Molecular analysis of the origin and genetic diversity of *Coffea arabica* L.: Implications for coffee improvement

P. Lashermes[1], M.C. Combes[1], P. Trouslot[1], F. Anthony[2] & A. Charrier[3]

[1]*ORSTOM, 911 Av. Agropolis BP 5045, F-34032, Montpellier, France*
[2]*CATIE, 7170 Turrialba, Costa Rica*
[3]*ENSAM, place Viala, F-34060, Montpellier, France*

Key words : Amphidiploid, Breeding, *Coffea arabica*, Diversity, DNA marker, Origin

Abstract

DNA-markers were used to investigate the hybrid origin of the allotetraploid species *Coffea arabica* and to evaluate its genetic variation. Work on the chloroplast genome provides strong support for the hypothesis that a species close to *C. eugenioides* donated the maternal genome of *C. arabica* while the analysis of the nuclear ribosomal DNA indicated that the paternal parent was most likely a species belonging to the canephoroid group (*C. canephora*, *C. congensis*, *C. brevipes*). The genetic diversity among cultivated and subspontaneous accessions of *C. arabica* was analysed using random amplified polymorphic DNA (RAPD) markers. The narrow genetic base of commercial cultivars was confirmed. On the other hand, a relatively large genetic diversity was observed within the germplasm collection demonstrating the importance of collecting missions. An East-West differentiation would exist in the primary centre of diversification of *C. arabica* These results are discussed in relation to coffee improvement.

Introduction

Coffee-trees belong to the genus *Coffea* in the family *Rubiaceae* (Bridson and Verdcourt 1988). While more than 100 distinct taxa have been so far identified in African and Madagascan intertropical forests, commercial coffee production relies mainly on two species: *Coffea arabica* L. and *Coffea canephora* Pierre. *C. arabica* is the only tetraploid species (2x = 44) in the genus and is self-fertile while other species are diploid and generally self-incompatible (Charrier and Berthaud 1985). Higher quality is associated with *C. arabica* and arabica coffee represents 70% of world production. On the other hand, many *Coffea* species form a valuable gene reservoir for breeding purposes. To date, *C. canephora* provides the main source of disease resistance traits not found in *C. arabica*, but other diploid species present considerable interests (Berthaud and Charrier 1988).

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. Their implications regarding coffee breeding are discussed.

Origin of *C. arabica*

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.

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. When several accessions from the same species were analysed, the average distance is reported.

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

Earlier attempts to determine the genetic origin of *C. arabica* relied on karyotyping (Bouharmont 1959), chromosome pairing in hybrids with diploid species (Krug and Mendes 1940; Monaco and Medina 1965; Chinnappa 1968; Kammacher and Capot 1972; Charrier 1978a) 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. However, the diploid progenitors were not unequivocally identified.

RFLP analysis using nuclear single-copy probes

A study was conducted to determine relationships among a series of *Coffea* species including *C. arabica* using restriction fragment length polymorphism (RFLP) markers. Nuclear genomic clones obtained from a *C. arabica* library and selected to be single-copy using doubled haploid genotypes of *C. canephora*, were used as probe. RFLP-based distances between *C. arabica* and a number of diploid species were estimated (Table 1).

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. *C. congensis*, *C. canephora* and *C. eugenioides* appeared to be the closest species to *C. arabica*. Comparable results were obtained using random amplified DNA markers (Lashermes et al. 1993).

Phylogenetic relationships inferred from chloroplast DNA variation

Analysis of chloroplast DNA (cpDNA) variation has proven to be extremely valuable for plant phylogenetic reconstruction (Olmstead and Palmer 1994). 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. 1996c).

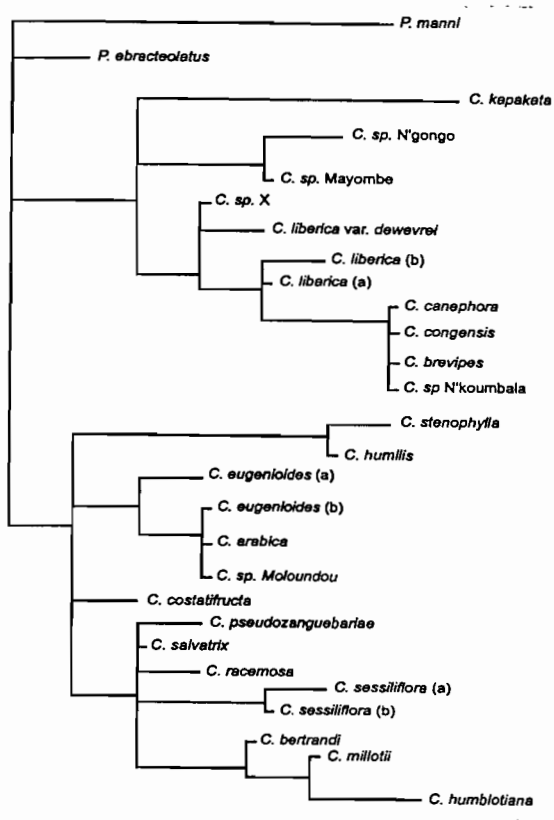

Figure 1 : Phylogeny of *Coffea* species generated by Wagner parsimony analysis based on the chloroplast DNA variation. Branch lengths are proportional to the number of supporting character states.

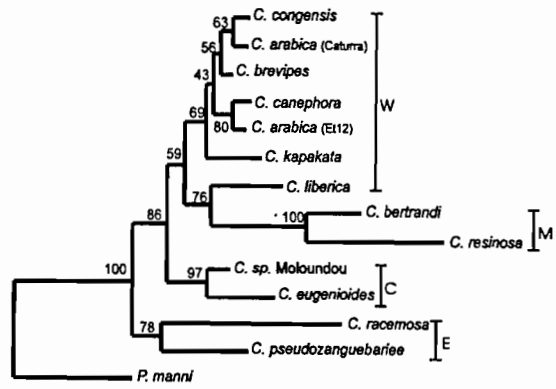

Figure 2 : Neighbor-joining tree resulting from sequence analysis of the internal transcribed spacer region (ITS2) in nuclear ribosomal DNA of 13 coffee species. Numbers on the branches are bootstrap values from 100 replicates of analyses. Major groups are indicated by letters: W (West Africa), C (Central Africa), E (East Africa) and M (Madagascar).

First study on the chloroplast genome in coffee was performed using total cpDNA RFLP analysis (Berthou et al. 1983). The reduced number of species analysed, did not allow phylogenetic reconstruction. More recently, cpDNA variations have been investigated in 27 coffee taxa representing the main species and undetermined taxa (Cros 1994). RFLP 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 RFLP analysis showed a low level of polymorphism while the intergenic sequence (*trn*L-*trn*F) appeared more polymorphic.

A phylogenetic analysis was performed despite the limited cpDNA variation evidenced (Figure 1). 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.

Nuclear Ribosomal DNA sequence variation

The 18-26S nuclear ribosomal DNA was studied (Lashermes et al. 1996b). Among nuclear gene regions, the rDNA repeat unit is particularly attractive for genetic studies (reviewed in Jorgansen and Cluster 1988). Complementary use of cpDNA and rDNA not only can identify the progenitors of allopolyploids but can also illustrate the direction of hybridisation. Preliminary observations showed significant polymorphism in the internal transcribed spacer ITS2, and this region was sequenced for a number of *Coffea* species including two genotypes of *C. arabica* (Caturra and Et 12).

Important ITS2 nucleotide sequence variations were observed between species. 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 *C. sp.*

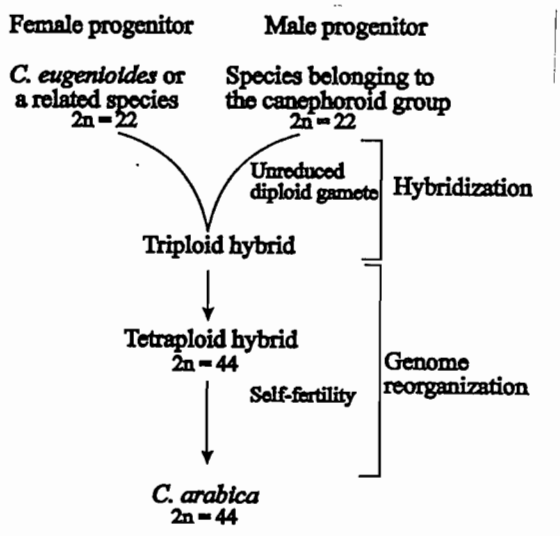

Figure 3 : Schematic representation of the proposed mode of speciation for *Coffea arabica*

Moloundou, and appeared almost identical to the sequences of canephoroid species (*C. canephora*, *C. congensis* and *C. brevipes*). Several groups are revealed which are to some extent consistent with the classical biogeographical grouping (i.e. Madagascar, East Africa, Central and West Africa). These results indicated that a species belonging to the canephoroid group was most likely the paternal ancestor of *C. arabica*. Although a limited sequence variation was detected between the two accessions, *C. arabica* genotypes showed only one major type of sequence (canephoroid type).

Proposed mode of speciation for C. arabica

Information's on the origin of *C. arabica* are summarised in the Figure 3. The allotetraploid origin of *C. arabica* is corroborated and specified by the different molecular analyses. A sequence 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 has been proposed (Demarly 1975). Steps have little to support them and should be considered with circumspection. It is noticeable that *C. sp.* Moloundou which has recently been observed to be self-compatible, was evidenced as a species related to the maternal ancestor of *C. arabica*. In addition, genome reorganisation is likely to have played a major role in the evolution of the archetype tetraploid to the present amphidiploid *C. arabica*. For instance, *C. arabica* exhibit only the ribosomal DNA of a single parent suggesting that homogenisation of rDNA or the elimination of a sequence type have occurred.

Genetic diversity of *C. arabica*

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; Carvalho 1988). 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. Arabica coffee was introduced for cultivation in Yemen from Ethiopia in earlier time by the Arabs (Smith 1985). The coffee trees from Yemen gave rise to two distinct botanical types (Krug et al. 1939) : 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.

Enlarging the genetic base has prompted several collecting missions. In particular, two expeditions were undertaken in 1964-65 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). This material may have been subjected to human interference and the truly wild or spontaneous origin of accessions is questionable as mentioned by the collectors; the term subspontaneous seems therefore more appropriate to describe this material. Observations of agromorphological characters indicated considerable phenotypic diversity among the collected material (Charrier 1978b; Anthony et al. 1993; Bouharmont and Montagnon 1995).

Evaluation of the genetic diversity through molecular markers

A extremely low diversity was observed in *C. arabica* using genetic markers. Isozyme markers has been shown to be useless for estimating genetic diversity between accessions (Berthou and Trouslot 1977). Similarly, a very low polymorphism was detected through RFLP (unpublished data). Only techniques such as random amplified polymorphic DNA (RAPD) and AFLP (Vos et al. 1995) allowed the detection of a significant level of polymorphism (Orozco-Castillo et al. 1994; Lashermes et al. 1993, 1996a). However, the assays performed until now failed to disclose variation between cultivars belonging to the same type, either *bourbon* or *typica*. It is therefore believed that the encountered agromorphological variation, which gave rise to so many cultivars, results from few major-gene spontaneous mutations conditioning plant, fruit and seed characters. The limited diversity detected in *C. arabica* could have different origins including the allotetraploid origin and mode of speciation, and the predominantly self-pollinating nature of *C. arabica*.

A differentiation in the centre of genetic diversity

RAPD markers were used to estimate the level of genetic diversity within the germplasm collection and the relatedness between cultivated and subspontaneous accessions of *C. arabica* (Lashermes et al. 1996a). 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.

Despite a low degree of variation, the RAPD method was effective in grouping germplasm (Figure 4). In particular, 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

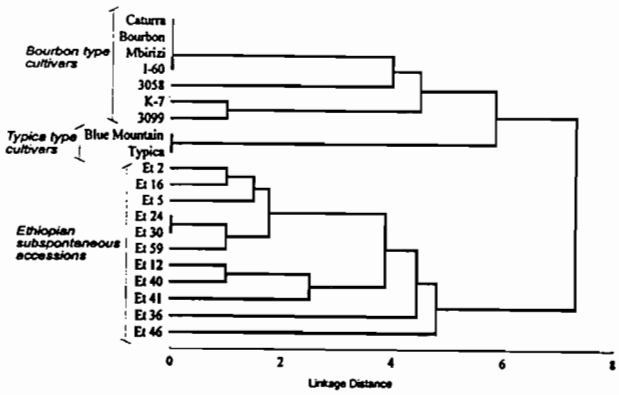

Figure 4 : Dendrogram of *Coffea arabica* accessions based on RAPD data and generated by single linkage cluster analysis.

the hypothesis that the arabica plants transferred to Yemen for cultivation by the Arabs originated from the south eastern part of the evergreen mountainous region of Ethiopia. Furthermore, an East-West differentiation would exist in the primary centre of diversification of *C. arabica*.

Conclusions

Results from the different molecular analyses have considerable implications regarding arabica breeding.

The DNA-based marker studies indicated clearly a relatively large genetic diversity within the arabica germplasm collected in Ethiopia and demonstrated the importance of collecting missions. This material provides a new genetic resource which is expected to be great value for coffee breeding. In addition, the genetic differentiation observed between the commercial cultivars and the germplasm collected in the south west highlands of Ethiopia, could be related to the heterosis effect which has been reported 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 subspontaneous Ethiopian accessions and *bourbon* type cultivars (Charrier 1978b). If confirmed, this result could have a considerable impact on the

development of improved cultivars.

The reduced diversity present in *C. arabica* and the possibility to intercross the different coffee species (reviewed in Berthaud and Charrier 1988) encourage the use of the diploid gene pool in arabica breeding (Carvalho 1988). In so doing, results of the different molecular phylogenetic reconstruction would suggest to focus efforts on the species closed to the maternal and paternal ancestors of *C. arabica*. While the utilisation of *C. canephora* has been considered to some extent, others taxa (*C. eugenioides*, *C. congensis*, *C. sp.* Moloudou...) have been neglected as source of desirable genes. Only few interspecific hybrids have been already developed (Le Pierrès 1995) and use of these species in crossing programme on a large-scale is required. Enlarging the current collection by additional accessions of the above species could justified supplementary prospecting missions in Central Africa. In addition, the recent advances in the development of a genetic map of the coffee genome (Paillard et al. 1996) offer new opportunities for the utilisation of exotic germplasm.

References

Anthony, F., Bertrand B., Dufour M. & J.V. Escalant, 1993. Evaluacion y caracterizacion de los recuros geneticos de café conservados en el germoplasma del CATIE. XVI Simposio sobre la caficultura latinoamericana, Managua (Nicaragua), IICA/PROMECAFE Ed., in press.

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 20: 91-96.

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

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

Berthou, F. & P. Trouslot, 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.

Berthou, F., Mathieu C. & F. Vedel, 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.

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

Bridson, D.M. & B. Verdcourt, 1988. Flora of tropical East Africa - Rubiaceae (part 2), Polhill R.M. (Ed), Balkema, Brookfield, Rotterdam.

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

Carvalho, A., 1988. Principles and practices of coffee plant breeding for productivity and quality factors: *Coffea arabica*. In: R. J. Clarke and R. Macrae (Eds) Coffee vol 4 : Agronomy, pp 129-165. Elsevier Applied Science, London.

Charrier, A., 1978a. La structure génétique des caféiers spontanés de la région malgache (Mascarocoffea). Mémoires ORSTOM (87), ORSTOM Ed., Paris, 218p.

Charrier, A., 1978b. 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 14: 4-10.

Charrier, A. & J. Berthaud, 1985. Botanical classification of coffee. In: M.N. Clifford and K.C. Wilson (Eds), Coffee: Botany, Biochemistry and production of beans and beverage, pp. 13-47. Croom Helm, London, Sydney.

Chinnappa, C.C., 1968. Interspecific hybrids of *Coffea canephora* and *Coffea arabica*. Current Science, 37: 676-677.

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.

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

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

Guillaumet, J.L. & F. Hallé, 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 14: 13-18.

Jorgansen, R.A. & P.D. Cluster, 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann. Missouri Bot. Garden 75: 1238-1247.

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

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

Krug, C.A. & A.J.T. Mendes, 1940. Cytological observations in *Coffea* - IV. Journal of Genetics 39: 189-203.

Krug, C.A., Mendes J.E.T. & A. Carvalho, 1939. Taxonomia de *Coffea arabica* L.. Descriçao das variedas e formas encontradas no estado de Sao Paulo. Boletim Tecnico 62, Instituto Agronomico do Estado (Campinas, Brazil) 57p.

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

Lashermes, P., Trouslot P., Anthony F., Combes M.C. & A. Charrier, 1996a. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. Euphytica 87: 59-64.

Lashermes, P., Combes M.C., Trouslot P. & A. Charrier, 1996b. Molecular phylogeny of the *Coffea* species based on nuclear rDNA (ITS2) sequences. Submitted.

Lashermes, P., Cros J., Combes M.C., Trouslot P., Anthony F., Hamon S. & A. Charrier, 1996c. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. Submitted.

Le Pierrès, D., 1995. Etude des hybrides interspécifiques tétraploïdes de première génération entre *Coffea arabica* L. et les caféiers diploïdes. Thèse, Université de Paris-XI Orsay, 298p.

Monaco, L.C. & D.M. Medina, 1965. Hibridaçoes entre *Coffea arabica* e *C. kapakata*. Analise citologica de um hibrido triploide. Bargantia 24: 191-201.

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

Orozco-Castillo, C., Chalmers K.J., Waugh R. & W. Powell, 1994. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor. Appl. Genet. 87: 934-940.

Paillard, M., Lashermes P. & V. Pétiard, 1996. Construction of a molecular linkage map in coffee. Theor. Appl. Genet. in press.

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, pp. 1-12. Croom Helm, London, Sydney.

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, pp. 48-97. Croom Helm, London, Sydney.

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

Vos, P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. & M. Zabeau, 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.

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.

Lashermes Philippe, Combes Marie-Christine, Trouslot Pierre, Anthony François, Charrier A. (1996).

Molecular analysis of the origin and genetic diversity of Coffea arabica L. : implications for coffee improvement.

In : Meeting on tropical plants = Réunion sur les plantes tropicales : communications et posters.

Montpellier : CIRAD ; MICAP, p. 23-29.

Meeting on Tropical Plants = Réunion sur les Plantes Tropicales, Montpellier (FRA), 1996/03/11-15.