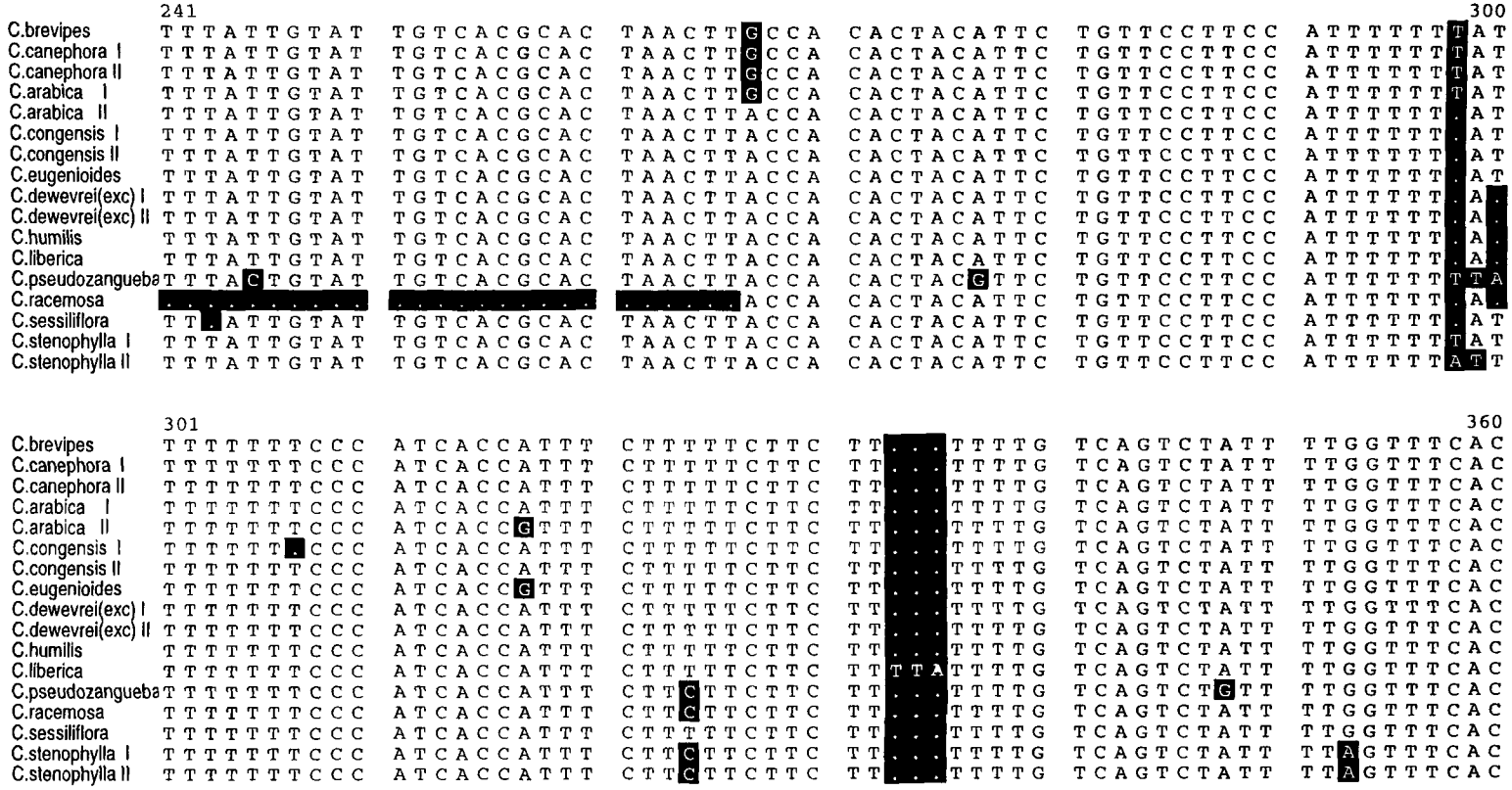


Fig. 5.10. Multiple alignment of *Coffea* species for the nuclear sequences, two different sequences for the same species means two distinct alleles. Shading shows sequence non-identity.

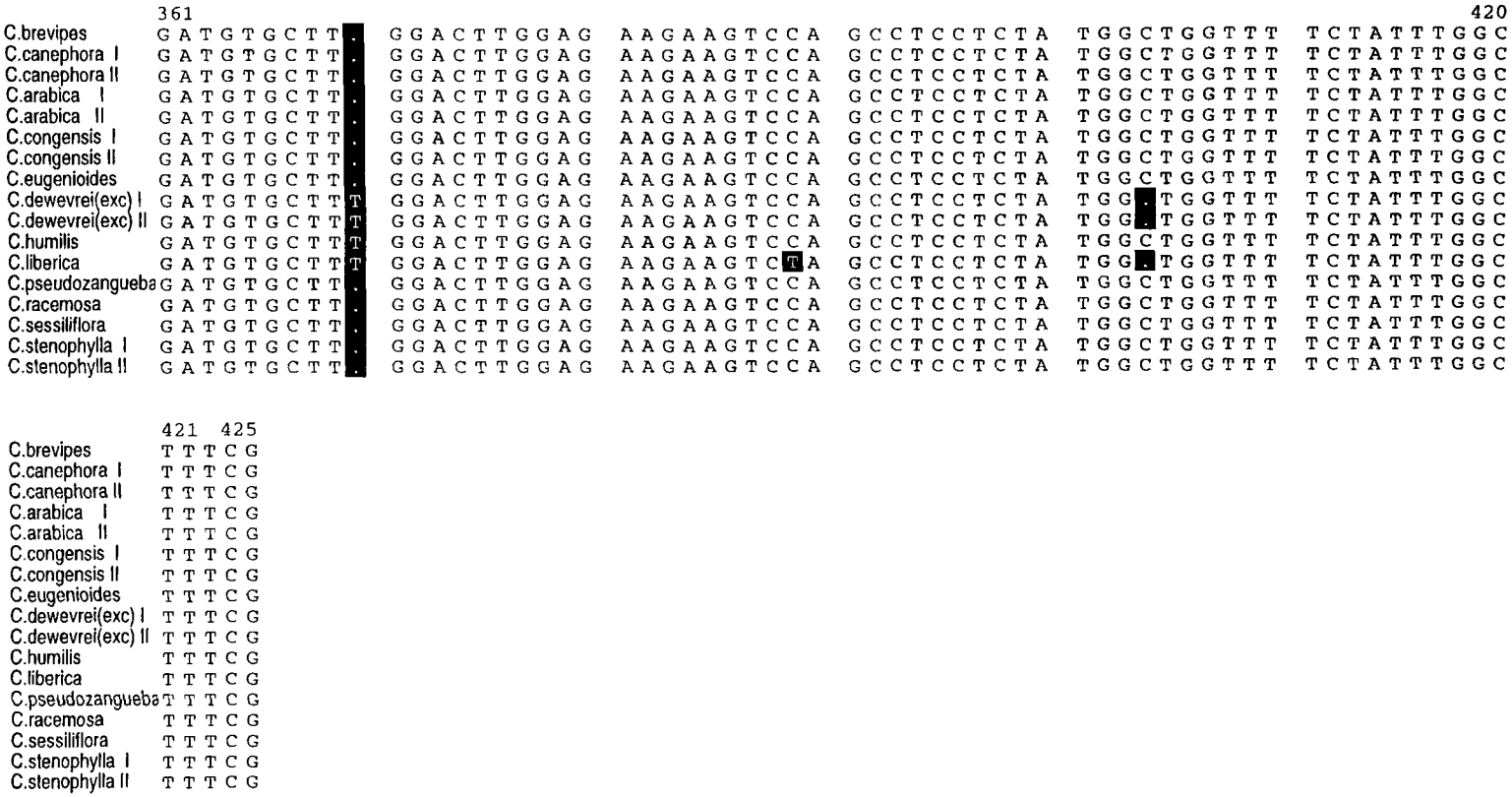
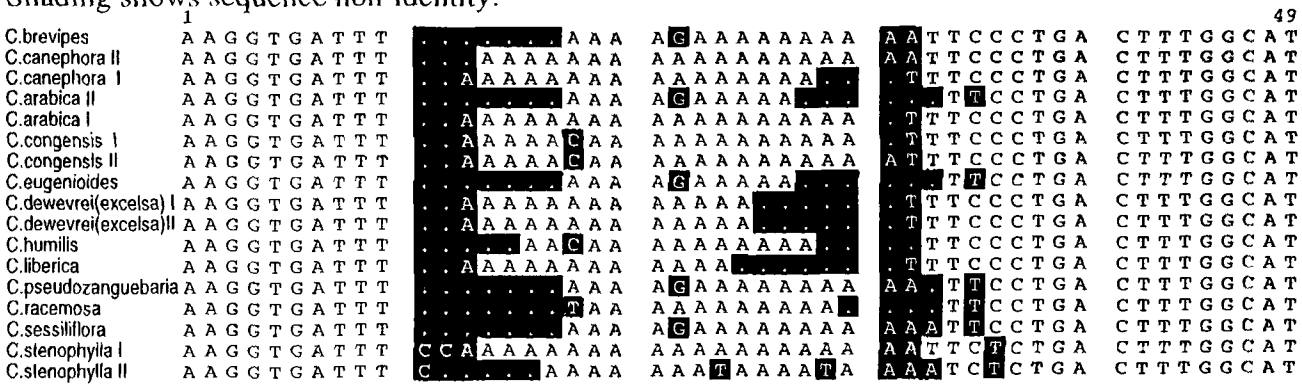


Fig. 5.11. Multiple alignment of *Coffea* spp. for the nuclear microsatellite region sequences. Shading shows sequence non-identity.



Two different sequences were obtained for each of *C. arabica*, *C. congensis*, *C. dewevrei* (excelsa), *C. stenophylla* and *C. canephora* accessions. A mononucleotide microsatellite repeat was also found in the accessions sequenced. An alignment of the microsatellite region is shown in Figure 5.11. Even though this small region was highly variable, it was excluded from the phylogenetic analysis because of the potential for Taq polymerase slippage during amplification leading to inflated genetic differences, however, the sequence of the microsatellite region tended to show similar separation of species to that of the major nuclear sequence region (data not shown), particularly for the separation of two sequences in each of *C. arabica*, *C. congensis*, *C. stenophylla* and *C. canephora* accessions.

Two points are particularly significant. First, the *C. canephora* accession I, *C. brevipes* and one of the *C. arabica* sequences are identical. Second, the other *C. arabica* sequence and that from *C. eugenioides* are also identical. The two *C. arabica* sequences differ in four bases (3 substitutions and 1 deletion). The two *C. canephora* sequences differ in two positions (both substitutions), the two *C. congensis* sequences are distinct in three positions (two substitutions and one deletion), the two *C. stenophylla* sequences vary in five positions (two substitutions and three deletions), and the two *C.*

dewevrei (excelsa) in 1 base substitution. *C. liberica* differs from *C. dewevrei* (excelsa) I and II genotypes in two and three base substitutions respectively.

5.3.7 GENETIC DISTANCES AND PHYLOGENETIC TREE OF COFFEE SPECIES BASED ON NUCLEAR SEQUENCES

Genetic distances between pairs of accessions are indicated in Table 5.7. A phylogenetic tree derived from the genetic distances is given in Figure 5.12. Only bootstraps higher than 50 % are shown in the tree. As indicated (Figure 5.10), *C. arabica* exhibited two different alleles which are separated into two different groups, *C. canephora* also had two different alleles, which fell in the same clade (Figure 5.12). Similar observations can also be made for *C. congensis*, *C. stenophylla* and *C. dewevrei* (excelsa). Even though a number of clones were also sequenced for each of the other coffee species, only one allele was revealed. It is important to notice that one sequence of *C. arabica* is in the same clade as *C. canephora* and *C. brevipes*, and a different sequence of *C. arabica* shares the same clade with *C. eugenioides* (Figure 5.12).

Table 5.7 Pairwise distance matrix for the coffee nuclear sequences. (Using Kimura's 2-parameter distance).

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
1. C. brevipes																
2. C. canephora I	0.0000															
3. C. canephora II	0.0048	0.0048														
4. C. arabica I	0.0000	0.0000	0.0048													
5. C. arabica II	0.0072	0.0072	0.0072	0.0072												
6. C. congensis I	0.0072	0.0072	0.0120	0.0072	0.0096											
7. C. congensis II	0.0072	0.0072	0.0120	0.0072	0.0096	0.0048										
8. C. eugenioides	0.0072	0.0072	0.0072	0.0072	0.0000	0.0096	0.0096									
9. C. dewevrei I	0.0096	0.0096	0.0096	0.0096	0.0072	0.0121	0.0121	0.0072								
10. C. dewevrei II	0.0121	0.0121	0.0121	0.0121	0.0096	0.0145	0.0145	0.0096	0.0024							
11. C. humilis	0.0096	0.0096	0.0096	0.0096	0.0072	0.0121	0.0120	0.0072	0.0048	0.0072						
12. C. liberica	0.0145	0.0145	0.0145	0.0145	0.0121	0.0170	0.0170	0.0121	0.0048	0.0072	0.0096					
13. C. pseudozanguebariae	0.0218	0.0218	0.0218	0.0218	0.0193	0.0243	0.0243	0.0193	0.0194	0.0218	0.0193	0.0244				
14. C. racemosa	0.0112	0.0112	0.0112	0.0112	0.0084	0.0141	0.0141	0.0084	0.0056	0.0084	0.0028	0.0113	0.0112			
15. C. sessiliflora	0.0145	0.0145	0.0145	0.0145	0.0120	0.0170	0.0169	0.0120	0.0097	0.0121	0.0096	0.0145	0.0268	0.0084		
16. C. stenophylla I	0.0120	0.0120	0.0120	0.0120	0.0096	0.0145	0.0144	0.0096	0.0072	0.0096	0.0072	0.0121	0.0193	0.0028	0.0120	
17. C. stenophylla II	0.0169	0.0169	0.0169	0.0169	0.0120	0.0169	0.0169	0.0120	0.0096	0.0121	0.0096	0.0145	0.0192	0.0056	0.0145	0.0048

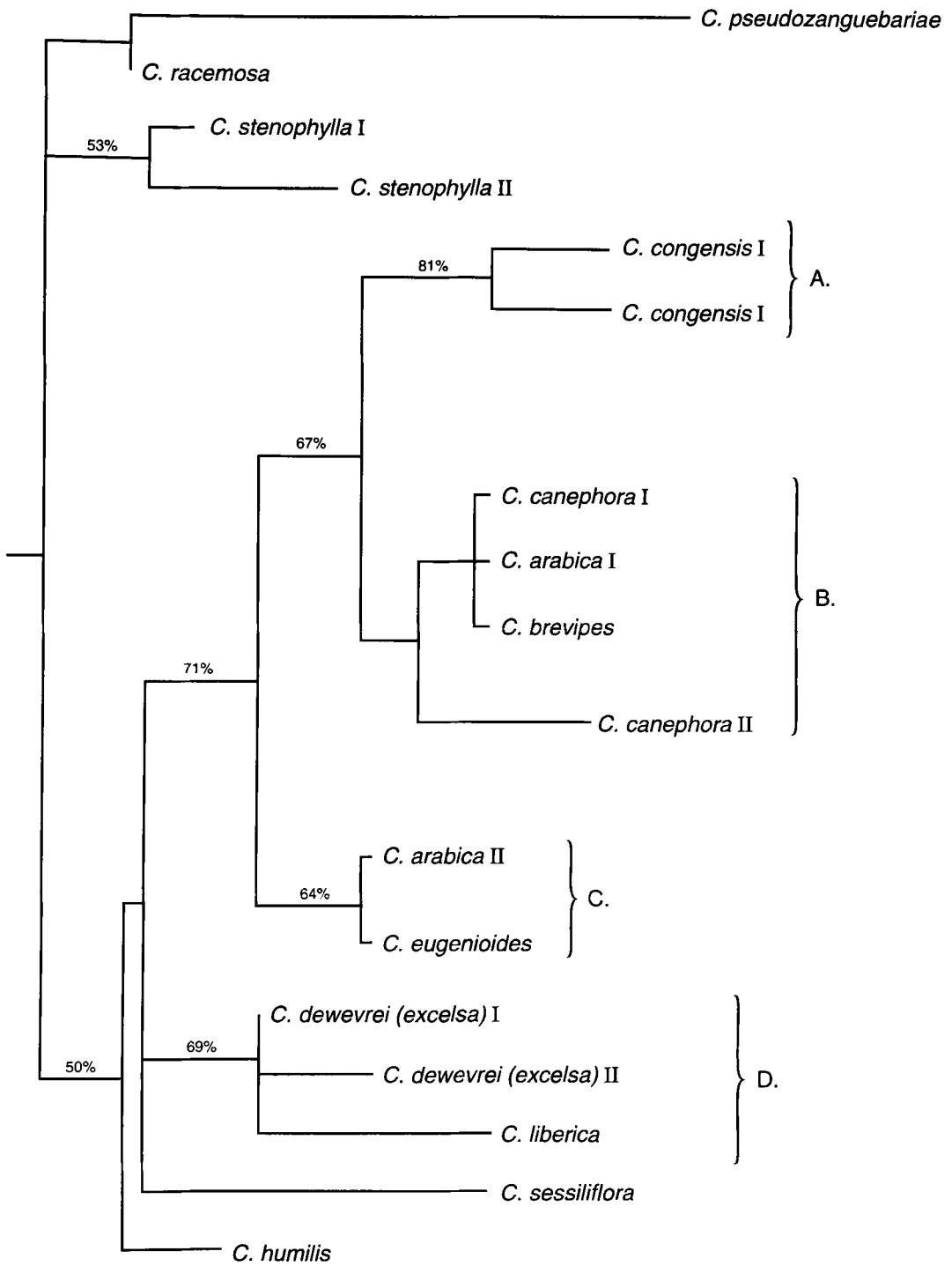


Figure 5.12. Phylogenetic tree (drawgram) of *Coffea* species derived from nuclear genomic sequences.

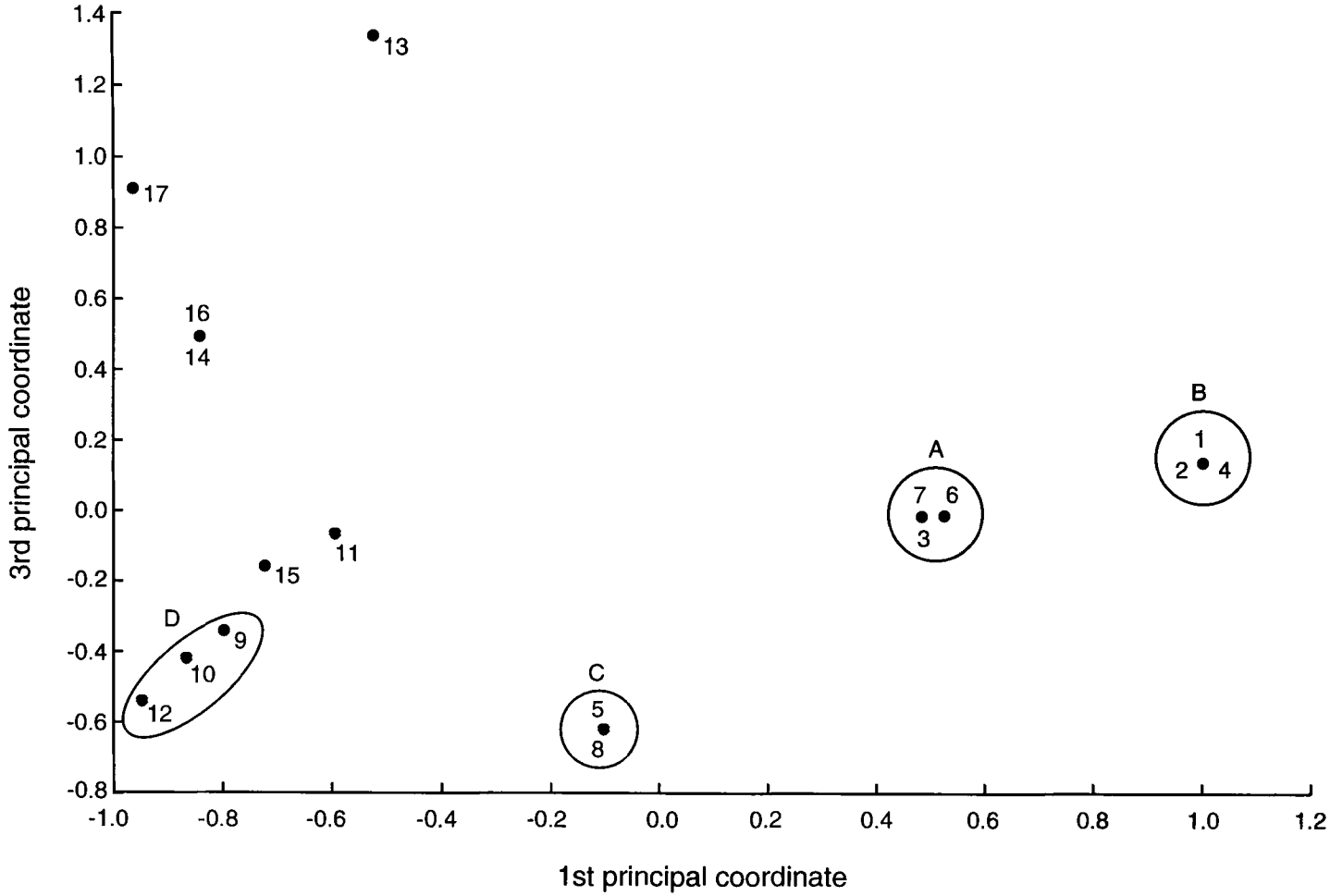


Figure 5.13. Principal coordinate analysis of 17 coffee clones representing 12 species for the nuclear sequences. The numbers in the figure correspond to the order number in Table 5.6.

Clades were formed for the *C.congensis* clones (Clade A), *C. arabica*, *C. canephora*, and *C. brevipes* (Clade B), *C. arabica* and *C. eugenioides* (Clade C), and *C. dewevrei* (excelsa) and *C. liberica* (Clade D). The alleles from the other species are clearly different from the groups already indicated and they do not form any particular group.

Principal Coordinate Analysis of the nuclear data is shown in Figure 5.13. The first and the third component displayed the best separation of the species and account for 55 % of the total variation. The principal coordinate analysis of nuclear data reflects broadly the same separation as in the phylogenetic tree.

5.3.8 CHLOROPLAST AND NUCLEAR GENOMIC SUBSTITUTION RATE COMPARISONS

The estimated value for the cp intergenic spacer/cp intron substitution rate was 1.5 and for the nuclear/chloroplast 2.2 (Table 5.8). Although the nucleotide substitution rate was not significantly different, the intergenic region seems to evolve faster than the intron. The nucleotide substitution rate in the nuclear region was, however, significantly different from that observed in the chloroplast, which suggests that the former evolves faster.

Table 5.8. Nucleotide substitution rates in sequences from chloroplast and nuclear genomic regions.

Comparison	Ratio	95% confidence Interval
Cp intergenic/intron	1.5	0.9-2.5
Nuclear/Chloroplast	2.2	1.3-2.8

5.4 DISCUSSION

In determining phylogenetic relationships, direct comparisons of the nucleotide sequence of an homologous DNA fragment displays the highest possible resolution, and is suitable for determining genetic relationships and evolution at different taxonomic levels (Olmstead and Palmer, 1994). Here, PCR-sequencing has been used in an attempt to determine phylogenetic relationships among *Coffea* species and to trace the possible evolution of the tetraploid species *C. arabica*.

5.4.1 ORGANELLAR SEQUENCES

As may have been expected, broadly similar relationships were obtained from both the intron and intergenic spacer sequence comparisons. The combined data was therefore considered to provide a higher resolution to the analysis. Four distinct clades were apparent (Figure 5.6). These clades are in congruence with the association of the species according to biogeographical distribution (Figures 5.7 and 1.2). The species in clade A originated from Central Africa, species in clade B from East Africa, Clade C, from West Africa and clade D from the Atlantic region of Africa. The genetic distances and principal coordinate analysis are also consistent and again reflect the same associations, suggesting that geographical factors may have played an important role in the speciation process.

C. racemosa, *C. sessiliflora* and *C. pseudozanguebariae* in clade B are well separated from the other clades (Figure 5.6). This is in agreement with the studies of Louarn (1993) who observed that interspecific crosses among members of clade B were interfertile, while a strong fertility barrier was observed between these and the other species. Cros et al. (1993) found that they have the lowest DNA content per nucleus (about 1 pg), suggesting that this fertility barrier could be related to genomic

size difference.

There is also a clear separation of *C. stenophyla* and *C. humilis* (93% bootstrap, Clade C) from the other species and this probably reflects the geographical isolation of these species.

Clade D includes *C. canephora*, *C. congensis*, *C. brevipes*, *C. liberica* and *C. dewevrei* (excelsa). This grouping is again consistent with that observed by Cros et al. (1993), who observed a major separation between a group containing these species and the others that were analyzed; even though the interspecific variation was very low at this level. In this study, *C. liberica* and *C. dewevrei* (excelsa) are assimilated into this group. However, although these two last species are in the same clade, the associations within the group are relatively loose.

C. liberica and *C. dewevrei* (excelsa) have been previously considered to represent a single species (Anthony et al., 1985). The results presented here (Table 5.4 and Figure 5.6) suggest that they are distinct and justify the species division. Both belong to the subsection *Pachycoffea*. My results are therefore in agreement with those of Berthou et al. (1980) who using isozyme analysis found that the genetic distance between *C. liberica* and *C. dewevrei* (excelsa) was of the same order as between *C. liberica* and

C.humilis.

Two nucleotide substitutions separated the two *C. canephora* accessions. This may be explained by the fact that *C. canephora* is an outbreeding species and highly polymorphic, and both accessions came from different collection sites. In contrast, all *C. arabica* accessions were identical, which presumably reflects their narrow genetic base and inbreeding nature.

An important finding was that, on the basis of their organellar sequence, *C. arabica* and *C. eugenioides* were indistinguishable. The results suggest that *C. eugenioides* could be the maternal progenitor of *C. arabica*. According to Carvalho (1958), Cramer (1957), and Berthou and Trouslot (1977), the ancestors of *C. arabica* seem to be *C. eugenioides* and *C. canephora*, *C. liberica* or *C. congensis*. Those results are consistent with these previous studies. They are also consistent with the results in chapter 4, based on RAPD and STS data. However, the results presented here go further in suggesting that *C. eugenioides* is the maternal ancestor of *C. arabica*.

The total sequence divergence in the chloroplast DNA (intergenic and intron) was 3.7%, which is higher than that obtained previously by RFLP on the overall chloroplast genome (Cros et al., 1993). This result indicates that the

data obtained by sequence analysis is as or more informative than that collected by RFLP. The size and frequency of additions and deletions observed in the sequences were similar to that found by Zurawski and Clegg (1987) and the substitution rate in the intergenic spacer tended to be higher than that found in the trnL intron. Previous observations suggest that changes occur more frequently in intergenic regions than in introns (Zurawski and Clegg, 1987; Clegg et al., 1991). This is perhaps not surprising as the trnL intron belongs to the Group I class of intron which exhibit distinct secondary structures and have autocatalytic activity (Taberlet et al., 1991). In contrast, Gielly and Taberlet (1994) observed that the trnL intron evolves at a rate close to that of the intergenic spacer.

For both the intron and intergenic spacer regions, the occurrence of nucleotide substitutions was far higher than the occurrence of insertion-deletions. This is in contrast with the results of Golenberg et al. (1993), who indicated that the rate of insertion/deletions is equal or higher than the rate of nucleotide substitutions within groups of closely related taxa in the grass family in noncoding regions between the chloroplast rbcL and atpB genes. However, the results presented here are in agreement with those of Zurawski et al. (1984) which indicated that short addition/deletions events occur approximately one quarter

as often as nucleotide differences in the noncoding region between rbcL and atpB in a barley-maize comparison.

5.4.2 NUCLEAR SEQUENCES

Nuclear DNA sequences have previously been used to assess genetic relationships in plants (Delseny et al., 1990; Furnier et al., 1990) and high substitution rates make them useful for genetic diversity studies. However, nuclear DNA is inherited from two distinct progenitors, and recombination occurs, potentially affecting sequence comparisons. This should therefore be considered when inferring phylogenetic relationships. Nevertheless, variation in nuclear DNA provides complementary information to that obtained by organellar analysis. In this study I thought that nuclear sequence analysis could be particular useful in tracing the progenitors of the tetraploid *C. arabica* from the diploid *Coffea* species studied.

The nuclear sequences show two distinct alleles for *C. arabica* (Figure 5.10). This may have been expected, since *C. arabica* is an allotetraploid (Carvalho, 1952; Grassias and Kammacher, 1975). In the derived tree (Figure 5.12) the two alleles are placed into different clades which are well supported (64 % and 67 % bootstrap respectively). Separation of the two alleles is also clearly observed in the principal coordinate analysis (Figure 5.13). In

addition, the highly heterozygous diploid species *C. canephora*, *C. congensis*, *C. stenophylla* and *C. dewevrei* (excelsa) also contained two different alleles. Being diploid outbreeders, the presence of two sequences is simply taken to represent heterozygosity at this locus.

It should be noted that, even though the microsatellite region was not included in the phylogenetic analysis, it also showed the same separation of alleles as the major nuclear sequence analysis (except for *C. dewevrei* (excelsa), which showed only one allele). The microsatellite region was unexpectively identical for *C. arabica* and *C. eugenioides.*

One of the two *C. arabica* sequences was identical to *C. canephora* and *C. brevipes.* and the other to that of *C. eugenioides.* These findings suggest that the tetraploid *C. arabica* was derived from *C. eugenioides* (the maternal progenitor), and either *C. canephora* or *C. brevipes.* In this respect, it is interesting to notice that *C. brevipes* and *C. canephora* also share identical chloroplast trnL intron sequences (Figure 5.5).

As indicated previously, the ancestors of *C. arabica* have often be assumed to be *C. eugenioides* and *C. canephora*, *C. liberica* or *C. congensis* (Carvalho, 1952; Cramer, 1957; Narasimhaswamy, 1962; Berthou et al., 1983). While my

results are in general agreement with this proposal they add *C. brevipes* as a potential paternal ancestor, and exclude *C. liberica* or *C. congensis*.

5.4.3 GENERAL COMPARISON AMONG CHLOROPLAST, MITOCHONDRIAL AND NUCLEAR SEQUENCES AS TOOLS FOR PHYLOGENETIC ANALYSIS AND EVOLUTION OF *COFFEA* SPECIES

Available DNA sequence data from plants now allow a detailed investigation of the rates of nucleotide substitutions in the three plant genomes (Wolfe et al., 1987), reconstruction of the phylogenetic relationships among some higher plants (Gielly and Taberlet, 1994), and comparison of evolutionary rates among lineages (Zurawski et al., 1984). I compared noncoding regions of chloroplast DNA, mitochondrial DNA, and nuclear DNA. Nuclear DNA variation was higher than that found in the chloroplast DNA (nuclear/cp substitution rate 2.2) and mitochondrial DNA variation zero. The limited data indicate that nuclear DNA rate of evolution is double that for the chloroplast genome, in agreement with Wolfe et al. (1987). Wolfe et al., (1987) also suggest that in angiosperms mtDNA evolves at least 5 times more slowly than nuclear sequences, and consider that mtDNA is more conservative than cpDNA, due to the higher synonymous substitution rate in chloroplast genes (3 times higher).

Nuclear sequences have been used less frequently for phylogeny reconstruction than cpDNA. Nevertheless, the information generated by nuclear sequences has been important in evaluating phylogenetic relationships (Barbier et al., 1991; Gepts and Clegg, 1989). For distantly related taxa, highly conserved coding sequences allow accurate assessment of character homology, therefore enabling distinct comparisons. For closely related taxa, rapidly evolving, noncoding sequences in the nucleus should provide informative nucleotide variation for phylogenetic analysis. In this work, nuclear sequences proved informative for tracing the evolution of the tetraploid *C. arabica*. The region sequenced was assumed to be noncoding as it contained only short ORFs (Open Reading Frame), showed no homology to known sequence genes in EMBL and Gen Bank and was presumed single copy.

The mitochondrial DNA sequences were uninformative reflecting the highly conservative nature of this DNA genome. The low rates of mutations make this molecule practically worthless for phylogenetic studies at lower taxonomic levels. However, the low rate of plant mtDNA substitutions suggest that comparative sequencing efforts will be more useful at higher phylogenetic levels than those for which cpDNA is now being applied.

In conclusion, the results presented here are consistent

with those inferred from classical taxonomical comparisons and reflect the origin and geographical distribution of coffee species. It has been demonstrated that *C. eugenioides* is the maternal progenitor of the tetraploid *C. arabica*, and strong evidence is provided for *C. canephora* being the paternal donor. However, the data do not exclude *C. brevipes* as a paternal donor. More detailed studies at the nuclear genomic level should help to resolve this issue.

CHAPTER 6. GENERAL DISCUSSION, CONCLUSIONS AND FUTURE OPPORTUNITIES

Coffee is the most valuable agricultural export commodity for many developing countries but has not yet benefited from many of the advances in biotechnology which are having an impact upon other major crops. It is, however, essential that these approaches be applied to coffee improvement. In this thesis, PCR-based genetic markers have been exploited, to address a number of important questions in *Coffea* biology. Using the Random Amplified Polymorphic DNA (RAPDs) technique, it has been demonstrated that this approach is useful for the genetic characterization of coffee germplasm, estimation of relatedness and identification (selection) of interspecific gene introgression.

The evaluation of plant germplasm and genetic diversity are essential to the effective management of coffee genetic resources and incorporation of novel genes into breeding programs. Methods for the maintenance of genetic resources vary according to several factors, including the species, its geographical distribution, breeding system and seed behaviour. Both the evaluation of diversity and the avoidance of duplication of accessions is of great importance in germplasm conservation, to provide the right material to breeders and to reduce the costs of

maintenance. Different techniques are available to perform such evaluations. Here, it has been shown that RAPD markers can be used to study genetic diversity in *Coffea* species, and that this approach could be applied to the evaluation of coffee genetic resources, regardless of the method of conservation. As an example of the impact of this work, (and of the overall E.U. project) a complete evaluation of the coffee germplasm maintained in the coffee germplasm collection in CATIE (Centro Agronomico Tropical de Investigacion y Ensenanza) in Turrialba, Costa Rica (Anthony, 1995, personal communication) will be conducted over the next few years. This evaluation is important because results from the work presented here, where accessions from the CATIE collection were used, have shown that the amount of variability present in this collection is limited, probably due to selective sampling of material from collections in the Ivory Coast. As this is considered an international resource collection for *Coffea* breeding, it is important to establish the extent of the variation present.

It was also demonstrated that RAPDs can detect natural and artificial introgression in the genus *Coffea*. As in other plants (Rieseberg and Seiler, 1990), natural introgression of genes is important in determining genetic relationships and tracing evolution of *Coffea* species. Artificial introgression implies that the particular markers detected

were associated with selective gene introgression from *C. canephora* into *C. arabica*. These genes could be, for example, of adaptive significance. If indeed such markers were linked to useful genes they could be immediately useful in marker assisted breeding schemes. However, prior to their employment, a test of linkage to specific characters should be carried out.

In practice, little attention has been paid to intraspecific hybridization in *C. arabica*, mainly because of the lack of knowledge of the original sources of genetic material within the species. However, it is known that all the varieties grown in Latin America were derived from a few common seeds, which explained the narrow genetic base represented in these populations. The limited diversity found was probably due to local adaptive mutation and selective maintenance. With the discovery and collection of new *C. arabica* genotypes in the centres of origin, breeding programmes are currently giving more emphasis and time to hybridization within this species. In addition, the observation of interspecific hybridization allows us to conclude that ploidy is not an absolute barrier to gene flow between *Coffea* species. Diploid species and *C. arabica* should therefore be considered a common gene pool; and it should be possible to improve *C. arabica* varieties as well as *C. canephora* using existing genetic resources.

In combination with specific chloroplast and mitochondrial markers it has also been demonstrated that RAPD markers allow an estimation of genetic variation and relationships within and among *Coffea* species. These results both complemented and reaffirmed earlier results using alternative or traditional approaches. The evaluation of genetic diversity and phylogenetic relationships is of great importance, since it allows quantification of the genetic variation present in the coffee gene pool. Based on similarity levels it may also suggest which combination of species could be more effectively used for hybridization and breeding purposes. Molecular markers can therefore potentially play a vital role in allowing coffee breeders to exploit the most suitable material for their breeding programs and to achieve their objectives more rapidly.

In the final chapter, the molecular technique with perhaps the highest possible resolution -DNA sequence analysis- has been applied to study phylogenetic relationships and evolution of *Coffea* species. The information derived from sequencing chloroplast genomic regions clearly differentiated the species in the genus *Coffea*. The topology of the phylogenetic tree reflected the origin and geographical distribution of the species. The sequence analysis revealed broadly the same distribution of species shown by RAPDs but was more powerful in its resolution. While RAPDs grouped together the coffee species from the

west and central atlantic African areas, sequence analysis allowed a clear separation of these two regional groups. The sequences revealed that of the species analysed, *C. eugenioides* was likely the maternal progenitor of the tetraploid *C. arabica*. These results were consistent with the RAPD results presented in Chapter 4. Examination of sequences from a nuclear DNA PstI clone also provided important information on the genetic relationships and evolution of *C. arabica*. This study revealed two distinct sequences (alleles) in *C. arabica* and some of the diploid clones. Importantly, the sequence of one of these alleles was identical to the sequence for *C. eugenioides*, and the other identical to a sequence from *C. canephora* and *C. brevipes*. While these results tend to confirm *C. eugenioides* as the maternal ancestor of *C. arabica*, they also suggested that *C. canephora* or *C. brevipes* were the paternal parent. From the breeding point of view, this could indicate that these species are more likely to have more similar genetic structure and possible chromosome homology with *C. arabica*, and therefore could be more useful in the crossing programs to transfer valuable genes into commercial coffee varieties. From an evolutionary perspective, these results give a new dimension to previous studies on evolution of coffee, providing evidence at the level of DNA information for the identification of the ancestors of the tetraploid *C. arabica*. Importantly, this information is consistent with data presented in chapters

3 and 4 based on RAPD markers and specific sequence tagged sites (STS).

The results also reveal valuable and new information on the genetic relationships within the genus *Coffea*, which could contribute to the development of improved coffee germplasm based on intra and interspecific hybridization programs. Currently, interspecific hybridization followed by the generation of synthetic amphidiploids is being conducted, and my results could contribute to the identification of appropriate gene pools to be included in the breeding programmes.

6.1 FUTURE WORK

An important objective of future work would be to establish a genetic linkage map of *Coffea*, which would facilitate the identification of markers linked to important genes that confer useful agronomic characteristics, such as disease and nematode resistance and high yield. Initially, this could be done using RAPD and RFLP markers. During my studies, a preliminary study was initiated on a F2 mapping population derived from an intraspecific cross between two *C. arabica* cultivars, N39 and VCE 1592 (Hybrid of Timor 1343). The level of polymorphism between the two parents was screened, using RFLPs and RAPDs. When RFLP markers were used, six probes from a coffee Pst I library (C5, C6, C7,

C24, C26, C31) and four restriction enzymes (Eco RI, Xba I, Bam HI, Eco RV) were utilized. The level of polymorphism was low, with only one polymorphism found in five probe/enzyme combinations from a total of 24. An example of the polymorphism is observed in Figure 6.1. With RAPD markers, from 52 primers 12 revealed polymorphism between the two parental lines. From these, only three bands segregated in the expected Mendelian fashion (3:1) in the F2 population. An example of the polymorphism observed between the parents and nine progeny plants is shown in Figure 6.2. The reason for the low level of polymorphism was possibly due to the inbreeding nature of *C. arabica* and the fact that the cultivars used as parents were highly homozygous and similar in genotype. This means, that to find enough molecular markers for mapping purposes, a great amount of work would need to be done, and unfortunately this was not possible in the time scale of the thesis.

A population derived from an interspecific cross would of course reveal more polymorphism, and would be much more suitable for genetic mapping. Interspecific crosses between *C. arabica* and *C. canephora* and the development of an F2 population are currently being assembled in CATIE, Costa Rica. The polymorphism already detected in my studies between the two species could be useful in the initiation of the mapping studies in this population. Among other

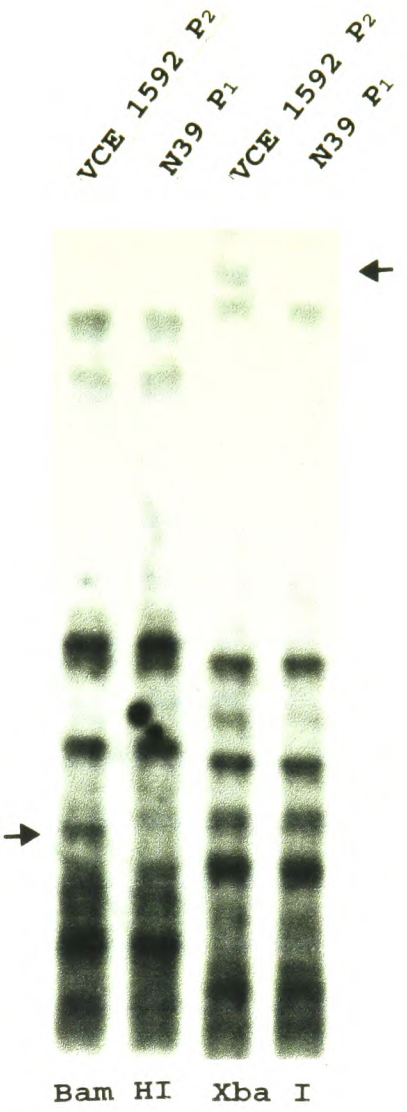


Figure 6.1. Polymorphism revealed by RFLP analysis in *C. arabica* accessions N39 and VCE1592 with enzymes Bam HI and Xba I and probe C5. Arrows indicate polymorphic bands. The coffee accessions used are the original parents of the F_2 population intended to be used for mapping studies.

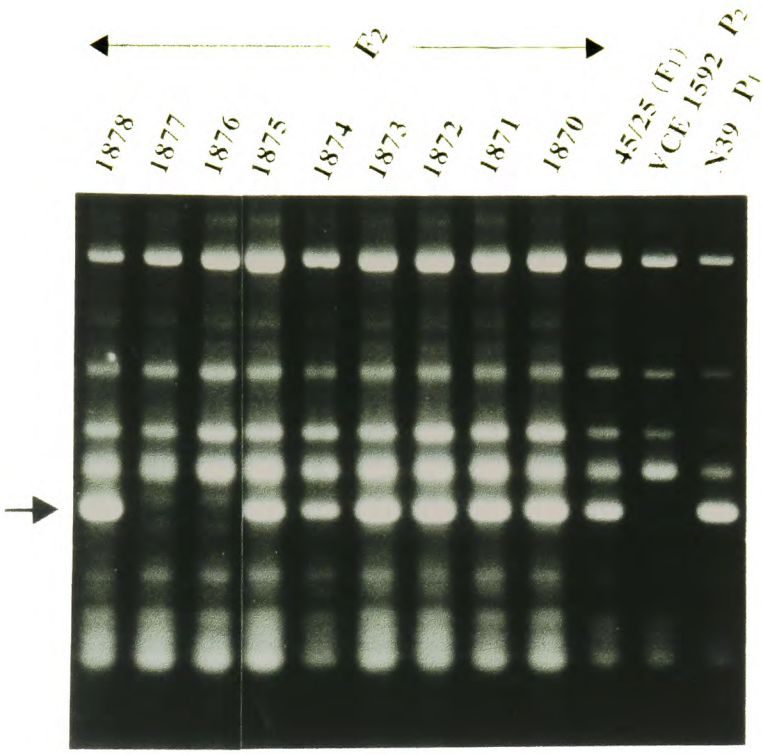


Figure 6.2. Segregation of RAPD markers in a F_2 progeny derived from an interspecific cross between *C. arabica* cultivars N39 and VCE1592.

alternatives for mapping is the development of double haploid lines. In barley, for example, double haploid lines and bulk segregant analysis were used to map and identify markers linked to qualitative and quantitative characters (Barua et al., 1993). In coffee, double haploid lines have been derived from a heterozygous clone IF200, which was generated from *C. canephora* cultivars. These double haploid lines are going to be a base population for mapping studies, based on RFLP and RAPD markers (Lashermes et al., 1993). Preliminary tests have shown that 40 % of the RAPD primers assayed detected polymorphism between the parental clone IF200 and the homozygous parent tester DH160-02, and 11 % within the DH population (Lashermes, 1995; personal communication). Segregation of polymorphic markers observed in the DH genotypes were in agreement with Mendelian segregations.

An important application of a coffee genetic linkage map would be to identify markers linked to important genes, such as those conferring resistance to leaf rust (*Hemileia vastatrix*), coffee berry disease (Colletotrichum coffeanum) and to nematodes (*Meloidogyne* spp. and *Pratylenchus* spp.). Subsequently, monogenic and polygenic characters controlling important agronomical traits could be evaluated. Linked or diagnostic markers could then be incorporated into breeding programs for indirect selection of superior coffee lines. This could, for example, allow

early selection of desirable progenies, which would save a significant amount of labour and time in this perennial crop, which using the traditional breeding requires 12 to 15 years to create a new variety. This accelerated molecular breeding approach should be possible in the near future.

An important practical application that could be started immediately is the genetic "fingerprinting" of commercial coffee varieties. The present work showed that some RAPD markers were diagnostic of specific species and of groups within the *C. arabica* accessions. By using an appropriate number of primers, it should be possible to assign bands diagnostic of specific commercial varieties within the *C. arabica* and *C. canephora* cultivars. After determining markers which are diagnostic for a particular species or group of specific accessions, such as the coffee Bourbon type, more robust PCR assays can be used. For example, bands specific to species or accessions could be cloned and sequenced and primers designed for using in a robust and informative genetic "fingerprint" assay.

One limitation attributed to the RAPD approach is its reliability, which can restrict its application. However, the results presented here demonstrated that the genetic relationships among *Coffea* species determined by RAPD markers were consistent with those derived from sequence

analysis. This demonstrates that reliability is not a limiting factor for using RAPDs to examine genetic diversity in plant populations.

Concerning new approaches, microsatellites and AFLP are the next generation of genetic markers. These could be productively used in examining coffee genetic diversity and constructing genetic maps. These new approaches are being used in other crops for similar purposes. For example microsatellites have been used for determining genetic variability in soybean and rice (Morgante and Olivieri, 1993; Zhao and Kochert, 1992; Powell et al., 1995), and mapping in potato populations (Milbourne, 1995; personal communication). In the present study, sequence analysis of several anonymous PstI clones identified two mcrosatellites regions and an example is showed in Figure 6.3. The other was in the centre of the fragment sequenced for the phylogenetic studies. Once microsatellite regions are identified, primers could be designed to flank them and their amplification products used in coffee genetic diversity analysis. AFLPs are being used for genetic mapping in barley (Waugh, 1995; personal communication), and tea (Wachira, 1995; personal communication). Given sufficient resources, Microsatellites and AFLPs could be used in the near future as molecular tools for conservation and evaluation of coffee genetic resources and accelerated, molecular assisted, coffee breeding programmes.

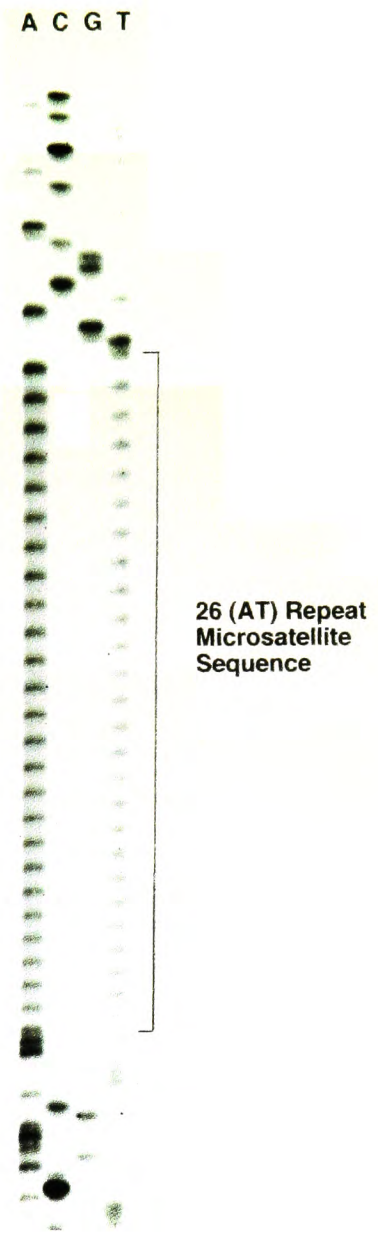


Figure 6.3. Microsatellite region in an anonymous clone sequence from a coffee Pst I library. The bracket indicates the single sequence repeats.

REFERENCES

Anonymous (1989) Coffee and biotechnology. Communique Rural Advancement Fund International, Pittsburg, USA, pp. 1-12.

Anonimous (1995a) Prices paid to growers. December 1989 to 1994. *EB 3440/95 (E)*. International Coffee Organization, London, England.

Anonimous (1995b) Supply of Coffee. Crop years commencing in 1987 to 1994 and coffee years 1988/89 to 1994/95. *EB 3535/95 (E)*. International Coffee Organization, London, England.

Anthony, F., Couturon, E. & de Namur, C. (1985) Les cafeiers sauvages du Cameroun: resultats d'une mission de prospection effectuée par l'ORSTOM en 1983. *11eme Colloque de I'A.S.I.C.*, Lomé (Togo).

Arnold, M.L, Buckner, C.M., & Robinson, J.J. (1991) Pollen mediated introgression and hybrid speciation in Louisiana irises. *Proceedings of the National Academy of Sciences of the USA*, **88**, 1398-1402.

Barbier, P., Morishima, H. & Ishihama, A. (1991) Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. *Theoretical and Applied Genetics*, **82**, 693-702.

Barret, J.W., Rajora, O.P., Yeh, F.C.H., Dancik, B.P. & Strobeck, C. (1992) Mitochondrial DNA variation and genetic relationships of *Populus* species. *Genome*, **36**, 87-93.

Barua, U.M., Chalmers, K.J., Hackett, C.A., Thomas, W.T.B., Powell, W. & Waugh, R. (1993) Identification of RAPD markers linked to a *Rhynchosporium secalis* resistance locus in barley using near-isogenic lines and bulked segregant analysis. *Heredity*, **71**, 177-184.

Beckman, J.S. & Soller, M. (1986a) Restriction fragment length polymorphism and genetic improvement of agricultural species. *Euphytica*, **35**, 111-124.

Beckman, J.S. & Soller, M. (1986b) Restriction fragment length polymorphisms in plant genetic improvement. *Oxford Surveys of Plant Molecular and Cell Biology*, **3**, 192-246.

Bell, C.J. & Ecker, J.R. (1994) Assignment of thirty microsatellite loci to the linkage map of *Arabidopsis*. *Genomics*, **19**, 137-144.

Berthaud, J. (1976) Etude cytogenétique d'un haploide de *Coffea arabica* L. *Cafe, Cacao, The*, **20**, 91-96.

Berthaud, J. (1977) Caracteristiques comparées des hybrides interspecifiques tetraploids et hexaploids *Coffea arabica* L. x *Coffea canephora* Pierre. In: *VIII Colloque Scientifique International sur la Chimie du cafe*. ASIC. Paris, pp. 393-397.

Berthaud, J. (1980) L'incompabilité chez *Coffea canephora*: methode de test et determinisme genétique. *Cafe, Cacao, The*, **24**, 267-274.

Berthaud, J. (1986) Les resources génetiques pour l'amelioration des cafeiers africains diploides: evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Consequences pour l'application. *Travaux et Documents de l'Orstom*, **188**, 372 p.

Berthaud, J. & Guillaumet, J.L. (1978) Les cafeiers sauvages en Central Afrique. *Cafe, Cacao, The*, **22**, 171-186.

Berthaud, J. & Charrier, A. (1988) Genetic resources of Coffea. In: Clarke, R.J. & Macre, R. (eds) *Coffee Vol. 4 Agronomy.* Elsevier Applied Science, London.

Berthou, F. (1975) Méthode d'obtention de polyploides dans le genre *Coffea* par traitements localisées de bourgeons a la colchicine. *Cafe, Cacao, The*, **19**, 197-202.

Berthou, F. & Trouslot, P. (1977) L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une méthode d'electrophorèse en serie, premiers resultats, *8eme Coloque Scientifique International sur le Cafe. 8th Conference of ASIC* Abidjan, Ivory Coast. 373-384.

Berthou, R., Mathieu, C. & Vedel, F. (1983) Chloroplast and mitochondrial DNA variation as indicators of phylogenetic relationships in the genus *Coffea* L. *Theoretical and Applied Genetics*, **65**, 77-84.

Berthou F., Trouslot, P., Hamon, S., Vedel, F. & Quetier, F. (1980) Analyze en electrophorèse du polymorphisme biochimique des cafeires, variation enzymatique dans dix-huit populations sauvages, variation de 1'ADN mitochondrial dans les éspéces *C. canephora, C. eugenioides* et *C. arabica. Cafe, Cacao, The,* **24**, 313-326

Berthouly, M., (1991) Informe final del Proyecto Desarrollo y Reproduccion de variedades con resistencia a la Roya del cafe: *Cultivo de tejidos.* Proyecto AID/Rocap (596-0090). 35 pp.

Bonierbale, M.W., Plaisted, R.I. & Tanksley, S.D. (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics*, **120**, 1095-1103.

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32**, 314-331.

Bouharmont, J. (1963) Somatic chromosomes of some *Coffea* species. *Euphytica*, **12**, 254-257.

Bridson, D. (1982) Studies in *Coffea* and *Psilanthus* (*Rubiaceae* subfam. *Cinchonoideae*) for part 2 of the 'Flora of Tropical East Africa: *Rubiaceae*'. *Kew Bulletin (Royal Botanic Garden)*, **36,** 817-859.

Bridson, D. (1987) Nomenclatural notes on *Psilanthus*, including *Coffea* sect. *Paracoffea* (*Rubiaceae* tribe *Coffeeae*). *Kew Bulletin*, (*Royal Botanic Garden*), **42**, 817-859.

Brown, A.H.D. (1989) The case for core collections. Part III. Size and structure of collections. In: Brown, A.H.D., Frankel O.H. & Marshall, D.R. (eds) *The Use of Plant Genetic Resources*, pp. 136-156, Cambridge University Press, Cambridge, UK.

Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C. & Stuber, C.W. (1988) Gene mapping with recombinant inbreds in maize. *Genetics*, **118**, 519-526.

Cardy, B.J. & Kannenberg, L.W. (1982) Allozymic variability among maize inbred lines and hybrids, applications for cultivar identification. *Crop Science*, **22**, 1016-1020.

Carlson, J.E., Tulsieram, L.K, Glaubitz, J.C, Luk, V.W.K, Kauffelde, C. & Ruthedge R. (1991) Segregation of random amplified DNA markers in F1 progeny of conifers. *Theoretical and Applied Genetics*, **83**, 194-200.

Carvalho, A. (1952) Taxonomia de *Coffea arabica* L. 6. Caracteres morfologicos dos haploides. *Bragantia*, ***12***, 201-212.

Carvalho, A. (1958) Advances in coffee production technology. Recent advances in our knowledge of coffee trees. *II Genetics. Coffee and Tea Industries and the Flavour Field*, **81**, 30-36.

Carvalho, A. (1985) Breeding of Arabica for productivity and quality. In: Clifford, M.N. & Wilson, K.C. (eds) *Coffee: Botany, Biochemistry and Production of Beans and Beverage*. Croom Held Ltd Publishers, Beckenham, UK.

Carvalho, A. & Krug, C.A. (1949) Genetica de *Coffea* XII. Hereditariedade da cor amarela da semente. *Bragantia*, **9**, 193-202.

Carvalho, A. & Monaco, L.C. (1968) Relaciones genéticas de especies seleccionadas de *Coffea*. *Cafe*, **4**, 3-19.

Carvalho, A., Ferwerda, F.P., Frahm-Leliveld, J.A., Medina, D.M., Mendes, A.J.T. & Monaco, L.C. (1969) Coffee (*Coffea arabica* L. and *C. canephora* Pierre ex Froehner). In: Ferwerda, F.P & Wit. F. (eds) *Perennial Crop Breeding in the Tropics*. Veenman & Zonen, Wageningen. pp. 189-241.

Chalmers, K.J., Sprent, J.I., Simons, A.J., Waugh, R. & Powell, W. (1992) Patterns of genetic diversity in a tropical tree legume (*Gliricidia*) revealed by RAPD markers. *Heredity*, **69**, 465-472.

Chalmers, K.J., Barua, U.M., Hackett, C.A., Thomas, W.T.B., Waugh, R. & Powell, W. (1993) Identification of RAPD markers linked to genetic factors controlling the milling energy requirement of barley. *Theoretical and Applied Genetics*, **87**, 314-320.

Chaparro, J., Wilcox, P., Grattapaglia, D., O'Malley, D., McCord, S., Sederoff, R., McIntyre, L. & Whetten, R. (1992) Genetic mapping of pine using RAPD markers: construction of a 191 marker map and development of half-sib genetic analysis. In: *Advances in gene technology: feeding the world in the 21st century.* Miami Winter Symposium, Miami, Fl.

Chapman, C.G.D. (1989) Collection strategies for the wild relatives of field crops. Part V. Wild relatives of crops. In: Brown, A.H.D., Frankel, O.H. & Marshall, D.R. (eds) *The Use of Plant Genetic Resources*, pp. 263-279, Cambridge University Press, Cambridge, UK.

Charrier, A. (1977) La structure génétique du genre *Coffea*, ses consequences pour l'amélioration des cafeiers cultives. *8th Conference of ASIC*, Abdijan, Ivory Coast. 407-410.

Charrier, A. (1978). La structure génétique des cafeiers spontanes de la region malgache *Mascarocoffea*. Leurs relations avec les cafeiers d'origine africaine *Eucoffea*. *Memoirs ORSTOM*, **87**, 223 p.

Charrier, A. & Berthaud, J. (1985) Botanical classification of coffee. In: Clifford, M.N & Wilson, K.C. (eds) *Coffee: Botany, Biochemistry and Production of Beans and Beverage*. Croom Held Ltd Publishers, Beckenham, UK.

Chevalier, A. (1947) Les cafeiers du globe. Systematique des cafeiers et des faux-cafeiers. *Encyclopedie Biologique XXVII*, Fas. III, P. Lechevalier, Paris 356 p.

Clegg, M.T. & Zurawski, G. (1992a) Chloroplast DNA and the study of plant phylogeny. In: Soltis, P.S., Soltis, D.E. & Doyle. J.J. (eds) *Molecular Systematics of Plants*, Chapman and Hall, New York, 1-13.

Clegg, M.T. & Zurawski, G. (1992b) Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: Soltis, P.S., Soltis, D.E. & Doyle J.J. (eds.), *Molecular Systematics of Plants*, 1-13. Chapman and Hall, New York, NY.

Clegg, M.T., Learn, G.H. & Golenberg, E.M. (1991) Molecular evolution of chloroplast DNA. In: Selander, R.K., Clark, A.G., & Whittam, T.S. (eds). *Evolution at the Molecular Level*, Sinauer Associates, Sunderland, pp 135-149.

Cramer, P.J.S. (1957) Review of literature on coffee research in Indonesia. Instituro Interamericano de Ciencias Agricolas, Turrialba, Costa Rica, (*Miscellaneous publication No. 15*).

Crawford, D.J. (1983) Phylogenetic and systematic inferences from electrophoretic studies. In: Tanksley S.D. & Orton, T.J. (eds) *Isozymes in Plant Genetics and Breeding*. Part A. pp. 257-285. Elsevier Science Publishers B.V., Amsterdam.

Cros, J., Lashermes, Ph., Marmey, Ph., Anthony, F., Hamon, S. & Charrier, A. (1993) Molecular analysis of genetic diversity and phylogenetic relationships in *Coffea*. In: *Proceedings of 15th International Conference on Coffee Science (ASIC)*, Montpellier, pp. 41-46.

Curtis, S.E. & Clegg M.T. (1984) Molecular evolution of chloroplast DNA sequences. *Molecular Biology and Evolution*, **1**, 291-301.

Dawson, I.K., Simons, A.J., Waugh, R. & Powell, W. (1995) Detection and pattern of interspecific hybridization between *Gliricidia sepium* and *G. maculata* in Meso-America revealed by PCR-based assays. *Molecular Ecology*, in press.

Debener, T., Salamini, F. & Gebhardt, C. (1990) Phylogeny of wild and cultivated *Solanum* species based on nuclear restriction fragment length polymorphisms (RFLPs). *Theoretical and Applied Genetics*, **79**, 360-368.

Delseny, M., McGrath, J.M., This, P., Chevre, A.M. & Quiros, C.F. (1990) Ribosomal RNA genes in diploid and amphidiploid *Brassica* and related species: organisation, polymorphism, and evolution. *Genome*, **33**, 733-744.

Demeke, T., Adams, R.P. & Chibbar, R. (1992) Potential taxonomic use of random amplified polymorphic DNA (RAPD): a case study in *Brassica*. *Theoretical and Applied Genetics*, **84**, 990-994.

Dietrich, J. (1992) A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics*, **131**, 423-447.

Digby, P.G.N. & Kempton, R.A. (1987) Multivariate analysis of ecological communities. Chapman and Hall, London, 125 pp.

D'Ovidio, R., Tanzarella, O.A. & Porceddu, E. (1990) Rapid and efficient detection of genetic polymorphism in wheat through amplification by polymerase chain reaction. *Plant Molecular Biology*, **15**, 169-171.

Downie, S.R. & Palmer, J.D. (1992) Restriction site mapping of the chloroplast DNA inverted repeat: a molecular phylogeny of the *Asteridae*. *Annals of the Missouri Botanical Garden*, **79**, 266-283.

Doyle, J.J., Doyle, J.L. & Brown, A.H.D. (1990) Chloroplast DNA phylogenetic affinities of newly described species in *Glycine* (Leguminosae: Phaseoleae). *Systematic Botany*, **15**, 466-471.

Edwards, K., Johnstone, C. & Thompson, C. (1991) A simple method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, **19**, 1349.

Engelmann, F., Jouve, L., Chabrillange, N., Dussert, S., Anthony, F. & Charrier, A. (1993) Cold sensitivity of *in vitro* microcuttings of *Coffea arabica* and *C. canephora* during storage at various temperatures. Evolution of sugar, proline, MDA and ethylene production. *15th. Conference of ASIC*. Montpellier, France.

Erikson, L.R., Straus, N.A. & Bewersdorf, W.D. (1983) Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphidiploids. *Theoretical and Applied Genetics*, **65**, 201-206.

Erlich, H.A., Gelfand, D. & Sininsky, J.J. (1991) Recent advances in the polymerase chain reaction. *Science*, **252**, 1643-1651.

Feinberg, A.P. & Vogelstein, B. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity (Addendum). *Annals of Biochemistry*, **137**,266-267.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368-376.

Felsenstein, J. (1993) Phylogeny Inference Package (PHYLIP) version 3.5. Available by anonymous ftp from the Internet address evolution.genetics.washington.edu.

Fenell, S.R. (1994) Molecular and genetical studies in *Arachis hypogaea* and *Vicia Faba*. PhD thesis. University of Dundee, Dundee, UK.

Furnier, G.R., Cummings, M.P. & Clegg, M.T. (1990) Evolution of the avocados as revealed by DNA restriction fragment variation. *Journal of Heredity*, **81**, 183-188.

Gawel, N.J. & Jarret, R.L. (1991) A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Molecular Biology Reports*, **9**, 262-266.

Gebhardt, C., Ritter, E., Debener, T., Schachtschabell, U., Walkemeier, B., Uhrig, H. & Salamini, F. (1989) RFLP-analysis and linkage mapping in *Solanum tuberosum*. *Theoretical and Applied Genetics*, **78**, 65-75.

Genstat 5 Committee. (1987) Genstat 5 Reference Manual. Clarendon Press, Oxford, UK.

Gepts, P. & Clegg, M.T. (1989) Genetic diversity in pearl millet (*Pennisetum glaucum* L.) at the DNA sequence level. *Journal of Heredity*, **80**, 203-208.

Gielly, L. & Taberlet, P. (1994) Chloroplast DNA polymorphism at the intrageneric level and plant phylogenies. *Comptes Rendus de l'Academie des Sciences, Paris, Life sciences, Evolution*, **317**, 685-692.

Golenberg, E.M., Clegg, M.T., Durbin, M.L., Doebley, J. & Din Pow, M. (1993) Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution*. Vol. **2**, No. 1, March, pp. 52-54.

Gotlieb, L.D. (1981) Electrophoretic evidence and plant populations. *Progress in Phytochemistry*, **7**, 1-46.

Grassias, M. & Kammacher, P. (1975) Observations sur la conjugaison chromosomique de *Coffea arabica* L. *Cafe, Cacao, The*, **19**, 177-190.

Hadrys, H., Balick, M. & Schierwater, B. (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology,* **1**, 35-63.

Hamilton A.C (1974) In: Lind, E.M. & Morrison, M.E.S. (eds) East African Vegetation. Longman, London, 1974. 188-209

Hart, G.E. (1979) Evidence for a triplicate set of glucosephosphate isomerase structural genes in hexaploid wheat. *Biochemical Genetics*, **17**, 585-598.

Hayashi, K. (1992) PCR-SSCP, a method for detection of mutations. *Genetic analysis, techniques and applications*, **3**, 73-79.

Hazan, J., Dubay, C., Pankowiak, M.P., Becuwe, N. & Wissenbach, J. (1992) A genetic linkage map of human chromosome-20 composed entirely of microsatellite markers. *Genomics*, **12**, 183-189.

Hedges, S.B. (1992) The number of replication needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Molecular Biology and Evolution*, **9**, 366-369.

Helentjaris, T. (1987) A genetic linkage map for maize based on RFLPs. *Trends in Genetics*, **3**, 217-221.

Helentjaris, T., King, G., Slocum, M., Siedenstrang, C. & Wegman, S. (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tool for applied plant breeding. *Plant Molecular Biology,* **5**, 109-118.

Helentjaris, R., Slocum, M., Wright, S., Shaefer, A. & Nienhuis, J. (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theoretical and Applied Genetics,* **72**, 761-769.

Higgins, D.G. (1992) Sequence ordination: a multivariate analysis approach to analysing large sequence data sets. *CABIOS*, **8**, pp 15-22.

Higgins, D.G. & Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene,* **73**, 237-244.

Hillis, D.M., Larson, A., Davis, S.K. & Zimmer, E.A. (1990) Nucleic acids III: sequencing. In: Hillis, D.M. & Mortiz, C. (eds) *Molecular systematics*. Sinauer, Sunderland, Massachusetts.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y, Li, Y.-Q., Kano, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura, M. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular General Genetics*, **217**, 185-194.

Hofling, J.F. & Oliveira, A.R. (1981) A serological investigation of some *Coffea* species with emphasis on the origin of *C. arabica*. *Ciencia e Cultura*, **33**, 66-72.

Hu, J. & Quiros, C.F. (1991) Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell Reports*, **10**, 505-554.

Hulbert, S.H., Liott, T.W., Legg, E.J., Lincoln, S.E., Lander, E.S. & Michelmore, R.W. (1988) Genetic analysis of the fungus *Bremia lactucae* using restriction fragment length polymorphism. *Genetics,* **120**, 947-958.

Ishii, T., Terachi, N., Mori, N. & Tsunewaki, K. (1993) Comparative study on the chloroplast, mitochondrial and nuclear genome differentiation in two cultivated rice species, *Oryza sativa* and *O. glaberrima*, by RFLP analysis. *Theoretical and Applied Genetics*, **86**, 88-96.

Janczewski, D.N., Yuhki, N., Gilbert, D.A. & Jefferson, T. (1992) Molecular phylogenetic inference from saber-toothed cat fossils of Rancho La Brea. *Proceedings of the National Academy of Sciences of the USA*, **89**, 9769-9773.

Johnson, E., Hosein, F. & Sirju-Charran, G. (1990) Use of isoperoxidases in the characterisation of *Theobroma cacao* L. germplasm. *Annual Report of the Cocoa Research Unit for 1990.*

Kangfu, Y. & Pauls, K.P. (1993) Rapid estimation of genetic relatedness among heterogeneous populations of alfalfa by random amplification of bulked genomic DNA samples. *Theoretical and Applied Genetics*, **86**, 788-794.

Kartha, K.K., Mroginski, L.A., Pahl, K. & Leung, N.L. (1981) Germplasm preservation of coffee (*Coffea arabica* L.) by in vitro culture of shoot apical meristems. *Plant Science Letters*, **22**, 301-307.

Kempton, R. & McNicol, J. (1990) Graphical Methods for Multivariate Data. Scottish Agricultural Statistics Service, pp 1-67.

Kimura, M. (1983) *The natural theory of molecular evolution.* Cambridge University Press, Cambridge.

King-Ying, T., Chiu-I, L., Shih-Tung, L. & Yu-Sun, C. (1993) Detection of point mutations in the chloroplast genome by single-stranded conformation polymorphism analysis. *The Plant Journal*, **3**, 183-186.

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. *Theoretical and Applied Genetics*, **81**, 661-667.

Kocher, T.D. (1992) PCR, direct sequencing, and the comparative approach. *PCR Methods and Application*, **1**, 217-221.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pablo, S., Villablanca, F.X. & Wilson, A.C. (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA*, **86**, 6196-6200.

Kresovich, S., Williams, J.G.K., McFerson, J.R., Routman, E.J. & Schaal, B.A. (1992) Characterization of genetic identities and relationships of *Brassica oleracea* L. via random amplified polymorphic DNA assay. *Theoretical and Applied Genetics*, **85**, 190-196.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Dayly, M.J., Lincoln, S.E. & Newburg, G. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174-181.

Lashermes, P., Paillard, M., Marmey, P., Gavalda, M.C., Couturon, E., Petiard, V. & Charrier, A. (1993) Toward the construction of a genetic map in coffee. *15th. Conference of ASIC*. Montepellier, France.

Leroy, J.F. (1967) Recherches sur les cafeiers. Sur la classification biologique des cafeiers et sur l'origine et l'aire du genre *Coffea*. *Comptes Rendus de l'Academie des Sciences*. Paris, **265**, 1043-1045.

Leroy, J.F. (1980) Evolution et taxogenese chez les cafeiers. Hypothèse sur leur origine. *Comtes Rendus de l'Academie des Sciences*, Paris, **291**, 593-596.

Leroy, J.F. (1982) L'origine kenyane du genre *Coffea* et la radiation des espèces a Madagascar. In: *10th International Colloquium on the Chemistry of Coffee*, ASIC, Paris, pp. 413-420.

Lobreau-Callen, D. & Leroy, J.F. (1980) Quelques données palynologiques sur le genre *Coffea* et autres genres du cercle des cafeiers. *9th International Colloquium on the Chemistry of Coffee*. ASIC, Paris, pp. 479-506.

Louarn, J. (1976) Hybrides interspecifiques entre *Coffea canephora* Pierre et *C. eugenioides* Moore. *Cafe, Cacao, The*, **20**, 433-452.

Louarn, J. (1978) Diversité comparée des descendances de *Coffea arabica* obtiennes en autofécondation en fécondation libre an Tonkoui 75-78 ORSTOM, Man, Ivory Coast.

Louarn, J. (1982) Bilan des hybridations interspecifiques entre cafeiers africains diploides en collection en Cote d'Ivoire. *10th Conference of ASIC*, Salvador, Brazil. pp 369-474.

Louarn, J. (1993) Structure genetique des cafeiers africains diploides basée sur la fertilité des hybrides interspecifiques. *Communication au XV congres de 1'ASIC*, Montpellier, France.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, New York.

Markert, C.L. & Moller, F. (1959) Multiple forms of enzymes: tissues, ontogenetic and species specific patterns. *Proceedings of the National Academy of Sciences of the USA*, **45**, 753-763.

Mazzafera, P. & Carvalho A. (1992) Breeding for low seed caffeine content (*Coffea* L.) by interspecific hybridization. *Euphytica* **59**, 55-60.

Menancio, D.I., Hepburn, A.G. & Hymowitz, T. (1990) Restriction fragment length polymorphism (RFLP) of wild perennial relatives of soybean. *Theoretical and Applied Genetics*, **79**, 235-240.

Mendez, A.J.T. (1958) Advances in coffee production technology. Recent advances in our knowledge of coffee trees. *3 - Cytology. Coffee and Tea Industries*, **81**, 37-42.

Mendez, A.J.T. & Biacchi, O. (1940) Observacoes citologicas sem *Coffea* V. Una variedade haploide (di-haploide) de *C. arabica* L. *Instituto Agronomico de Campinas, Boletin Tecnico*, 77.

Michelmore, R.W., Paran, I. & Kesseli, R.V. (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the USA*, **88**, 9828-9832.

Miller, J.C. & Tanksley, S.D. (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theoretical and Applied Genetics*, **80**, 437-448.

Monaco, L.C. & Carvalho, A. (1964) Genetica de *Coffea*. Hibridos entre especies (Genetics of *Coffea*. Interspecific hybrids) *Ciencia e Cultura*, S. Paulo, **16(2)** 144.

Morden, C.W., Doebley, J.F. & Schwartz, K.F. (1990) Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (*Poaceae*). *Theoretical and Applied Genetics*, **80**, 296-304.

Moreno, G. (1989) Etude du polymorphisme de l'hybride de Timor en vue de l'amélioration du cafeier arabica. These de Docteur-ingenieur, ENSA Montpellier, France, 127 p.

Morgante, M. & Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant Journal*, **3**, 175-182.

Narasimhaswamy, R.L. (1962) Some thoughts on the origin of *Coffea arabica*. *Coffee*, **4(12)**, 1-5 Turrialba, Costa Rica.

Nei, M., & Li, W-H. (1979) Mathematical model for studying genetical variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the USA*, **74**, 5267-5273.

Newton, K.J. (1988) Plant mitochondrial genomes: organization, expression and variation. *Annual Review of Plant Physiology and Plant Molecular Biology*, **39**, 503-532.

Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L. & Landegren, U. (1990) Automated DNA diagnostics using an Elisa-based oligonucleotide ligation assay. *Proceedings of the National Academy of Sciences of the USA*, **87,** 8923-8927.

Nienhuis, J., Helentjaris, M., Slocum, M., Ruggero, B. & Schaefer, A. (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. *Crop Science*, **27**, 797-803.

O'Brien, S. (1993) Genetic Maps. Locus Maps of Complex Genomes. Book 6, Plants, sixth edition, Cold Spring Harbor Laboratory Press, NY.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. & Ohyama, K. (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. *Journal of Molecular Biology*, **223**, 1-7.

O'Donell, K. (1992) Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*). *Current Genetics*, **22**, 213-220.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. & Ozeki, H. (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, **322**, 572-574.

Olmstead, R.G. & Palmer, J.D. (1994) Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany*, **81(9)**, 1205-1224.

Olsen, G.J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994) FastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applied Biosciences*, **10**, 41-48.

Orozco-Castillo, C., Chalmers, K.J., Waugh, R. & Powell W. (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics*, **87**, 934-940.

Osborn, T.C., Alexander, D.C. & Forbes, J.F. (1987) Identification of restriction fragment length polymorphism linked to genes controlling soluble solid content in tomato fruit. *Theoretical and Applied Genetics*, **73**, 350-356.

Palmer, J.D. (1985a) Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntyre, R.J. (ed.) *Molecular Evolutionary Genetics*. Plenum Press, New York, pp. 131-240.

Palmer, J.D. (1985b) Comparative organization of chloroplast genomes. *Annual Review of Genetics*, **19**, 325-354.

Palmer, J.D. (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist*, **130**, S6-S29 (Supplement).

Palmer, J.D. (1992) Comparison of chloroplast and mitochondrial genome evolution in plants. In Herrmann, R.G. (ed.), *Cell Organelles*, 99-133. Springer-Verlag, Berlin.

Palmer, J.D. & Shields, C.R. (1984) Tripartite structure of the *Brassica campestris* mitochondrial DNA plasmid in the genus *Brassica*. *Nature*, **301**, 437-440.

Palmer, J.D. & Stein, D.B. (1986) Conservation of chloroplast genome structure among vascular plants. *Current Genetics*, **10**, 823-833.

Palmer, J.D., Shields, C.R., Cohen, D.B. & Orton, T.J. (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. *Theoretical and Applied Genetics*, **65**, 181-189.

Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.K. & Manhart, J.R. (1988) Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden*, **75**, 1180-206.

Paran, I. & Michelmore, R.W. (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics*, **85**, 895-993.

Pefley, E.B. & Orozco-Castillo, C. (1987) Polymorphism of isozymes within plant introductions of *Allium cepa* L. and *A. fistulosum* L. *Horticultural Science*, **22(5)**, 956-957.

Powell, W., Phillips, M.S., McNicol, J.W. & Waugh, R. (1991) The use of DNA markers to estimate the extent and nature of genetic variability in *Solanum tuberosum* cultivars. *Annals of Applied Biology*, **118**, 423-432.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. & Rafalski, A.J. (1995) The utility of RFLP, RAPD, AFLP and SSRP (microsatellite) for germplasm analysis. In preparation.

Pradhan, A.K., Prakash, S., Mukhopadhyay, A. & Pental, D. (1992) Phylogeny of *Brassica* and allied genera based on variation in chloroplast and mitochondrial DNA patterns: molecular and taxonomic classifications are incongrous. *Theoretical and Applied Genetics*, **85**, 331-340.

Ragot, M. & Hoisington, D.A. (1993) Molecular markers for plant breeding. Comparisons of RFLP and RAPD genotyping costs. *Theoretical and Applied Genetics*, **86**, 975-984.

Reed, K.C. & Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Research*, **13**, 7207-7221.

Reiter, R.S., Williams, J.G.K., Feldmann, K.A., Rafalski, J.A., Tingey, S.V. & Scolnik, P.A. (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbreds and random amplified polymorphic DNAs. *Proceedings of the National Academy of Sciences of the USA*, **89**, 1477-1481.

Reyner, J.F., Pernes, J. & Chaume, R. (1978) Diversité observée sur les descendances issues de pollinisation libre au Tonkoui. In: Charrier, A. (ed.) Etude de la structure et de la variabilite génétique des cafeiers. *IFCC Bulletin No. 14*, pp. 69-74.

Rick, C.M. & Tanksley, S.D. (1981) Genetic variation in *Solanum pennellii*: comparisons with two other sympatric tomato species. *Plant Systematics and Evolution*, **139**, 11-45.

Rieger, T., Michaelis, A. & Green, M.M. (1976) Glossary of genetics and cytogenetics, Classical and Molecular. Springer-Verlag, Berlin, Heidelberg, New York.

Rieseberg, L.H., & Seiler, G. (1990) Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus* L.). *Economic Botany*, **445**, 79-91.

Rieseberg, L.H., Soltis, D.E. & Palmer, J.D. (1988) A molecular reexamination of introgression between *Helianthus annuus* and *H. bolanderi* (*Compositae*). *Evolution*, **42**, 227-238.

Rieseberg, L.H., Beckstrom-Sternberg, S. & Doan, K. (1990) *Helianthus annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences of the USA*, **87**, 593-597.

Ritter, E., Gebhardt, C. & Salamini, F. (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics*, **125**, 645-654.

Rodrigues, C.J. Jr, Bettencourt, A.J. & Rijol, L. (1975) Races of the pathogen and resistance to coffee rust. *Annual Review of Phytopathology*, **13**, 49-70.

Roy, A., Frascaria, N., Mackay, J. & Bousquet, J. (1992) Segregating random amplified polymorphic DNA (RAPDs) in *Betula alleghaniensis*. *Theoretical and Applied Genetics*, **85**, 173-180.

Russell, J.R. (1994) Molecular variation in *Theobroma* species. PhD thesis. University of Reading, Reading, U.K.

Russell, J.R, Hosein, F., Johnson, E., Waugh, R. & Powell W. (1993) Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. *Molecular Ecology*, **2**, 64-70.

Saitou, N. & Imanishi, T. (1989) Relative efficiencies of the Fitch-Margoliash, maximum parsimony, maximum-likelihood, minimum-evolution and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biology and Evolution*, **6**, 514-525.

Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406-425.

Sarfatti, M., Katan, J., Fluhr, R. & Zamir, D. (1989) An RFLP marker in tomato linked to the *Fusarium oxysporum* resistance gene I2. *Theoretical and Applied Genetics*, **78**, 755-759.

Sederoff, R.R. 1987. Molecular mechanisms of mitochondrial-genome evolution in higher plants. *American Naturalist*, **130**, S30-S45 (Supplement).

Serikawa, T., Kuramoto, T., Hilbert, P., Mori, M., Yamada, J., Dubay, C.J., Lindpainter, K., Ganten, D., Guenet, J.L., Lathrop, G.M. & Beckman, J.S. (1992) Rat gene mapping using PCR-analysed microsatellites. *Genetics*, **131**, 701-721.

Shattuck-Eidens, D.M., Bell, R.N., Neuhausen, S.L. & Helentjaris, T. (1990) DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. *Genetics*, **126**, 207-217.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W. & Stone, E.M. (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics*, **16**, 325-332.

Shinozaki, K., Ohme, M. Tanaka, M., Wakasugi, T., Haysida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kata, A., Tohdoh, N., Shimada, H. & Sugiura, M. (1986) The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Molecular Biology Reports,* **4**, 110-147.

Singh, R.B. & Williams, J.T. (1984) Maintenance and multiplication of plant genetic resources. In: Holden, J.W. & Williams, J.T. (eds) *Crop Genetic Resources: Conservation and Evaluation.* Allen and Unwin, London. pp 120-127.

Smith, R. (1988) History of Coffee. In: Clifford, M.N. & Wilson, K.C. (eds) *Coffee: Botany, Biochemistry and Production of Beans and Beverage*. Croom Held Ltd Publishers, Beckenham, UK.

Smith, R.L. & Sytsma, K.J. (1990) Evolution of *Populus nigra* (sect. *Aigeiros*): introgressive hybridization and the chloroplast contribution of *Populus alba* (sect. *Populus*). *American Journal of Botany*, **77**, 1176-1187.

Soller, M. & Beckmann, J.S. (1983). Genetic polymorphism in varietal identification and genetic improvement. *Theoretical and Applied Genetics*, **67**, 25-33.

Song, K.M., Osborn, T.C. & Williams, P.H. (1988) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 2. Preliminary analysis of subspecies within *B. rapa* (sym. *campestris*) and *B. oleracea*. *Theoretical and Applied Genetics*, **76**, 593-600.

Sugiura, M. (1989) The chloroplast chromosomes in land plants. *Annual Review of Cell Biology*, **5**, 51-70.

Sybenga, J. (1960) Genetics and cytology of coffee. A literature review. *Bibliographia Genetica*, **19**, 217-316.

Sytsma, K.J. & Schaal, B.A. (1985) Phylogenetics of *Lisanthius skinneri (Gentianaceae)* species complex in Panama utilising DNA restriction fragment analysis. *Evolution*, **39**, 594-608.

Taberlet, P., Gielly, L. Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105-1109.

Tanksley, S.D. & Orton, T.J. (Eds) (1983) *Isozymes in Plant Genetics and Breeding* 1A. Elsevier, Amsterdam, 516 pp.

Tanksley, S.D., Bernatsky, R., Lapitan, N.L. & Prince, J.P. (1988) Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of the National Academy of Sciences of the USA*, **85**, 6419-6423.

Tanksley, S.D., Young, N.D., Paterson, A.H. & Bonierbale, M.W. (1989) RFLP mapping in plant breeding: New tools for an old science. *Biotechnology*, **7**, 257-263.

Thomas, A.S. (1942) 'The wild arabica coffee of the Boma Plateau Anglo-Egyptian Sudan'. *Empire Journal of Experimental Agriculture*, **10**, 207-212.

Thompson, J.D., Higgins, D.G., & Gibson T.J. (1994) Clustal-W-Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, **22**, 4673-4680.

Todd, J.A. (1992) La carte des microsatellites est arrivée! *Human Molecular Genetics*, **1**, 663-666.

Tragoonrung, S., Kanazin, V., Hayes, P.M. & Blake, T.K. (1992) Sequence-tagged-site facilitated PCR for barley genome mapping. *Theoretical and Applied Genetics,* **84**, 1002-1008.

Van der Vossen, H.A.M. (1974) Plant breeding. *Coffee Research Foundation of Kenya, Annual Report, 1973-74*. pp. 40-51.

Van der Vossen, H.A.M. (1985) Coffee selection and breeding. In: Clifford, M.N. & Wilson, K.C. (eds), *Coffee: Botany, Biochemistry and Production of Beans and Beverage*. Croom Held Ltd Publishers, Beckenham, UK.

Van der Vossen, H.A.M. & Walyaro, D.J. (1980) Breeding for resistance to coffee berry disease in *Coffea arabica* L. II. Inheritance of the resistance. *Euphytica*, **29**,777-791.

Vishveshwara, S. (1963) A preliminary report on meiotic irregularities in diploids and diploid interspecific hybrids in coffee. *Caryologia*, **16**, 535-539.

Vosberg, H.P. (1989) The polymerase chain reaction: an improved method for the analysis of nucleic acids. *Human Genetics*, **83**, 1-15.

Wachira, F.N., Waugh R., Hackett, C.A. & Powell W. (1995) Detection of genetic diversity in tea (*Camelia sinensis*) using RAPD markers. *Genome*, **38**, 201-210.

Wakasugi, T., Tsudsuki, J., Ito, S., Nakashima, K., Tsudsuki, T. & Sugiura, M. (1994) Loss of all NDH genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thumbergii*. *Proceedings of the National Academy of Sciences of the* USA, 21, 9794-9798.

Walters, T.W., Posluszny, U. & Kevan, P.G. (1989) Isozyme analysis of grape (*Vitis* I.) A practical solution. *Canadian Journal of Botany*, **67**, 2894-2899.

Waugh, R. & Powell, W. (1992) Using RAPD markers for crop improvement. *Trends in Biotechnology*, **10**, 186-191.

Weber, J. & May, P.E. (1989) Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, **44**, 388-396.

Weeden, N.J. & Lamb, R.C. (1985) Identification of apple cultivars by isozyme phenotypes. *Journal American Society of Horticultural Sciences*, **110**, 509-515.

Weining, S. & Langridge, P. (1991) Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. *Theoretical and Applied Genetics*, **82**, 209-216.

Welsh, J. & McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, **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 Research*, **20**, 303-306.

Wendel, J.F. & Weeden, N.F. (1990) Visualisation and interpretation of plant isozymes. In: Soltis, D.E. & Soltis, P.S. (eds), *Isozymes in Plant Biology*. Chapman and Hall, London, UK.

White, T.J., Arnheim, N. & Erlich, H.A. (1989) The polymerase chain reaction. *Trends in Genetics*, **5**, 185-89.

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

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**, 6531-6535.

Wolfe, K.H., Li, W.-H. & Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Sciences of the USA*, **84**, 9054-9058.

Wolfe, K.H., Morden, C.W. & Palmer J.D. (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences of the USA*, **22**, 10648-10652.

Wright, F. (1993) Sequence analysis course SA3: Phylogenetic trees from molecular sequences. Scottish Agricultural Statistics Service, Edinburgh University.

Xiong, B. & Kocher, T.D. (1991) Comparisons of mitochondrial DNA sequences of seven morphospecies of black flies (*Diptera*: *Simuliidae*). *Genome*, ***34***, 306-311.

Yang, X. & Quiros, C. (1993) Identification and classification of celery cultivars with RAPD markers. *Theoretical and Applied Genetics*, **86**, 205-212.

Zabeau, M. & Voss, P. (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application 92402629.7.

Zhao, X. & Kochert G. (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.) *Molecular General Genetics*, **231**, 353-359.

Zurawski, G., Clegg, M.T. & Brown A.H.D. (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics*, **106**, 735-749.

Zurawski, G. & Clegg, M.T. (1987) Evolution of higher-plant chloroplast DNA encoded genes: implications for structure-function and phylogenetic studies. *Annual Review of Plant Physiology*, **38**, 391-418.

APPENDIX

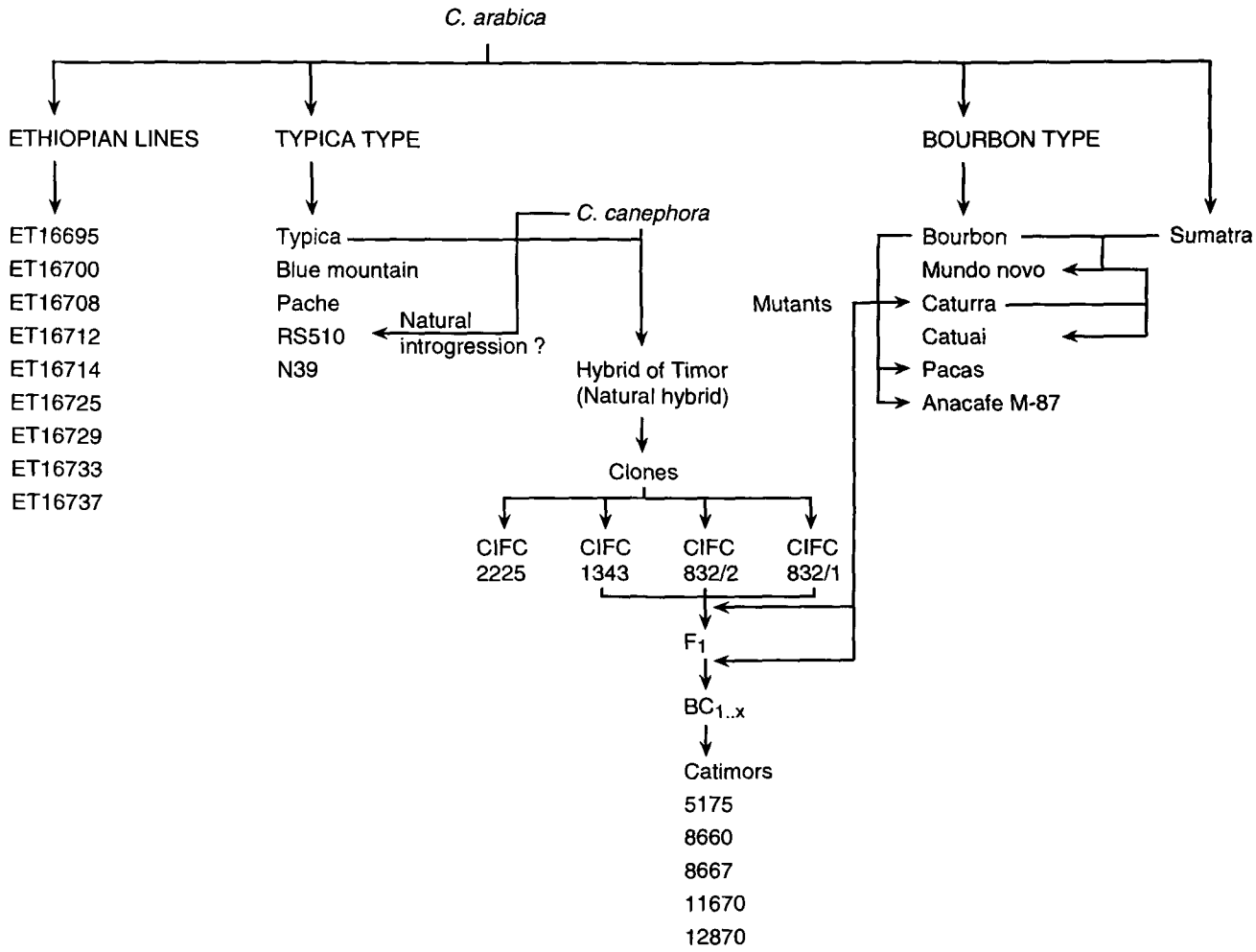


Figure A. Relationships between *C. arabica* accessions used in the study on molecular diversity.

Table A. Description and relationships of *Coffea arabica* germplasm used in the study on molecular diversity

1. TYPICA TYPE

1.1. Typica

Prototype variety for the description of *C. arabica*, as a comparison model with other varieties. The general characteristics are as follow: Plant height from 3.5 to 4 meters, conical shape, usually an individual trunk, with orthotropic branches originated from the nodes, prolific plagiotropic branches originated from the nodes of the former branches, which form angles between 50 and 70 ° in relation to the central vertical axe. The leaves have oblong-elliptical shape, with acute base and apex and glabrous surface. The young leaves have brown colour. The leaf-blade and border are flat and slightly undulated, with size from 8 to 18 cm length by 2 to 7 cm width, depending on the age and height of the plant. The inflorescence consists of very short lateral axes with short pedicellate flowers disposed in axillar glomerules, with generally three to five in each glomerule. The coffee fruit is a drupe normally containing two seeds. The coffee seed is elliptical, plane-convex, possessing a longitudinal furrow on the plane surface. The typica type is considered to be lower yielding in comparison to the other coffee varieties.

1.2. Blue mountain

Cultivar developed from a Typica population in Jamaica, with most of the characteristics of the typica type but higher yielding.

1.3. Pache

Dwarf cultivar originated in Guatemala, and was identified as a mutation of the Typica variety. It is characterized by its lower growth with high number of orthotropic and plagiotropic branches, shorter internodes and abundant foliage.

1.4. RS-510

Variety derived from seed collected from wild *C. arabica* growing in the Boma Plateau, Rume Valley, South East Sudan, where this species co-exist with *C. canephora*. One of the particular characteristic of RS-510 is that it carries the dominant R - and the recessive K - genes for resistance to coffee berry disease (Van der Vossen and Walyaro, 1980).

1.5. N-39

Variety developed in Tanzania, with the characteristics similar to that of the typica type, but with higher yield and quality.

2. BOURBON TYPE

2.1. Bourbon

The original variety originated from coffee seeds introduced in South America from the Island of La Réunion, formerly called Bourbon. In comparison to the Typica variety, Bourbon is a larger plant which has more orthotropic and plagiotropic branches, at a more reduced angle in relation to the central vertical axis, shorter internodes and a higher amount of axillary glomerulus; the young leaves are green, the leaves wider and the leaf margins more undulated. The fruit is smaller, as it is the seed. The better vigour and conformation, as well as the higher number of buds, give Bourbon a higher yield than Typica.

2.2. Caturra

Variety derived form a dwarf mutant (Ct) plant found in a Bourbon population in Minas Gerais, Brazil. It is characterized by shorter height, with a thicker principal trunk, with prolific orthotropic and plagiotropic branches, and short internodes. It is a high yielding variety under appropriate management.

2.3. Mundo novo

This variety is considered to be derived from selection carried out within a natural segregating population from hybridization between the *C. arabica* cultivars Bourbon and Sumatra. It originated in Urupes, São Paulo, Brazil.

2.4. Catuai

This variety originated from a hybridization between *C. arabica* cultivars Mundo Novo and Caturra. The development of this variety was conducted in the Campinas Agricultural Research Institute, São Paulo, Brazil, in 1949. It is a low-growing plant with cylindrical shape.

2.5. Pacas

A Bourbon mutant similar to Caturra, which originated in El Salvador. It is a low-growing plant, with short internodes and prolific secondary branches and abundant foliage.

2.6. Anacafe M-87

A Bourbon mutant found in Guatemala, with similar characteristics to that of Caturra and Pacas.

3. ETHIOPIAN LINES

The Ethiopian lines were obtained from the field collection in CATIE, Costa Rica and the description in the germplasm record book is as follow:

Introuccion Number	Description	Location in the field	Source	Origin
16695	IRCC-204 ET 6	A2	IRCC, France	Ethiopia
16700	IRCC-210 ET 11C	A7	IRCC, France	Ethiopia
16708	IRCC-219 ET 19	A3	IRCC, France	Ethiopia
16712	IRCC-225 ET 25	A4	IRCC, France	Ethiopia
16714	IRCC-227 ET 27	A8	IRCC, France	Ethiopia
16725	IRCC-239 ET 41	A7	IRCC, France	Ethiopia
16729	IRCC-244 ET 47	A4	IRCC, France	Ethiopia
16733	IRCC-249 ET 52	A2	IRCC, France	Ethiopia
16737	IRCC-255 ET 57	A2	IRCC, France	Ethiopia

4. HYBRIDS

4.1. Hybrid of Timor

This is a tetraploid hybrid from a natural cross between *C. arabica*, 'Typica' type and *C. canephora*, found in the Island of Timor. Different Hybrid of Timor clones have been used in breeding programs, such as CIFC 832/1, CIFC 832/2, CIFC 1343, and CIFC 2252. Hybrid of Timor possesses the dominant T- gene conferring resistance to Coffee Berry Disease (*Colletotrichum coffeanum*) and the genes for resistance to all known races of leaf rust (*Hemileia vastatrix*)(Rodriguez et al., 1975).

The hybrid of Timor plant used in the study on molecular diversity of coffee was derived from seed obtained from the field collection in CATIE and was identified as Hybrid of Timor 2252/28.

The hybrid of Timor accession used in the preliminary study on a F2 mapping population was identified as VCE1592 (HT CIFC 1343) and was obtained from LARI (Lyamungu Agricultural Research Institute), Moshi, Tanzania.

4.2. Catimors

The Catimor accessions were obtained from the field collection in CATIE (Centro Agronomico Tropical de Investigacion y Enseñanza), Costa Rica and ANACAFE (Asociacion Nacional del Cafe), Guatemala. Catimors were derived from segregating crosses between *C. arabica*, variety Caturra, and different Hybrid of Timor clones. The detailed description of the Catimor material is described below.

4.2.1. Catimor 5175

CIFC HW-26/13 19/1 Caturra x 832/1 Hybrid of Timor/13.

4.2.2. Catimor 8660

Catimor UFV 2762 19/1 Caturra X 832/1 Hybrid of Timor F_5 UFV 1359-45 F_4 UFV 386-45 F_3 IIAA 857-3 F_2 CIFC HW-26/5 F_1.

4.2.3. Catimor 8667

Catimor UFV 3005 19/1 Caturra X 832/1 Hybrid of Timor F_5 UFV 1348-150 F_4 UFV 386-58 F_3 IIAA 857-3 F_2 CIFC HW-26/5 F_1.

4.2.4. Catimor 11670

CCC-135 yellow Caturra CV 1 X CCC 48-1574 CV 2 Hybrid of Timor CIFC 1343 F_1 PTAS 1321 A 1330 F_3 PTAS 1321 a 1324.

4.2.5. Catimor 12870

UFV 4716 (2047-788EP1 UFV) Catimor CENICAFE F_4 x UFV 180/139

Table B.1. Pairwise distance matrix for the chloroplast intergenic spacer trnL-trnF sequences. (Using Kimura's 2-parameter distance).

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
1. C. brevipes																
2. C. canephora 3751	0.0051															
3. C. canephora 3753	0.0025	0.0025														
4. C. arabica	0.0077	0.0128	0.0102													
5. C. congensis	0.0051	0.0102	0.0077	0.0129												
6. C. eugenioides	0.0077	0.0128	0.0102	0.0000	0.0129											
7. C. dewevrei	0.0079	0.0079	0.0053	0.0159	0.0132	0.0159										
8. G. thumburgia	0.0239	0.0293	0.0266	0.0267	0.0267	0.0267	0.0302									
9. C.humilis	0.0129	0.0180	0.0154	0.0102	0.0181	0.0102	0.0213	0.0321								
10. I. findlaysoniana	0.0489	0.0545	0.0517	0.0518	0.0546	0.0518	0.0562	0.0416	0.0575							
11. C. liberica	0.0000	0.0051	0.0025	0.0077	0.0051	0.0077	0.0079	0.0239	0.0129	0.0489						
12. M. axillare	0.0266	0.0320	0.0293	0.0293	0.0321	0.0293	0.0302	0.0217	0.0348	0.0359	0.0266					
13. C. racemosa	0.0051	0.0102	0.0077	0.0025	0.0103	0.0025	0.0132	0.0240	0.0077	0.0490	0.0051	0.0266				
14. C. pseudozanguebariae	0.0103	0.0155	0.0129	0.0077	0.0155	0.0077	0.0186	0.0268	0.0129	0.0548	0.0103	0.0322	0.0051			
15. C. sessiliflora	0.0103	0.0154	0.0129	0.0077	0.0155	0.0077	0.0186	0.0267	0.0129	0.0547	0.0103	0.0321	0.0051	0.0051		
16. C. stenophylla	0.0154	0.0206	0.0180	0.0128	0.0207	0.0128	0.0240	0.0349	0.0025	0.0547	0.0154	0.0321	0.0102	0.0155	0.0154	
17. V. edulis	0.0567	0.0597	0.0567	0.0598	0.0627	0.0598	0.0616	0.0490	0.0658	0.0512	0.0567	0.0522	0.0568	0.0630	0.0628	0.0688

Table B.2. Pairwise distance matrix for the chloroplast intron trnL sequences. (Using Kimura's 2-parameter distance).

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. C. brevipes													
2. C. canephora 3751	0.0019												
3. C. canephora 3753	0.0000	0.0019											
4. C. arabica	0.0019	0.0038	0.0019										
5. C. congensis	0.0038	0.0057	0.0038	0.0057									
6. C. eugenioides	0.0019	0.0038	0.0019	0.0000	0.0057								
7. C. dewevrei	0.0019	0.0038	0.0019	0.0038	0.0057	0.0038							
8. C.humilis	0.0057	0.0076	0.0057	0.0038	0.0095	0.0038	0.0076						
9. I. findlaysoniana	0.0332	0.0352	0.0332	0.0352	0.0373	0.0352	0.0352	0.0352					
10. C. liberica	0.0076	0.0095	0.0076	0.0095	0.0114	0.0095	0.0095	0.0133	0.0392				
11. C. pseudozanguebariae	0.0057	0.0076	0.0057	0.0038	0.0095	0.0038	0.0076	0.0076	0.0353	0.0133			
12. C. racemosa	0.0057	0.0076	0.0057	0.0038	0.0095	0.0038	0.0076	0.0076	0.0393	0.0133	0.0038		
13. C. sessiliflora	0.0038	0.0057	0.0038	0.0019	0.0076	0.0019	0.0057	0.0057	0.0373	0.0114	0.0019	0.0019	
14. C. stenophylla	0.0038	0.0057	0.0038	0.0019	0.0076	0.0019	0.0057	0.0019	0.0332	0.0114	0.0057	0.0057	0.0038

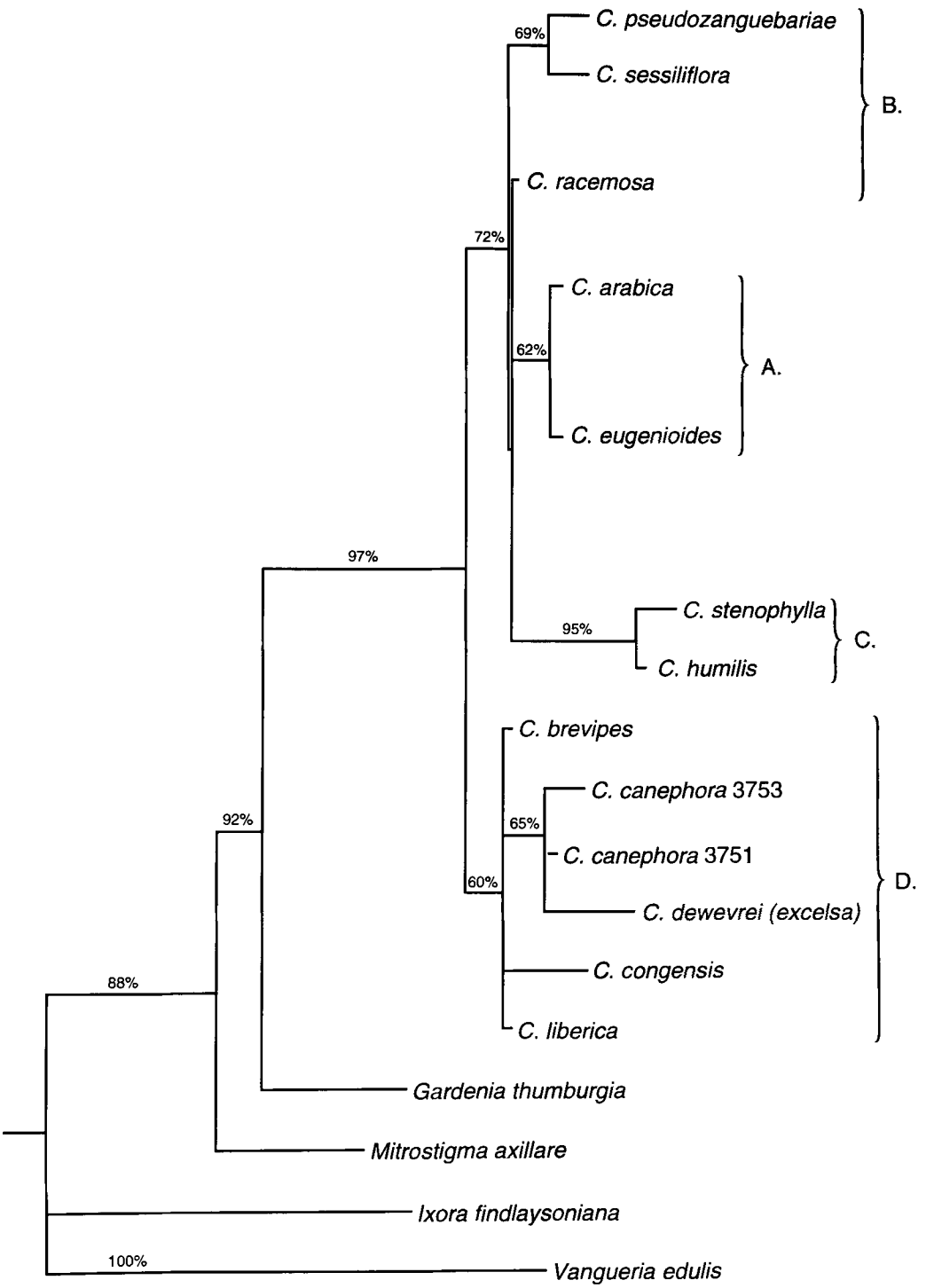


Figure B.1. Phylogenetic tree (drawgram) of *Coffea* species and four outgroup species derived from chloroplast trnL-trnF intergenic spacer sequences.

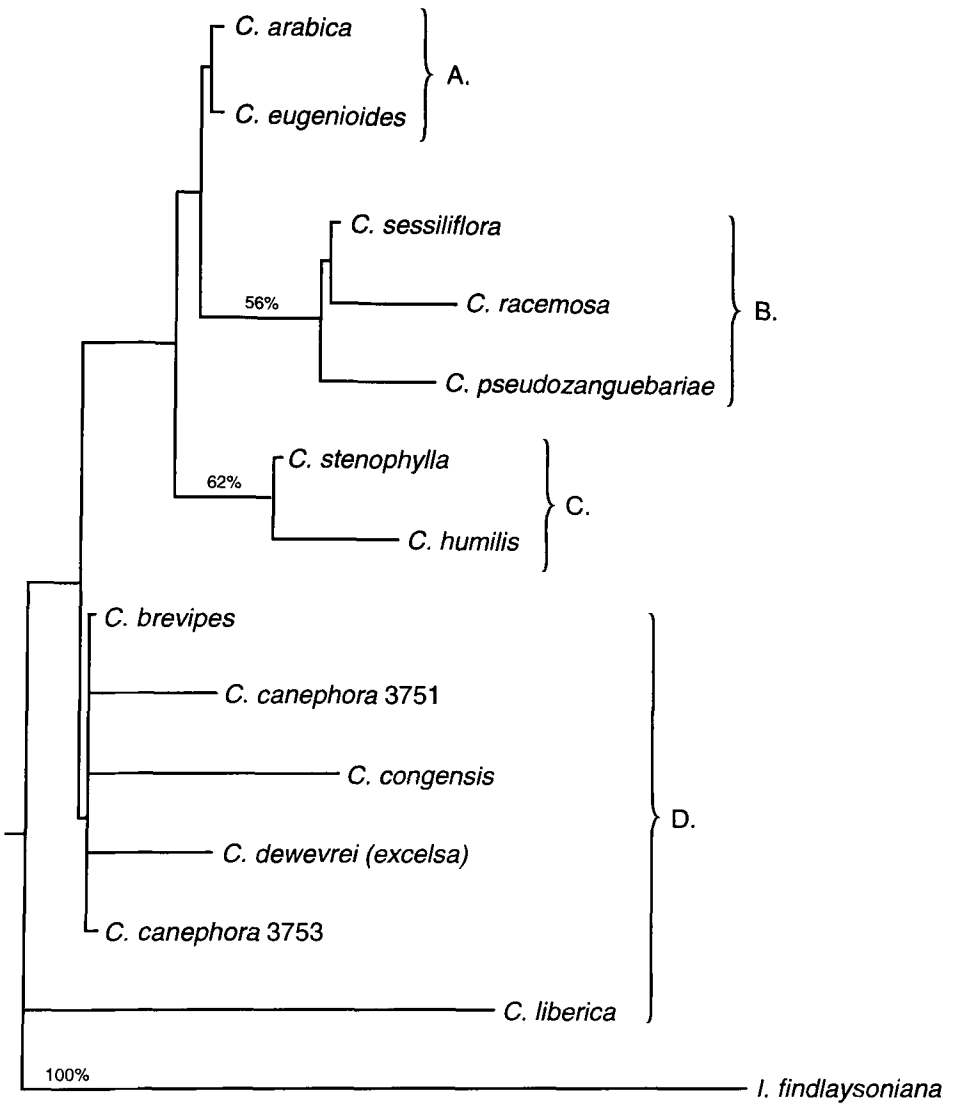


Figure B.2. Phylogenetic tree (drawgram) of *Coffea* species and one outgroup species derived from chloroplast trnL intron sequences.

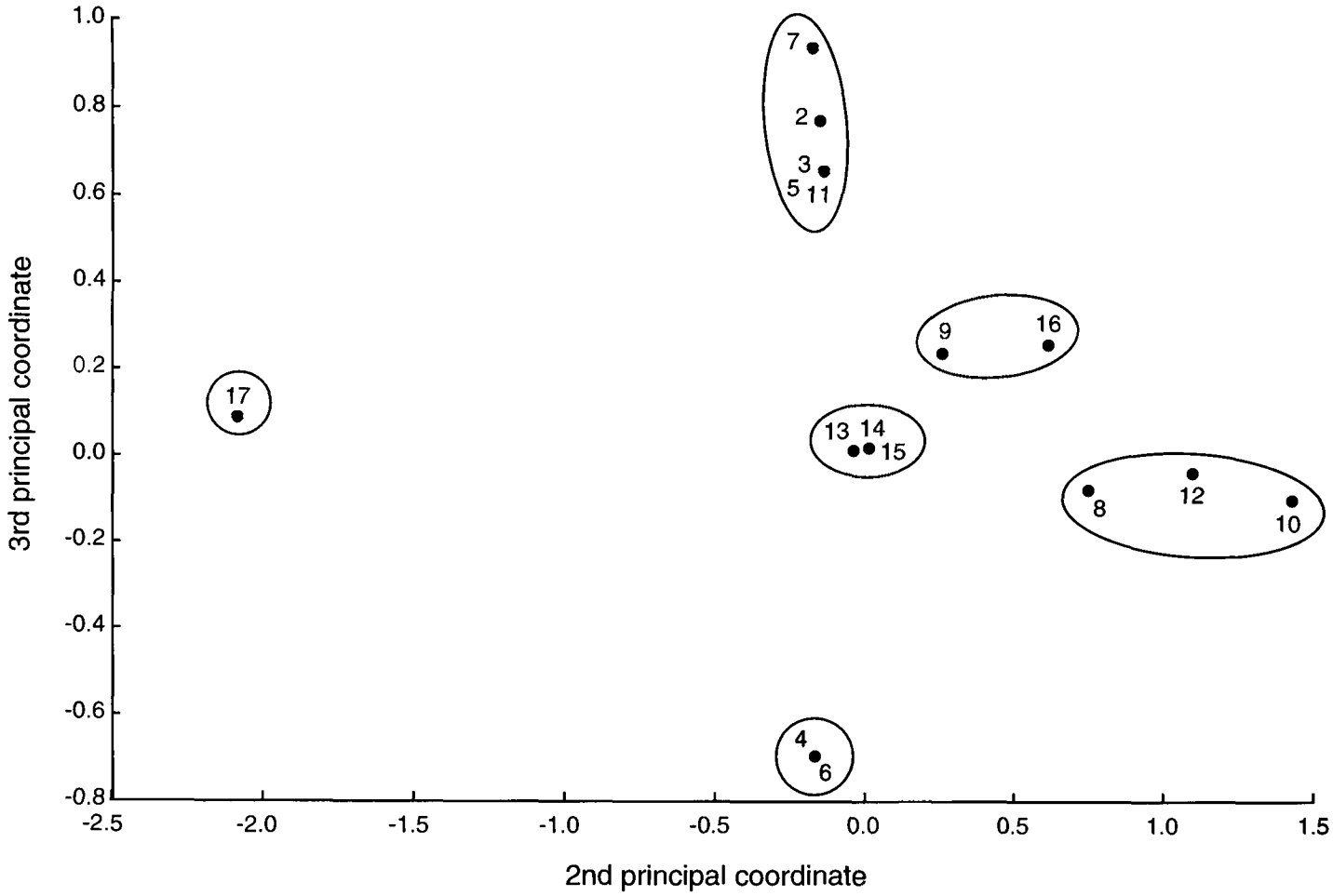


Figure C.1. Principal coordinate analysis of 14 coffee accessions representing 12 species and four outgroup species for the chloroplast trnL-trnF intergenic spacer sequences. The numbers in the figure correspond to the order number in Table 5.2.

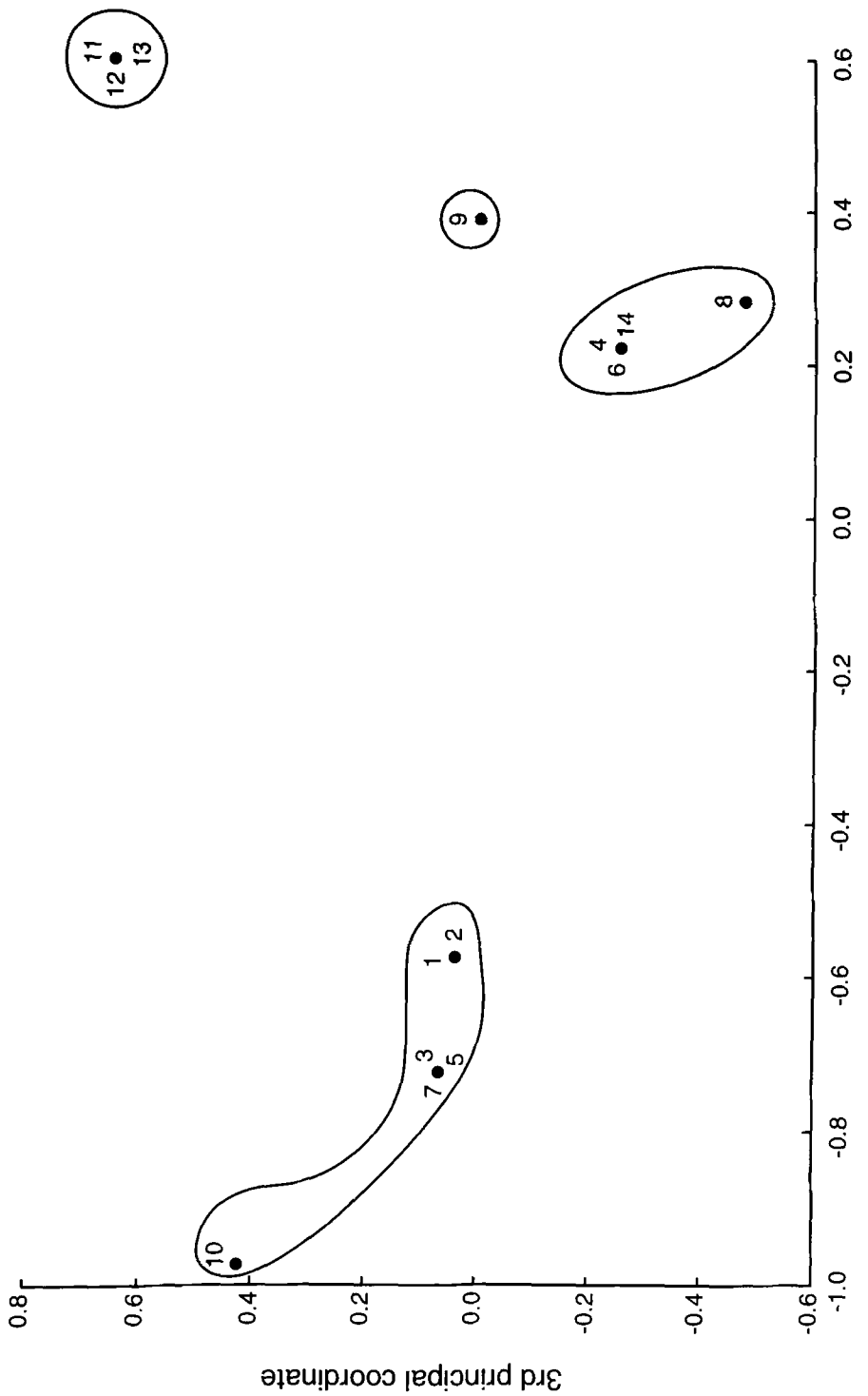


Figure C.2. Principal coordinate analysis of 14 coffee accessions representing 12 species and one outgroup species for the chloroplast trnL intron sequences. The numbers in the figure correspond to the order number in Table 5.3.

Theor Appl Genet (1994) 87:934–940

THEORETICAL AND APPLIED GENETICS

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

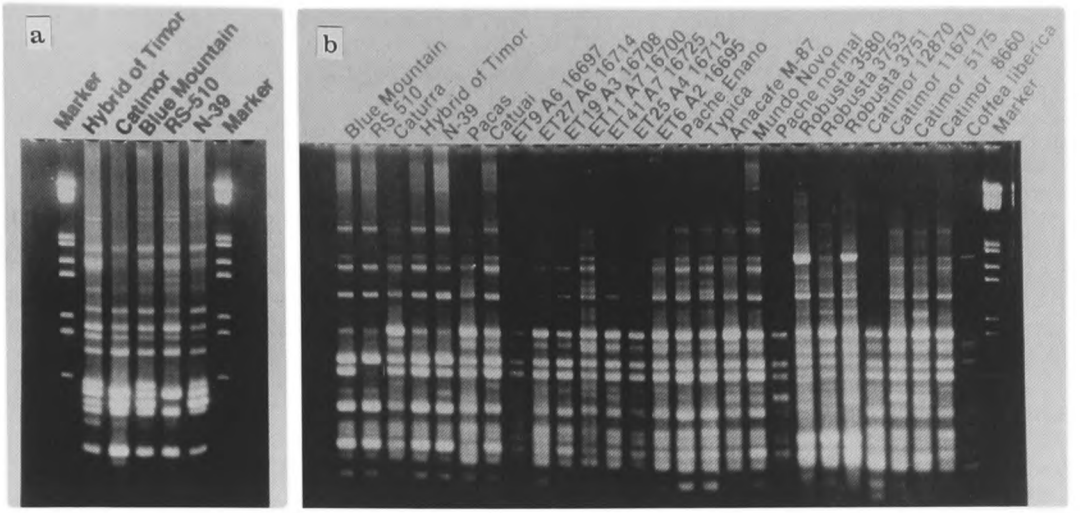


Fig. 1. **a** RAPD profiles of five *Coffea arabica* genotypes with primer SC10-33. **b** RAPD profiles of the 27 *Coffea* genotypes listed in Table 1 after amplification with primer SC10-15

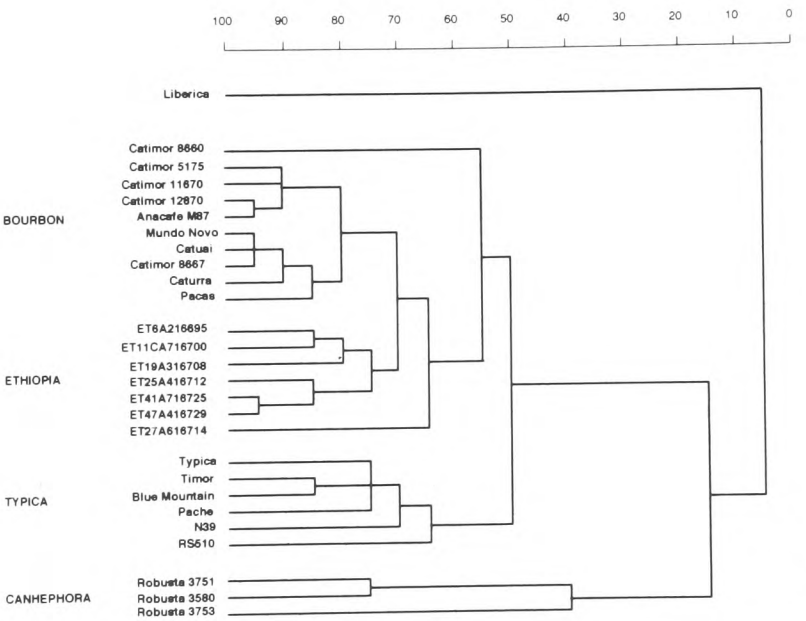


Fig. 2. Dendrogram of the *Coffea* accessions listed in Table 1 based on single linkage cluster analysis. Only 41 the informative polymorphic amplification products were used in the analysis.

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

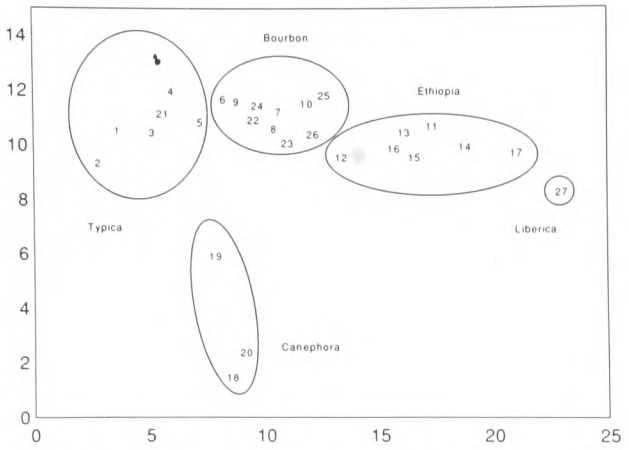


Fig. 3. Principle co-ordinate analysis of the 27 coffee accessions analysed using the data set derived from 41 polymorphic markers

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

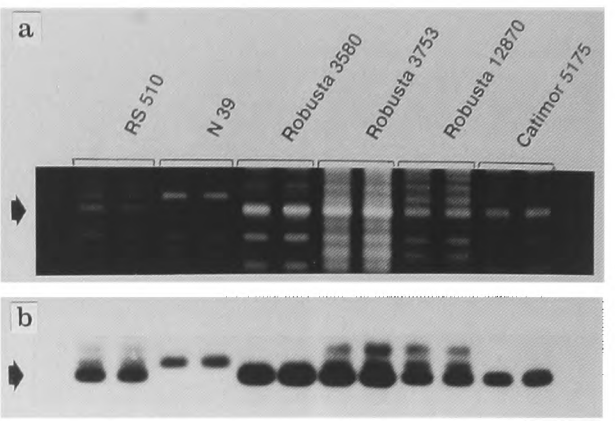


Fig. 5. a Amplification products from RS510, N39, Robusta 3753, 3751, 12870 and Catimor 5175 generated with primer SC10-30. The product diagnostic of interspecific gene flow is *arrowed.* **b** The arrowed band was used to probe a Southern-blot of amplified fragments generated by SC10-30. Hybridisation 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