ORIGIN AND GENETIC DIVERSITY OF *COFFEA ARABICA* L. BASED ON DNA MOLECULAR MARKERS

P. LASHERMES [1], M.C. COMBES [1], J. CROS [1],P. TROUSLOT [1], F. ANTHONY [2], A. CHARRIER[1]

[1] ORSTOM, 911 Av. Agropolis BP 5045, F-34032, Montpellier cedex, France
[2] CATIE, 7170 Turrialba, Costa Rica

Coffee-trees belong to the tribe *Coffeae* in the family *Rubiaceae* (Bridson and Verdcourt, 1988). 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 (Berthaud and Charrier, 1988) 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.

1-Evaluation of genetic diversity between cultivated and wild accessions of *Coffea arabica* through random amplified polymorphic DNA analysis

The large number of named varieties and selections of arabica coffee belies the actually very narrow genetic diversity of the base populations from which they were selected (van der Vossen, 1985). Historical evidence indicates that these base populations all descended from the few trees that survived various efforts to spread arabica coffee from Southern Arabia, now Yemen, into the main coffee producing areas in Latin America, East Africa and Asia. The coffee trees from Yemen gave rise to two distinct botanical types : 1) *C. arabica* var.

typica Cramer, which was the earliest grown coffee in Asia and Latin America, and 2) *C. arabica* var. *bourbon* (B. Rodr.) Choussy, which came to South America through the island of La Réunion, formerly called Bourbon. The genetic uniformity within these populations is further enhanced by the predominantly self-pollinating nature of *C. arabica*. Enlarging the genetic base has become a priority for further crop improvement and has prompted several collecting missions. In particular, two expeditions to explore and collect arabica coffee materials were undertaken in 1964 to South West Ethiopia under the auspices of the FAO (FAO, 1968), and in 1966 by ORSTOM in the Illubabor and Kaffa provinces of Ethiopia (Guillaumet and Hallé, 1978).

Random amplified polymorphic DNA (RAPD) analysis (Welsh and McClelland, 1990 ; Williams et al., 1990) was performed to estimate the level of genetic diversity within the germplasm collection and the relatedness between cultivated and subspontaneous accessions of *C. arabica*. Six varieties representing the two distinct cultivated coffee types (*typica* and *bourbon*), the cultivar K-7 resulting from a selection work in Kenya (Walyaro, 1983), 11 samples representing the different collecting sites of the ORSTOM mission in Ethiopia, and two accessions collected in Kenya (Berthaud et al., 1980), were included in this study.

As previously reported (Berthou and Trouslot, 1977 ; Lashermes et al., 1993), a low molecular diversity is detected in *C. arabica*. However, the RAPD method appeared to be effective in resolving genetic variation in arabica coffee and in grouping germplasm.

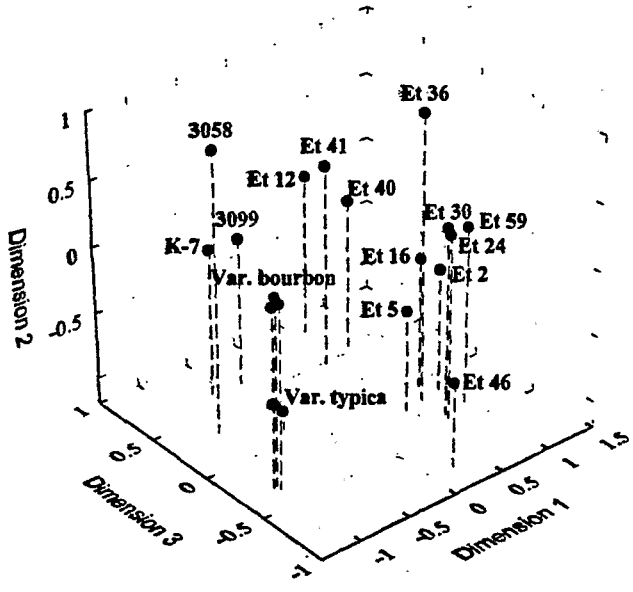

Figure 1. Non-metric multidimensional scaling applied to the matrix of RAPD-based genetic distances between 20 accessions of *C. arabica*.(Var. *bourbon* represents the accessions Caturra, Bourbon, Mbirizi and I-60 ; Var. *typica* represents the accessions Typica and Blue mountain).

This study indicated (Figure 1) a relatively large genetic diversity within the arabica germplasm collection and demonstrated the importance of collecting missions. As expected from their origin, we were not able to distinguished the cultivars belonging to the same type, either *bourbon* or *typica*. On the other hand, both *bourbon* and *typica* types showed important differences. The cultivar K7, which has been grown on a large scale in Kenya (Walyaro, 1983), appeared closely related to one of the accessions collected in the north of Kenya (Marsabit Mountain).

A clear separation was observed between the Ethiopian germplasm collected in the south west highlands of Ethiopia (Illubabor and Kaffa provinces), and the cultivated material spread world-wide from Yemen and the accessions collected in North Kenya. This result supports the hypothesis that the arabica plants transferred to Yemen for cultivation by the Arabs (Smith, 1985) originated from the south eastern part of the evergreen mountainous region of Ethiopia (Sidamo and Harar provinces). An east-west differentiation may exist in the primary centre of diversification of *C. arabica*. Similar observations have been reported from agro-morphological data (Bouharmont and Montagnon, 1995). Such differentiation may explain the large heterosis effect which has been noted in F1 hybrids resulting from crosses between indigenous cultivars from the south western and south eastern parts of Ethiopia (Bayetta-Bellachew et al., 1993), and between spontaneous Ethiopian accessions and *bourbon* type cultivars (Charrier, 1978). Therefore, the possibility of employing RAPD-based genetic distance measures for predicting hybrid performance should be considered.

2- 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., in preparation).

CpDNA variations have been investigated in 27 coffee taxa representing the main species and undetermined taxa (Cros, 1994). 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 2) using Wagner parsimony was performed.

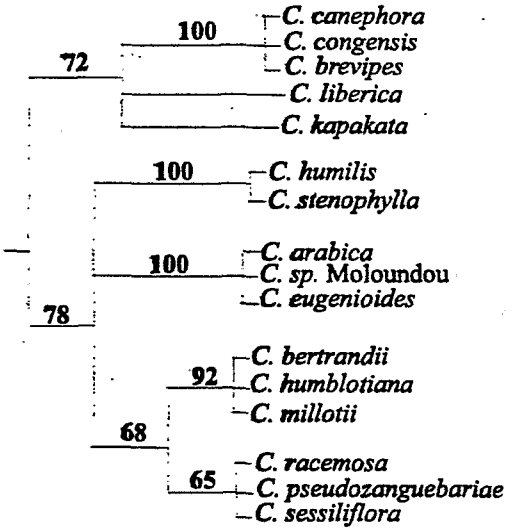

Figure 2. Phylogenetic tree of *Coffea* species based on chloroplast DNA variation. Strict consensus of the most parsimonious Wagner trees is represented. Values (%) on branches are bootstrap indices of support.

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 *C.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.

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. congensis	West and Central Africa	0.70
C. eugenioides	Central Africa	0.74
C. canephora	West and Central Africa	0.77
C. humilis	West Africa	0.77
C. sp. X	Unknown	0.80
C. sp. Moloundou	Central Africa	0.81
C. brevipes	Central Africa	0.87
C. kapakata	Central Africa	0.87
C. liberica	West and Central Africa	0.88
C. salvatrix	East Africa	0.88
C. stenophylla	West Africa	0.89
C. racemosa	East Africa	0.95
C. farafanganensis	Madagascar	1
C. humblotiana	Comores islands	1
C. millotii	Madagascar	1
C. pseudozanguebariae	East Africa	1

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.

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 type of sequence although important ITS2 nucleotide sequence variations were observed between species.

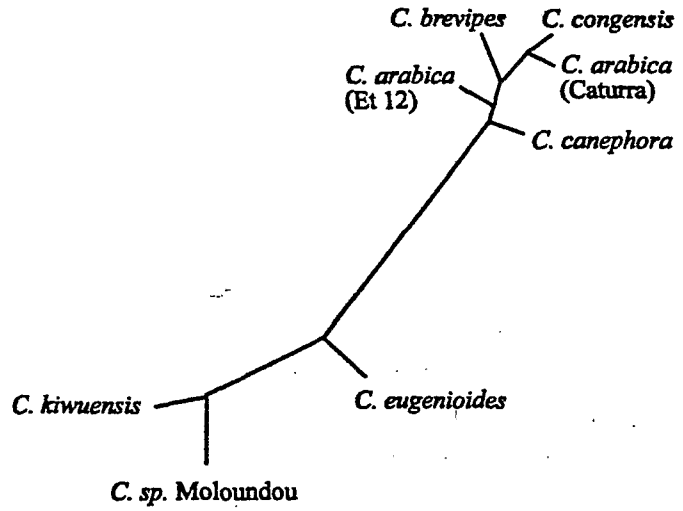

Figure 3. Parsimony analysis of ITS2 sequences of nuclear ribosomal DNA among *Coffea* species as putative ancestors of *C. arabica*. Branch lengths correspond to numbers of informative mutations.

Analysis for a restricted number of species showed (Figure 3) that the ITS2 region of *C. arabica* diverged markedly from the sequences of *C. eugenioides* and its sister-group (*C. kiwuensis* and *C. sp.* Moloundou), and appeared almost identical to the sequences of canephoroid species (*C. canephora*, *C. congensis* and *C. brevipes*).

3- Discussion of the origin of *C. arabica*

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 extent 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 *trnL-trnF* 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.

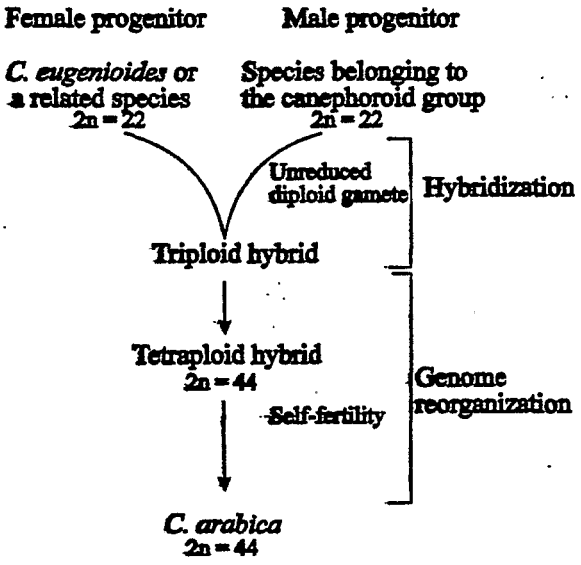

Figure 4. Proposed mode of speciation for *Coffea arabica*

Demarly (1975) has proposed sequences of events leading to the formation of *C. arabica*, such as unreduced gamete formation, breakdown of the self-incompatibility system, and adaptation to new habitats. These steps have little to support them and the described mode of speciation should be considered as one possibility. In particular, a mode of speciation involving association of unreduced gametes from both parental species and direct formation of a tetraploid hybrid, cannot be discarded.

C. arabica did not exhibit additivity of the ITS2 sequences of putative progenitors suggesting that homogenisation of rDNA or the elimination of a locus may have occurred. Similar results were observed for several RFLP loci (data not shown). Genome recombination could have been important among the various mechanisms which are likely to have played a major role in the progressive diploidisation from the archetype tetraploid to the present amphidiploid *C. arabica*. It seems evident that low, continued natural selection was necessary. The degree to which the ancestral genomes have recombined in the amphidiploid needs to be established by further experimentation.

References

Bayetta-Bellachew, Mesfin-Ameha, Dejene-Mekonnen, 1993. Heterosis and combining ability in coffee (*Coffea Arabica* L.). 15th Conference of ASIC, Montpellier (France), 234- 242.

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

Berthaud, J., Charrier A., 1988. Genetics resources of *Coffea*. In: R J Clarke and R Macrae (eds) Coffee vol 4: Agronomy. Elsevier Applied Science, London, 1-42.

Berthaud, J., Guillaumet J.L., Le Pierres D., Lourd M., 1980. Les caféiers sauvages du Kenya: prospection et mise en culture. Café-Cacao-Thé, 24: 101-112.

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.

Berthou, F., Trouslot, P., 1977. L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une méthode d'électrophorèse en série. 8th Conference of ASIC, Abidjan (Côte d'Ivore), 373-383.

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

Bouharmont P., Montagnon C., 1995. Diversité phenotypique de *Coffea arabica* observée en collection au Cameroun. 16th Conference of ASIC, Kyoto (Japan), In press

Bridson D.M., Verdcourt B., 1988. Flora of tropical east Africa - Rubiacae (part 2), Polhill R.M. (eds), 227p.

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

Charrier, A. 1978. Synthèse de neuf années d'observation et d'expérimentation sur les *Coffea arabica* collectés en Ethiopie par une mission ORSTOM. In: A. Charrier (ed.) Etude de la structure et de la variabilité génétique des caféiers, IFCC (Paris), Bulletin no 14, pp.4-10.

Cros J., 1994. Implications phylogénétiques des variations de l'ADN chloroplastique chez les caféiers (genres *Coffea* L. et *Psilanthus* Hook. f.). Thèse de doctorat. Université Montpellier II. 137p.

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

F.A.O., 1968. FAO coffee mission to Ethiopia 1964-1965. Report, FAO ed., Rome (Italy), 200 p.

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

Guillaumet J.L., Hallé F., 1978. Echantillonnage du materiel *Coffea arabica* récolté en Ethiopie. In: A. Charrier (ed.) Etude de la structure et de la variabilité génétique des caféiers, IFCC (Paris), Bulletin no 14, pp.13-18.

Jorgansen R.A., Cluster P.D., 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.. ASIC (Paris), 9e colloq., London, 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 C.A., Mendes A.J.T., 1940. Cytological observations in *Coffea* - IV. Journal of genetics, 39: 189-203.

Lashermes P., Cros J., Marmey P., Charrier A., 1993. Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. Genetic Resources and Crop Evolution. 40: 91-99.

Olmstead R.G., Palmer J.D., 1994. Chloroplast DNA systematics: a review of methods data analysis. Am. J. Bot., 81: 1205-1224.

Smith F.R., 1985. A history of coffee. In: M.N. Clifford and K.C. Wilson (eds) Coffee, Botany, Biochemistry and production of beans and beverage. Croom Helm, London & Sydney, pp 1-12.

Van der Vossen, H.A.M., 1985. Coffee selection and breeding. In: M.N. Clifford and K.C. Wilson (eds) Coffee, Botany, Biochemistry and production of beans and beverage. Croom Helm, London & Sydney, 48-97.

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

Walyaro D.J.A., 1983. Considerations in breeding for improved yield and quality in arabica coffee (*Coffea arabica* L.). Doctoral thesis, Wageningen, Netherlands, 122 p.

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

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

Abstract

Phylogenetic relationships between 25 *Coffea* species, including the only tetraploid species *Coffea arabica*, were inferred using DNA restriction length polymorphisms (RFLP) and DNA sequencing of both nuclear and chloroplastic genomes. The allotetraploid origin of *C. arabica* is corroborated by the extent 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.

In addition, 20 coffee accessions of *C. arabica* representing the major types of cultivar and sub-spontaneous genotypes collected in Ethiopia were compared by RAPD (random amplified polymorphic DNA) analysis. This method appeared to be effective in resolving genetic variation in arabica coffee and grouping germplasm. An east-west differentiation may exist in the primary centre of diversification of *C. arabica*.

Résumé

Les relations phylogénétiques de 25 espèces de caféiers dont l'espèce tétraploïde *C. arabica* ont été abordées au niveau de leur ADN génomique nucléaire et chloroplastique par l'utilisation de la variation de la longueur des fragments de restriction (RFLP) et la comparaison de leur séquences. L'origine allotétraploïde de *C. arabica* est supportée par l'importance du polymorphisme RFLP. L'étude du génome chloroplastique indique *C. eugenioides* ou une espèce proche de *C. eugenioides* comme parent femelle de *C. arabica*. L'étude de l'unité nucléaire codant pour l'ARN ribosomique conduit à une espèce du groupe des canephoroides (*C. canephora, C. congensis, C. brevipes*) comme parent mâle. L'origine de *C. arabica* apparaît comme un événement récent de la fin du quaternaire et les conditions de sa formation sont en discussion.

De plus, 22 souches de *C. arabica* représentant les 2 grands groupes de cultivars - Typica et Bourbon - et des caféiers subspontanés collectés en Ethiopie ont été comparées par l'analyse du polymorphisme obtenu après amplification de fragments aléatoires d'ADN (RAPD). Elle met clairement en évidence une différenciation Est-Ouest dans le centre primaire de diversification en Ethiopie.