PHYLOGENETIC RELATIONSHIPS OF COFFEE SPECIES AND ORIGIN OF *Coffea arabica* TETRAPLOID GENOME

Lashermes P.[1], Combes M.C.[1], Topart P.[2] & Anthony F.[2]

[1] IRD, BP 5045, 34032 Montpellier cedex 1, Francia. [2] CATIE, 7170 Turrialba, Costa Rica.

Key words: *Coffea arabica*, coffee, phylogenetic relationships, cpDNA polymorphism, rDNA polymorphism

Introduction

Coffee-trees belong to the tribe *Coffeae* in the family *Rubiaceae*. The subgenus *Coffea* consists of approximately 100 taxa so far identified in African and Madagascan intertropical forests. *Coffea arabica* L. is both the most widely cultivated species of *Coffea* and the only tetraploid species ($2x = 44$) in the genus. Arabica coffee has its primary centre of genetic diversity in the highlands of South West Ethiopia and the Boma Plateau of Sudan. Populations of *C. arabica* have been also reported in Mount Imatong (Sudan) and Mount Marsabit (Kenya). Carvalho (1952) suggested an allotetraploid origin since *C. arabica* presents a diploid meiotic behaviour and a centre of genetic diversity situated outside the distribution area of the diploid coffee species. According to Grassias and Kammacher (1975), and based on cytogenetic observation, *C. arabica* has to be considered as a segmental allotetraploid.

In recent years, DNA-based genetic markers have been developed which offer new potential in analysis of genetic diversity and in elucidating the evolutionary history of plants. In this report, recent results obtained with *C. arabica* are presented.

Molecular genetic characterisation of *C. arabica*

Phylogenetic relationships inferred from chloroplast DNA variation

The low frequency of structural changes in the chloroplast molecule (cpDNA) together with a conservative rate of sequence evolution (Olmstead and Palmer, 1994) make it an ideal target for plant phylogenetic study. Maternal inheritance of cpDNA in coffee has been established in interspecific hybrids between *C. arabica* and *C. canephora* (4x) and in an intraspecific progeny of *C. canephora* (Berthou et al., 1983 ; Lashermes et al., 1996).

CpDNA variations have been investigated in the main coffee species and undetermined taxa (Cros et al., 1998). RFLP (restriction fragment length polymorphism) analysis of cpDNA using homeologous probes from lettuce (*Lactuca sativa*) was accomplished. In addition, the sequence of the *trn*L-*trn*F intergenic region was established. The overall chloroplast genome showed a low level of polymorphism while the intergenic sequence (*trn*L-*trn*F) appeared more polymorphic. A phylogenetic analysis (Figure 1) using Wagner parsimony was performed.

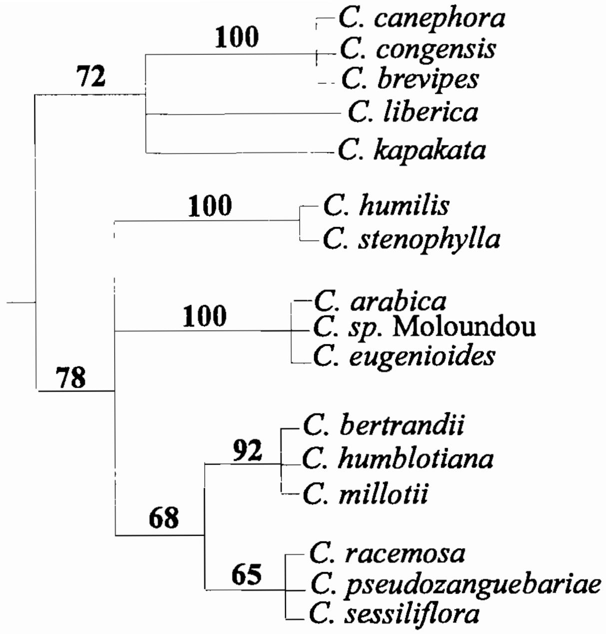

Figure 1. Phylogenetic tree of *Coffea* species based on chloroplast DNA variation. Strict consensus of the most parsimonious Wagner trees is represented. Values (%) on branches are bootstrap indices of support.

Several clades are revealed which are to some extent consistent with the classical biogeographical grouping (i.e. Madagascar, East Africa, West Africa). Results confirmed a monophyletic origin of *Coffea* species. CpDNA from *C. arabica* appeared similar to cpDNA from *C. eugenioides* and *Coffea sp.* Moloundou, suggesting that *C. arabica* could have diverged maternally from a species related to those species. Chloroplast genomes from *C. canephora* and *C congensis* were found to be identical, as previously reported by Berthou et al. (1983) following a total cpDNA RFLP analysis.

RFLP analysis using single-copy nuclear probes

A study was conducted to determine relationships among a series of *Coffea* species including *C. arabica* by comparing restriction fragment patterns. Probes from nuclear genomic arabica and arabusta libraries were selected to be single-copy using doubled haploid genotypes of *C. canephora*. RFLP-based distances between *C. arabica* and a large number of species were estimated (Table 1). When several accessions from the same species were analysed, the average distance is reported. *C. congensis*, *C. canephora* and *C. eugenioides* seemed to be the closest species to *C. arabica.* All distance values were higher than the expected one if *C. arabica* was an autotetraploid resulting from the duplication of one of the diploid species studied.

Table 1. Distances (complement of the Jaccard index) between *C. arabica* and a representative panel of diploid *Coffea* species based on RFLP data obtained using nine nuclear single-copy probes.

Species	**Distribution area**	**Distance to *C. arabica***
C. canephora	West and Central Africa	0.70 - 0.80
C. congensis	West and Central Africa	0.70 - 0.80
C. eugenioides	Central Africa	0.70 - 0.80
C. humilis	West Africa	0.70 - 0.80
C. sp. Moloundou	Central Africa	0.80 - 0.85
C. sp. X	Unknown	0.80 - 0.85
C. brevipes	Central Africa	0.85 - 0.90
C. kapakata	Central Africa	0.85 - 0.90
C. liberica	West and Central Africa	0.85 - 0.90
C. salvatrix	East Africa	0.85 - 0.90
C. stenophylla	West Africa	0.85 - 0.90
C. farafanganensis	Madagascar	0.90 - 1
C. humblotiana	Comores islands	0.90 - 1
C. millotii	Madagascar	0.90 - 1
C. pseudozanguebariae	East Africa	0.90 - 1
C. racemosa	East Africa	0.90 - 1

Nuclear ribosomal DNA sequence analysis

Among nuclear gene regions, the rDNA repeat unit is attractive for phylogeny reconstruction and genetic studies because of its ubiquity in all organisms, rapid concerted evolution, and the diverse rates of evolution observed within and among component subunits and spacers (reviewed in Jorgansen and Cluster, 1988). The internal transcribed spacer region ITS2 of 18-26S nuclear ribosomal DNA was sequenced for a number of *Coffea* species, including two genotypes of *C. arabica* (Caturra and Et 12).

No evidence of ITS length variants or major sequence variants within arabica accessions was found. *C. arabica* genotypes showed only one major type of sequence although important ITS2 nucleotide sequence variations were observed between species (Lashermes et al., 1997).

Analysis for a restricted number of species showed (Figure 2) that the ITS2 region of *C. arabica* diverged markedly from the sequences of *C. eugenioides* and its sister-group (*C. kiwuensis* and *Coffea sp.* Moloundou), and appeared almost identical to the sequences of canephoroid species (*C. canephora*, *C. congensis* and *C. brevipes*).

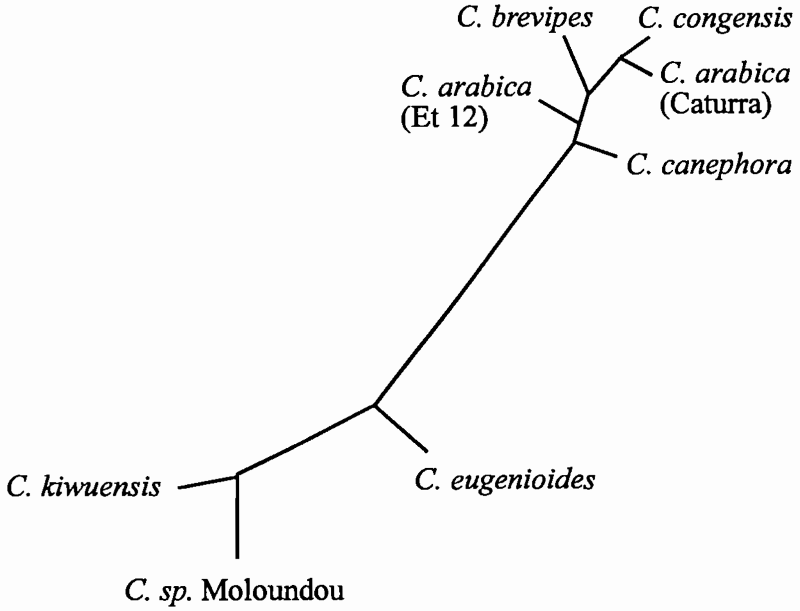

Figure 2. Parsimony analysis of ITS2 sequences of nuclear ribosomal DNA among *Coffea* species as putative ancestors of *C. arabica*. Branch lengths correspond to numbers of informative mutations.

In situ *hybridisation*

In situ hybridisation was carried out on chromosome preparations of *C. arabica* using digoxigenin-labelled total DNA from *C. canephora* as one probe and biotin-labelled total DNA from *C. eugenioides* as the second (Lashermes et al., 1999). The use of a specific filter to detect either FITC or Texas Red showed that two groups of chromosomes were clearly differentiated in *C. arabica* (Figure 3). Twenty two chromosomes showed a predominant yellow fluorescence suggesting a stronger affinity with the total genomic DNA probe from *C. canephora.* The remaining 22 chromosomes appeared in red-orange suggesting that the red coloration due to the total genomic DNA from *C. eugenioides* was more intense than the yellow-green one. Both genomic DNA hybridised strongly to the centromeric regions with weaker hybridisation along the chromosome arms.

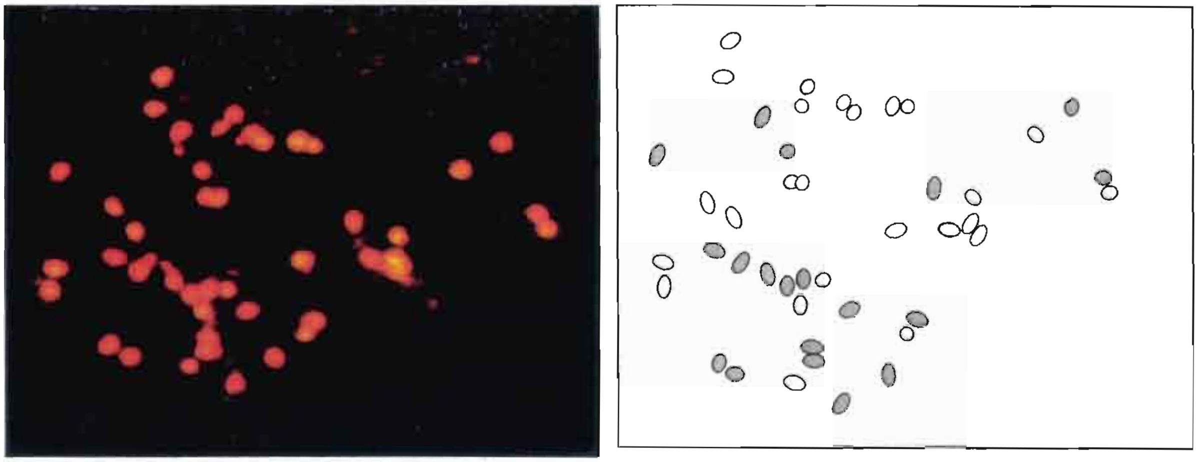

Figure 3. a) Preparation following simultaneous *in situ* hybridisation with digoxigenin-labelled total DNA from *C. canephora* and biotin-labelled total DNA from *C. eugenioides*. The two signals were superimposed by double exposure. **b)** Schematic representation of a) showing the specific origin of *C. arabica* chromosomes. The stippled ovals represent chromosomes that exhibit stronger hybridisation with the *C. eugenioides* probe.

Origin of *C. arabica* tetraploid genome

Earlier attempts to determine the genetic origin of *C. arabica* relied on analysis of meiotic behaviour of *C. arabica*, karyotyping (Bouharmont, 1959), chromosome pairing in hybrids with diploid species (Krug and Mendes, 1940; Kammacher and Capot, 1972) and in dihaploid plants of *C. arabica* (Vishveshwara, 1960; Berthaud, 1976; Kammacher, 1980). These studies have revealed marked chromosome affinity and the absence of substantial chromosome differentiation between the two constitutive genomes of *C. arabica,* and between *C. arabica* and the diploid *Coffea* species. The normal diploid behaviour of *C. arabica* is thought to be due to a genetic system (Grassias and Kammacher, 1975). Investigation of the origin of *C. arabica* can be based on the results of the different DNA sequence evolution studies.

The allotetraploid origin of *C. arabica* is corroborated by the analysis of polymorphism observed by RFLP. In addition, hypothese involving intergeneric combination or association of distant *Coffea* species are improbable. Work on the chloroplast genome strongly supports the notion that a species close to *C. eugenioides* donated the maternal genome of *C. arabica.* Analysis of rDNA showed that the paternal parent was a species from the canephoroid group *(C. canephora*, *C. congensis*). The very low divergence between ITS2 sequences of canephoroid species and *C. arabica*, as well as the similarity of the chloroplastic *trn*L-*trn*F intergenic sequences from *C. arabica*, *C. eugenioides* and *C. sp.* Moloundou, clearly indicate that formation and speciation of *C. arabica* are recent events and most likely occurred during the late quaternary period. Information on the origin of *C. arabica* is summarised in Figure 4.

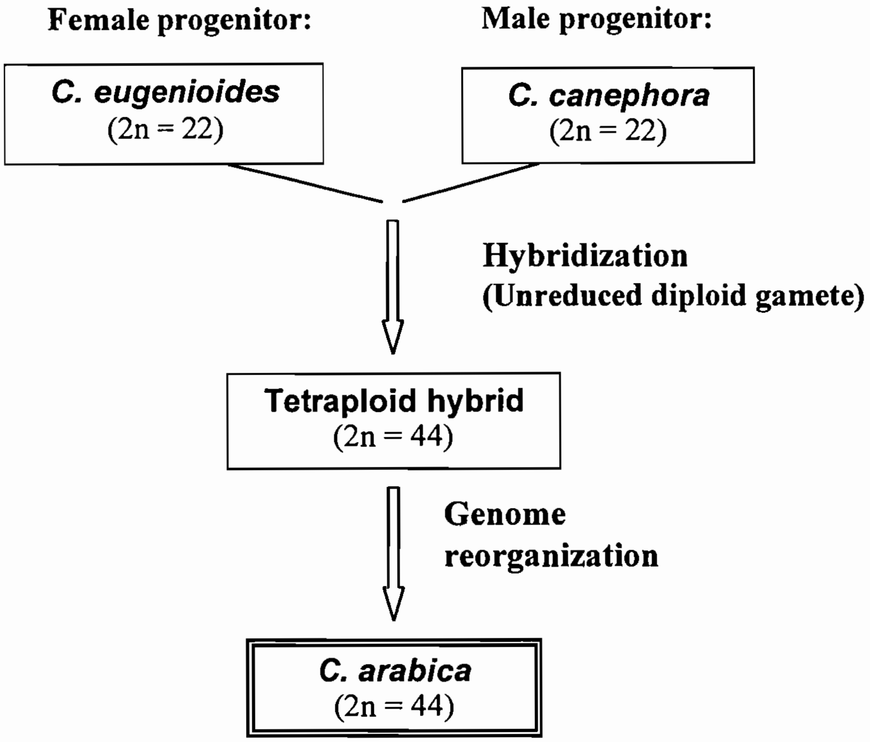

Figure 4. Proposed mode of speciation for *Coffea arabica*

Arabica speciation process

Results clearly suggested that *C. arabica* is an amphidiploid formed from the hybridisation between *C. eugenioides* used as female parent and *C. canephora* or ecotypes related to those species. Polyploidisation resulting into fertile hybrid types could have arisen by different ways such as chromosome doubling of a diploid interspecific hybrid or via backcrossing of a spontaneous triploid (Harlan and de Wet, 1975). Although the mode of origin remains obscure, it is likely to involve unreduced gametes. Such an event is easily conceivable in coffee trees. In particular, cold treatment has been reported in *C. canephora* and *C. liberica* to induce abnormal pollen development, including uninuclei microspore formation (Lanaud and Parvais, 1980). Later on, forms combining self-fertility and regular meiosis could have been retained during the evolutionary process. Breakdown of the self-incompatibility system occurring in most diploid species has often been observed in coffee interspecific hybrids (Charrier, 1978). The presence of pairing regulating genes has been recently established (Lashermes et al., 2000). However, further work is required to determine the genome reorganisation that occurred during the evolution of the tetraploid archetype to the present amphidiploid *C. arabica*.

References

Berthaud J (1976). Etude cytogénétique d'un haploïde de *Coffea arabica*. Café Cacao Thé (Paris) 20: 91-96.

Berthou F, Mathieu C, Vedel F (1983). Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. Theor. Appl. Genet. 65: 77-84.

Bouharmont J (1959). Recherche sur les affinités chromosomiques dans le genre *Coffea*. INEAC (Brussels), Série Sci. 77, 94p.

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

Charrier A (1978). La structure génétique des caféiers spontanés de la région malgache *Mascarocoffea*. leurs relations avec les caféiers d'origine africaine (*Eucoffea*). Mémoire ORSTOM 87, ORSTOM ed., Paris.

Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P (1998). Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis. Molecular Phylogenetics and Evolution 9: 109-117.

Demarly Y (1975). Amélioration des caféiers liée aux progrès génétiques. 7th Conference of ASIC (Paris), pp. 423-435.

Grassias M, Kammacher P (1975). Observations sur la conjugaison chromosomique de *Coffea arabica* L.. Café Cacao Thé (Paris) 19: 177-190.

Harlan JR, deWet JMJ (1975). On a winge and a prayer: the origins of polyploidy. The botanical review 41: 361-390.

Jorgansen RA, Cluster PD (1988). Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Annals of the Missouri Botanical Garden 75: 1238-1247.

Kammacher P (1980). Sur le comportement méiotique des dihaploïdes de *Coffea arabica* L.. 9th Conference of ASIC (Paris), pp. 717-724.

Kammacher P, Capot J (1972). Sur les relations caryologiques entre *Coffea arabica* et *C. canephora*. Café Cacao Thé (Paris) 16: 289-294.

Krug CA, Mendes AJT (1940). Cytological observations in *Coffea* - IV. Journal of genetics 39: 189-203.

Lanaud C, Parvais JP (1980). Observations, avant mise en culture, des divisions anormales des noyaux de grains de pollen de caféier induits au froid. 9th conference of ASIC, London (UK), pp 547-554.

Lashermes P, Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A (1996). Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L.. Theor. Appl. Genet. 93: 626-632.

Lashermes P, Combes MC, Trouslot P, Charrier A (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theor. Appl. Genet. 94: 947-955.

Lashermes P, Combes MC, Robert J, Trouslot P, D'hont A, Anthony F, Charrier A (1999). Molecular characterisation and origin of the *Coffea arabica* L. genome. Molecular & General Genetics 261: 259-266.

Lashermes P, Paczek V, Trouslot P, Combes MC, Couturon E, Charrier A (2000). Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C. arabica* x *C. canephora*. J. of Heredity 91: 81-85.

Olmstead RG, Palmer JD (1994). Chloroplast DNA systematics: a review of methods data analysis. Am. J. Bot. 81: 1205-1224.

Vishveshwara S (1960). Occurrence of a haploid in *Coffea arabica* Kents. Indian Coffee (Chikmagalure) 24 (3): 123-124.

Lashermes Philippe, Combes Marie-Christine, Topart Patrick, Anthony François. (2000)

Phylogenetic relationships of coffee species and origin of *Coffea arabica* tetraploid genome

In : Anthony François (ed.), Rodriguez E. (ed.). *Mejoramiento sostenible del cafe Arabica por los recursos geneticos, asistido por los marcadores moleculares, con enfasis en la resistencia a los nematodos*

Turrialba (CRI) ; Paris : CATIE ; IRD, p. 17-24

Memorias del Taller, Turrialba (CRI), 2000/08/29-30.