Genetic Diversity and Introgression Analyses in Coffee (*Coffea arabica* L.) Using Molecular Markers

F. ANTHONY[1], M. C. COMBES[2], J. C. HERRERA[2], N. S. PRAKASH[3], B. BERTRAND[4], P. LASHERMES[2]

[1]CATIE-IRD, Turrialba, Costa Rica
[2]IRD, Montpellier, France
[3]Regional Coffee Research Station, Andhra Pradesh, India
[4]IICA/PROMECAFE-CIRAD, San José, Costa Rica

SUMMARY

DNA markers (AFLP, RAPD, RFLP, SSR) were recently used to assess the genetic diversity among wild and cultivated *C. arabica* accessions, and to detect introgressions from *C. canephora* and *C. liberica* into *C. arabica* genome. The results allowed for the definition of breeding strategies using the whole genetic diversity that are conserved in field genebanks and for the control of alien gene transfer to improve arabica cultivars.

Almost all polymorphism was generated by the Ethiopian material. The southwestern Ethiopian accessions were grouped separately from the southeastern Ethiopian accessions. The cultivars were classified according to their genetic origin (i.e. Typica or Bourbon). The Yemen cultivars were grouped with the Typica-derived accessions, confirming the Yemen origin of the coffee plant cultivated in Amsterdam and Paris at the beginning of the 18th century and later known as Typica.

Introgressions of *C. canephora* and *C. liberica* were identified in derivatives from natural interspecific hybrids (i.e. Timor Hybrid and S.26). The introgressed genotypes were distinguished from the *C. arabica* accessions by additional bands (i.e. introgressed markers) and missing bands (i.e. markers related to introgression process). The missing bands might be associated with the stabilization process of introgressed fragments over the generations.

Segregation of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* x *C. canephora*) was studied using a complete linkage map of *C. canephora*. The chromosomes segregated at random in the tetraploid hybrid, indicating the absence of preferential pairing of the four sets of chromosomes. Recombination in the tetraploid hybrid was not significantly restricted by the genetic differentiation of chromosomes belonging to the different genomes.

RÉSUMÉ

Des marqueurs de l'ADN (AFLP, RAPD, RFLP, SSR) furent récemment utilisés pour évaluer la diversité génétique présente chez les caféiers (*C. arabica*) sauvages et cultivés, et pour détecter les introgressions de *C. canephora* et *C. liberica* dans le génome *C. arabica*. Les résultats permettent de définir des stratégies d'amélioration qui utilisent l'ensemble de la diversité génétique conservée dans les collections en champ et de contrôler les transferts de gènes pour améliorer les cultivars *C. arabica*.

Presque tout le polymorphisme fut généré par le matériel d'Ethiopie. Les accessions du sud ouest de l'Ethiopie se sont classées séparément des accessions du sud est. Les cultivars se sont

regroupés selon leur origine génétique (Typica ou Bourbon). Les cultivars du Yémen furent associés avec les accessions dérivées du Typica, ce qui confirme l'origine yéménite du caféier cultivé à Amsterdam et Paris au début du XVIIIe siècle et connu plus tard comme Typica.

Les introgressions de *C. canephora* et *C. liberica* furent identifiées dans des descendances d'hybrides interspécifiques naturels (Hybride de Timor et S26). Les génotypes introgressés se sont distingués des accessions *C. arabica* par la présence de bandes additionnelles (marqueurs introgressés) et l'absence de bandes (marqueurs liés au processus d'introgression). Les bandes manquantes pourraient être associées au processus de stabilisation des fragments introgressés au cours des générations.

Les ségrégations du génome *C. canephora* furent étudiées chez l'hybride interspécifique tétraploïde (*C. arabica* x *C. canephora*), en utilisant une carte de liaison de *C. canephora*. Les chromosomes ségrègent au hasard chez l'hybride tétraploïde, ce qui indique l'absence d'appariements préférentiels des quatre ensembles de chromosomes. Les recombinaisons chez l'hybride tétraploïde ne sont pas significativement limitées par la différentiation génétique des chromosomes appartenant aux différents génomes.

INTRODUCTION

Coffea arabica L. is an amphidiploid species (2n=4x=44) (Lashermes et al., 1999) native to the highlands of South West Ethiopia (Sylvain, 1955), the Boma Plateau of Sudan (Thomas, 1942) and Mount Marsabit of Kenya (Anthony et al., 1987). It is the only polyploid coffee species and is self-fertile at approximately 90% (Carvalho et al., 1991) while other coffee species are generally self-incompatible. Arabica coffee has been cultivated in Yemen for at least five centuries but spread to South East Asia about 1700. In the early 18th century, progenies of a single plant from Indonesia, cultivated in Amsterdam and Paris, were spread to Latin America (Chevalier and Dagron, 1928). Other introductions followed in the late 18th century from Yemen to Brazil, via Bourbon Island (now Réunion) (Haarer, 1956). These base populations gave rise to many cultivars and were described as two distinct varieties, respectively *C. arabica* var. *arabica*, usually called *C. arabica* var. *typica* Cramer, and *C. arabica* var. *bourbon* (B. Rodr.) Choussy, commonly called Typica and Bourbon respectively (Krug et al., 1939; Carvalho et al., 1969). The cultivars present an homogeneous agronomic behaviour, characterised by a high susceptibility to many pests and diseases (Bertrand et al., 1999).

Enlarging the genetic base and improvement of arabica cultivars have become priorities. Spontaneous accessions collected in the primary centre of diversity as well as wild relative *Coffea* species constitute a valuable gene reservoir for breeding purposes (Anthony et al., 1999). Genes from diploid species can be transfer into *C. arabica* cultivars exploiting natural and controlled interspecific hybrids. However, transferring various resistance genes without reducing coffee quality appears as a very difficult task in an acceptable time-frame through traditional breeding approaches.

In recent years, DNA-based genetic markers have gained widespread applications in many fields of plant genetics and breeding. Several results related to *C. arabica* genetics and based on molecular markers utilisation have been already reported in ASIC conferences on: genetic diversity and phylogenetic relationships in *Coffea* (Cros et al., 1993), the origin of *C. arabica* genome (Lashermes et al., 1995), the use of molecular markers for assisting selection (Lashermes et al., 1997a) and the development of microsatellite markers (Mettulio et al., 1999) which constitute powerful markers for breeding and genetic mapping. The results presented here concern the genetic diversity available in wild and cultivated *C. arabica*

accessions, and an analysis of introgressions from *C. canephora* and C. *liberica* into *C. arabica* genome.

GENETIC DIVERSITY ANALYSIS

Genetic diversity of wild coffee

The genetic diversity was studied using Random Amplified Polymorphic DNA (RAPD) markers among 119 coffee individuals representing 88 accessions derived from spontaneous and subspontaneous trees in Ethiopia, 6 cultivars grown locally in Ethiopia and 2 Typica- and Bourbon-derived accessions (Anthony et al., 2001). The sampling could be considered representative of the FAO (1968) and ORSTOM (Guillaumet and Hallé, 1978) material conserved in the CATIE field genebank. Only 16 of the 150 10-mer oligonucleotides used in the study (10.7%) detected polymorphism between accessions. This result confirmed the low polymorphism observed in the species *C. arabica* at the level of the rDNA (Lashermes et al., 1997b) and cpDNA (Cros et al., 1998). The Ethiopian accessions tended to form groups according to their origin (Figure 1). Ethiopian 1 was composed of 78 accessions and 2 Ethiopian cultivars (Anfilo, Dalle). It comprised all accessions from Gojjam, Ilubabor and Shoa provinces, all accessions except three from Kefa, one from Harerge and two from Sidamo. Except for 1 accession from Kefa province, all accessions classified in the other groups (Ethiopian 2, 3, 4) originated from Harerge province in the South East and Sidamo province in the South. Most of the detected diversity was found in accessions classified as Ethiopian 1. They presented 28 of the 29 identified markers whereas the accessions classified in other groups presented only 5 to 16 markers. The Typica- and Bourbon-derived accessions presented only 3 and 7 markers, respectively. This should increase interest in spontaneous and subspontaneous coffee for enlarging the genetic base of cultivars.

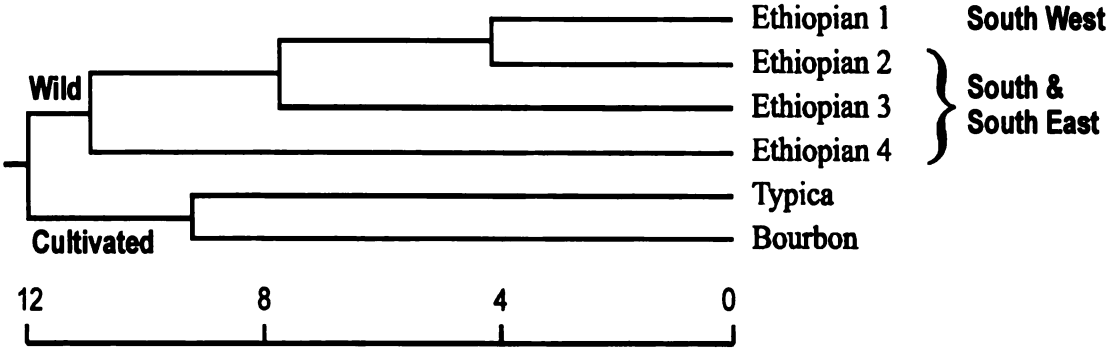

Figure 1. Structure of the genetic diversity in spontaneous and subspontaneous coffee (*C. arabica*) collected in Ethiopia, using RAPD markers (Anthony et al., 2001)

The distinction between the southwestern and southern Ethiopian coffee trees is not a consequence of their genetic isolation due to the presence of the tectonic break "the Great Rift Valley", which crosses Ethiopia from North East to South West, since the molecular characterisation of *C. arabica* genome suggested a recent origin of the species (Lashermes et al., 1999). The genetic distance estimated by RAPD markers showed that southern and southeastern coffee trees presented a low differentiation from southwestern coffee trees. This supports the hypothesis that southern and southeastern coffee trees were not selected from wild coffee growing locally but introduced from the South West where Lejeune (1958) situated the first cultivation of coffee. Moreover, no references mention the existence of wild coffee on the east side of the tectonic break.

Genetic diversity of cultivated coffee

Amplified Fragment Length Polymorphism (AFLP) markers were used to assess polymorphism among 8 accessions derived from the Typica and Bourbon genetic bases, 2 accessions of cv. Catuai ((Typica x Bourbon) x Bourbon), 4 cultivars growing in the Popular Democratic Republic of Yemen (Eskes, 1989) and 11 subspontaneous-derived accessions (Anthony et al., in press). A total of 107 AFLP polymorphic markers were used to construct a dendrogram using genetic distances between accession pairs (Figure 2). The Typica- and Bourbon-derived accessions were classified in 2 distinct groups according to their genetic origin, each group being supported by a high bootstrap value (89%). The 2 accessions of cv. Catuai were classified closer to the Typica-derived accessions than to the Bourbon-derived accessions. The subspontaneous accessions were clearly separated from the cultivated accessions, confirming the classification based on RAPD data. They did not constitute a structured group, but rather chains, with poor bootstrap values. The subspontaneous accession from South Ethiopia (E-238), which was classified in the Ethiopian 4 group by RAPD data, was separated from the southwestern Ethiopian and Sudan accessions. The 4 Yemen cultivars were associated with the Typica-derived accessions. This result was in accordance with the historical data about the diffusion of the Typica genetic base from Yemen to Amsterdam, via Java. The estimation of the genetic distances showed that the cultivars were closer to the subspontaneous coffee of the west side of the Great Rift Valley than to the east side. If the coffee of Yemen came from seeds harvested in the Harar region, as affirmed by Wellman (1961), it was an intermediate step in the introduction to Yemen.

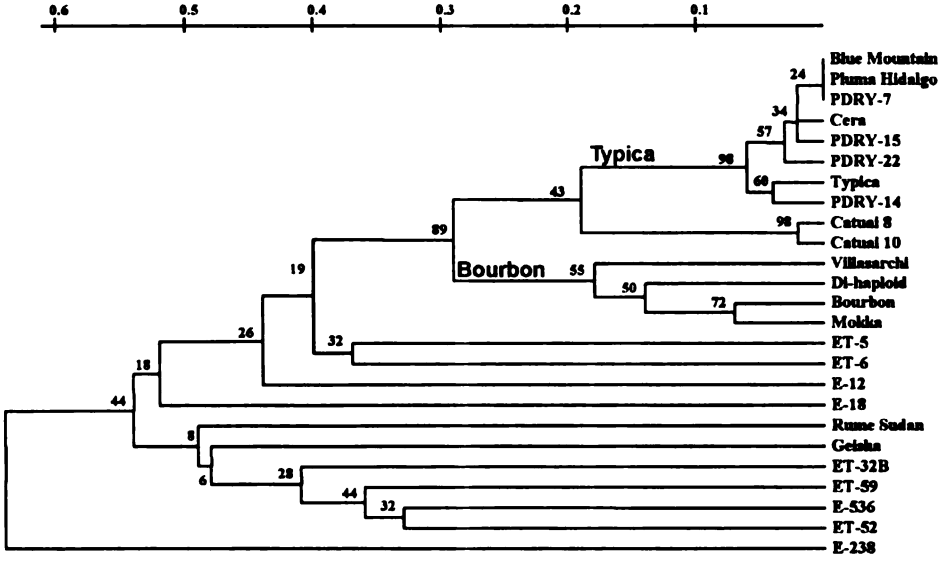

Figure 2. Dendrogram generated after UPGMA using AFLP-based genetic distance. Numbers on the branches are bootstrap values (%) obtained from 200 replicate analyses (Anthony et al., in press)

The number of polymorphic markers varied significantly within the cultivated and wild accessions. Whereas 94 AFLP markers were polymorphic within wild coffee, only 7 and 14 were respectively polymorphic within the Typica-derived accessions and the Yemen cultivars, and within the Bourbon-derived accessions. The mode of diffusion of coffee and the selection that followed have strongly reduced the genetic diversity present in the wild coffee. The higher polymorphism observed in the Bourbon group indicated that the genetic base of Bourbon was constituted by the descendants of several individuals and not from one single

individual as for the genetic base of Typica. This result confirmed the historical data given by Haarer (1956) in which several introductions took place from Yemen to the Reunion Island.

INTROGRESSION ANALYSIS

Introgression from *C. canephora*

Twenty-one Timor Hybrid-derived accessions were analysed for the introgression of *C. canephora* genetic material using AFLP markers (Lashermes et al., 2000). They were compared to 23 *C. arabica* accessions and 8 *C. canephora* accessions. The Timor Hybrid-derived accessions were distinguished from the *C. arabica* accessions by 178 markers consisting of 109 additional bands and 69 missing bands. The additional bands corresponded to introgressed fragments whereas a part of the missing bands might be associated with the stabilization process of introgressed fragments over the generations. The number of additional and missing bands varied respectively from 18 to 59 and from 0 to 32 among the Timor Hybrid-derived accessions (Figure 3). The introgressed fragments were estimated to represent from 8% to 27% of the *C. canephora* genome. Assuming a unique genotype of *C. canephora* was involved in the formation of the Timor Hybrid, the overall 109 introgressed fragments identified in the Timor Hybrid-derived accessions were estimated to represent 51% of the *C. canephora* genome. Most of the introgressed chromosome segments were not eliminated or counter-selected during the process of selfing and selection. These results should justify the development of adapted breeding strategies.

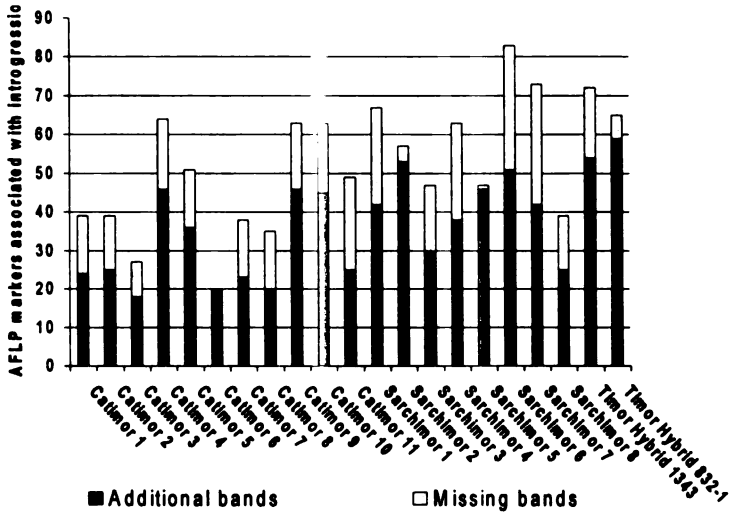

Figure 3. Number of AFLP polymorphic bands attributable to introgression detected in Timor Hybrid-derived accessions (Lashermes et al., 2000)

Behaviour of the *C. canephora* genome and its interaction with the *C. arabica* genome were investigated in tetraploid hybrids (*C. arabica* x *C. canephora* 4x) called arabusta hybrids (Herrera et al., in press). Segregation and co-segregation of Restriction Fragment Length Polymorphism (RFLP) and microsatellite loci-markers were studied in two back-cross (BC1) populations of 28 and 45 individuals. The presence of specific *C. canephora* markers were scored for 11 RFLP and 13 microsatellite loci, distributed on at least 7 of the 11 linkage groups identified in *C. canephora* by Lashermes et al. (in press). The segregation of *C. canephora* alleles in the BC1 plants conformed to the expected values of a theoretical binomial distribution assuming a random chromosome segregation (Figure 4). The recombination rate of *C. canephora* chromosome segments estimated in the arabusta hybrid

was found to be similar to the recombination rate observed in *C. canephora*. The recombination in the tetraploid hybrids appeared therefore not to be affected significantly by the genetic differentiation between chromosomes belonging to the different genomes. The arabusta hybrids appeared to be particularly favourable to intergenomic recombinations. Genes of *C. canephora* might be more readily introgressed into *C. arabica* genome than originally believed.

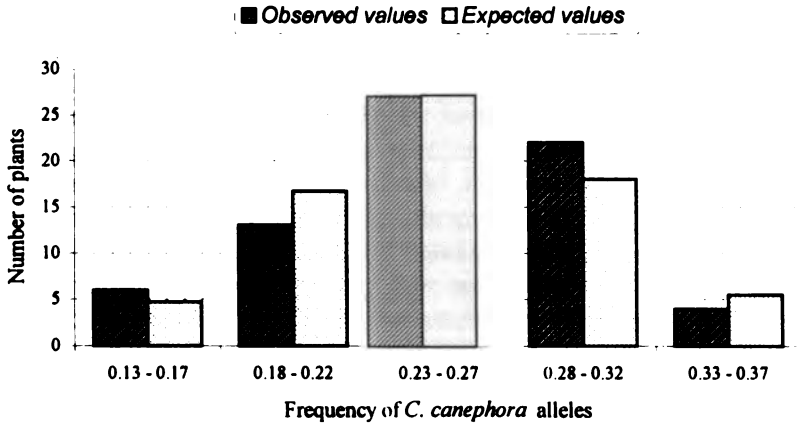

Figure 4. Frequencies of *C. canephora* alleles observed in BC_1 hybrids (*C. arabica* x *C. canephora*) and expected values for a theoretical binomial distribution assuming a random segregation at all loci (Herrera et al., in press)

Introgression from *C. liberica*

The offspring S.288 of a putative spontaneous hybrid (*C. arabica* x *C. liberica*), and 17 introgression lines derived from the cross (S.288 x Kent) were evaluated for introgression of

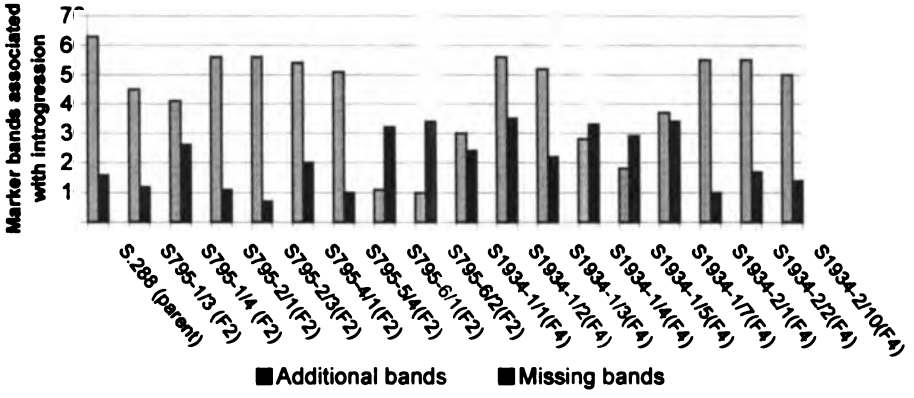

Figure 5. Numbers of AFLP polymorphic bands attributable to introgression in S.288 parent and introgressed lines (Prakash et al., in press)

C. liberica genetic material, using AFLP markers (Prakash et al., in press). The AFLP profiles of introgression lines were compared to 5 accessions each of *C. arabica* and *C. liberica*. The introgression lines were distinguished from the *C. arabica* accessions by 102 markers consisting of 65 additional bands and 37 missing bands. Large variation was observed in the number of additional bands (10 to 56) and missing bands (7 to 35) among the introgression lines (Figure 5). The differences in the level of introgression between introgressed parents, F2

and F4 progenies was not pronounced. The alien genetic material appeared to be fixed and not eliminated or counter-selected over generations. The limited number of introgressed markers in general indicated that the introgression was restricted to few chromosome segments. Considering the 36 AFLP primer combinations common to this study and to the analysis of *C. canephora* introgression (Lashermes et al., 2000), the number of polymorphic bands attributed to introgression was found less in the *C. liberica* introgressed lines than in the Timor Hybrid-derived lines.

CONCLUSION

Molecular markers appeared particularly relevant to fingerprint coffee accessions, to reveal the structure of genetic diversity present in wild and cultivated accessions, and to detect chromosome segments introgressed from diploid relative species. They were also used successfully for characterising mechanisms of introgression in interspecific hybrids between *C. arabica* and diploid relative species. The results allow for the definition of breeding strategies using the whole genetic diversity that are conserved in field genebanks. Wild genitors can be chosen based on the diversity structure revealed by molecular markers and on field characterisation data.

Efforts should now be concentrated on the identification and localisation of resistance genes available in the genetic resources for *C. arabica* breeding. The development of a genetic map would constitute a powerful tool for a molecular control of alien gene transfer. The development of a method of selection assisted by molecular markers would increase the efficiency of coffee breeding programs by 1) allowing for selection at early stage and on a large number of breeding lines, 2) reducing the number of backcross cycles required to restore the quality of the Typica and Bourbon cultivars, and 3) selecting in one-step for various traits or resistance genes.

ACKNOWLEDGMENTS

The research was supported by IRD (Institut de recherche pour le développement, ex ORSTOM, France) and the European Community through the International Scientific Co-operation Programme (INCO-DC contract ERBIC18CT970181). N.S. Prakash was funded by a post-doctoral Grant from the government of India.

REFERENCES

Anthony F., Berthaud J., Guillaumet J.L. and Lourd M., 1987. Collecting wild *Coffea* species in Kenya and Tanzania. Plant Genet. Resources Newsl. **69**: 23-29.

Anthony F., Astorga C. and Berthaud J., 1999. Los recursos genéticos: las bases de una solución genética a los problemas de la caficultura latinoamericana. In: B. Bertrand & B. Rapidel (Eds.), Desafíos de la caficultura centroamericana, pp. 369-406. IICA, San José, Costa Rica.

Anthony F., Bertrand B., Quiros O., Lashermes P., Berthaud J. and Charrier A., 2001. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. Euphytica **118**: 53-65.

Anthony F., Combes M.C., Astorga C., Bertrand B., Graziosi G. & Lashermes P., in press. The origin of cultivated coffee (*Coffea arabica* L.) revealed by molecular markers. Theor. Appl. Genet.

Bertrand B., Aguilar G., Santacreo R. and Anzueto F., 1999. El mejoramiento genético en América Central. In: B. Bertrand & B. Rapidel (Eds.), Desafíos de la caficultura centroamericana, pp. 407-456. IICA, San José, Costa Rica.

Carvalho A., Ferwerda F.P., Frahm-Leliveld J.A., Medina P.M., Mendes A.J.T. and Monaco L.C., 1969. Coffee. In: Ferwerda F.P. and Wit F. (Eds.), Outlines of perennial crop breeding in the Tropics, pp. 189-241. Veenman and Zonen NV, Wageningen.

Carvalho A., Medina Filho H.P., Fazuoli L.C., Guerreiro Filho O. and Lima M.M.A., 1991. Aspectos genéticos do cafeeiro. Rev. Brasil. Genet. **14**: 135-183.

Chevalier A. and Dagron M., 1928. Recherches historiques sur les débuts de la culture du caféier en Amérique. Communications et Actes de l'Académie des Sciences Coloniales, Paris.

Cros J., Lashermes P., Marmey P., Anthony F., Hamon S. and Charrier A., 1993. Molecular analysis of genetic diversity and phylogenetic relationships in *Coffea*. 15th international conference on coffee science, 6-11 June 1993, Montpellier, ASIC, pp. 41-46.

Cros J., Combes M.C., Trouslot P., Anthony F., Hamon S., Charrier A. and Lashermes P., 1998. Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis. Mol. Phylogenet. Evol. **9**: 109-117.

Eskes A.B., 1989 Identification, description and collection of coffee types in P.D.R. Yemen. IPGRI, Rome.

FAO, 1968. FAO coffee mission to Ethiopia 1964 - 1965. FAO, Rome.

Guillaumet J.L. and Hallé F., 1978. Echantillonnage du matériel récolté en Ethiopie. Bull. IFCC **14**: 13-18.

Haarer A.E., 1956. Modern coffee production. Leonard Hill (books) Limited, London.

Herrera J.C., Combes M.C., Anthony F., Charrier A. and Lashermes P., in press. Introgression into the allotetraploid coffee (*Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* x *C. canephora*). Theor. Appl. Genet.

Krug C.A., Mendes J.E.T. and Carvalho A., 1939. Taxonomia de *Coffea arabica* L. Bolétim Técnico n° 62. Instituto Agronômico do Estado, Campinas, Brazil.

Lashermes P., Combes M.C., Cros J., Trouslot P. and Anthony F., 1995. Origin and genetic diversity of *Coffea arabica* L. based on DNA molecular markers. 16th international conference on coffee science, 10-14 April 1995, Kyoto, ASIC, pp. 528-536.

Lashermes P., Agwanda C.O., Anthony F., Combes M.C., Trouslot P. and Charrier A., 1997a. Molecular marker-assisted selection: a powerful approach for coffee improvement. 17th international conference on coffee science, 21-25 July 1997, Nairobi, ASIC, pp. 474-480.

Lashermes P., Combes M.C., Trouslot P. & Charrier A., 1997b. 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 M.C., Robert J., Trouslot P., D'Hont A., Anthony F. and Charrier A., 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. Mol. Gen. Genet. 261 (2): 259-266.

Lashermes P., Andrzejewski S., Bertrand B., Combes M.C., Dussert S., Graziosi G., Trouslot P. and Anthony F., 2000. Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). Theor. Appl. Genet. **100** (1): 139-146.

Lashermes P., Combes M.C., Prakash N.S., Trouslot P., Lorieux M. and Charrier A., in press. Genetic linkage map of *Coffea canephora*: effect of segregation distorsion and analysis of recombination rate in male and female meiosis. Genome.

Lejeune J.B.H., 1958. Rapport au Gouvernement Impérial d'Ethiopie sur la production caféière. FAO, Rome.

Mettulio R., Rovelli P., Anthony F., Anzueto A., Lashermes P. and Graziosi G., 1999. Polymorphic microsatellites in *Coffea arabica*. 18th international conference on coffee science, 2-6 August 1999, Helsinki, ASIC, pp. 344-347.

Montagnon C. and Bouharmont P., 1996. Multivariate analysis of phenotypic diversity of *Coffea arabica*. Genet. Res. Crop Evol. **43**: 221-227.

Prakash N.S., Combes M.C., Somanna N. & Lashermes P., in press. AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. Euphytica.

Sylvain P.G., 1955. Some observations on *Coffea arabica* L. in Ethiopia. Turrialba **5**: 37-53.

Thomas A.S., 1942. The wild arabica coffee on the Boma Plateau, Anglo-Egyptian Sudan. Empire J. Expt. Agric. **10**: 207-212.

Wellman F.L., 1961. Coffee: botany, cultivation and utilization. Leonard Hill Books, London.