Chapter 7

MOLECULAR BREEDING IN COFFEE (*COFFEA ARABICA* L.)

P. Lashermes[1], M. C. Combes[1], P. Topart[2], G. Graziosi[3], B. Bertrand[4], F. Anthony[2]

[1]IRD, Montpellier BP 5045, 34032 Montpellier Cedex I France; [2]CATIE, Turrialba, Costa Rica; [3]Università deli Studi di Trieste, Italy; [4]IICA/PROMECAFE, Costa Rica

Running title: Molecular breeding in coffee

1. Introduction

The cultivated coffee *Coffea arabica* L. (2n=4x=44) is an allotetraploid species native to Africa, containing two diploid genomes that originated from two different diploid wild ancestors (2n=2x=22), *C. canephora* and *C. eugenioides* or ecotypes related to those species (Lashermes *et al.*, 1999). It is the only polyploid species in the genus and is self-fertile while other *Coffea* species are diploid and generally self-incompatible. *C. arabica* is characterised by a very low genetic diversity (Fig. 1), which is attributable to its origin, reproductive biology, and evolution. In addition, most cultivars are derived from the few trees which survived various efforts to spread arabica growing world-wide (Van der Vossen, 1985; Lashermes *et al.*, 1996a). It is believed that the encountered agro-morphological variation which gave rise to so many named varieties, results from few major-gene spontaneous mutations conditioning plant, fruit and seed characters (Carvalho, 1988). The cultivars, therefore, present a homogeneous agronomic behaviour characterised by a high susceptibility to many pests and diseases, and very low adaptability (Bertrand *et al.*, 1999).

Enlarging the genetic base and improvement of arabica cultivars have become priorities for researchers. 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).

T. Sera et al. (eds.), Coffee Biotechnology and Quality, 101–112.

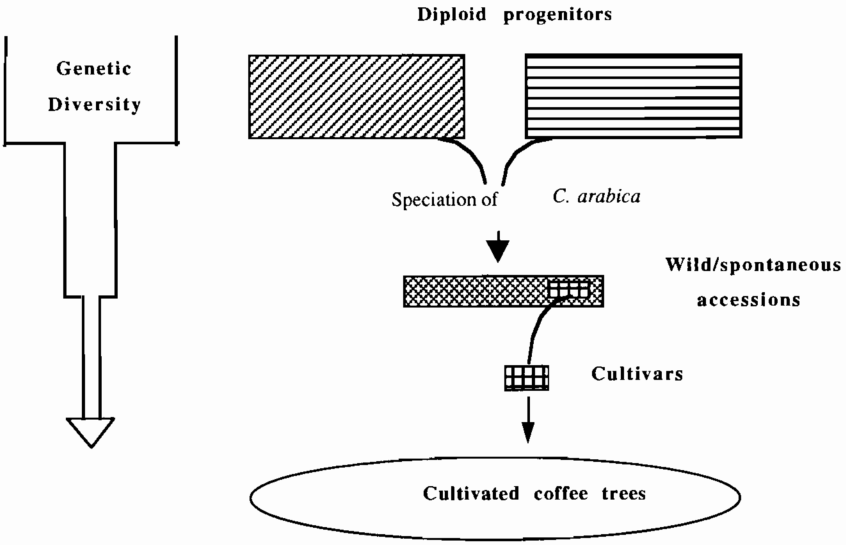

Figure 1. Genetic base of the cultivated Arabica coffee trees.

To date, *C. canephora* provides the main source of disease and pest resistance traits not found in *C. arabica* including coffee leaf rust (*Hemileia vastatrix*), Coffee Berry Disease (CBD) caused by *Colletotrichum kahawae* and resistance to root-knot nematode (*Meloidogyne sp.*). Likewise, other diploid species present considerable interests in this respect. For instance, *C. racemosa* constitutes a promising source of resistance genes to coffee leaf miner (Guerreiro Filho *et al.*, 1999).

The Timor Hybrid is an atypical tree which was identified in a *C. arabica* field, planted in 1927, on the island of Timor (Bettencourt, 1973). Based on information relating to the coffee germplasm introduced into Timor at the beginning of the century, the limited fertility of the original plant, characteristics of disease resistance, and preliminary molecular investigations, it is believed that the Timor Hybrid originated from a spontaneous inter-specific cross between *C. arabica* and *C. canephora* (Bettencourt, 1973; Goncalves and Rodrigues, 1976; Lashermes *et al.*, 1993). Progenies of the Timor Hybrid have been distributed world-wide, and, when observed, showed 2n = 44 chromosomes (Rijo, 1974). In recent decades, they have been used intensively in coffee breeding programmes as the main source of resistance to pests and diseases (Charrier and Eskes, 1997). Exploitation of Timor Hybrid populations has so far relied on conventional procedures in which a hybrid is produced with an outstanding arabica genotype, and the progeny is sledded (or, back-crossed) and selected over at least 3-4 generations. Undesirable genes from the resistance-donor parents are expected to be gradually eliminated. However, conventional coffee breeding methodology faces considerable difficulties in so doing. In particular, strong limitations are due to the long generation time of coffee tree (five years), the high cost of field trial, and the lack of accuracy of current strategy. One can estimate that a minimum of 25 years after hybridisation (five back-cross generations) is required to restore the genetic background of the recipient cultivar and there by ensure good quality of the improved variety. Combining various genes of resistance without reducing coffee quality appears therefore 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. In particular, the development of marker-assisted selection (MAS) programmes promises to overcome present limitations of conventional coffee breeding. General principle of MAS is that if a gene(s) conferring a trait of interest is linked to an easily identifiable molecular marker, it may be much more efficient to select for the marker than for the trait itself.

2. Molecular analysis of arabica coffee introgression lines

Introgressed arabica genotypes derived from the Timor Hybrid were analysed for the presence of *C. canephora* genetic material using the amplified fragment length polymorphism (AFLP) approach (Vos *et al.*, 1995). In order to gain insights into the mechanism of introgression in *C. arabica*, Lashermes *et al.* (2000) estimated the amount of introgression present in such material. The Timor Hybrid-derived genotypes included in this analysis consisted of two accessions representing two different progenies of the Timor Hybrid (progenies 832-1 and 1343), and 19 introgression arabica lines (BC_1F_4) derived from different hybrids between accessions of various Timor Hybrid progenies (832-1, 832-2, and 1343) and commercial arabica cultivars. The Timor Hybrid-derived genotypes were evaluated using 42 different AFLP primer combinations, and compared to 23 accessions of *C. arabica* and 8 accessions of *C. canephora*.

A total of 1062 polymorphic fragments were scored among the 52 accessions analysed (Fig. 2). The number of polymorphic bands was much higher within the canephora accessions (i.e. 945) than within the accessions of *C. arabica* (i.e. 109). The group constituted by the Timor Hybrid-derived genotypes was distinguished from the accessions of *C. arabica* by 178 markers consisting of 109 additional bands (i.e. introgressed markers) and 69 missing bands. AFLP markers identified in Timor Hybrid-derived genotypes were considered as introgressed markers when not detected in any of the accessions of *C. arabica* and observed in at least one of the canephora accessions analysed in this study.

AFLP, therefore, seemed to be an extremely efficient technique for DNA marker generation in coffee trees as well as for introgression detection in *C. arabica.* The genetic diversity observed in the Timor Hybrid-derived genotypes appeared approximately double to that in *C. arabica.* Although representing only a small proportion of the genetic diversity available in *C. canephora,* the Timor Hybrid obviously constitutes a considerable source of genetic diversity for arabica breeding. Analysis of genetic relationships among the Timor Hybrid-derived genotypes suggested that introgression was not restricted to chromosome substitution but also involved chromosome recombination. Furthermore, the Timor Hybrid-derived genotypes varied considerably in the number of AFLP markers attributable to introgression (Fig. 3).

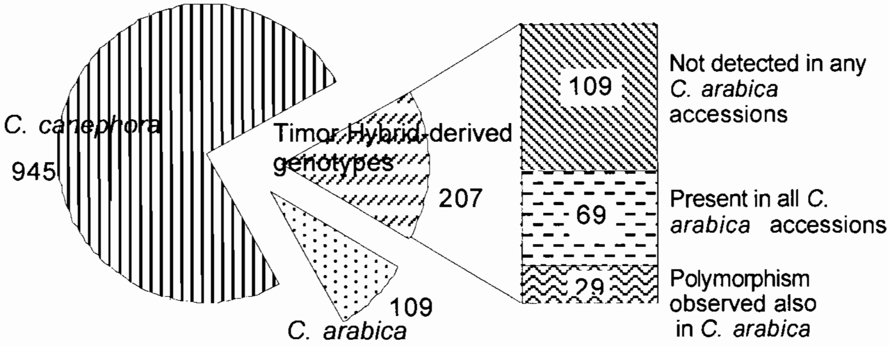

Figure 2. Pie charts depicting the numbers of polymorphic AFLP bands observed among individuals within each group constituted by the accessions of *C. arabica*, *C. canephora* and introgressed Timor Hybrid-derived genotypes, respectively. For the introgressed material, the polymorphic markers either attributable to the Arabica parent or associated with the introgression of *C. canephora* chromosome segments were distinguished (Lashermes *et al.* 1999).

In this way, the introgressed markers identified in the arabica coffee introgressed genotypes were estimated to represent from 8% to 27% of the *C. canephora* genome. Nevertheless, the amount of alien genetic material in the introgression arabica lines remains always substantial and should justify the development of adapted breeding strategies.

3.1.QUARANTINED PATHOGENES

If a virulent pathogen does not occur naturally in the test environment, artificial inoculation is prohibited for safety reasons. For instance, CBD is still restricted to the continent of Africa, and the availability of markers linked to the resistance gene(s) as reported for the *T* gene (Agwanda *et al.*, 1997), could allow pre-emptive breeding in countries (Asia, Latin America) where quarantine barriers are still effective.

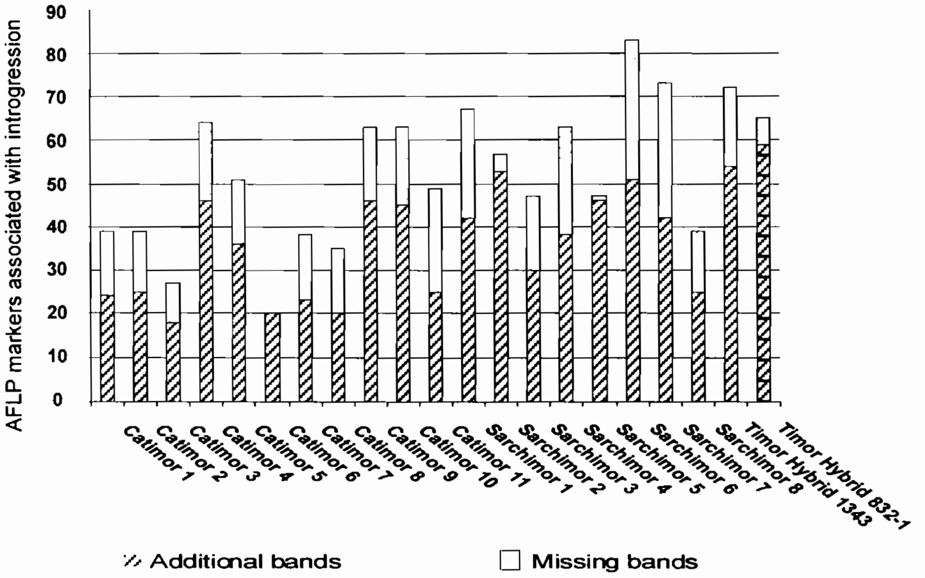

Figure 3. Numbers of AFLP polymorphic bands attributable to introgression detected in Timor Hybrid-derived genotypes (Lashermes *et al.* 1999).

3.2. RELIABILITY/LIMITATION OF DIRECT TESTING

Conventional selection progress could be hampered by the difficulty to ensure reliable test for the resistance trait. Seedling test could also present strong inconvenient. For instance, the present test for evaluation for root-knot nematode is destructive leading to important difficulties in the utilisation of identified plant resistance sources

(Bertrand *et al.*, 1997). In addition, expression of many resistance genes can be strongly influenced by environmental conditions.

3.3. DEVELOPMENTALLY REGULATED CHARACTER

Early selection based on the marker genotype of young seedlings would be particularly beneficial for late expressed traits.

3.4.TRANSFER OF RECESSIVE RESISTANCE GENES

The classical procedure of transferring a recessive resistance gene includes a progeny test after each back-cross generation to determine the presence of the desired allele. With MAS, the transfer can be accomplished without interruptions leading to an important time saving.

3.5. PYRAMIDING OF RESISTANCE GENES/COMBINING VALUABLE TRAITS

Pyramiding of resistance genes has been suggested as a strategy to provide durable resistance (i.e. coffee leaf rust). However, conventional breeding is complicated by the fact that, it is difficult or often impossible to distinguish the various resistance genotypes. Once the different genes conferring resistance to the same pathogen are tagged by tightly linked marker, they could be relatively easily be accumulated into a single genotype via marker-facilitated selection. Comparable advantages versus conventional are procured when trying to combine simultaneously resistance genes to different disease/pests.

4. Molecular-assisted back-cross breeding

Repeated back-crossing simultaneously accomplishes two essential goals: 1) allows segregation to remove donor parent chromosomes unlinked to the target gene, and, 2) allow recombination to remove donor parent segments which are linked to the target gene. Both objectives could be considerably facilitated by the use of molecular markers.

4.1.GENOME SELECTION

Beside the target trait, it is important to consider the complete genome of individuals. Chromosomal segments are segregating within back-cross progenies and the individuals show various contents of the desired parental genome (Fig. 4). A genome selection could, therefore, be performed by the use of markers scattered throughout the genome, resulting in a reduction in number of back-cross generations required to restore the genetic background of the recipient cultivar.

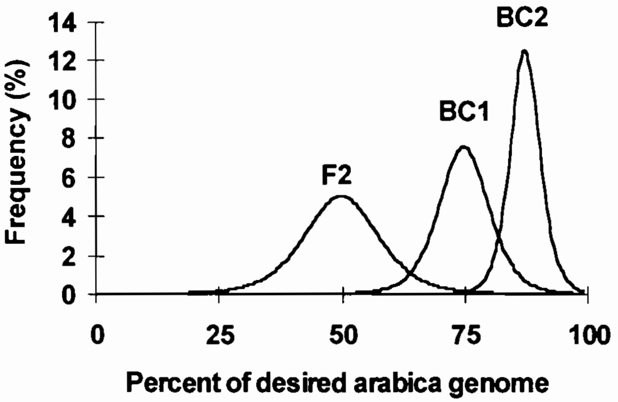

Figure 4. Frequency of individuals in F2, BC1, and BC2 having various contents (%) of the desired parental genome.

Values were estimated (Fig. 5) for a hypothetical arabica genome of 22 chromosome pairs of, on average, 100 cM each (Total genome of 2200 cM), using equations developed by Hillel *et al.* (1990) and Hospital *et al.* (1992). In the absence of selection, parental donor DNA was only removed by a factor of two in each generation. Simulations were given for MAS programme in which the either 10, or 2% best (in terms of percent recurrent parent genome) individuals in each generation were used as the parent for the next generation. Results equivalent to BC5 generation without selection was obtained after only two marker-assisted BC generations allowing a considerable time saving.

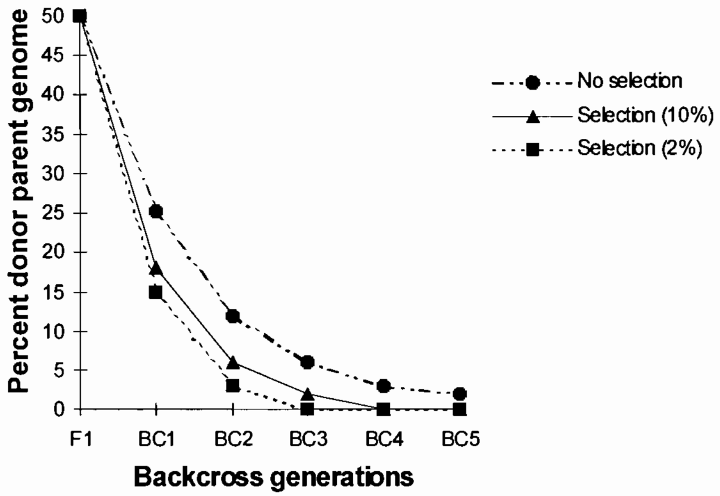

Figure 5. Average content (%) of the donor parental genome in back-cross generations under various intensities of genomic selection.

Table 1 presents the results of various levels of selection imposed on BC1 and BC2. Apparently the moderate levels of selection in both BC1 and BC2 resulted in individual, almost identical to the recipient variety. Achieving a similar result by a unique genome selection in one of these two generations requires extremely intensive selection effort.

Table 1. Percentage of the recipient arabica genome under various intensities selection in BC_1 and BC_2.

Selection in BC_2	Selection in BC_1 (%)			
(%)	100	30	10	2
100	87.5	89.7	91.0	92.0
30	90.9	92.4	93.2	94.1
10	92.5	93.6	94.5	95.1
2	94.1	94.9	95.5	96.1

4.2. REDUCING LINKAGE DRAG

Removing of the linked donor segment could take many generations (Stam and Zeven 1981). Many examples of "linkage drag" are known in which undesirable traits that are closely linked to a target gene, are carried out along during breeding programme (Zeven *et al.*, 1983, Young and Tanksley, 1989). For instance, in Arabica, even after 6 back-cross generations, a region of 32cM flanking a target gene is expected to persist (Fig. 6).

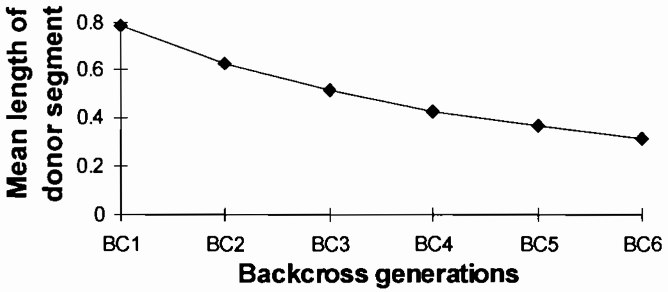

Figure 6. Mean length of donor segment surrounding the target gene after various numbers of back-cross generations The length is expressed as a proportion of the carrier chromosome (chromosome of 100 cM long) (after Stam and Zeven 1981).

In most plant genomes 32cM is enough DNA to contain hundreds of genes. DNA markers can be used to eliminate, or at least significantly reduce, linkage drag by allowing the identification of rare recombinant individuals, which are usually only selected by chance in classical breeding (Paterson *et al.*, 1991). In approximately 150 back-cross plants, there could be 95% chance that at least one plant would have experienced a crossover within 1 cM on one side, or, the other of the gene being selected. With one additional back-cross generation of 300 plants, there would be a 95% chance of a crossover within 1 cM of the other side of the gene, generating a segment surrounding the target gene of less than 2 cM.

5. Genetic mapping

Some utilization of MAS presupposes the existence of a detailed linkage map, which represents the relative order of genetic markers, and their relative distances from one to another along each chromosome of an organism (Paterson *et al.*, 1991).

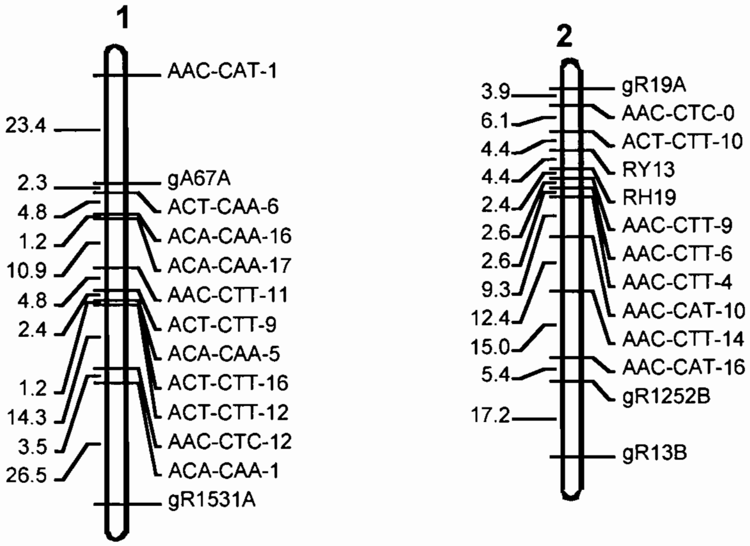

Figure 7. Linkage groups 1 and 2 of the genetic map of *C. canephora*. Map distances in cM are indicated on the left side of linkage groups and marker locus name is on the right side.

A first genetic map of the coffee genome was established using a random population derived from the clone IF200 of *C. canephora* (Paillard *et al.*, 1996). So far, more than 150 markers have been placed on the 11 linkage groups/chromosomes (Fig. 7). Additional linkage maps on different coffee species including *C. arabica* are being constructed. In so doing, the development of microsatellite PCR-based markers will be particularly useful (see chapter by Rovelli *et al.* in this book). These genetic linkage maps are considered useful in providing important information on coffee genome and chromosomal organisation. In particular, one might use it to map important genes as done for the S-locus controlling self-incompatibility (Lashermes *et al.*, 1996b).

6. Conclusions and prospects

The development of molecular markers for coffee trees has opened a new perspective in breeding. The conventional selection of self-, or back-crossed coffee tree progenies for further breeding is extremely laborious and time-consuming. The implementation of MAS could, therefore, be very promising. In particular, the integration of MAS in coffee breeding promises to drastically increase the efficiency of breeding programmes by:

1) allowing for selection at an early stage and on a large number of breeding lines,
2) reducing the number of back-cross cycles required to restore the quality of the traditional cultivars,
3) combining in one-step, selection for various traits or genes of resistance.

Furthermore, new findings from genome research indicate that there is tremendous genetic potential locked up in wild and cultivated germplasm resources that can be released only by shifting the paradigm from searching for phenotypes to searching for superior genes with the aid of molecular linkage maps (Tanksley and McCouch, 1997). In addition, ongoing technological developments, including automation, allele-specific diagnostics and DNA chips, will make MAS approaches based on large-scale screening much more powerful and effective.

7. Acknowledgements

This work was supported in part by the European Community through the International Scientific Co-operation Programme (Inco-Dc Contract Erbic 18CT970181).

8. References

Agwanda, C., Lashermes P., Trouslot P., Combes M.C., Charrier, A. (1997), Identification of RAPD markers for resistance to Coffee Berry Disease, *Colletotrichum kahawae*, in Arabica coffee. *Euphytica*, 97:241-248

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

Bertrand, B., Aguilar G., Bompard E., Rafinon A., Anthony, F. (1997), Comportement agronomique et résistance aux principaux déprédateurs des lignées de Sarchimor et Catimor au Costa Rica. *Plantation, Recherche, Développement*, 5:312-321

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

Bettencourt, A. (1973), Cosideraçoes gerais sobre o "Hibrido de Timor". Instituto Agronomico de Campinas, Circular No. 31 (285p), Campinas, Brazil

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

Charrier A., Eskes A.B. (1997), Les caféiers. In: *L'amélioration des plantes tropicales*, Charrier A, Jacquot M, Hamon S, Nicolas D (eds.). Collection Repères, CIRAD et ORSTOM, Paris, pp 171-196

Goncalves, M. M., Rodrigues, M. L. (1976), Estudos sobre o café de Timor. II. Nota sobre as posibilidades de producao do hibrido de Timor no seu habitat natural. Lisboa, Portugal. Missao de Estudos Agronomicos do Ultramar, Portugal. Comunicacoes, 86:31-72

Guerreiro Filho, O., Silvarolla M. B., Eskes, A. B. (1999), Expression and mode of inheritance of resistance in coffee to leaf miner *Perileucoptera coffeella*. *Euphytica*, *105*: 7-15

Hillel, J., Schaap T., Haberfeld A., Jeffreys A.J., Plotzky Y., Cahaner A., Lavi, U. (1990), DNA fingerprints applied to gene introgression in breeding programs. Genetics, 124: 783-789

Hospital, F., Chevalet C., Mulsant, P. (1992), Using markers in gene introgression breeding programs. *Genetics*. *132*: 1199-1210

Lande, R., Thompson, R. (1990), Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, *124*: 743-756

Lashermes, P., Cros J., Marmey P., Charrier, A. (1993), Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. *Genet. Res. Crop. Evol.*, *40*: 91-99

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

Lashermes, P., Couturon E., Moreau N., Paillard M., Louarn, J. (1996b), Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. *Theor. Appl. Genet.*, 93: 458-462

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*:259-266

Lashermes, P., Andrzejewski, A., 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*: 139-146

Melchinger, A.E. (1990), Use of molecular markers in breeding for oligogenic disease resistance. Plant *Breed.*, 104: 1-19

Paillard, M., Lashermes P., Pétiard, V. (1996), Construction of a molecular linkage map in coffee. Theor. *Appl. Genet.*, *96*: 41-47

Paterson, A.H., Tanksley S.D. and Sorrells, M. E. (1991), DNA markers in plant improvement. *Adv. Agronomy*, *46*: 39-90

Rijo, L. (1974), Observações cariologicas no cafeeira "Hibrido de Timor". *Portugaliae Acta Biologica*, serie A, *vol. XIII*: 157-168

Stam, P., Zeven, A. C. (1981), The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by back-crossing. *Euphytica*, *30*: 227-238

Tanksley, S.D., McCouch, S. R. (1997), Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, *277*: 1063-1066

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

Vos, P., Hogers R., Bleeker M., Reijans M., Van der Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. and Zabeau, M. (1995), AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.*, *23*: 4407-4414

Young, N.D., Tanksley, S.D. (1989), RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during back-cross breeding. *Theor. Appl. Genet.*, 77: 353-359

Zeven A.C., Knott D.R., Johnson R. (1983), Investigation of linkage drag in near isogenic lines of wheat by testing for seedling reaction to races of stem rust, leaf rust and yellow rust. *Euphytica*, 32: 319-327

Lashermes Philippe, Combes Marie-Christine, Topart Patrick., Graziosi G., Bertrand B., Anthony François. (2000).

Molecular breeding in coffee (Coffea arabica L.). In : Sera T. (ed.), Soccol C.R. (ed.), Pandey A.(ed.), Roussos Sevastianos (ed.). Coffee biotechnology and quality.

Dordrecht : Kluwer, p. 101-112.

International Seminar on Biotechnology in the Coffee Agro-Industry, 3., Londrina (BRA), 1999/05/24-28.

ISBN 0-7923-6852-8.