TESFAHUN ALEMU SETOTAW

GENETIC DIVERSITY AND GENOME INTROGRESSION IN COFFEE

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2009

TESFAHUN ALEMU SETOTAW

GENETIC DIVERSITY AND GENOME INTROGRESSION IN COFFEE

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 11 de Dezembro de 2009

Pesq. Eveline Teixeira Caixeta (Co-Orientadora)	Prof. Cosme Damião Cruz (Co-Orientador)
Pesq. Antonio Alves Pereira	Pesq. Antonio Carlos Baião de Oliveira

Prof. Ney Sussumu Sakiyama
(Orientador)

Dedicated to my mother Yelfwaga Ayelegn Bimerew

ACKNOWLEDGMENT

I would like to thank the Postgraduate program of Genetics and Breeding at the Universidade Federal de Viçosa, for giving the opportunity to study my doctor science. The professors of the postgraduate program of Genetics and Breeding for the knowledge they shared me and support for the last four years.

Special thanks also for my mother Yelfwaga Ayelegne, my brother Fikruadis Abate, and my sisters Firehiwote Aylegn and Maré Ayelegn for their unlimited support for me through my life and for their passion. I thank Adane Chofeir and Endeshaw Terefe for their friendship.

I also would like thank my advisor Prof. Ney Sussumu Sakiyama for his guidance and support from the first day of my life here in Brazil until this moment.

My special thanks also go to colleagues of BioCafé/BIOAGRO for their help at all moment in my stay in Brazil. Especial thanks also for Eveline Teixeira Caixeta and Eunize Zambolim for their technical assistance and their advice.

My special thanks also go to TWAS/Cnpq for the financial support.

To my roommates Jose Osmar da Silva costa, Thiago, Rodrigo and my friends Telma, Kátia, Juliana, Robson for their love, support and good times I passed with them.

I want to pass my deep thanks for others I didn't describe their name here but they contributed for my achievement.

BIOGRAPHY

Tesfahun Alemu was born in 1975 in Wogera Awraja, Gondar Administrative Region, Ethiopia. He had his elementary education at Woreta Elementary School, from 1980 to 1986. He studied in Junior Secondary School from 1986 to 1991 and had high school education in Woreta Secondary School at Woreta. He joined Alemaya University of Agriculture in 1991 and graduated in 1995 with B.Sc. degree in Plant Sciences. After graduation, he was employed in Commission of Sustainable Agriculture and Environment Rehabilitation for Amhara Region (Co SAERAR) at Bahirdar. After serving more than one year, he joined Ethiopian Agricultural Research Organization (EARO) in February, 1997 at Kulumsa Research Center under Barley Improvement Program.

He served in Barley Improvement Program of EARO at Kulumsa Research Center until he joined the School of Graduate Studies for M.Sc. in Plant Breeding at Alemaya University in September 1999 and graduated in 2001. After graduated he returned back to EARO to work as a researcher.

He joined the department of postgraduate program in Genetics and Breeding of Universidade Federal de Viçosa in August 2005 and submitted his thesis to defend in December 2009.

TABLE OF CONTENT

RESUMO vii

ABSTRACT viii

GENERAL INTRODUCTION 1

CHAPTER 1 3

GENETIC DIVERSITY PATTERNS IN *Coffea arabica* L. CULTIVARS GROWN IN BRAZIL BASED ON COEFFICIENT OF PARENTAGE. 3

PADRÃO DA DIVERSIDADE GENÉTICA EM CULTIVARES DO *Coffea arabica* L. LANÇADOS NO BRASIL BASEADO COEFICIENTE DE PARENTESCO. 4

1. INTRODUCTION 5

2. MATERIALS AND METHODS 8

3. RESULTS 14

4. DISCUSSION 20

5. CONCLUSIONS 23

6. REFERENCES 24

CHAPTER 2 26

GENOME INTROGRESSION OF HÍBRIDO DE TIMOR AND ITS RELATIONSHIP WITH OTHER COFFEE SPECIES. 26

INTROGRESSÃO DO GENOMA DO HÍBRIDO DE TIMOR E RELAÇÃO COM OUTRAS ESPÉCIES DO CAFÉ. 27

1. INTRODUCTION 28

2. MATERIALS AND METHODS 30

3. RESULTS 33

4. DISCUSSION 42

5. CONCLUSIONS 46

6. REFERENCES 47

CHAPTER 3 51

GENETIC DIVERSITY AND BREEDING POTENTIAL OF HÍBRIDO DE TIMOR. 51

DIVERSIDADE GENÉTICA E POTENCIAL DO HÍBRIDO DE TIMOR NO MELHORAMENTO DE CAFÉ. 52

1. INTRODUCTION 53

2. MATERIALS AND METHODS 55

3. RESULT 59

4. DISCUSSION 65

5. CONCLUSIONS 69

6. REFERENCES 70

RESUMO

SETOTAW, Tesfahun Alemu, D.Sc., Universidade Federal de Viçosa, Dezembro de 2009. **Diversidade genética e introgressão do genoma em café.** Orientador: Ney Sussumu Sakiyama. Co-orientadores: Eveline Teixeira Caixeta e Cosme Damião Cruz.

Para estudar a diversidade genética e o padrão de melhoramento entre cultivares de *C. arabica* lançados entre 1939 e 2009, foi utilizado um total de 110 cultivares. A fim de avaliar a introgressão do genoma do Híbrido de Timor e sua relação com outras espécies de café, foram utilizados cinco acessos de *C. arabica*, dez de *C. canephora* var conilon, quinze de *C. canephora* var robusta e quarenta e seis de Híbrido de Timor, com os marcadores moleculares AFLP, RAPD e SSR. O coeficiente de parentesco estimado entre cultivares de *C. arabica* foi usado para o estudo da diversidade genética e do padrão de melhoramento do café arábica no Brasil, mostrando uma baixa diversidade genética. O padrão de melhoramento de *C. arabica* no Brasil foi definido pelas treze linhagens ancestrais. Entre elas, Bourbon Vermelho, Sumatra e Híbrido de Timor contribuíram com mais de 80.00% dos genes para os cultivares de *C. arabica*. As duas primeiras progênies, Mundo Novo e Icatu Vermelho, contribuíram com 87.56% dos genes para os cultivares de *C. arabica* no Brasil. A diversidade genética entre cultivares de *C. arabica* lançados recentemente foi aumentada com a introdução de novas linhagens parentais no programa de melhoramento. O estudo de relação genética entre Híbrido de Timor e outras espécies mostrou alta similaridade genética entre Híbrido de Timor e *C. arabica.* A análise de introgressão do genoma entre Híbrido de Timor CIFC 4106 com *C. arabica* e *C. canephora* var robusta mostrou 18.9% de introgressão do genoma de *C. canephora.* A mesma análise, considerando todos os acessos de Híbrido de Timor, foi de 10.00%, o que confirma a baixa introgressão de *C. canephora*. Este resultado confirma que Híbrido de Timor não é planta F_1, mas que é proveniente de, no mínimo, dois retrocruzamentos com *C. arabica*. Além disso, este estudo mostrou a existência de alta diversidade genética entre acessos de Híbrido de Timor, o que é importante no melhoramento de *C. arabica* no Brasil e no mundo, uma vez que Híbrido de Timor é usado como fonte de resistência para doenças e pragas no café.

ABSTRACT

SETOTAW, Tesfahun Alemu, D.Sc., Universidade Federal de Viçosa, December, 2009. **Genetic diversity and genome introgression in coffee.** Advisor**:** Ney Sussumu Sakiyama. Co-advisors: Eveline Teixeira Caixeta and Cosme Damião Cruz.

To study the genetic diversity among cultivars of *C. arabica* released from 1939 to 2009 and to study the breeding pattern, a total of 110 cultivars were included. Five *C. arabica,* ten *C. canephora* var conilon, fifteen *C. canephora* var robusta and forty six accessions of Híbrido de Timor were included to study the genome introgression of *C. canephora* into Híbrido de Timor and its relationship with other coffee species. To study the genome introgression and the genetic diversity within Híbrido de Timor, AFLP, RAPD and SSR molecular markers were used. The estimated coefficient of parentage among cultivars of *C. arabica* was used to study the genetic diversity and the breeding pattern of *C. arabica* in Brazil. The study showed low genetic diversity among Brazilian *C. arabica* cultivars. The breeding pattern of cultivars of *C. arabica* was defined by 13 ancestral lines. Among them, Bourbon Vermelho, Sumatra and Híbrido de Timor contributed more than 80% of the gene to *C. arabica* cultivars in Brazil. Mundo Novo and Icatu Vermelho, the first progenies, contributed 87.65% of the gene to Brazilian *C. arabica* cultivars. The genetic diversity among cultivars released in recent years increased due to the introduction of new parental lines in the breeding program of *C. arabica*. The genetic relationship study between Híbrido de Timor and other species showed high genetic similarity between Híbrido de Timor and *C. arabica*. The genome introgression analysis between Híbrido de Timor CIFC 4106 with *C. arabica* and C. canephora showed 18.9% of the genome of Híbrido de Timor introgressed from *C. canephora*. The mean genome introgression of *C. canephora* into Híbrido de Timor considering all accessions of Híbrido de Timor was 10.00%, which confirmed Híbrido de Timor is not an F_1 plant instead at least two times backcrossed with *C. arabica*. In addition, the study demonstrated the existence of high genetic diversity among accessions of Híbrido de Timor, which is important for the future breeding program of arabica coffee, since it is used as source for resistance gene for diseases and pests.

GENERAL INTRODUCTION

The first coffee seed was introduced in Brazil in 1727 from Guiana Francesa to the north of Brazil and rapidly distributed all over the country in the direction of north to south (Eccardi and Sandalj 2003). The first cultivar introduced in Brasil was *C.arabica var typica.* Currently Brazil is the major producer and exporter of coffee for the international coffee market. The two principal species of coffee cultivated in Brazil are *C. arabica* L. and *C. canephora* Pierre. *C. arabica* occupied the large portion of the area planted in São Paulo, Minas Gerais and Parana. *C. arabica* L is a true allotertraploid species with chromosome number of 2n=4x=44 (Clarindo and Carvalho, 2008) and autogamous with variable rate of out crossing reach up to 15 % (Carvalho 1988).

The existed genetic variability in *C. arabica* is considered narrow and this trend is true here in Brazil too. One of the principal cause for the narrow genetic base in Brazil is most of the coffee cultivars developed were derived from small number of mother plant. Since genetic variability is an important base for the success of any breeding program, the coffee breeding program of Brazil introduced *C. arabica* accessions from different countries in deferent periods. After the first introduction of *C. arabic var typica*, introduced Bourbon Vermelho in 1852 from the Union Island and coffee Sumatra from the Island of Sumatra in 1896 (Carvalho, 1957). After this period introductions were done to increase the genetic base of the coffee Arabica breeding program in Brazil (Carvalho et al., 1989 Bettencourt 1973, 1968).

From these centers coffee cultivars were released for commercial production but the genetic background these cultivars were similar. To avert this situation the breeding programs introduced the interspecific cross such as, Icatu, and Híbrido de Timor (Carvalho et al., 1989 Bettencourt 1973, 1968). The Hibrido de Timor resulted from the natural crossing between *C. arabica* and *C. canephora* species and used as source of gene resistance in the breeding program of coffee arabica (Carvalho et al., 1989 Bettencourt 1973, 1968). The detail description about coffee breeding program in Brazil and principal coffee varieties is presented by Carvalho and Fazuoli (1988).

C. arabica production was greatly affected by diseases and pests which reduce its productivity, due to lacks of resistance genes for the major diseases and pests. Currently the breeding program of C. arabica is using Híbrido de Timor as a source of gene for resistance to disease and pests. Híbrido de Timor was first found in plantation of cultivar Típica in Timor Island in 1912 (Bettencourt 1973) and used as source of resistance gene for economically important diseases and pests of coffee such as coffee

leaf rust (*Hemilia vastatrix*), coffee berry disease (CBD) caused by *Colletotrichum Kahawae*, root knot nematode (M*eloidgyne exigua*) and bacteriosis caused by *Pseudomonas syringae* pv garçae (Bertrand et al. 2003). Using Híbrido de Timor as source of resistance gene, cultivars were released for production in Kenya, Brazil, Colombia and Costa Rica (Lasheremes et al. 2000; Charries and Eskes 1997; Bertrand et al. 2003; Perreira et al. 2005). The first introduction of Híbrido de Timor accessions to Brazil date back from 1976 via vegetative propagation and seeds from CIFC (Centero da Investigação de Ferrugem do Café, Portugal), IIAA (Instituto Investigação Agronomia de Angola) and ERU (Estação Rgional de Uige) Perriera et al. (2002). These materials were used extensively in the breeding program of coffee for resistance to diseases and pests in Brazil.

Genome introgression of *C. arabica* and *C. canephora* into accession of Híbrido de Timor was studied using AFLP molecular marker technique and found that introgression of *C. canephora* genome ranged from 8% (in Catimor 3) to 25% (in Sachimor) (Lashermes et al. 2000). Since Brasilian germplasm bank has several accession of Híbrido de Timor available and they are important sources of disease resistance gene and used in large extent in the breeding program of coffee in Brazil and the world, understanding the genome introgression from their origin (*C. arabica* and *C. canephora*) and their relation to others species is important for the breeding program of coffee. In addition it is also important to study the genetic diversity existed among accessions of Híbrido de Timor which helps to exploit for the future breeding programs of coffee.

So this work was done with the following objectives:

1) Analyze the coefficient of parentage among *C. arabica* cultivars and the ancestral lines, study the genetic diversity among cultivars, estimate the genetic contribution of each ancestral line for each cultivar, and study the breeding pattern of the breeding programs.
2) Molecular characterization of accession of Híbrido de Timor and its relationship with other coffee species (such as *C. arabica* and *C. canephora*); and Knowing the contribution of *C. arabica* and *C. canephora* in genome of Híbrido de Timor accessions.
3) Investigating the existed genetic diversity among the accessions of Híbrido de Timor using RAPD, AFLP and SSR molecular markers.

CHAPTER 1

GENETIC DIVERSITY PATTERNS IN *Coffea arabica* L. CULTIVARS GROWN IN BRAZIL BASED ON COEFFICIENT OF PARENTAGE.

ABSTRACT

The genetic diversity analysis of 110 cultivars of *C. arabica* released from 1939 to 2009 was done based on the coefficient of parentage (COP). In addition, the genetic contribution of each ancestral line for each cultivar and the breeding pattern of the breeding programs were studied. The low genetic diversity was observed within the *C. arabica* cultivars. The genetic base of the 110 cultivars was defined by 10 ancestors. The seven ancestors contributed 98.76 % of the gene in *C. arabica* cultivars. Bourbon Vermelho contributed 47.59 % for the genetic pool of the *C. arabica* cultivars followed by Sumatra (20.26 %) and Icatu (13.23%). The 98.76 % of the genetic base of *C. arabica* cultivars constituted by seven ancestors indicated the narrow genetic base of the cultivars. The increase in the genetic diversity among Brazilian *C. arabica* cultivars was observed in recent decades with the introduction of new parental lines with diverse genetic base. But still in Brazil the mean COP value among cultivars of *C. arabica* is very high when compared with other crops studied. The 110 cultivars clustered into four cluster groups based on COP. The distributions of genotypes over the cluster groups showed the effect of parental line contribution. The result demonstrated the importance of understanding the genetic base of the *C. arabica* cultivars and planning the future breeding programs to develop cultivars with different genetic background.

Key words: analysis multivariate, genetic contribution, genetic variability, clustering analysis.

PADRÃO DA DIVERSIDADE GENÉTICA EM CULTIVARES DO *Coffea arabica* L. LANÇADOS NO BRASIL BASEADO COEFICIENTE DE PARENTESCO.

RESUMO

Analise de diversidade genética de 110 cultivares do *C. arabica* lançadas entre anos 1939 a 2009 foi feito com base no coeficiente de parentesco (COP). A contribuição genética de cada linhagem ancestral e o padrão do melhoramento do programas foram estudados. Baixa diversidade genética foi observada entre os cultivares de *C. arabica*. A base genética dos 110 cultivares foi definido pelo treze ancestrais. Sete ancestrais contribuíram 98.76% dos genes do cultivares do *C. arabica* no Brasil. Bourbon Vermelho contribui 47.56 % dos genes para cultivares de *C. arabica* seguido pelo Sumatra (20.26%) e Icatu (13.23%). A contribuição dos 98.76% dos genes do cultivares dos *C. arabica* pelos sete ancestrais confirma a base genética das cultivares no Brasil é estreita. O aumento na diversidade genética dos cultivares lançados nas décadas recentes ocorre pela introdução das novas linhagens parentais com diversos backgrounds genéticos no melhoramento do café. No Brasil, o valor médio do COP ainda é alto em *C. arabica* em comparação a outras culturas estudadas. A análise do agrupamento baseada COP dividiu 110 cultivares de *C. arabica* em quatro grupos. A distribuição dos cultivares nos grupos formados mostrou o efeito da contribuição das linhagens parentais. Esse resultado demonstrou a importância de se conhece a base genética de cultivares de *C. arabica* para planejar programas de melhoramento futuros, possibilitando o desenvolvimento de novos cultivares com background genético diferente.

Palavras chave: análise multivariada, contribuição genética, variabilidade genética, análise de agrupamento

1. INTRODUCTION

The first coffee seed was introduced in Brazil in 1727 from Guiana Francesa to the north of Brazil and rapidly distributed all over the country in the direction of north to south (Eccardi and Sandalj 2003). The first cultivar introduced in Brasil was *C.arabica* var typica. Currently Brazil is the major producer and exporter of coffee to the international market. The two principal species of coffee under cultivation in Brazil are *C. arabica* L. and *C. canephora* Pierre. *C. arabica* occupied the large portion of the planted area of coffee in São Paulo, Minas Gerais and Parana. *C. arabica* L is a true allotertraploid species with chromosome number of 2n=4x=44 (Clarindo and Carvalho, 2008) and autogamous with variable rate of out crossing reach up to 15 % (Carvalho 1988).

The existed genetic variability in *C. arabica* is considered narrow and this trend is true here in Brazil too. The principal cause of the narrow genetic base in *C. arabica* in Brazil is most of the coffee cultivars were derived from small number of mother plant. Since genetic variability is an important base for the success of any breeding program, the Brazilian coffee breeding programs introduced *C. arabica* accessions from different countries in different periods. After the first introduction of *C. arabic* var typica, Bourbon Vermelho was in 1852 from the Island of the Reunions, and coffee Sumatra from the Island of Sumatra in 1896 (Carvalho, 1957). After this period introductions were done to increase the genetic base of the arabica coffee breeding program in Brazil (Carvalho 1988, Bettencourt 1973, Bettencourt and Carvalho 1968).

The first organized coffee breeding program in Brazil was started in 1927 by the Instituto Agronômico de Campinas (IAC) São Paulo. From this period the coffee research programs developed potential high yielding and disease resistant varieties for commercial production. Currently, the coffee breeding research is carried out in Minas Gerais, Paraná, Bahia and Espírito Santos. From these institutes coffee cultivars were released for commercial production but their genetic background were similar. To avert this situation the breeding programs introduced the interspecific cross such as, Icatu, and Híbrido de Timor (Carvalho et al., 1989 Bettencourt 1973, 1968). The Hibrido de Timor was resulted from the natural crossing between *C. arabica* and *C. canephora* species and used as source of gene resistance in the breeding program of coffee arabica (Bettencourt and Carvalho1968, Bettencourt 1973, Carvalho et al. 1989). The detail

description about coffee breeding program in Brazil and principal coffee varieties is presented by Carvalho and Fazuoli (1993).

The genetic diversity available in the global germplasm collection is higher than the genetic diversity explored in applied plant breeding. The reduction in genetic diversity is the potential problem in the future breeding programs and it will result in genetic vulnerability (Zhou et al. 2002). This problem is also evident in *C. arabica* which has low genetic diversity among cultivars. In any breeding program diversifying the genetic base and increasing the number of varieties released for production with different genetic composition is vital, which helps to reduce looses due to disease out break and other constraints.

The success of any breeding program depends on the complete knowledge and understanding of the genetic diversity of the available germplasm. For this reasons the breeder try to study the genetic diversity within the base population to select parents for the crossing programs. In many crops, the genetic improvement for yield generally was accompanied by a loss in genetic diversity among the cultivars (Walsh, 1981). To avoid the lose in genetic diversity understanding the genetic base population and cultivars using the coefficient of parentage will be crucial and may give a direction during the selection of parents for the future crossing programs. Estimating the genetic diversity between plants is useful in studying the evolution of plant populations or species and planning crosses for hybrid or homozygous cultivar development (Cox et al. 1985). One of the methods used to understand the relationship between genotypes or cultivars within the base population and study the genetic diversity is the use of the coefficient of parentage (COP), which explains the genetic and parental relationship between each cultivar. COP between two cultivars is defined as the probability that a random allele in one cultivar is identical by descent to a random allele at same locus in the other cultivar (Kempthorne, 1957 and Falconer and Mackay, 1996). The pedigree analyses use the family relationship among cultivars to quantify the probability of having identical genes at a random locus commonly referred to as COP (Malécot 1948). So understanding this situation within the base population will help to select the more divergent parents and develop varieties with different genetic background.

Different studies were conducted to study the genetic diversity in different crops. The genetic diversity study on sugarcane based on AFLP molecular marker and COP showed high genetic correlation between AFLP genetic similarity and COP (COP=0.42, $P<0.001$) (Lima et al. 2002). COP was used to study the genetic diversity and to understand the breeding pattern in soybean (Gizlice et al. 1993; 1994;1996; Cox et al.

1985; Zhou et al., 2002; Cui et al. 2000), wheat (Cox et al. 1985) and barley (Graner et al. 1994).

This work was done with the following objectives: i) analyze the coefficient of parentage among *C. arabica* cultivars and the ancestral lines, ii) study the genetic diversity among cultivars based on COP iii) estimate the genetic contribution of each ancestral line for each cultivar, and iv) study the breeding pattern of the Brazilian coffee breeding programs.

2. MATERIALS AND METHODS

Cultivars of C. arabica Studied

For this study a total of 110 cultivars of arabica coffee and 22 ancestral lines were included. The cultivars included in this study were released from 1937 to 2009 by IAC (Instituto Agronômico de Campinas), Epamig/UFV (Empresa de Pesquisa Agropecuária de Minas Gerais/ Universidade Federal de Viçosa), Funtec (Fundo de Apoio Tecnológico à cafeicultura or Fundação Pro Café) and IAPAR (Instituto Agronômico do Paraná).

Table 1: The 110 cultivars of *C. arabica* L. released from 1937 -2009, with pedigree, and year of release

Code	Name of the variety	Parent 1	Parent 2	Year of release
1	Bourbon Vermelho IAC 662	Bourbon Vermelho		1939
2	Bourbon Amarelo IAC J10	Bourbon Vermelho	Amarelo de Botucatu	1952
3	Bourbon Amarelo IAC J19	Bourbon Vermelho	Amarelo de Botucatu	1952
4	Bourbon Amarelo IAC J2	Bourbon Vermelho	Amarelo de Botucatu	1952
5	Bourbon Amarelo IAC J20	Bourbon Vermelho	Amarelo de Botucatu	1952
6	Bourbon Amarelo IAC J22	Bourbon Vermelho	Amarelo de Botucatu	1952
7	Bourbon Amarelo IAC J24	Bourbon Vermelho	Amarelo de Botucatu	1952
8	Bourbon Amarelo IAC J9	Bourbon Vermelho	Amarelo de Botucatu	1952
9	Ibairi IAC 4061	Bourbon Vermelho	Mokka	1999
10	Mundo Novo 515-20	Sumatra	Bourbon Vermelho	1977
11	Mundo Novo 374-19	Sumatra	Bourbon Vermelho	1977
12	Mundo Novo IAC 376 - 4	Sumatra	Bourbon Vermelho	1977
13	Mundo Novo IAC 379 - 19	Sumatra	Bourbon Vermelho	1977
14	Mundo Novo IAC 382 - 14	Sumatra	Bourbon Vermelho	1977
15	Mundo Novo IAC 388 - 17	Sumatra	Bourbon Vermelho	1977
16	Mundo Novo IAC 388 - 17 - 1	Sumatra	Bourbon Vermelho	1977
17	Mundo Novo IAC 388 - 6	Sumatra	Bourbon Vermelho	1977
18	Mundo Novo IAC 464 - 12	Sumatra	Bourbon Vermelho	1977
19	Mundo Novo IAC 467 - 11	Sumatra	Bourbon Vermelho	1977
20	Mundo Novo IAC 480 - 6	Sumatra	Bourbon Vermelho	1977
21	Mundo Novo IAC 501 - 5	Sumatra	Bourbon Vermelho	1977
22	Mundo Novo IAC 502 - 1	Sumatra	Bourbon Vermelho	1977
23	Mundo Novo IAC 515 - 11	Sumatra	Bourbon Vermelho	1977
24	Icatu Vermelho IAC 2941	Tetraplóide de C.	Bourbon Vermelho	1992
25	Icatu Vermelho IAC 2942	Tetraplóide de C.	Bourbon Vermelho	1992
26	Icatu Vermelho IAC 2945	Tetraplóide de C.	Bourbon Vermelho	1992
27	Icatu Vermelho IAC 4040	Tetraplóide de C.	Bourbon Vermelho	1992
28	Icatu Vermelho IAC 4041	Tetraplóide de C.	Bourbon Vermelho	1992
29	Icatu Vermelho IAC 4043	Tetraplóide de C.	Bourbon Vermelho	1992
30	Icatu Vermelho IAC 4045	Tetraplóide de C.	Bourbon Vermelho	1992
31	Icatu Vermelho IAC 4046	Tetraplóide de C.	Bourbon Vermelho	1992
32	Icatu Vermelho IAC 4228	Tetraplóide de C.	Bourbon Vermelho	1992
33	Caturra Vermelho (IAC 477)	Bourbon Vermelho		1951

Table 1 Continued…

Code	Name of the cultivar	Parent 1	Parent 2	Year of release
34	Obatã IAC 1669-20	Villa Sarchi	Híbrido de Timor CIFC 832/2	1996
35	Tupi IAC 1669-33	Villa Sarchi	Híbrido de Timor CIFC 832/2	1996
36	Tupi RN IAC 1669-13	Villa Sarchi	Híbrido de Timor	1996
37	Oeiras MG 6851	Caturra Vermelho (CIFC 19/1)	Híbrido de Timor CIFC 832/1	1999
38	IAPAR 59	Villa Sarchi CIFC 971/10	Híbrido de Timor CIFC 832/2	1992
39	Pau Brasil MG1	Catuaí Vermelho IAC 141	Híbrido de Timor UFV 442-34	2004
40	Sacramento MG1	Catuaí Vermelho IAC 81	Híbrido de Timor UFV 438-52	2004
41	Paraiso MG H419-1	Catuaí Amarelo IAC 30	Híbrido de Timor UFV 445-46	2002
42	Araponga MG1	Catuaí Amarelo IAC 86	Híbrido de Timor UFV 446-08	2004
43	Catigua MG1	Catuaí Amarelo IAC 86	Híbrido de Timor UFV 440-10	2004
44	Catigua MG2	Catuaí Amarelo IAC 86	Híbrido de Timor UFV 440-10	2004
45	Catigua MGS3	Catuaí Amarelo IAC 86	Híbrido de Timor UFV 440-10	2007
46	Laurina IAC 870	*C. arabica*	*C. mauritiana*	1999
47	IBC- Palma 1	Catuaí Vermelho IAC 81	Catimor UFV 353	2000
48	IBC- Palma 2	Catuaí Vermelho IAC 81	Catimor UFV 353	2000
49	Sabiá	Catimor UFV 386	Acaiá	2000
50	Canário	Catuaí amarelo	Híbrido de Timor	2000
51	Siriema	*Coffea racemosa*	*C. arabica* Blue Mountain	2000
52	IPR 97	Villa Sarchi CIFC 971/10	Hibrido de Timor CIFC 832/2	2001
53	IPR 98	Villa Sarchi CIFC 971/10	Híbrido de Timor CIFC 832/2	2001
54	IPR 99	Villa Sarchi CIFC 971/10	Hibrido de Timor CIFC 832/2	2001
55	IPR 104	Villa Sarchi CIFC 971/10	Hibrido de Timor CIFC 832/2	2001
56	Saíra	Catuaí Amarelo IAC	Catindu (UFV 374, cv 643),	2004
57	Acaiá IAC 474 - 1	Mundo Novo		1977
58	Acaiá IAC 474 - 19	Mundo Novo		1977
59	Acaiá IAC 474 - 20	Mundo Novo		1977
60	Acaiá IAC 474 - 4	Mundo Novo		1977
61	Acaiá IAC 474 - 6	Mundo Novo		1977
62	Acaiá IAC 474 - 7	Mundo Novo		1977
63	Acaiá Cerrado MG 1474	Mundo Novo		1989
64	Mundo Novo Amarelo IAC 4266	Bourbon Amarelo	Mundo Novo	1951
65	Caturra amarelo IAC 476-11	Caturra Vermelho		1951
66	Catuaí Vermelho IAC 144	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
67	Catuaí Vermelho IAC 15	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
68	Catuaí Vermelho IAC 24	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
69	Catuaí Vermelho IAC 44	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
70	Catuaí Vermelho IAC 51	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
71	Catuaí Vermelho IAC 72	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
72	Catuaí Vermelho IAC 81	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972
73	Catuaí Vermelho IAC 99	Caturra Amarelo, IAC 476-11	Mundo Novo IAC 374-19	1972

Table 1: Continued…

Code	Name of the cultivar	Parent 1	Parent 2	Year of released
74	Catuaí Amarelo IAC 100	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
75	Catuaí Amarelo IAC 17	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
76	Catuaí Amarelo IAC 32	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
77	Catuaí Amarelo IAC 39	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
78	Catuaí Amarelo IAC 47	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
79	Catuaí Amarelo IAC 62	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
80	Catuaí Amarelo IAC 74	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
81	Catuai Amarelo IAC 86	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
82	Catuai Amarelo IAC 30	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
83	Catuai Amarelo IAC 70	Caturra Amarelo IAC 476-11	Mundo Novo IAC 374-19	1972
84	Rubi MG 1192	Catuaí	Mundo Novo	1985
85	Topázio MG 1190	Catuaí Amarelo	Mundo Novo,	1988
86	Ouro Verde IAC H5010-5	Catuaí Amarelo IAC 70	Mundo Novo IAC 515-20	2000
87	Ouro Bronze IAC 4925	Catuaí Amarelo IAC 70	Mundo Novo IAC 515-20	2000
88	Ouro Verde Amarelo IAC 4397	Catuaí Amarelo IAC 70	Mundo Novo IAC 515-20	2000
89	Icatu Amarelo IAC 2907	Bourbon amarelo	Icatu Vermelho	1992
90	Icatu Amarelo IAC 2944	Bourbon amarelo	Icatu Vermelho	1992
91	Icatu Amarelo IAC 3686	Bourbon amarelo	Icatu Vermelho	1992
92	Icatu Precoce: IAC 3282	Bourbon amarelo	Icatu Vermelho	1992
93	Catucaí Amarelo 2SL	Icatu	Catuaí	2000
94	Catucaí Amarelo 3SM	Icatu	Catuaí	2000
95	Catucaí Amarelo Multilínha	Icatu	Catuaí	2000
96	Catucaí Vermelho 19/8	Icatu	Catuaí	2000
97	Catucaí Vermelho 20/15	Icatu	Catuaí	2000
98	Catucaí Vermelho 24/137	Icatu	Catuaí	2000
99	Catucaí Vermelho 36/6	Icatu	Catuaí	2000
100	Catucaí Vermelho Multilínha	Icatu	Catuaí	2000
101	Obatã amarelo IAC 4739	Obatã IAC 1669-20	Catuaí Amarelo	1996
102	Tupi amarelo IAC 5167	Tupi IAC 1669-33,	Catuaí Amarelo	1996
103	Acauã	Mundo Novo IAC 388-17	Sarchimor IAC 1668	2000
104	IPR 102	Catuaí	Icatu	2001
105	IPR 103	Catuaí	Icatu	2001
106	IPR 105	Mundo Novo 374-19	Caturra amarelo IAC 476-11	2001
107	IPR 106	Catuaí	Icatu	2001
108	IPR 107	IAPAR 59	Mundo Novo IAC 376-4	2001
109	IPR 108	IAPAR 59	Catucaí	2001
110	Travessia	Catuaí Amarelo	Mundo Novo	2009

Data analysis

Analysis I- Estimation of Coefficient of parentage:

a) Estimation coefficient of parentage between cultivars of C. arabica

The parentage information of each cultivars were obtained from its respective center, journal article published and from the Ministry of Agriculture. This pedigree information was used to estimate the coefficient of parentage between cultivars. The coefficient of parentage (COP) between cultivars estimated for all possible combination among cultivars and ancestral lines included in this study. COP between two individuals is defined as the probability that a random allele at a random locus in one individual is identical by descent to a random allele at the same locus in another individual (Malécot 1948; Kempthorne 1957). COP values were computed by the formula $COP_{XY} = 1/2\left(COP_{XA} + COP_{XB}\right)$, where Y is a genotype, A and B are the parents of Y, and X is a second genotype that is not a descendent of genotype Y (Kempethone 1957; Zhou et al., 2002).

During the estimation of the COP the following assumptions were considered in case when it is necessary.

- The assumption made by Cox et al. (1985) was considered here to estimate COP. The COP value between a cultivar/ancestor and a reselection from it equaled 0.75. The value COP between two selections from the same cultivar or ancestors was $(0.75)^2 = 0.56$
- If the genotype P and Q are derived from different crosses the COP between P and Q is unaffected by inbreeding.
- If the progenitors of the genotype Z are unknown then set F_Z to 1 for self fertilizing crops and to 0 for out crossing crops.
- If the number of self generation is unknown for a breeding line set F_Z to 15/16 equivalent to four times selfing.
- If P and Q are sister lines derived from the same cross (Z) their Coefficient of parentage is affected by selfing up to their most recent common ancestor Z and $COP_{PQ} = \left(1 + F_Z\right)/2$

b) Estimation of COP among era of release

To analyze the consequence of the breeding program on the diversity of cultivars released along the years, the cultivars were grouped according to the year of release as: up to 1959, 1960-1979, 1980-1999 and 2000-2009. The mean COP among year of released estimated based on the method elaborated by Gizlice et al. (1996).

c) Estimation of COP among research institute

The cultivars grouped based on the research institute released for production. So the cultivars grouped in to four groups as cultivars from IAC, cultivars from Epamig/UFV, Funtec or ProCafé and IAPAR. The mean COP was estimated with similar manner as indicated above.

Analysis II- Multidimensional scaling and cluster analysis

The approach employed here is similar to that described by Gizlice et al. (1996). The COP is a measures of similarity (i.e 0= unrelated and 1= identity). It does not represent coordinates in Euclidean space, a prerequisite for a cluster analysis (Cui et al. 2000). Thus, the first step is in this analysis was to generate a set of Euclidean coordinates for each cultivar by multidimensional scaling (MDS) (SAS institute, 2007). The appropriate number of dimensions was determined based on the stress value which measures the correlation of the new geometrical representation with the original COP matrix (Johnson and Wichern, 1992 and Kruskal 1964). The R^2 was calculated from the comparison of input COP data with predicted values derived from the MDS coordinates. The options used in MDS analysis was (SIMILAR= 1, COEF= IDENTITY, and LEVEL = ABSOLUTE) as described by Gizlice et al. (1996).

The nonhierarchical cluster analysis among the cultivars investigated in this study was done using FASTCLUS procedure of SAS statistical package using the MDS Euclidean coordinates as source data (SAS Institute, 2007). The PROC MEAN procedure was used to calculate mean COP within and among clusters for each analysis.

Analysis III- Estimation of the genetic contribution

The relative contribution of each ancestor and the first progenies to the cultivars of *C. arabica* were estimated according to the method elaborated by Gizlice et al., 1994.

Estimation of the genetic contribution of ancestor lines

The method described by Gizlice et al. (1994) was used to estimate the genetic contribution of the ancestor lines to the cultivars of *C. arabica*. According to the method to estimate the genetic contribution of the ancestors to the cultivars, the COP among the ancestors was considered zero so that the COP can be used directly to estimate the genetic contribution of each ancestral lines to the cultivars. Burbon Vermelho, Bourbon Amarelo, Mundo Novo, Icatu Vermelho, Catuai Vermelho, Acaiá IAC 474, Caturra Amarelo IAC 476, Catuai Vermelho, Catuai Amarelo, Icatu Amarelo, Catucai Amarelo and Catucai Vermelho were considered as parent for some of cultivars when necessary.

3. RESULTS

Genetic diversity among year of release

A total of 121 cultivars of *C. arabica* were released from 1939 to 2009. Among these cultivars about 52% were released after 1980 (Table 1). The genetic diversity was increased among those cultivars released after 1980 (Table 2). The COP was increased among cultivars released before 1959 (0.839) to cultivars released in 1960-1970 (0.902) (Table 2). After this period the mean COP was reduced substantially for the last four decades with minimum value of 0.463 which indicated the introduction of new parental lines in the breeding programs like Híbrido de Timor, Icatu and others.

Table 2: The mean coefficient of parentage of the Brazilian *C. arabica* L cultivars within and between periods of released.

Year of release	Up to 1959	1960-1979	1980-1999	2000-2009
Ate 1959	0.839			
1960-1979	0.825	0.902		
1980-1999	0.536	0.55	0.507	
2000-2009	0.560	0.601	0.450	0.463
Number	17	42	27	38

Genetic diversity among research institutes

Among 121 *C. arabica* cultivars released in Brazil 66 % of them were released by the Instituto de Agronômico Campinas (IAC) (Table 3). The mean COP among research centers indicated high mean COP among cultivars released by IAC. The lowest mean COP was recorded by *C. arabica* cultivars released by IAPAR which is 0.435 (Table 3).

Table 3: The within and between mean coefficient of parentage of the Brazilian *C. arabica* cultivars released by different research institutes.

Research institute	IAC	EPAMIG/UFV	Funtec	IAPAR
IAC	0.719			
EPAMIG/UFV	0.519	0.468		
Funtec	0.573	0.419	0.514	
IAPAR	0.513	0.436	0.437	0.435
Number	80	13	17	11

IAC= Instituto Agronômico de Campinas

EPAMIG= Empresa de Pesquisa Agropecuária de Minas Gerais

UFV= Universidade Federal de Viçosa

IAPAR= Instituto Agronômico do Paraná

Funtec= Fundo de Apoio Tecnológico à cafeicultura ou ProCáfe

Multivariate and cluster analysis

The analysis of PROC MDS of SAS with 20 dimension produced an excellent Euclidian representation of the COP matrix with R^2=0.99 and stress = 0.01. The 20 dimensional coordinates from the MDS analysis were used to produce the best cluster groups using the FASTCLUS analysis. The cluster groups ranging from four to 12 was produced to select the best cluster group (SAS, 2002). The acceptable cluster group was selected based on the criteria set by (Gizlice et al. 1996; Cui et al. 2000 and Zhou et al., 2002). Based on the cluster analysis using MDS coordinates, four acceptable clusters were obtained out seven clusters. The three clusters were considered unacceptable because it has one member in each cluster. The first and the second cluster have the highest mean COP, 0.822 and 0.703, respectively. These clusters also shared the same genetic base (Table 4). For the Cluster I the two ancestors Bourbon Vermelho and Sumatra were contributed more than 90% of the genetic base (Table 5). This result is supported by the highest mean value of COP between cluster groups.

Table 4: The mean coefficient of parentage between clusters formed based on COP among 121 varieties of *C. arabica* cultivars.

Cluster	I	II	III	IV	V
I	0.822				
II	0.565	0.703			
III	0.376	0.257	0.437		
IV	0.442	0.314	0.375	0.436	
V	0.182	0.123	0.094	0.187	0.00
No	74	27	7	10	3

Table 5: The most important ancestors and their relative genetic contribution (GC) to 4 nonhierarchical clusters of Brazilian *C.arabica* cultivars

Cluster	Ancestor name	% GC
I	Bourbon Vermelho	63.18
	Sumatra	31.30
	Amarelo de Botucatu	6.45
II	Bourbon Vermelho	71.05
	Tetraploid *C. canephora*	21.45
	Sumatra	7.49
III	Híbrido de Timor	42.85
	Villa Sarchi	57.14
IV	Bourbon Vermelho	47.88
	Híbrido de Timor	52.12

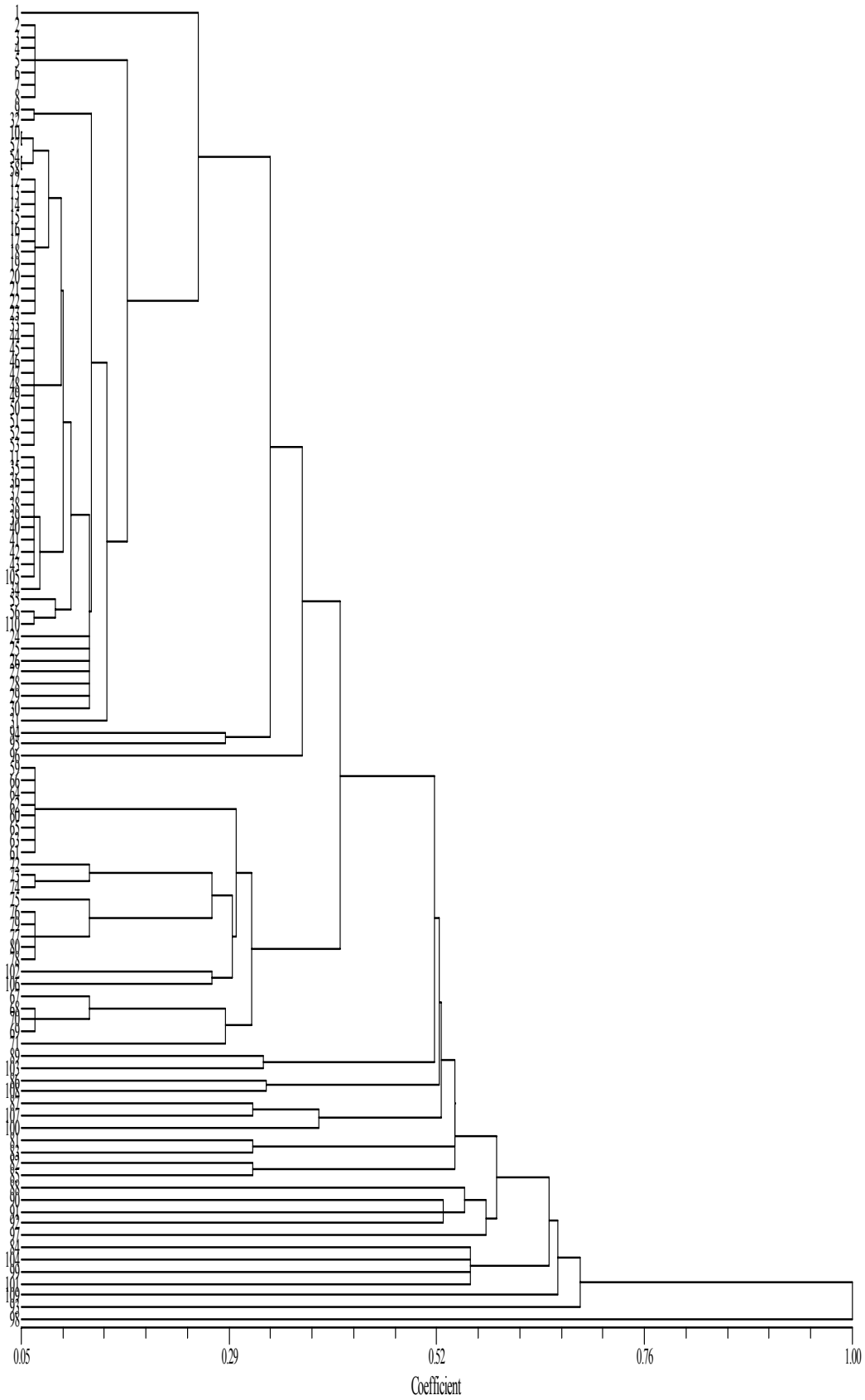

Figure 1: The dendrograma produced for 121 cultivars of *C. arabica* based on (1-coeffecient of parentage) matrix using UPGMA clustering method.

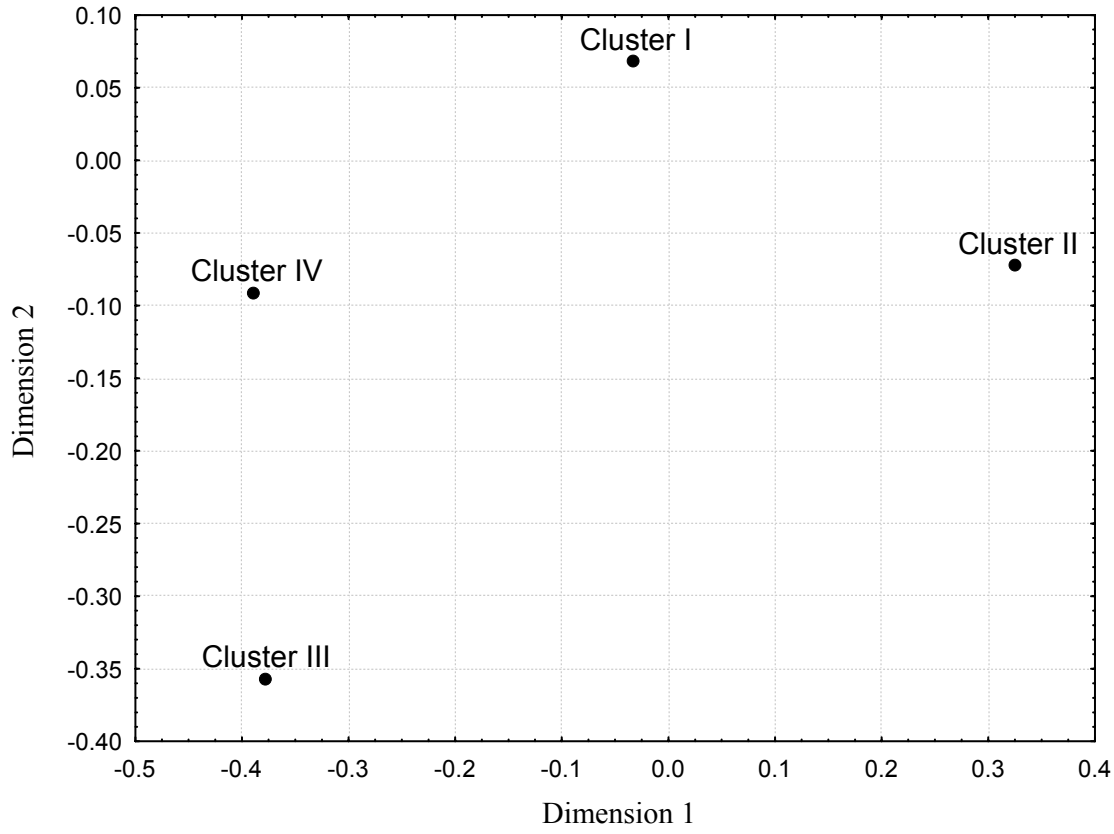

Figure 2: plot of four cluster representing 110 cultivars of *C. arabica* released in Brazil from 1939-2009. The score was obtained by averaging multidimensional scaling. The complement of the linear distance (1-distance) between any two cultivars estimates the coefficient of parentage between them. Distance ≥ 1 indicate no relationship.

Genetic contribution of Ancestor lines for the Brazilian C. arabica cultivars

The genetic base of 121 cultivars released in Brazil between 1939 and 2009 was defined by 10 ancestors. The contribution of the ancestors for the genetic base of *C. arabica* in Brazil ranged from 0.413 % to 47.59 % (Table 6). The seven ancestors contribute 98.76 % of the gene to the *C. arabica* cultivars in Brazil. This indicates the low genetic diversity among them. The genetic contribution of the principal ancestors to cultivars of *C. arabica* showed Bourbon Vermelho contributed 47.59 % of the gene to the *C. arabica* cultivars followed by Sumatra (20.25 %) and Icatu (13.23%).

Table 6: Relative genetic contribution of the ancestral lines for the 121 cultivars of *C. arabica* released in Brazil from 1939 until 2009 and their first progeny.

Code	Name of cultivar	Genetic contribution		
			%	Accumulated
1	Bourbon Vermelho	57.56	47.596	47.596
2	Sumatra	24.50	20.259	67.855
3	Icatu	16.00	13.230	81.085
4	Híbrido de Timor	9.00	7.442	88.527
5	Villa Sarchi	6.25	5.168	93.695
6	Amarelo de Botucatu	5.62	4.651	98.346
7	Blue Mountain	0.50	0.413	98.760
8	Moka	0.50	0.413	99.173
9	*C. racemoça*	0.50	0.413	99.587
10	*C. mauritina*	0.50	0.413	100.000
		120.935	100.000	

4. DISCUSSION

Genetic diversity study

Genetic diversity among year of release

The genetic diversity among Brazilian *C. arabica* cultivars increased in recent decades (Table 2). The increase in genetic diversity among cultivars was accompanied with the introduction of new parental lines in the breeding programs as a parent for the cultivars developed after 1980. The high mean COP among cultivars released before 1959 (0.839) and 1960-1979 (0.902) (Table 2) is due to small number of parental lines involved in the development of cultivars. The low mean COP (0.463, Table 2) for the cultivars released after 2000 showed the involvement of different parental lines in the development of new cultivars which diversify the genetic base of the *C. arabica* cultivars released during this period. This fact consequently increased the genetic diversity among *C. arabica* cultivars. Even if a lot of work is done to diversify the genetic base of *C. arabica* cultivars in Brazil still the mean COP value among released cultivars is very high when compared with other crops studied such as soybean (Cox et al 1985, Gizlice et al 1993, Cui et al 2002, Zhou et al 2002), barley (Graner et al 1994) and bread wheat (Cox et al 1985)).

Genetic diversity between research institutes

Among the research institutes involved in coffee research, IAC is the oldest and the prominent research institute well known for its research work especially on *C. arabica*. IAC released 66 % of the cultivars of *C. arabica* in Brazil. The highest mean COP was recorded from the cultivars released by IAC (0.719, Table 3), which indicated the low genetic diversity among cultivars. The basic reason for the high mean COP is most of the cultivars of *C. arabica* released by IAC are sister lines (cultivars of Mundo Novo, Catuai, and Bourbon) which has the same genetic composition. Other research institutes (EPAMIG/UVF, Funtec, and IAPAR) were released cultivars with different genetic background resulting in low mean COP (Table 3). Among the research institutes IAPAR recorded the lowest mean COP (0.435, Table 3) showing the high genetic diversity among cultivars released by the center. The low mean COP value by IAPAR, EPAMIG/UFV and Funtec in relation to IAC showed the possibility of increasing the genetic diversity among cultivars of *C. arabica*.

The highest value of mean COP among research institutes (Table 3) indicated the existence of high germplasm exchange among them. This result also demonstrated the possibility of increasing the genetic diversity among cultivars released without losing the productivity and the quality. The genetic diversity study among 115 accessions from Ethiopia, Eretria, Yemen and Brazilian *C. arabica* accessions showed the existence of high genetic variability among accessions found in IAC (Silvestrini et al 2007). This research result showed the potential of the germplasm collection existed in Brazil for development of new cultivars and diversify the genetic base of the new *C. arabica* cultivars.

Multivariate and cluster analysis

The cluster analysis performed based on the 20 dimensional scales produced by Proc MDS perfectly classified the 121 cultivars into four cluster groups. The cultivars released by IAC grouped in Cluster I and Cluster II. These two cluster showed low genetic diversity which explained by high mean COP within clusters (0.822 and 0.703, respectively, Table 4). The low genetic diversity observed in Cluster I and Cluster II can be explained by the analysis of the genetic contribution of ancestors for the cluster (Table 5). The analysis of the genetic contribution of the parental lines (ancestors) showed that Bourbon Vermelho contributed 63.18% and 71.05% for the gene to cultivars in cluster I and cluster II, respectively (Table 5). High mean COP between these clusters indicated high genetic similarity between cultivars. The low mean COP in Cluster III and Cluster IV showed the existence of higher genetic diversity within clusters. The high genetic diversity between cultivars in cluster III and cluster IV was resulted from the incorporation of new parental lines in the breeding programs of *C. arabica*.

In addition the use of COP to study the genetic diversity among cultivars, it is a valuable tool for breeder to understand the current situation of their cultivars and planning a better breeding program to increase the genetic diversity in the future. Understanding the real breeding pattern of cultivars also helps to incorporate new parental lines to diversify the genetic base of the future cultivars to be released.

Genetic contribution of ancestors for the Brazilian C. arabica cultivars

The 94 % of the genetic base of *C. arabica* cultivars was constituted by seven ancestors indicated the narrow genetic base of the cultivars (Table 6). Bourbon Vermelho which contributed 52.76 % of the gene to cultivars because it involved in most of the crossing programs due to its productivity and good cup quality (Carvalho 1957, Carvalho and Fazuoli 1993). This indicated the low genetic diversity among *C. arabica* cultivars released in Brazil. To reverse this situation and increase the genetic base of the cultivars in coffee introducing new ancestor lines will be very crucial. On this line the introduction of Híbrido de Timor, the interspecific hybrid between *C. arabica* and *C. canephora* var Robusta played an important role to diversify the genetic base in coffee arabica breeding in Brazil.

Híbrido de Timor involved in crossing programs to develop new cultivars of *C. arabica* after confirmed it has resistance genes for the coffee leaf rust (*Hemileia vastatrix*) and other diseases. For this reason most of the *C. arabica* cultivars released in recent years contain Híbrido de Timor in their genetic background. So, Híbrido de Timor had great contribution in diversifying the genetic base of the *C. arabica* cultivars in Brazil. In this study the research institutes also showed it is possible to developed cultivars with different genetic background which has good productivity and quality.

5. CONCLUSIONS

Low genetic diversity was observed among cultivars of *C. arabica* released in Brazil until 2009. The reason for low genetic diversity is most of the cultivars derived from the same parental lines.

The low genetic diversity was among cultivars was observed for those released before 1980 and after this period the genetic diversity among cultivars was increased substantially due to the introduction of new parental lines in the breeding program. Low genetic diversity also observed among cultivars of *C. arabica* released by Instituto Agronômico de Campinas (IAC) since most of the cultivars released by the center are sister lines originated from the same crosses.

The genetic base of Brazilian *C. arabica* cultivars defined by 10 ancestors and among them seven ancestors contributed 98.76 % of the gene, which showed the genetic base of cultivars of C. arabica is very narrow. Among ancestral lines Bourbon Vermelho, Sumatra and Icatu Vermelho contributed 47.59 %, 20.259 and 13.230 %, respectively of the gene to cultivars of *C. arabica* released in Brazil.

The genetic diversity of cultivars released in recent years increased due to the introduction of new ancestral lines in the crossing program of *C. arabica* breeding in the country, specially the involvement of Híbrido de Timor in the crossing program as source of gene for resistance for diseases and pests.

6. REFERENCES

Bettencourt AJ, Carvalho (1968) Melhoramento visando a resistência do cafeeiro à ferrugem. Bragantia 27(4): 35-68

Bettencourt AJ (1973) Considerações sobre o Hibrido de Timor. Campinas, Instituto Agronoico. 20p. (Circular 23).

Carvalho A (1957) Melhoramento do cafeeiro: considerações gerais sobre os métodos de melhoramento empregados. In: INSTITUTO AGRONÔMICO. I Curso de Cafeicultura. 3.ed. São Paulo, p.59-68.

Carvalho A (1988) Principles and Practice of Coffee Plant Breeding for Productivity and Quality factors: *Coffea arabica.* Coffee: Agronomy. Ed. R.J. Clarke. New York: Elsevier Applied Science

Carvalho A, Fazuoli LC (1993) Café. In: O Melhoramento de Plantas no Instituto Agronômico, v.1, FURLANI, A. M. C.; VIÉGAS, G.P. Ed. Campinas: Instituto Agronômico, p.29-76.

Carvalho A, Fazuoli LC, Da Costa WM (1989) Melhoramento do cafeeiro: XLI. Produtividade do Híbrido de Timor, de seus derivados e de outras fontes de resistência a *Hemileia vastatrix*. Bragantia 48 (1): 73-86

Clarindo WS, Carvalho CR (2008) First Coffea arabica karyogram showing that this species is a true allotetraploid. Plant Systematic and Evolution 274: 237-241

Cox TS, Lookhar GL, Walker DE, Harrell LG, Albers LD, Rodgers DM (1985) Genetic relationship among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polycrimide gel electrophoretic patterns. Crop Sci 25: 1058-1063

Cox TS, Kiang YT, Gorman MB, Rodgers DM (1985) Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci 25: 529-532

Cui Z, Thomas E, Carter Jr, Burton JW (2000) Genetic diversity patterns in Chinese soybean based on coefficient of parentage. Crop Sci 40: 1780-1793

Delannay X, Rodgers DM, Palmer RG (1983) Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Sci 23: 944-949

Eccardi F, Sandal JV (2003) *O café*: ambientes e diversidade. Rio de Janeiro:Casa da Palavra

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, Ed 4. Longmans Green, Harlow, Essex, UK

Gizlice Z, Carter Jr. TE, Burton JW (1993) Genetic diversity in North American soybean: I-Multivariate Analysis of founding stock and relation to coefficient of parentage. Crop Sci 33: 614-620

Gizlice Z, Carter Jr. TE, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci 34: 1143-1151

Gizlice Z, Carter Jr. TE, Gerig TM, Burton JW (1996) Genetic diversity patterns in North Americans public soybean cultivars based on coefficient of parentage. Crop 36: 753-765

Graner A, Ludwing WF, Melchinger AE (1994) Relationships among European barley germplasm: comparison RFLP and pedigree data. Crop Sci 34: 1199-1205.

Johnson RA, Wichern DW (1992) Applied Multivariate Statistical Analysis. Prentice-Hall, New Jersey. 3rd ed.

Kempthorne O (1957) An Introduction to Genetic Statistics. Iowa State University Press. Ames, IA.

Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29: 115-129

Malécot G (1948) Les mathematiques de lhérédité. Masson, Paris. English translation. The mathematics of heredity. 1969. W.H. Freeman and Co., San Francisco, CA.

Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A. (1999) Molecular characterization and origin of the *Coffea arabica* L. genome. Mol. Gen. Genet. **261**: 259–266.

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, De Souza CLJr, De Souza AP, (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor Appl Genet 104: 30-38

SAS Institute. (2007) SAS user's guide. SAS Inc.

Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O, Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. Genet Resour Crop Evol 54:1367-1379

Walsh J (1981) Germplasm Resources Are Losing Ground. *Science* 214: 421-423

Zhou X, Carter Jr.TE, Cui Z, Miyazaki S, Burton JW (2002) Genetic diversity patterns in Japanese soybean cultivars based on coefficient of parentage. Crop Sci 42: 1331-1342

CHAPTER 2

GENOME INTROGRESSION OF HÍBRIDO DE TIMOR AND ITS RELATIONSHIP WITH OTHER COFFEE SPECIES.

ABSTRACT

Seventy seven coffee accessions comprised of five *C. arabica*, fifteen *C. canephora* var. Robusta, ten *C. canephora* var Conilon, one *C. eugenoids* and forty six Híbrido de Timor were characterized using AFLP marker. The data were used for genome introgression analysis of Híbrido de Timor and to study its genetic relationship with other coffee species. To understand the genetic relationship with other coffee species, multidimensional scaling analysis (MDS) and principal coordinate analysis (PCoA) based on 1-Jaccard coefficient, and the model based Bayesian clustering analysis were used. The PCoA and MDS showed the clear differentiation among coffee species and high genetic similarity of Híbrido de Timor with *C. arabica*. The analysis of AMOVA partitioned the total variation within population (39.05%) and among populations (60.95%), which confirmed high genetic differentiation among coffee species. The pairwise F_{ST} analysis proved the existence of high genetic similarity between *C. arabica* and Híbrido de Timor. The results from model based Bayesian clustering analysis using Structure program confirmed high genetic similarity of Híbrido de Timor with *C. arabica*. They grouped in the same cluster with shared ancestral probability greater than 0.92. The CIFC 4106 considered the original Híbrido de Timor plant grouped together with *C. arabica* with shared ancestral probability 0.92, which proved the existence of high genetic similarity between *C. arabica* and Híbrido de Timor. The analysis of genome introgression of *C. arabica* and *C. canephora* var Robusta into CIFC 4106 showed lower contribution of the *C. canephora* genome (18.9%). The genome introgression analysis, the distance based genetic diversity study and the model based Bayesian clustering analysis confirmed the high genetic similarity between *C. arabica* and CIFC 4106 which supported the hypothesis that Híbrido de Timor is not an F_1 plant instead at least two times backcrossed with *C. arabica*.

Key words: *C. arabica*, Molecular markers, AFLP, genetic diversity, multivariate analysis, Bayesian model

INTROGRESSÃO DO GENOMA DO HÍBRIDO DE TIMOR E RELAÇÃO COM OUTRAS ESPÉCIES DO CAFÉ.

RESUMO

Setenta e sete acessos de café, sendo cinco *C. arabica*, quinze *C. canephora* var. Robusta, dez *C. canephora* var Conilon e quarenta e seis Híbrido de Timor foram caracterizados usando marcador molecular AFLP. Os dados foram usados para análise de introgressão do genoma do Híbrido de Timor e estuda relação com outras espécies do café. Para entender a relação genética do Híbrido de Timor com outras espécies de café, a análise de coordenadas principais, agrupamento baseado na dissimilaridade genética de Jaccard e agrupamento baseado no modelo Bayesiano foram usados. A análise de coordenadas principais mostrou diferenciação entre as espécies de café e alta similaridade genética entre *C. arabica* e Híbrido de Timor. Análise de AMOVA mostrou a variação total dividida entre populações (60.95%) e dentro de populações (30.05), que confirma a alta diferenciação genética entre as espécies de café. O F_{ST} entre populações comprovou a existência de similaridade genética entre Híbrido de Timor e *C. arabica*. O resultado da análise de agrupamento baseado no modelo de Bayesiano usando o programa Structure confirmou a alta similaridade genética entre Híbrido de Timor e *C. arábica.* E grupados no mesmo grupo com probabilidade do maior que 0.92. O CIFC 4106 considerado planta original do Híbrido de Timor, foi grupado com *C. arabica* com probabilidade de 0.90, o que demonstra alta similaridade genética entre Híbrido de Timor e *C. arabica*. A análise de ingrogressão do genoma de C. canephora var Robusta e *C. arabica* com Híbrido de Timor provou a existência de baixa introgressão do genoma de *C. canephora* no Híbrido de Timor (18.9%). A análise introgressão do genoma, o estudo da diversidade genética baseado distancia genética e análise de agrupamento baseada no modelo Bayesiano confirmou a alta similaridade genética entre Híbrido de Timor e *C. arabica*, o que suportando a hipótese de que o Híbrido de Timor não é planta F_1 e sim resultante de pelo menos de dois retrocruzamento com *C. arabica*.

Palavras chave: *C. arabica*, marcador molecular, AFLP, diversidade genética, analise multivariada, modelo de Bayesiano.

1. INTRODUCTION

C. arabica L. (2n=2x=44) is a true allotetraploid species (Clarindo and Carvalho 2008) native to Africa. *Coffea arabica* L. and *Coffea canephora* Pierre are the two most cultivated and commercialized coffee species in the world. Among them *C. arabica* L. has more than 70% contribution in world coffee market. It is originated in the south western Ethiopia and produce high cup quality. Even if the world coffee production and consumption depend on *C. arabica*, its production was greatly affected by diseases and pests which reduce its productivity, due to lack of resistance genes for the major diseases and pests. Due to this the breeding programs of *C. arabica* has been used Híbrido de Timor as a source of gene for resistance to diseases and pests.

Híbrido de Timor is the interspecific hybrid between *C. arabica* and *C. canephora*. Híbrido de Timor was first found in plantation of cultivar Typica in Timor Island in 1917 (Bettencourt 1973) and used as source of resistance gene for economically important diseases and pests of coffee such as coffee leaf rust (*Hemileia vastatrix*), coffee berry disease (CBD) caused by *Colletotrichum Kahawae*, root knot nematode (M*eloidgyne exigua*) and bacteriosis caused by *Pseudomonas syringae* pv garçae (Bertrand et al. 2003). Using Híbrido de Timor as source of resistance gene, cultivars were released for production in Kenya, Brazil, Colombia and Costa Rica (Lasheremes et al. 2000; Charries and Eskes 1997; Bertrand et al. 2003; Pereira et al. 2005).

The small portion of the genome of *C. canephora* was introgressed into Híbrido de Timor which gave resistance to coffee leaf rust and other disease. Researcher also confirmed this fact using different molecular marker. Genome introgression of *C. arabica* and *C. canephora* into accession of Híbrido de Timor was studied using AFLP molecular marker technique and found that introgression of *C. canephora* genome ranged from 8% (in Catimor 3) to 25% (in Sachimor) (Lashermes et al. 2000). Bertrand et al. (2003) evaluated the effect of genome introgression of *C. canephora* on cup quality of lines derived from Híbrido de Timor and showed the possibility of finding lines resistance to disease (coffee leaf rust) and nematodes combined with good quality as *C. arabica* cultivars.

The relationship of different coffee species can be assessed using molecular markers such as Amplified Fragment Length Polymorphism (AFLP), Simple Sequence

Repeat (SSR) and Randomized Amplified Polymorphic DNA (RAPD). AFLP was more preferred in the study of genetic diversity and genetic relationship among population because it has a capacity of screening of many different DNA regions distributed randomly throughout the genome (Mueller and Wolfenbarger 1999). The advantage of AFLP markers in relation to SSR is it does not requires prior sequence information and relatively low star-up cost. The relationship of different coffee species was done by different authors suing molecular markers. High genetic relationship of Híbrido de Timor with *C. arabica* was reported by Lashermes et al. (1993, 1999 and 2003).

Since Brazilian germplasm bank has several accession of Híbrido de Timor available and they are important sources of disease resistance gene and used in large extent in the breeding program of coffee in Brazil and the world, understanding the genome introgression from their origin (*C. arabica* and *C. canephora*) and their relation to others species is important for the breeding program of coffee. So this work was done with the objectives: 1) to characterize the accession of Híbrido de Timor using AFLP marker and its relationship with other coffee species (such as *C. arabica* and *C. canephora*); and 2) to know the contribution of *C. arabica* and *C. canephora* in genome of Híbrido de Timor accessions.

2. MATERIALS AND METHODS

Genetic materials

Seventy five coffee accessions which include 5 *C. arabica*, 15 *C. canephora* var Robusta, 10 *C. canephora* var Conilon and 46 Híbrido de Timor accessions (Table 1) were used for the study of genome introgression of Híbrido de Timor and genetic relationship between Híbrido de Timor and other species of coffee.

Table 1: List of Coffee accessions used for genetic relationship study and genome introgression of Híbrido de Timor (HT)

Code	Genotype Name	Description	Code	Genotype Name	Description
A1	Catuaí VerUFV2144	*C. arabica*	HT27	UFV 427-01	HT
A2	Catuaí IAC44	*C. arabica*	HT28	UFV 427-09	HT
A3	Típica UFV 2945	*C. arabica*	HT29	UFV 427-15	HT
A4	Bourbon UFV 2952	*C. arabica*	HT30	UFV 427-22	HT
A5	Bourbon UFV535-1	*C. arabica*	HT31	UFV 427-55	HT
HT6	CIFC 832/1	HT	HT32	UFV 427-56	HT
HT7	CIFC 832/2	HT	HT33	UFV 427-65	HT
HT8	CIFC 4106	HT	HT34	UFV 427-90	HT
HT9	CIFC 1343/269	HT	HT35	UFV 438-52	HT
HT10	UFV376-01	HT	HT36	UFV 439-02	HT
HT11	UFV 376-04	HT	HT37	UFV 440-22	HT
HT12	UFV 376-05	HT	HT38	UFV 442-108	HT
HT13	UFV 376-35	HT	HT39	UFV 443-03	HT
HT14	UFV 376-37	HT	HT40	UFV 446-08	HT
HT15	UFV 376-52	HT	HT41	UFV 445-46	HT
HT16	UFV 376-57	HT	HT42	UFV 428-04	HT
HT17	UFV 376-79	HT	HT43	UFV 432-07	HT
HT18	UFV 377-01	HT	HT44	UFV 437-06	HT
HT19	UFV 377-02	HT	HT45	UFV 441-03	HT
HT20	UFV 377-23	HT	HT46	UFV 447-48	HT
HT21	UFV 377-24	HT	HT47	UFV 448-69	HT
HT22	UFV 377-34	HT	HT48	UFV 450-61	HT
HT23	UFV 379-07	HT	HT49	UFV 451-41	HT
HT24	UFV 408-18	HT	HT50	UFV 440-10	HT
HT25	UFV 408-26	HT	Co51	Encapa 03	Conillon
HT26	UFV 408-28	HT	Co52	Encapa 04	Conillon

Table 1, Cont…

Code	Genotype Name	Description
Co53	Encapa 05	Conillon
Co54	Encapa 06	Conillon
Co55	Encapa 07	Conillon
Co56	Encapa 08	Conillon
Co57	Encapa 09	Conillon
Co58	Conillon 66-1	Conillon
Co59	Conillon 66-2	Conillon
Co60	Conillon 66-3	Conillon
Ro61	Robusta 3751	Robusta
Ro62	Robusta 3580	Robusta
Ro63	Guarini 513	Robusta
Ro64	Guarini 514	Robusta
Ro65	Robusta C2258	Robusta
Ro66	Robusta2257-2	Robusta
Ro67	Robusta2257-1	Robusta
Ro68	Robusta 640-1	Robusta
Ro69	Robusta 640-2	Robusta
Ro70	Robusta 640-2	Robusta
Ro71	Apoatã-1	Robusta
Ro72	Apoatã-2	Robusta
Ro73	Apoatã -3	Robusta
Ro74	Guarini-1	Robusta
Ro75	Guarini-2	Robusta

Extraction of DNA

The DNA of the genotypes was extracted according to the method described by Diniz et al. (2005) from young green leaves collected from each genotype. The DNA concentration was quantified using Spectrophotometre Smart Spec of BioRad. The extracted DNA was diluted in TE (Tri-HCL 10mM, EDTA 1mM, pH 8.0) to concentration of 50 ηg/µl for AFLP analysis.

AFLP (Amplified Fragment Length Polymorphism) analysis

The AFLP genotyping of coffee accessions were done according to the method described by Brito et al. (2010). The primer combinations MSEI-AGC/ECORI-CGT and *MseI*-AGC/ECORI-CTC were used to genotype the coffee accessions in this study.

Data analysis

The gels of AFLP were scored by visual inspection for presence (1) or absence (0) of of specific AFLP-bands. Only distinct major bands were scored. To analysis the

AFLP data using Structure population genetic analysis software (Pritchard et al.2000) the data matrix was coded according to Falush et al. (2007). The AFLPdata statistical package (Ehrich 2007) was used to manage the data conversion from txt format to Structure format.

To study the genetic relationship between Híbrido de Timor and other coffee species, the distance based and model based clustering analysis was performed. For the distance based clustering analysis, first the Jaccard similarity coefficient (Jaccard 1908) was estimated using NTSYS-pc software (Version 2.10L; Rohlf, 2000). The clustering analysis and the dendrogram were generated from the similarity matrix by the UPGMA method. Reliability of clusters in each dendrogram was tested by bootstrap analysis (Felsenstein, 1985) with 1000 replications using Treecon.

The principal coordinate analysis (PCoA) was done among accessions based on genetic dissimilarity matrix (1-Jaccard similarity coefficient) using GenAlex 6.2 population genetic analysis software (Peakall and Smouse 2006). Nei genetic diversity index (Nei 1973), Shannon´s Information and percent polymorphic bands (P %) with in populations were estimated using POPGENE statistical software version 1.3 (Yeh and Boyle 1997). The pairwise F_{ST} analysis to understand the relationship between Híbrido de Timor with other coffee species was determined by AFLPsurv (Vekemans et al. 2002).

The model based Bayesian clustering analysis was done using Structure 2.2 population genetic analysis software (Pritchard et al. 2000) to group the accessions of coffee species into its respective groups applying admixture model. The number of populations (k) was varied from two to twelve with twenty replicate runs per each assumed k value, using a burning period length of 5000 runs and a postburning sampling by Markov Chain Monte Carlo of 50,000 runs to estimate the number of subpopulations for each of the k values. The appropriate number of cluster was determined according to Evanno et al. (2005).

To understand the genome introgression of *C. canephora* and *C. arabica* into Híbrido de Timor CIFC 4106, the one considered the original plant obtained in the Timor Island the percent of shared band with *C. arabica* and *C. canephora* var Robusta was estimated. In addition the percent of shared band was estimated with other accessions of Híbrido de Timor.

3. RESULTS

Genetic relationship between Híbrido de Timor and other coffee species

The individual population diversity measure was estimated for all the populations except *C. eugenoides* (because it was represented by one accession). High genetic diversity was observed within Híbrido de Timor and *C. canephora* var Robusta (Table 2). The percent polymorphic loci (P%) was high for Híbrido de Timor (P=67.59%) followed by *C. canephora* var Robusta. The principal coordinate analysis (Figure 1) and the dendrograma obtained based on genetic dissimilarity matrix using UPGMA clustering method (Figure 2) showed the clear differentiation among coffee species and high similarity between Híbrido de Timor and *C. arabica* accessions.

Table 2: The Nei gene diversity measure, Shannon´s information index and percent polymorphic loci (P%) produced using 108 AFLP bands for coffee genotypes.

Coffee species	Number of accessions	Nei gene diversity	Shannon´s Information	P(%)	No Band
C. arabica	5	0.0156	0.0236	4.63	5
Híbrido de Timor	46	0.1716	0.27	67.5	73
C. canephora var Conillon	10	0.1313	0.2034	46.3	50
C. canephora var Robusta	15	0.1853	0.2864	66.6	72
Over all mean	75	0.2497	0.3853	100	108
Standard diversion		0.1851	0.2429		

The AMOVA (Analysis of Molecular Variance) among coffee species showed high genetic differenciatiion among coffee species (Table 3). The total variation observed was partitioned in 60.95 % among population and 39.05 within populations (Table 3).The overall Fst value (0.609) demonstrated the existence of high genetic differenciation among coffee species.

The pairwise F_{ST} analysis between Híbrido de Timor and other coffee species showed high genetic similarity between *C. arabica* and Híbrido de Timor incontrast they showed high dissmilarity genetic with *C. canephora* and *C. eugenoids* (Table 4).

Table 3: AMOVA of genetic variation using AFLP markers

Source of Variation	*df*	Sum of squares	Variance component	Percent of total component variance
Among populations	4	498.62	10.4919	60.95
Within populations	72	484.10	6.7237	39.05
Total	76	982.72	17.2156	100
Fixation Index Fst = 0.609				

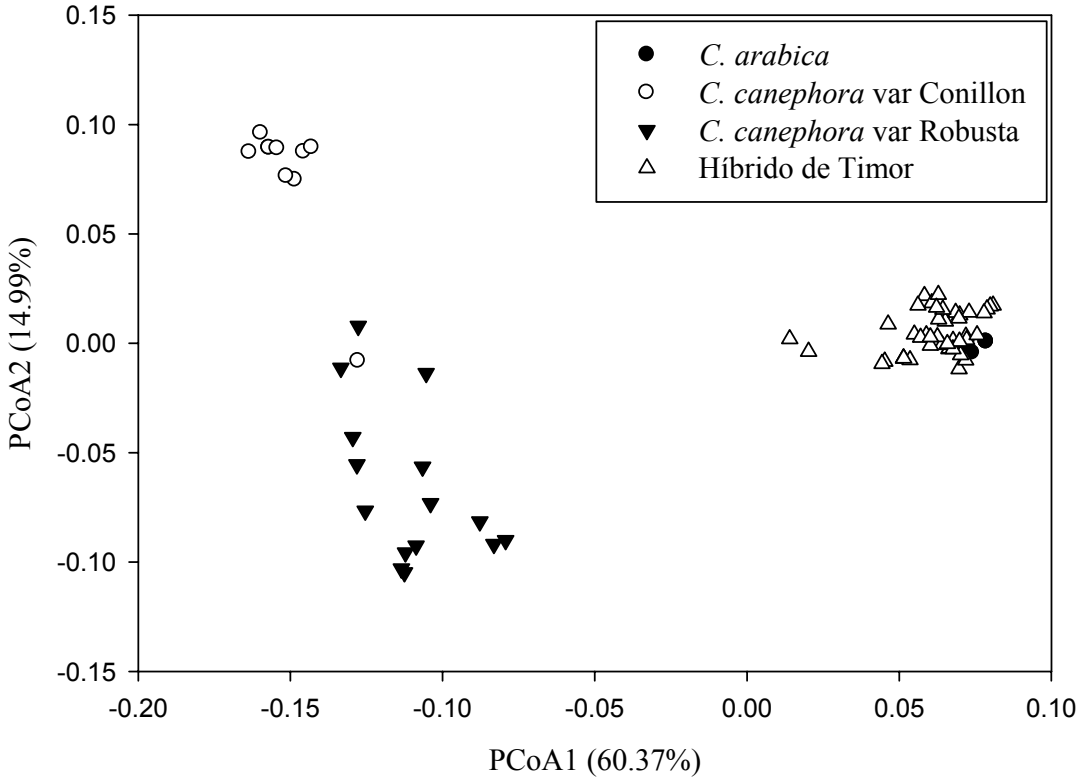

Figure 1: Principal coordinate analysis of AFLP diversity among coffee species.

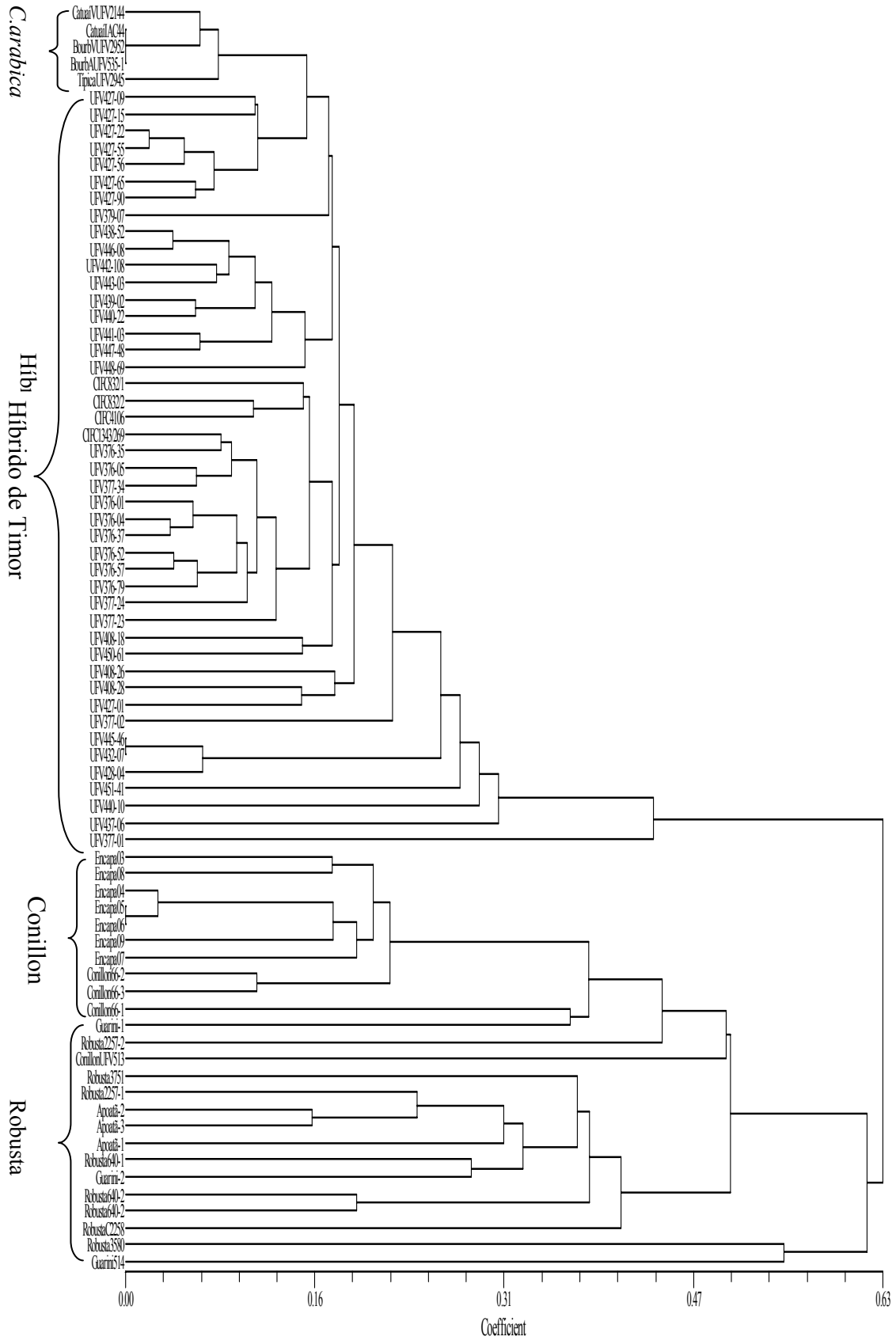

Figure 2: The dengrogram produced using clustering method UPGMA based on DS (1-Jaccard similarity coefficient). The dendrogram was produced by NTSYS genetic analysis software.

Table 4: Pairwise F_{ST} (allele frequency used was estimated by square root method) between coffee species (Phylip format) investigated in this study.

	C. arabica	*C. canephora* var Robusta	Hibrido de Timor	*C. canephora* var Conillon
C. arabica	0	0.6346	0.184	0.8038
C. canephora var Robusta	0.6346	0	0.4312	0.343
Hibrido de Timor	0.184	0.4312	0	0.6148
C. canephora var Conillon	0.8038	0.343	0.6148	0

F_{ST} over all populations = 0.608

The model based population structure analysis based on Bayesian statistics (Pritchard et al. 2000) produced the shared ancestral probability for each population and individuals to the respective cluster groups inferred. This helps to assign genotypes into proper group without any difficult and unambiguously. The individual was assigned to respective cluster when it has shared ancestral probability greater than 0.81. This analysis grouped the 75 accessions into five clusters (Table 5 and 6). Most of the accessions grouped maintained its original population.

The shared ancestral probability of each population and individual to the new cluster formed was presented on Table 5 and 6. The *C. arabica* and Híbrido de Timor were grouped in Cluster III with shared ancestral probability greater than 0.92. Only 5 accessions of Híbrido de Timor showed some type of admixture with Cluster II, III and Cluster IV. The *C. canephora* accessions distributed into Cluster I and Cluster V. Cluster V contains 9 of the 10 accessions of *C. canephora* var Conilon with ancestral shared probability >0.95 (Table 6). The high admixture proportion was observed with the accessions of *C. canephora* var Robusta which showed admixture with Cluster I, II, IV and Cluster V (Table 6). From the Robusta accessions 53% were grouped in Cluster I with shared ancestral probability more than 0.81 and the high admixture was observed from Robusta accessions obtained from the Epamig/UFV breeding program. The highest F_{ST} value was observed by Cluster V (F_{ST} =0.8) followed by Cluster III (F_{ST} = 0.77). Cluster V include all the accessions (clones) *C. canephora* var Conilon and Cluster II include all accessions of *C. arabica* and 90% of the accessions of Híbrido de Timor, which showed low genetic diversity among them.

Table 5: Proportion of membership of each pre-defined population in each of the 5 clusters inferred.

Given Population	Inferred Clusters					Number of Individuals
	I	II	III	IV	V	
C. arabica	0.002	0.002	0.991	0.004	0.001	5
Híbrido de Timor	0.005	0.02	0.923	0.049	0.003	46
C. canephora var conilon	0.061	0.006	0.002	0.002	0.929	10
C. canephora var robusta	0.716	0.165	0.003	0.013	0.103	15

Table 6: Inferred ancestral probability of individual accession for each cluster

code	Name of accession	Inferred cluster				
		I	II	III	IV	V
A1	Catuaí verUFV2144	0.002	0.006	0.989	0.002	0.001
A2	Catuaí IAC44	0.002	0.003	0.993	0.001	0.001
A3	Típica UFV 2945	0.002	0.005	0.988	0.004	0.001
A4	Bourbon UFV 2952	0.002	0.002	0.994	0.001	0.001
A5	Bourbon UFV535-1	0.002	0.002	0.994	0.001	0.001
HT6	CIFC 832/1	0.004	0.003	0.988	0.004	0.002
HT7	CIFC 832/2	0.005	0.004	0.949	0.038	0.004
HT8	CIFC 4106	0.043	0.004	0.924	0.024	0.005
HT9	CIFC 1343/269	0.002	0.002	0.992	0.002	0.001
HT10	UFV376-01	0.002	0.003	0.991	0.002	0.002
HT11	UFV 376-04	0.003	0.002	0.99	0.002	0.003
HT12	UFV 376-05	0.005	0.003	0.97	0.008	0.014
HT13	UFV 376-35	0.003	0.003	0.991	0.002	0.001
HT14	UFV 376-37	0.002	0.002	0.992	0.002	0.002
HT15	UFV 376-52	0.002	0.003	0.992	0.002	0.001
HT16	UFV 376-57	0.003	0.003	0.99	0.002	0.002
HT17	UFV 376-79	0.003	0.003	0.991	0.001	0.002
HT18	UFV 377-01	0.006	0.587	0.39	0.013	0.003
HT19	UFV 377-02	0.008	0.028	0.956	0.004	0.005
HT20	UFV 377-23	0.001	0.004	0.992	0.002	0.001
HT21	UFV 377-24	0.002	0.002	0.992	0.002	0.001
HT22	UFV 377-34	0.002	0.003	0.992	0.002	0.001
HT23	UFV 379-07	0.004	0.003	0.97	0.008	0.014
HT24	UFV 408-18	0.008	0.019	0.965	0.003	0.005
HT25	UFV 408-26	0.004	0.024	0.964	0.002	0.006
HT26	UFV 408-28	0.014	0.002	0.972	0.007	0.005
HT27	UFV 427-01	0.013	0.012	0.964	0.009	0.002
HT28	UFV 427-09	0.002	0.007	0.974	0.016	0.001
HT29	UFV 427-15	0.002	0.011	0.983	0.003	0.001
HT30	UFV 427-22	0.001	0.002	0.995	0.001	0.001
HT31	UFV 427-55	0.001	0.002	0.995	0.001	0.001

Table 6 contd…

Code	Name of accession	Inferred cluster				
		I	II	III	IV	V
HT32	UFV 427-56	0.001	0.002	0.994	0.001	0.001
HT33	UFV 427-65	0.003	0.002	0.992	0.001	0.002
HT34	UFV 427-90	0.002	0.003	0.992	0.001	0.002
HT35	UFV 438-52	0.004	0.003	0.99	0.002	0.002
HT36	UFV 439-02	0.003	0.003	0.984	0.003	0.008
HT37	UFV 440-22	0.004	0.003	0.98	0.002	0.011
HT38	UFV 442-108	0.004	0.003	0.972	0.016	0.004
HT39	UFV 443-03	0.003	0.002	0.99	0.002	0.003
HT40	UFV 446-08	0.002	0.002	0.992	0.002	0.002
HT41	UFV 445-46	0.008	0.013	0.971	0.006	0.003
HT42	UFV 428-04	0.005	0.031	0.955	0.007	0.002
HT43	UFV 432-07	0.01	0.013	0.969	0.005	0.003
HT44	UFV 437-06	0.006	0.004	0.603	0.378	0.009
HT45	UFV 441-03	0.004	0.003	0.988	0.002	0.002
HT46	UFV 447-48	0.003	0.004	0.989	0.002	0.002
HT47	UFV 448-69	0.002	0.003	0.908	0.086	0.002
HT48	UFV 450-61	0.007	0.009	0.979	0.002	0.002
HT49	UFV 451-41	0.004	0.004	0.762	0.226	0.003
HT50	UFV 440-10	0.002	0.404	0.588	0.003	0.002
Co51	Encapa 03	0.004	0.002	0.004	0.012	0.979
Co52	Encapa 04	0.002	0.002	0.001	0.001	0.995
Co53	Encapa 05	0.002	0.002	0.001	0.001	0.994
Co54	Encapa 06	0.002	0.002	0.001	0.001	0.994
Co55	Encapa 07	0.034	0.002	0.005	0.018	0.941
Co56	Encapa 08	0.002	0.003	0.003	0.003	0.989
Co57	Encapa 09	0.002	0.002	0.001	0.001	0.994
Co58	Conillon 66-1	0.557	0.003	0.002	0.008	0.431
Co59	Conillon 66-2	0.002	0.001	0.001	0.001	0.995
Co60	Conillon 66-3	0.006	0.004	0.001	0.01	0.979
Ro61	Robusta 3751	0.97	0.008	0.003	0.017	0.003
Ro62	Robusta 3580	0.075	0.003	0.001	0.919	0.002
Ro63	Guarini 513	0.229	0.011	0.002	0.47	0.287
Ro64	Guarini 514	0.07	0.021	0.001	0.902	0.006
Ro65	Robusta C2258	0.902	0.004	0.02	0.072	0.002
Ro66	Robusta2257-2	0.437	0.13	0.002	0.037	0.394
Ro67	Robusta2257-1	0.99	0.002	0.001	0.004	0.002
Ro68	Robusta 640-1	0.869	0.002	0.001	0.005	0.124
Ro69	Robusta 640-2	0.775	0.002	0.001	0.003	0.218
Ro70	Robusta 640-2	0.982	0.003	0.003	0.003	0.009
Ro71	Apoatã-1	0.99	0.004	0.001	0.002	0.003
Ro72	Apoatã-2	0.992	0.002	0.001	0.002	0.003
Ro73	Apoatã -3	0.988	0.002	0.004	0.002	0.004
Ro74	Guarini-1	0.539	0.002	0.002	0.035	0.422
Ro75	Guarini-2	0.931	0.002	0.001	0.003	0.063

Genome introgression analysis of Hibrido de Timor

The result showed that Híbrido de Timor shared shared much more AFLP bands (26.15%) with *C. canephora* var Robusta than var Conilon (Table 7). The scanned image of gel which included Híbrido de Timor, Robusta and Conilon (Figure 4) produced by AFLP molecular marker showed the band unique to var Robusta always shared by Híbrido de Timor.

Table 7: The number of AFLP alleles (presence of the band in the AFLP loci) shared by Híbrido de Timor CIFC 4106 with Robusta and Conilon.

	Number of AFLP alleles	%
Alleles shared by HT, Robusta and Conillon	44	69.412
Alleles shared by HT and Robusta	17	26.15
Alleles shared by HT and Conillon.	4	6.15
Total number of allele of *C. canephora*	65	100.00

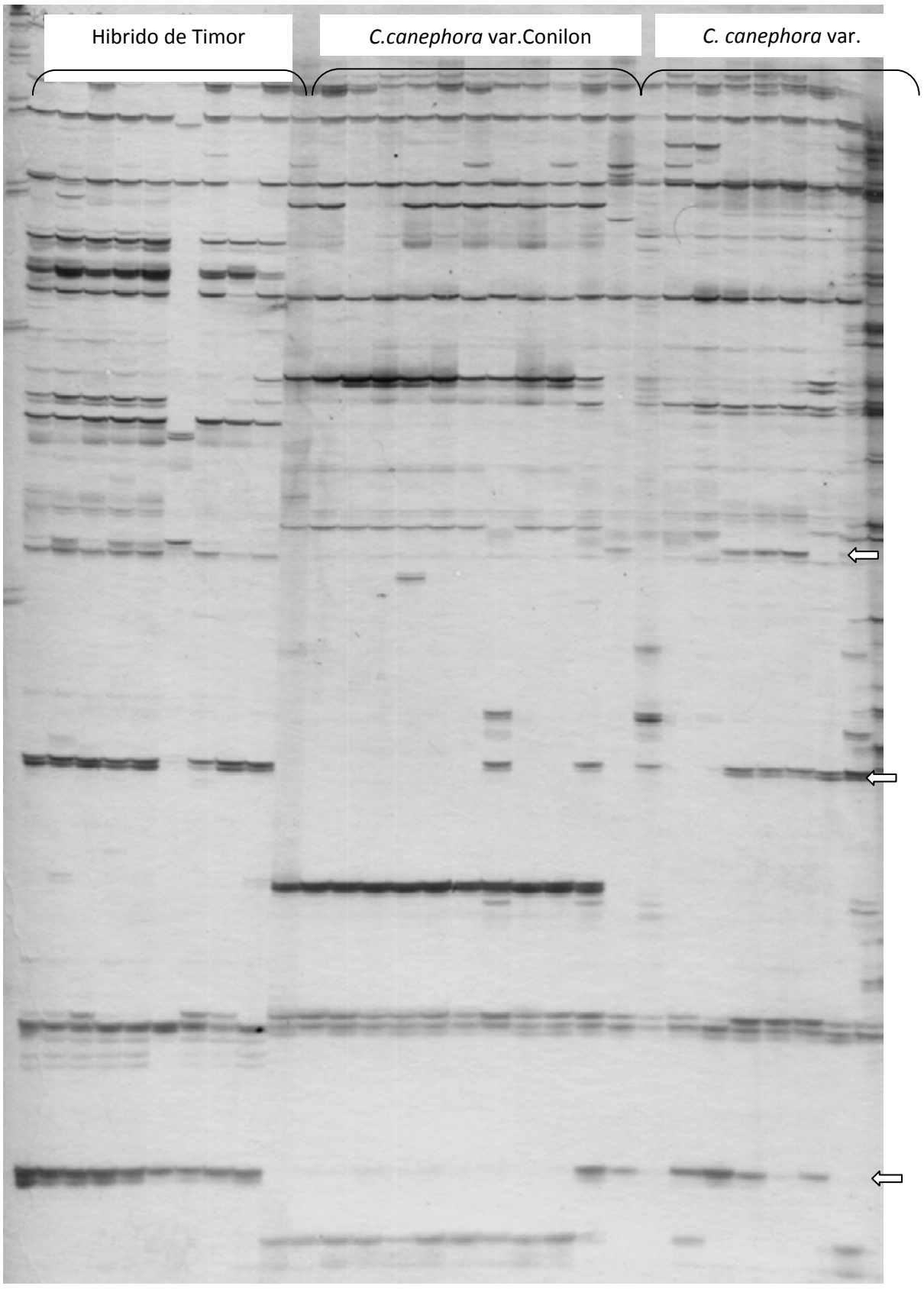

Figure 3: The AFLP gel which include accessions of Híbrido de Timor, *C canephora* var Conilon and *C. canephora* var Robusta

The results obtained with AFLP suggested that *C. canephora* var Robusta considered as one of the parent during the formation of the Híbrido de Timor Therefore, the 15 Robusta accessions were used for genome introgression analysis. The analysis of genome introgression showed that Híbrido Timor CIFC 4106 shared only 18.9% of the AFLP band with *C. canephora* var Robusta (Table 8 and 9). The result also showed that CIFC 832/1, CIFC 832/2 and CIFC 1343-269, introduced by seed from Timor Island to CIFC showed low level genome introgression in relation to the CIFC 4106 (Table 8). The mean genome introgression value of *C. canephora* var Robusta observed when all Híbrido de Timor accessions considered together is 10.00% , is not significantly different from the value expected from any backcrossed plant (Table 9).

Table 8: Genome introgression analysis of Híbrido de Timor CIFC 4106: the percentage of band shared by *C. arabica* and *C. canephora* with Híbrido de Timor.

	Number of AFLP alleles	%
Present in all *C. arabica* accessions	41	77.3
Polymorphism also observed in *C. arabica*	2	3.8
Not detected in any *C. arabica* accession	10	18.9
Total	53	100

Table 9: The number of bands shared by the *C. arabica*, *C. canephora* var Robusta with lines of Híbrido de Timor.

	Mean of All H.Timores		CIFIC 832/1		CIFIC 832/2		CIFIC 1343/269		CIFC 4106	
	No bands	%	No bands	%	No bands	%	No bands	%	No bands	%
Present in all *C. arabica* accessions	1878	84.20	42	89.4	42	76.4	41	83.7	41	77.3
Polymorphism also observed in *C. arabica*	128	5.74	2	4.2	3	5.5	4	8.2	2	3.8
Not detected in any *C. arabica* accession	223	10.00	3	6.4	10	18.1	5	10.1	10	18.9
Total	72	100	47		55	100	50	100	53	100
χ^2 BC_2 (*df*=1, 0.05)=3.84	0.285*		1.60*		1.62*		0.285*		1.96*	

*not significant at P=0.05

4. DISCUSSION

Genetic relationship of Híbrido de Timor with other coffee species

The individual population diversity measures estimated showed that *C. canephora* has the highest genetic diversity followed by Híbrido de Timor. The same result was observed by Lashermes et al. (2000), Orozco-Castillo (1994), Lashermes et al. (1993) and Silvestrini et al. (2008). The availability of high genetic diversity within *C. canephora* and Híbrido de Timor has a great advantage for the breeding program. The low genetic diversity among *C. arabica* accessions was also reported by Orozco-Castillo 1994, Lashermes et al. (1993, 1996 and 2000) and Silvestrini et al. (2008). Lashermes et al. (1995) pointed out that the low genetic diversity of *C. arabica* was related to the genesis of this allotetraploid species, reproductive biology and evolution. In Brazil the low genetic diversity among cultivars of *C. arabica* was related to the origin of the coffee cultivars, originated from the same plant (Carvalho 1945). Even if the low genetic diversity was observed within cultivars of *C. arabica* some studies showed the existence of potential resource for the breeding program of arabica coffee (Silverstrini et al. 2007). Silvestrini et al. (2007) studied *C. arabica* accessions from different origin and concluded the existence high genetic variability among accessions from Ethiopia which can be used in the breeding program.

The other important result obtained in this study was the existence of high genetic diversity among Híbrido de Timor accession. Since these genotypes are important coffee group used extensively in the breeding programs as a source of resistance gene for coffee leaf rust (*Hemilia vastatrix*), coffee berry disease (*Colletotrichum kahawae)*, root knot nematode (M*eloidgyne exigua*) and bacteriosis caused by *Pseudomonas syringae* pv garçae (Bettencourt 1973, Charries and Eskes 1997, Bertrand et al. 2003, Pereira et al. 2005 and Sera et al. 2005).

The low genetic differentiation between *C. arabica* and Híbrido de Timor indicated the high gene flow between these two groups of coffee (Lashermes et al. 1993, 2000). The high genetic similarity between *C. arabica* and Híbrido de Timor demonstrated by different methods of statistical analysis (Table 2-6 and Figure 1-3) resulted from the selection of Híbrido de Timor accessions having similar architecture and agronomic characters with *C. arabica* in the breeding programs. In addition high genetic similarity of Híbrido de Timor with arabica coffee backs to its origin where it is has more backcross with *C. arabica*. The same result also reported by Lashermes et al. (1993). The directional selection and backcrossing with *C. arabica* reduced the

introgressed alien genetic material from *C. canephora* but maintained the genes for resistance to disease and pests which are important for the breeding programs of arabica coffee.

The overall G_{ST} (that measure of genetic differentiation among population) value estimated in this study is 0.608 using different methods of statistical analysis gave similar results as reported by Silvestrini et al. (2007). The AMOVA analysis using AFLP molecular marker revealed most of the variation was partitioned between coffee species than within species. Similar results also reported using RAPD molecular marker in Coffee species (Silvestrini et al. 2008). This showed high genetic differentiation between coffee species. The pairwise F_{ST} estimated between coffee species demonstrated that Híbrido de Timor is more related to *C. arabica*, which is in accordance with result reported by Lashermes et al. (1993, 1995 and 2000). The different multivariate statistical analysis (Multidimensional Scaling, Principal coordinate analysis), clustering analysis using UPGMA method and population genetic analysis statistics clearly showed similar results and proved the existence of well defined population structure among coffee species. In addition it showed high genetic similarity between *C. arabica* and Híbrido de Timor.

The model based Bayesian clustering analysis using Structure program (Pritchard et al. 2000) was frequently used to study the population structure and genetic diversity in different crops (Holsinger et al 2004, Kwak et al. 2009, López-Gartner et al. 2009). This analysis grouped the seventy seven accessions into five clusters. The *C. arabica* and Híbrido de Timor accessions grouped in Cluster III with shared ancestral probability >0.92 (Table 5) which proved the existence of high genetic similarity between them. Similar result was presented by Lashermes et al. 1993, 2000. This result also supported by analysis multidimensional and the principal coordinate analysis (Figure 1, 2 and 3) and pairwise Fst analysis (Table 4).

The CIFC 4106 considered the first plant obtained in Timor Island (Pereira et al. 2005) grouped together with C. *arabica* with shared ancestral probability 0.949, which showed high similarity among them. CIFC 4106 showed high flowering, low fruiting capacity, produced fruit type of Moca and showed high self incompatibility under the Viçosa soil and climatic condition (Pereira et al. 2008). In addition, it didn't produce fruit when used as female parent. These characteristics suggest that CIFC 4106 is an interspecific hybrid as reported by Pereira et al. 2008. The grouping of CIFC 4106 with the same cluster with *C. arabica* with shared ancestral probability 0.94 proved this plant is not an F_1 plant instead one or more time backcrossed with *C. arabica*.

Cluster V includes the clones of Conillon with shared ancestral probability >0.95 which indicated high uniformity among this clones. The highest admixture probability was observed within accessions of *C. canephora* var Robusta which indicated the existence of high genetic diversity within this species (Lashermes et al. 1993, 1996, 2000, and Silvestrini et al. 2008,). The highest cluster F_{ST} was observed by Cluster III and Cluster V which accessions grouped with more than 0.90 shared ancestral probability. The study showed the model based clustering analysis based on Bayesian statistical analysis proved an efficient method in assigning genotypes in its respective group as reported by López-Gartner et al. (2009) in accessions of Arabica coffee and demonstrated the genetic relationship of Híbrido de Timor with other coffee species.

Genome introgression analysis of CIFC 4106

The result obtained in this study supported the hypothesis that the *C. canephora* var Robusta was the one crossed with *C. arabica* and resulted Híbrido de Timor. The historical evidences also showed that at the time when Híbrido de Timor found in Timor Islands of the only *C. canephora* grown in the Island was Robusta (http://www.sca-indo.org/history-of-indonesia/). From the historical evidence and the evidence found in this study the the Híbrido de Timor found in the Timor Island resulted from the interspecific cross between *C. canephora* var Robusta and *C. arabica*.

The genome introgression analysis CIFC 4106 with *C. arabica* and *C. canephora* var Robusta showed that high band sharing between *C. arabica* and Híbrido de Timor. High genetic similarity between *C. arabica* and Híbrido de Timor similar results was also reported by Lashermes et al. (1993 and 2000). The mean introgression percentage of shared band between *C. canephora* var Robusta with Híbrido de Timor is 10.00 % when all accessions of Híbrido de Timor considered together and is not significantly different from the value expected in BC2(backcross two) plant. Lashermes et al. (2000) using also AFLP maker suggested that Híbrido de Timor is F_1 plant, with was not agreed with our finding. Even if in this study we included large number of Híbrido de Timor accession in relation to their study (only included two accessions of Híbrido de Timor). In addition here we compared the CIFC 4106 the original Híbrido de Timor plant obtained in Island of Timor (Pereira et al. 2005).

The accession of Híbrido de Timor CIFC 4106 only showed 18.9 % of alien genetic material introgressed from *C. canephora* which is lower than expected from any F_1 plant. This value was not differ from the value expected in BC_1plant and not differs statistically from the value expected in BC_2 plant. So this confirmed the original plant of

Híbrido de Timor is not an F_1 plant instead one or more time backcrossed with *C. arabica*. The low genetic material introgression from *C. canephora* to CIFC 832/1, CIFC 832/2 and CIFC 1343/269 introduced by seed from the Timor Island by CIFC indicated the existence of directional selection of the plants type *C. arabica*.

The introgression of *C. canephora* genome into individual Híbrido de Timor accessions ranged from zero to 18.9. The high genetic similarity observed between *C. arabica* and Híbrido de Timor also reported by Lashermes et al. (1993 and 2000). The authors suggested that the Híbrido de Timor resulted from interspecific hybridization followed by several backcrossing with *C. arabica* (Lashermes et al. 1993).

Even if Híbrido de Timor showed high similarity with *C. arabica* it has considerable genetic diversity than the *C. arabica* as reported by Lashermes et al. 2000. This result showed the importance of Híbrido de Timor in the future Arabica coffee breeding in Brazil and the world. The results obtained widen the information on Híbrido de Timor and their predecessors which has great importance in the breeding program of coffee. The low percentage alien gene introgression reported here demonstrated that the Híbrido de Timor accessions maintained its Arabica character which is good to produce cultivars with good cup quality and disease resistance. Bertrand et al. 2003 reported the possibility of developing new cultivars with good cup quality and disease resistance from *C. canephora* introgressed lines. The high genetic variability of Híbrido de Timor with its resistance to disease and pests demonstrated the future potential of Híbrido de Timor for developing cultivars do *C. arabica* (Pereira et al 2005, Bertrand et al. 2002). In addition Híbrido de Timor will continue as a vehicle for transferring the gene resistance for coffee leaf rust and other diseases to *C. arabica*.

This research was done using four CIFC materials (CIFC 4106, CIFC 832/1, CIFC 832/2 and CIFC 1343/269) and 42 segregating accessions of Híbrido de Timor which showed the well representation of Hibrido de Timor derivatives in relation to past works. So the experiment was more informative in the genome introgression analysis of Híbrido de Timor in relation to *C. arabica* and *C. canephora* and its relationship with other coffee species. The above evidences clearly supported the hypothesis that CIFC 4106 is not an F_1 plant instead at least two times backcrossed with *C. arabica*.

5. CONCLUSIONS

High genetic diversity was observed among coffee species studied. Among them C. canephora and Híbrido de Timor showed high genetic diversity within population. Low genetic diversity was observed among accessions of *C. arabica.*

High genetic similarity also observed between *C. arabica* and Híbrido de Timor which indicated these two coffee groups shared more genetic material among them. The genome introgression analysis of *C. canephora* and *C. arabica* into Híbrido de Timor showed low percentage of genome introgression of *C. canephora* which indicated Híbrido de Timor mentain its *C. arabica* nature that is important in relation to C. arabica breeding.

The *C. canephora* genome introgression among accessions of Híbrido de Timor ranged from 0 to 18.9% with mean of 10.0%, which indicated Híbrido de Timor is not an F1 plant instead at least two times backcrossed with *C. arabica*. CIFC 4106 considered the original Híbrido de Timor plant found in Timor Island also showed similar genome introgression of *C. canephora*.

6. REFERENCES

Anthony F; Combes M. C; Astorga C.; Bertrand B; Graziosi G; Lashermes P. (2002) The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. Theor Appl Genet 104:894–900

Anthony F, Quiros O, Topart P, Bertrand B and Lashermes P (2002) Detection by simple sequence repeat markers of introgression from *Coffea canephora* in *Coffea arabica* cultivars. Plant Breeding 121: 542—544

Bertrand B, Guyot B, Anthony F, Lashermes P (2003) Impact of the Coffea canephora gene introgression on beverage quality of *C. Arabica.* Theor Appl Genet.107 : 387-394.

Bettencourt A (1973) Considerações gerais sobre o 'Hibrido de Timor'. Instituto agronomico de Campinas, Circular No. 23 (20p), Campinas, Brasil.

Bussell JD (1999) The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of Isotoma petraea (Lobeliaceae). Molecular Ecology, 8, 775–789.

Carvalho A (1945) Distribuição geográfica e classificação botânica do gênero Coffea com referência especial à espécie Arabica. Bol. Superint. Serv. Café. 21:174-180.

Charrier A, Eskes A (1997) Coffee. *In* : Tropical Plant Breeding. Eds. Charrier A, Jacquot S.,Hamon, D. Nicolas Coll. Repères, Cirad, France, pp. 128-152.

CLARINDO W R ; CARVALHO CR (2008) First Coffea arabica karyogram showing that this species is a true allotetraploid. Plant Systematics and Evolution, v. 274, p. 237-241.

Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P(2000) Characterization of microsatellites loci in Coffea arabica and related coffee species. Mol Ecol 9:1171–1193

Cubry P; Musoli P, Legnaté H, Pot D, Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2007) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome 51: 50-63

Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB, Zambolim EM, Loureiro ME, Pereira AA, Zambolim L (2005) Analysis of AFLP marker associated

to the *Mex-1* resistance locus in Icatu progenies. Crop Breeding and Applied Biotechnology 5**:** 387-393.

Ehrich, D (2006) AFLPdat: a collection of R functions for convenient handling of AFLPdata. Mol. Ecol. Notes, 6, 603-604.

Evanno, G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620

Excoffier L, Laval G and Schneider S (2006) Arlequin ver. 3.1: An integrated software package for population genetics data analysis. Available at http://cmpg.unibe.ch/software/arlequin3/arlequin31.pdf (verified 19 Apr. 2009).

Falush D, Stephens M and Pritchard JK (2007) Inference of population structure using multilocus genotype data: Dominant markers and null alleles. Mol. Ecol. Notes 7:574–578

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bulletin de la Societé Vanddoise des Sciences Natureles, v. 44, p. 223-270.

Kwak M, Kami JA, Gepts P (2009) The putative Mesoamerican domestication center of Phaseolus vulgaris is located in the Lerma-Santiago basin of Mexico. Crop Sci 49:554-563

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

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

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

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

López-Gartner G, Cortina H, McCouch SR, Moncada MDP (2009) Analysis of genetic structure in a sample of coffee (Coffea arabica L.) using fluorescent SSR markers. Tree Genetics & Genomes 5:435–446

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. TREE vol. 14, no. 10 : 389-394

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, USA 70, 3321–3.

Nei M (1978) Estimation of average hetrozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590

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

Peakall, R and Smouse P (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research *Molecular Ecology Notes*, Vol. 6, No. 1: 288-295.

Periera AA, Moura WM, Zambolim L, Sakiyama NS, Geraldo MC (2002) Melhoramento Genético do Cafeeiro no Estado do Minas Gerais – Cultivares Lançados e em Fase de obtenção. In: O Estado da Arte de Tecnologias na Produção de Café. Ed .ZAMBOLIM L. , Viçosa, UFV. pp. 253-296

Periera AA, Sakiyama NS, Zambolim L, Moura WM, Zambolim EM, Caixeta ET (2005) Identification and use of sources of durable resistance to coffee leaf rust in the UFV/EPAMIG breeeding program. In: Durable resistance to coffee leaf rust. Eds. ZAMBOLIM L., ZAMBOLIM E. M., VÁRZEA V. M. P., Viçosa, UFV. pp. 215-232.

Pereira AA, Oliveira ACB, Sakiyama NS (2008) Híbrido de Timor como fonte de resistência a doenças e de qualidade da bebida do cafeeiro. In: Núcleo de Estudos em Fitopatologia. (Org.). Manejo fitossanitário da cultura do cafeeiro. Brasília: Sociedade Brasileira de Fitopatologia, v. , p. 13-24.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pritchard JK, Wen W (2003) *Documentation for STRUCTURE software: Version 2*. Available from http://pritch.bsd.uchicago.edu.

Rohlf FJ (2000) NTSYS-pc Numerical taxonomy and multivariate analysis system Version 2.1. Setauket, New York: Exeter Software.

Rovelli P, Mettulio R, Anthony F, Anzueto F, Lashermes P, Graziosi G (2000) Microsatellites in Coffea arabica L. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) Coffee biotechnology and quality. Proceedings of the 3rd International Seminar on Biotechnology in the Coffee. Agroindustry, Londrina, Brazil. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 123–133

Sera T, Sera GH, Ito DS, Doi DS (2005) Coffee breeding for durable resistance to leaf rust disease at Instituto Agronômico do Paraná. In: Durable resistance to coffee leaf rust. Eds. ZAMBOLIM L., ZAMBOLIM E. M., VÁRZEA V. M. P., Viçosa, UFV. pp. 187-214

Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O,Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian Coffea arabica L. accessions using microsatellites markers. Genet Resour Crop Evol 54:1367–1379

Silvestrini M, Maluf M, Silvarolla MB, Guerreiro-Filho O, Medina-Filho HP. Vanini MMT. Oliveira AS.Gaspari-Pezzopane C, Fazuol LC (2008) Genetic diversity of a Coffea Germplasm Collection assessed by RAPD markers. Genet Resour Crop Evol 55:901–910

StatSoft, Inc. (2001). STATISTICA (data analysis software system), version 6. www.statsoft.com.

Steiger DL, · Nagai C, Moore PH, Morden CW, Osgood RV, Ming · R (2002) AFLP analysis of genetic diversity within and among Coffea arabica cultivars. Theor Appl Genet (2002) 105:209–215

Vekemans X, Beauwens T, Lemaire M, Roldan–Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Mol. Ecol. 11, 139–151.

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belgian J. Bot. 129:157.

CHAPTER 3

GENETIC DIVERSITY AND BREEDING POTENTIAL OF HÍBRIDO DE TIMOR.

ABSTRACT

AFLP, RAPD and SSR molecular marker were used for genetic diversity study among accessions of Híbrido de Timor. The principal coordinate analysis (PCoA), and clustering analysis based on dissimilarity genetic (1-Jaccard similarity coefficient) was done to understand the genetic diversity pattern of Híbrido de Timor. Percentage of polymorphic loci, Shannon's information index, Nei gene diversity were employed to assess the existed genetic variability within populations of Híbrido de Timor. The model based Bayesian clustering analysis also used to study the population structure and assign accession to its respective group. The percentage polymorphic loci, Shannon's information index and Nei gene diversity index showed high genetic diversity among accessions of Híbrido de Timor. The result from PCoA analysis demonstrated high genetic diversity and clear grouping pattern among accessions using RAPD and SSR molecular markers. The first two principal coordinates explained 68.62 % and 67.80% of the total variation produced by RAPD and SSR molecular markers, respectively. The lowest genetic diversity was detected using SSR molecular marker. The model based Bayesian clustering analysis done using Structure program formed four clusters for AFLP, RAPD and SSR molecular markers. The CIFC 4106 considered the original plant showed high admixture with four clusters formed by RAPD variation, which indicate it gave the origin of other accessions of Híbrido de Timor. Our study confirmed the existence of considerable genetic variability among accessions of Híbrido de Timor using AFLP, RAPD and SSR molecular markers. The model based clustering analysis using Structure program proved more efficient for the study of the population structure and assigning genotypes without any ambiguity based on shared ancestral probability.

Key word: AFLP, RAPD, SSR, molecular markers, Bayesian model, principapal coordinate analysis, population structure

DIVERSIDADE GENÉTICA E POTENCIAL DO HÍBRIDO DE TIMOR NO MELHORAMENTO DE CAFÉ.

RESUMO

Marcadores moleculares AFLP, RAPD e SSR foram usados para estudo da diversidade genética entre acessos do Híbrido de Timor. Análise de coordenadas principais (PCoA), agrupamento baseado na dissimilaridade genética (1-coeficiente de similaridade de Jaccard) e agrupamento baseado no modelo de Bayesiano foram usados para entender o padrão da diversidade genética do Híbrido de Timor. A percentagem de locos polimórficos, o índice de Shannon´s e índice de diversidade gênica de Nei foram empregados para acessar variabilidade genética entre acessos de Híbrido de Timor. O agrupamento baseado no modelo Bayesiano também foi usado para estudo da estrutura populacional e agrupar acessos nos grupos apropriados. A percentagem de locos polimórficos, índice de Shannon´s e índice de diversidade gênica de Nei mostraram alta diversidade genética entre acessos de Híbrido de Timor. Resultados da análise de PCoA demonstrou alta diversidade genética entre acessos de Híbrido de Timor e o padrão do agrupamento bem definido foi gerado usando marcadores moleculares RAPD e SSR. As primeiras coordenadas principais geradas usando conjunto dos dados RAPD, AFLP e SSR explicaram 66.83 % da variação total. Baixa diversidade genética entre acessos de Híbrido de Timor foi detectada pelo marcador SSR. A análise de agrupamento baseada no modelo do Bayesiano formou três grupos usando conjunto dos dados. O CIFC 4106 considerado com planta original do Híbrido de Timor compartilhou genoma com três grupos formados pelo marcador RAPD, o que indica que essa planta pode ser deu origem do outros acessos do Híbrido de Timor. Nosso estudo mostrou existência de alta diversidade genética entre acessos do Híbrido de Timor usando marcadores moleculares RAPD, AFLP e SSR. Agrupamento baseado no modelo de Bayesiano foi eficiente para estudo da estrutura populacionais do Híbrido de Timor e coloca genótipos no grupo apropriado baseado probabilidade compartilhada dos ancestrais.

Palavras chave: AFLP, RAPD, SSR, marcadores moleculares, modelo Bayesiano, análise principais coordenadas, estrutura populacionais.

1. INTRODUCTION

Híbrido de Timor, the interespefic hybrid between *C. arabica* and *C. canephora* was first found in plantation of cultivar Tipica in Timor Island in 1917 (Bettencourt 1973). The derivatives of this coffee tree have been used as source of resistance gene for economically important diseases and pests of coffee such as coffee leaf rust (*Hemileia vastatrix*), coffee berry disease (CBD) caused by *Colletotrichum Kahawae*, root knot nematode (M*eloidgyne exigua*) and bacteriosis caused by *Pseudomonas syringae* pv garçae (Bettencourt 1973, Bertrand et al. 2003). Varieties derived from Híbrido de Timor were released for production in Kenya, Brazil, Colombia and Costa Rica (Charries and Eskes 1997, Lasheremes et al. 2000; Bertrand et al. 2003; Pereira et al. 2005).

The first introduction of Híbrido de Timor accessions to Brazil date back from 1976 via vegetative propagation and seeds from CIFC (Centero da Investigação de Ferrugem do Café, Portugal), IIAA (Instituto Investigação Agronomia de Angola) and ERU (Estação Rgional de Uige) Periera et al. (2002). These materials were used extensively in the breeding program of coffee for resistance to diseases and pests in Brazil.

Since Híbrido de Timor is an important source of gene for resistance to diseases and used in large extent in the breeding program of coffee in the world, understanding the existed genetic variability will be very important and it will help in designing the breeding strategies to develop varieties which combine good cup quality with resistance to disease. In addition, it is important to plan appropriate conservation strategy to maintain and conserve the accessions of Híbrido de Timor. The study of genetic diversity among accessions also helps to identify the duplicated accessions in the germplasm bank and planning appropriate selection to form the core collection. The existence of considerable genetic diversity within Híbrido de Timor was reported by Lashermes et al. (2000) but he included only two accessions. So it is important to study the genetic variability including more accessions of Híbrido de Timor. To study the genetic variability, morphological, isoenzyme and molecular techniques can be used. Among them the DNA molecular technique is the best and able to detect differences at the genome level. Among the molecular technique Simple Sequence Repeats (SSR), Random Amplified Polymorphic DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP) are more frequently used for genetic diversity study and proved efficient to establish the core collection. Lashermes et al. (2000), Orozco-Castillo et al.

(1994), Silvestrini et al. (2007, 2008) used AFLP, RAPD and SSR techniques to study the genetic diversity of coffee species. Maluf et al. (2005) used RAPD, AFLP and SSR molecular markers to study the genetic diversity of arabica coffee and detected significant polymorphism among lines.

Since Híbrido de Timor is an important coffee group in the breeding program of *C. arabica* understanding the genetic diversity is important to exploit in the future breeding program. So this work was realized with the objective of investigating the existed genetic diversity among the accessions of Híbrido de Timor using RAPD, AFLP and SSR molecular markers.

2. MATERIALS AND METHODS

Genetic materials

Forty eight accessions of Híbrido de Timor from the germplasm bank of UFV (Universidade Federal de Viçosa)/EPAMIG (Empresa de Pesquisa Agropecuária de Minas Gerais) were analysed (Table 1).

Table 1: Accessions of Híbrodo de Timor investigated in this study.

Code	Genotype Name	Code	Genotype Name
1	CIFIC 832/1	25	UFV 427-22
2	CIFIC 832/2	26	UFV 427-55
3	CIFIC 4106	27	UFV 427-56
4	CIFIC 1343/269	28	UFV 427-65
5	UFV 376-01	29	UFV 427-90
6	UFV 376-04	30	UFV 438-52
7	UFV 376-05	31	UFV 439-02
8	UFV 376-35	32	UFV 440-22
9	UFV 376-37	33	UFV 442-108
10	UFV 376-52	34	UFV 443-03
11	UFV 376-57	35	UFV 446-08
12	UFV 376-79	36	UFV 445-46
13	UFV 377-01	37	UFV 428-04
14	UFV 377-02	38	UFV 432-07
15	UFV 377-23	39	UFV 433-11
16	UFV 377-24	40	UFV 435-11
17	UFV 377-34	41	UFV 437-06
18	UFV 379-07	42	UFV 441-03
19	UFV 408-18	43	UFV 447-48
20	UFV 408-26	44	UFV 448-69
21	UFV 408-28	45	UFV 449-20
22	UFV 427-01	46	UFV 450-61
23	UFV 427-09	47	UFV 451-41
24	UFV 427-15	48	UFV 440-10

Extraction of DNA

The DNA of the genotypes was extracted according to the method described by Diniz et al. (2005) from young green leaves collected from each genotype. The DNA concentration was quantified using Spectrophotometre Smart Spec of BioRad. The extracted DNA was diluted with TE (Tri-HCL 10mM, EDTA 1mM, pH 8.0) to concentration of 10 ηg/µl, 25 ηg/µl and 50 ηg/µl for analysis of RAPD, SSR and AFLP, respectively.

RAPD molecular marker

A total of 52 RAPD oligonucleotide *primers* were analyzed. Each reaction constituted total volume of 25ml with the following components: 2.5 µl of 10 ng/µl genomic DNA, 1.25 µl of *Taq* DNA polymerase, 0.25 µl of each dNTP, 1 µl of *primer*, 2.5 µl of KCl, 2.5 µl of Tris HCl pH 8.3, 2 mM of $MgCl_2$ and 12.5 µl of H_2O MillQ. The PCR reaction was done using thermo-cycle Gene Amp PCR System 9600. For the amplification, the DNA was desnatured at 95 ^{0}C for 1 minute followed by 40 cycle of 94 ^{0}C for 15 second to denature DNA, 35 ^{0}C for 30 second to annealing the primer, 72 ^{0}C for 1 minute extension of the primer followed by final extension to complete the process of amplification at temperature of 72 ^{0}C for 7 minutes and stored at 4 ^{0}C until used. The amplified DNA was separated using Agarose gel electrophoreses 1.2 %. The gel then immersed on the solution of Ethidium Bromide and visualized using Eagle Eye II Still Video System.

Microsatellite (SSR) marker

Eighteen microsatellite primers obtained from Combes et al. (2000) and Rovelli et al. (2000) were analyzed. The PCR reaction was realized in total volume of 20µl containing 50ηg of DNA, 0,6 unit of *Taq* DNA polimerase, buffer 1x, 1mM of $MgCl_2$, 150µM of each dNTP and 0.1µM of each *primer*. The amplification was done using the procedure of *touchdown*-PCR that consisted of desnaturation at 94°C for 2 minutes, followed by 13 cycle of desnaturation at 94°C for 30 seconds, primer annealing at 67°C to 55°C for 30 seconds, reducing 1°C of each cycle and extension of primer at 72°C for 30 second. This step followed by more 30 cycles of desnaturation at 94°C for 30 second, primer annealing at 55°C for 30 seconds and primer extension at 72°C for 30 seconds. The final extension was done at 72°C for 8 minutes. The PCR products were separated on a 6% denaturing polyacrylamide gel and visualized by silver staining solution

AFLP (Amplified Fragment Length Polymorphism) analysis

The AFLP genotyping of coffee accessions was done according to the method described by Brito et al. (2010). The following primer combination (*MseI*-AGC with *EcorI*-CGT, and *MseI*-AGC with *EcorI*-CTC) were used to genotype the forty eight accessions of Híbrido de Timor in this study.

Data analysis

The gels of RAPD, SSR and AFLP were scored for the presence and absence of clearly amplified fragments. To analysis the AFLP, RAPD and SSR data using Structure population genetic analysis software it was coded according to Flaush et al. (2007) to fit for analysis of Structure assuming a correct model of dominance. AFLP data statistical package (Ehrich 2007) was used to manage the data conversion from txt format to Structure and Arlequin data format.

To study the genetic diversity, the Jaccard genetic distance (1- Jaccard similarity coefficient) (Jaccard 1908), Nei genetic diversity index (Nei 1973), Shannon´s Information Index and percent polymorphic bands (P%) were estimated, using POPGENE statistical software version 1.3 (Yeh and Boyle 1997). Analysis of molecular variance (AMOVA, Excoffier et al. 2005) for AFLP, RAPD and SSR molecular markers was estimated using Arlequin 3.1 software and used to determine the intra and inter population genetic differentiation.

To define the population structure of the Híbrido de Timor accessions, the Bayesian model-based clustering method of Pritchard et al. (2000), implemented in the Structure 2.1 software (http://www.pritch.bsd.uchicago.edu/) to estimate population admixture through inferred ancestry. For each of the K=2 to K=12 settings, 20 independent simulation were performed using the admixture model and 5000 replicate for burn-in and postburning sampling by Markov Chain Monte Carlo of 50,000 runs to estimate the number of subpopulations for each of the k values. The appropriate number of cluster (K) was determined based on the ad hock statistic ΔK. The ΔK was determined according to Evanno et al. (2005).

To estimate the pairwise F_{ST} between population software AFLPsurv (Vekemans et al. 2002) was used. Multidimensional scaling was used to understand the genetic diversity pattern among accessions of Híbrido de Timor based on Jaccard genetic distance and the two dimensional graph was plotted to visualize the genetic diversity. To do this STATISTICA software (StatSoft Inc. 2001) was used. Principal coordinate

analysis (PCoA) was performed using the GenAlex 6.2 population genetic analysis software (Peakall and Smouse 2006).

3. RESULT

Genetic diversity analysis

The AFLP, RAPD and SSR molecular markers produced 108, 157 and 76 bands in accessions of Híbrido de Timor. The individual population diversity parameters estimated on Híbrido de Timor using AFLP, RAPD and SSR was presented on Table 2. These genetic diversity measures showed that Híbrido de Timor has considerable genetic diversity. The SSR molecular marker produced the lowest information in relation to RAPD and AFLP molecular markers.

Table 2: The Nei gene diversity measure, Shannon´s information index and percent polymorphic bands (P%) obtained using AFLP RAPD and SSR molecular markers for Híbrido de Timor accessions.

	Nei gene diversity	Shannon´s Information Index	P (%)	No Bands
AFLP	0.1349	0.2204	56.48	61
RAPD	0.1534	0.2277	46.50	73
SSR	0.059	0.0922	25	19

The estimated Genetic dissimilarity coefficient (1-Jaccard similarity coefficient) was used in principal coordinate analysis (PCoA). PCoA using RAPD and SSR markers presented well defined clusters (Figure 1 and 2). The PCoA based on the combined data of AFLP, RAPD and SSR (Figure 1) showed clear grouping pattern among accessions of Híbrido de Timor. In addition, the clustering analysis using UPGMA method based on the combined data of the three molecular markers formed three principal cluster groups (Figure 2). The CIFC4106 considered the first plant found in the Timor Island (Pereira et al. 2005) grouped together with CIFC 832/1, CIFC 832/2, and CIFC 1343/269 introduced from Timor Island by seed (Betencourt 1973) and send to Brazil via vegetative propagation. These groups also include three accessions of Híbrido de Timor introduced by seed to UFV from CIFC.

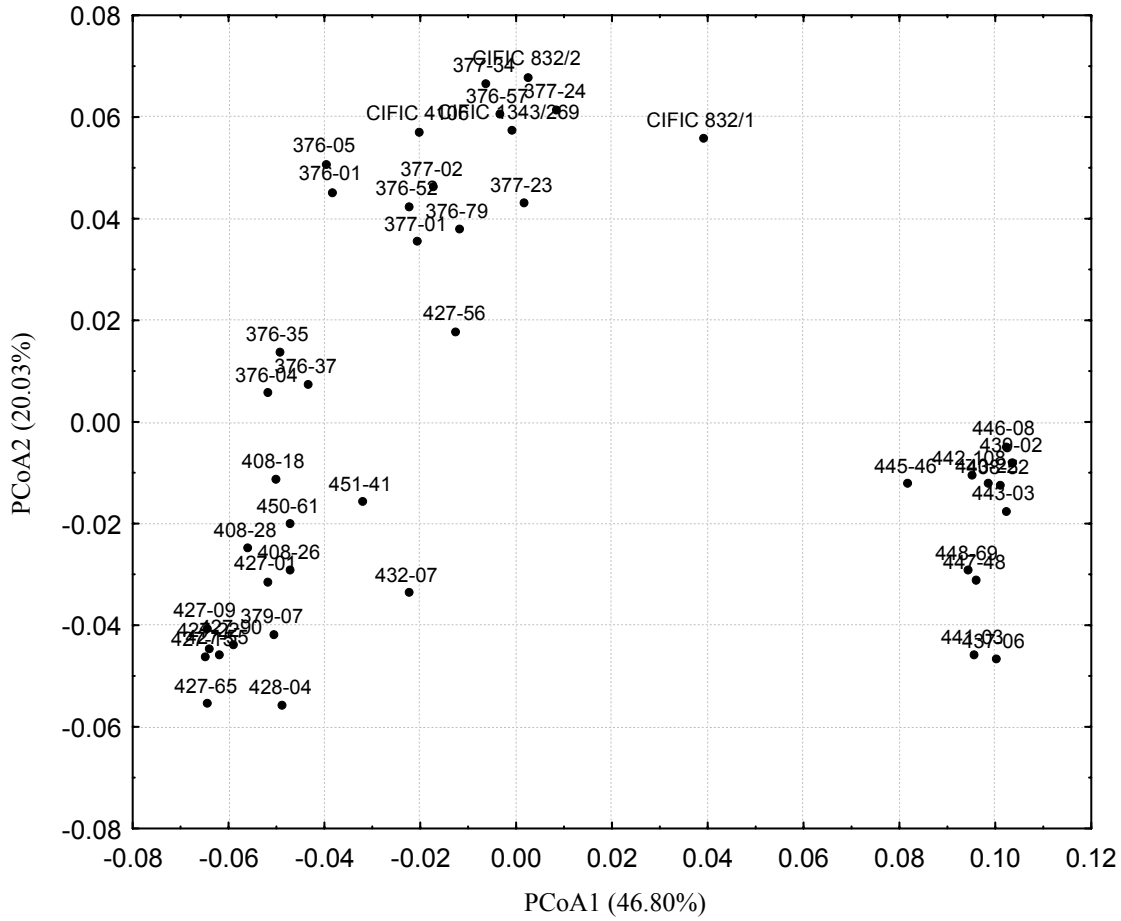

Figure 1: Principal coordinate analysis of RAPD diversity among accessions of Híbrido de Timor.

Genetic Structure

For the population genetic structure analysis we grouped the accessions into two classes. The first group comprise CIFC 832/1, CIFC 832/2, CIFC 1343/269 and CIFC 4106 and the second group comprised others accessions of Híbrido de Timor. The analysis of molecular variance (AMOVA) demonstrated the partitioned of the total variation within population (>80%) than among population (<20%) for AFLP, RAPD and SSR molecular markers (Table 3). The genetic differentiation detected by AFLP molecular marker was lower than the other two molecular markers. The highest genetic differentiation among population was detected by RAPD (16.22 % among populations).

Table 3: AMOVA of genetic variation using AFLP, RAPD and SSR markers

Marker	Source of Variation	*df*	Sum of squares	Variance component	Percent of total component variance
AFLP	Between populations	1	11.828	0.615	7.76
	Within populations	46	336.568	7.316	92.24
	Total	47	348.396	7.931	
RAPD	Between populations	1	21.504	1.722	16.22
	Within populations	45	400.326	8.896	83.78
	Total	46	421.830	10.618	
SSR	Between populations	1	5.505	0.357	10.99
	Within populations	44	127.321	2.893	89.01
	Total	45	132.826	3.251	

To understand the genetic structure of the accessions of Híbrido de Timor the model based clustering analysis based on Bayesian statistical analysis was performed using Structure program (Pritchard et al. 2000). The program gave the shared ancestral probability for each group and individual which helps unambiguous assignment of the accession in their respective group. Those individuals gave >80% shared ancestral probability were assigned in its respective group. Based on this criteria using AFLP molecular marker, about 78% of the accessions grouped in one cluster which proved AFLP not able to differentiate accessions as showed in PCoA (Figure 3). RAPD and SSR molecular markers grouped accessions into four clusters. For RAPD molecular marker 85% of accessions of Híbrido de Timor grouped in respective cluster with shared ancestral probability higher than 0.81. The rest of the genotypes showed some type of admixture. CIFC 4106 has shared ancestral probability with four clusters with >0.10. The SSR molecular marker grouped CIFC 4106, 832/1, 832/2 and CIFC 1343/269 in the same group with shared ancestral probability greater than 0.81 (result not presented). More than 80% of the accessions were grouped in to thier respective group with shared ancestral probability greater than 0.81.

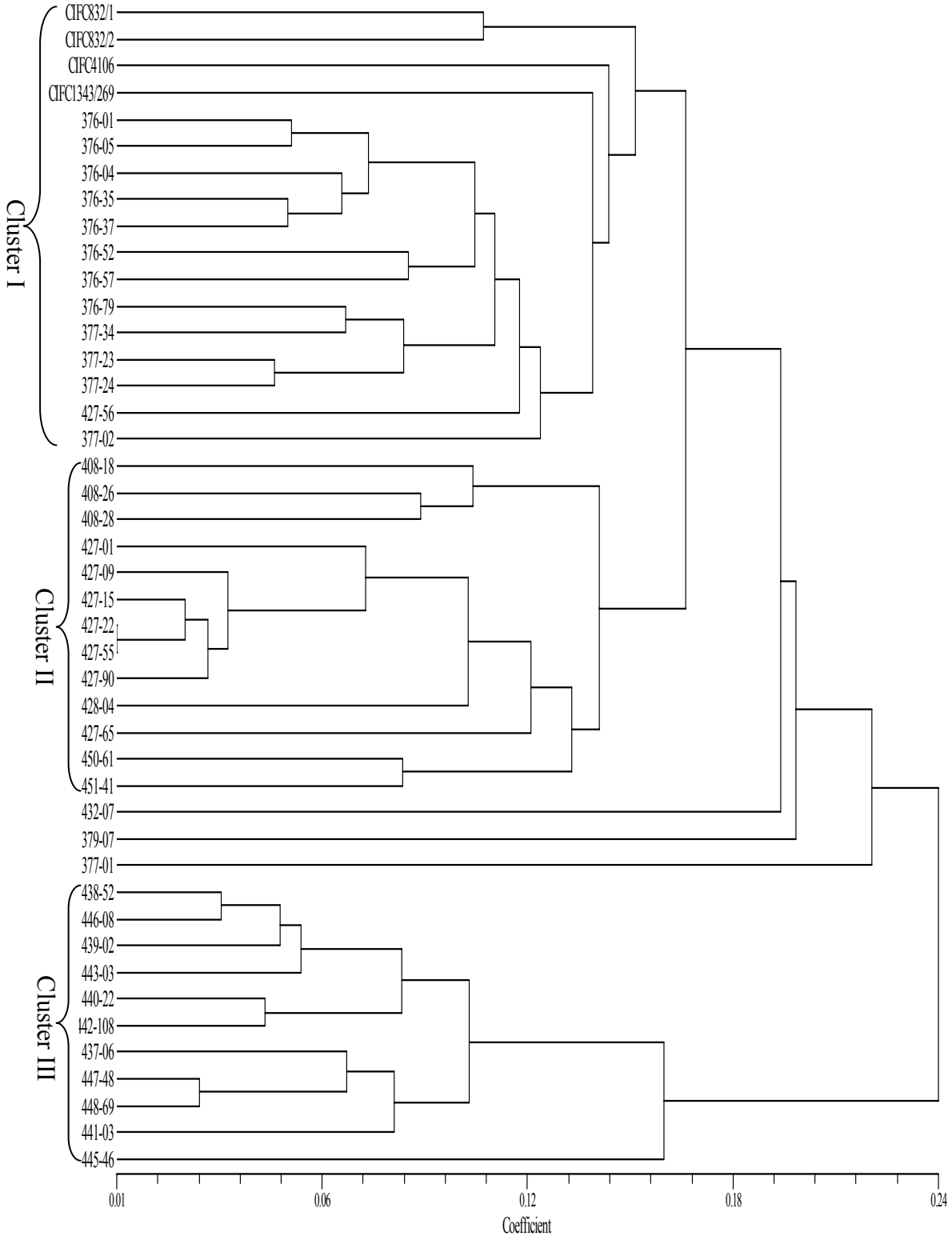

Figure 2: The dendrogram produced based on Genetic dissimilarity estimated from the combined AFLP, RAPD and SSR data suing cluster method UPGMA by NTSYS pc program.

CIFC materials grouped in Cluster II with shared ancestral probability greater than 0.902 (Table 4). The other accessions of Híbrido de Timor grouped into three clusters formed (Table 4). The clustering analysis using the combined data of molecular marker formed three cluster groups in agreement (Table 5). The CIFC 832/1, CIFC 832/2, CIFC 1343/269 and CIFC 4106 grouped together in Cluster III with shared ancestral probability more than 0.689. All the accessions derived from UFV 377 and three accessions derived from UFV 376 (UFV 376-01, UFV 376-05, UFV 376-52, UFV 376-58 and UFV 376-79) were grouped together in Cluster II with shared ancestral probability greater than 0.81. The rest of the accessions derived from UFV 376 shared some level of allele with group II and III. These accessions were originated from the accession received from CIFC and registered as CIFC 2234 and CIFC 2235, respectively. UFV 376 and UVF 377 were selected from IIAA 808/5 and IIAA 811/7, respectively in Tanzania. The accessions derived from UFV 408, UFV 427, UFV 428, UFV 432, UFV 450 and UFV 451 were grouped in Cluster III. UFV 427 and UFV 450 were selected at the Institute of Coffee, Angola. UFV 437, UFV 438, UFV 439, UFV 440, UFV 441, UFV 442, UFV 443, UFV 445, UFV 446 and UFV 447 were grouped in Cluster I. These accessions are derived from CIFC 2570. UFV 408, UFV 450 and UFV 451 were grouped in Cluster III. These accessions are derived from CIFC 1590/9. In general the grouping pattern using the model based Bayesian analysis technique (Pritchard et al. 2000) was in agreement with the origin of accessions.

Table 4: Proportion of membership of each pre-defined population in each of the 4 clusters

	Inferred Clusters			Number of individuals
GivenPop	I	II	III	
CIFC	0.092	0.902	0.008	4
Other Híbrido de Timor	0.282	0.306	0.412	43
Fst	0.552	0.293	0.296	

Table 5: Inferred shared ancestral probability of each accession in relation to 4 clusters inferred using RAPD variation.

Code	Name of accession	Inferred cluster		
		I	II	III
1	CIFC832/1	0.308	0.689	0.002
2	CIFC832/2	0.028	0.968	0.004
3	CIFC4106	0.009	0.972	0.019
4	CIFC1343/26	0.017	0.978	0.005
5	UFV376-01	0.002	0.966	0.031
6	UFV376-04	0.003	0.481	0.516
7	UFV376-05	0.002	0.989	0.01
8	UFV376-35	0.003	0.55	0.447
9	UFV376-37	0.009	0.541	0.451
10	UFV376-52	0.004	0.971	0.026
11	UFV376-57	0.005	0.992	0.003
12	UFV376-79	0.01	0.961	0.029
13	UFV377-01	0.004	0.818	0.179
14	UFV377-02	0.004	0.943	0.053
15	UFV377-23	0.01	0.98	0.01
16	UFV377-24	0.011	0.986	0.003
17	UFV377-34	0.005	0.992	0.003
18	UFV379-07	0.007	0.004	0.989
19	UFV408-18	0.003	0.046	0.951
20	UFV408-26	0.003	0.008	0.989
21	UFV408-28	0.002	0.011	0.987
22	UFV427-01	0.005	0.018	0.978
23	UFV427-09	0.002	0.004	0.994
24	UFV427-15	0.002	0.005	0.994
25	UFV427-22	0.002	0.004	0.994
26	UFV427-55	0.002	0.004	0.994
27	UFV427-56	0.084	0.75	0.165
28	UFV427-65	0.002	0.002	0.996
29	UFV427-90	0.005	0.004	0.991
30	UFV438-52	0.996	0.002	0.002
31	UFV439-02	0.997	0.002	0.001
32	UFV440-22	0.996	0.002	0.002
33	UFV442-108	0.992	0.006	0.002
34	UFV443-03	0.997	0.002	0.001
35	UFV446-08	0.996	0.002	0.001
36	UFV445-46	0.933	0.015	0.052
37	UFV428-04	0.004	0.003	0.993
38	UFV432-07	0.17	0.014	0.816
39	UFV437-06	0.997	0.002	0.001
40	UFV441-03	0.997	0.001	0.002
41	UFV447-48	0.997	0.002	0.002
42	UFV448-69	0.994	0.004	0.002
43	UFV450-61	0.006	0.029	0.966
44	UFV451-41	0.027	0.121	0.853

4. DISCUSSION

Genetic diversity analysis

Híbrido de Timor, the interspesfic hybrid between *C. canephora* and *C. arabica,* first found in Timor Island (Bettencourt 1973) is an important source of gene for disease and pest resistance in Arabica coffee breeding. Híbrido de Timor used as a principal source for resistance gene for coffee leaf rust (*Hemileia vastatrix*), coffee berry disease (CBD) caused by *Colletotrichum Kahawae*, root knot nematode (M*eloidgyne exigua*) and bacteriosis caused by *Pseudomonas syringae* pv garçae (Bettencourt 1973, Chaves 1976 Goncalves and Pereira 1998, Carvalho et al. 1989 and Pereira et al. 2005). The existence of considerable genetic diversity among accessions of Híbrido de Timor was reported by Lashermes et al. (2000). The availability of genetic diversity within Híbrido de Timor is very important for the future arabica coffee breeding program. The principal coordinate (Figure 1) and clustering analysis using UPGMA (Figure 2) confirmed the existence of high genetic diversity within accessions of Híbrido de Timor. The high genetic variability was detected by RAPD molecular marker. The high variability detected by RAPD also observed by Maluf et al. (2005). According to these authors this may be due to the difference in amplified regions detected by this marker which generate differences among accessions (Maluf et al. 2005). The combined data from AFLP, RAPD and SSR grouped the accessions of Híbrido de Timors into three groups. The grouping pattern followed its origin and its historical background (Figure 1 and 2). The dendrograma produced based on the combined data using UPGMA clustering method produced three principal clusters (Figure 2). The cluster grouped genotype from the same origin in the same cluster.

Genetic structure of Híbrido de Timor

The analysis of AMOVA showed the total variation was partitioned within (>80%) and among population (<20%), which is expected in out crossed populations (Brussel 1999), which showed high heterozygosity among accessions of Híbrido de Timor. The result obtained different from the expected result for the self pollinated crops. This may be due to high heterozygosity observed among accessions and most of the alleles still did not fixed as expected in most self pollinated crops. In addition most of the accessions of Híbrido de Timor are not pure lines. Both the genetic distance based analysis of diversity and population based statistics (percent polymorphic loci, Nei gene

diversity index and Shanno Index) demonstrated the efficiency of RAPD molecular marker in detecting genetic diversity among accessions of Híbrido de Timor.

The model based clustering method using structure program (Pritchard et al. 2000) considering the data from the three molecular markers clustered the accessions into three cluster groups. CIFC 4106, CIFC 832/1, CIFC 832/2 and CIFC 134/269 grouped together in cluster III which showed high similarity among them. These accessions are the first genetic materials are received from Island of Timor by CIFC. The accessions were grouped according to their origin. The model based Bayesian clustering analysis classified accessions of Híbrido de Timor in its appropriate groups.

CIFC 832/1, CIFC 832/2, CIFC 4106 and CIFC 1343/269 grouped in Cluster II with shared ancestral probability more than 0.689. These accessions were introduced to CIFC from Timor Island CIFC (Bettencourt 1973). Among them CIFC 832/1 and CIFC 832/2 were evaluated for their resistance to coffee leaf rust and found resistance for all of the races identified at that time and distributed all over the world (Bettencourt 1973). Still these accessions were used for the development of arabica coffee cultivars resistance for coffee leaf rust all over the world (Bettencourt 1973, Periera et al. 2005). CIFC 832/1 and CIFC 832/2 used as main source for resistance gene for coffee leaf rust (*H. vastatrix*) and using this genotypes as donor parent a lot of cultivars released in different part of the world (Bettencourt 1973Pereira et al. 2005, and Zambolim et al. 2005).

In addition clustering analysis was done for each molecular marker (result not presented). Even if the four clusters is formed using AFLP variation more than 80% the individuals assigned in one cluster which confirmed the low differentiation detected among accessions using AFLP molecular markers. This may be happen most of the accessions showed high heterozigoisity for the alleles of AFLP (Table 2).

The CIFC 4106 showed some level of admixture with three clusters using RAPD marker variation. This result demonstrated that CIFC 4106 is probably the original plant found in Timor Island that gave the origin of other accessions of Híbrido de Timor. Under Viçosa soil and climatic condition, CIFC 4106 have high flowering capacity but produced small amount of fruits. All the fruits contain the seed type Moca which indicated having a series problem of auto incompatibility during meiosis and gamete formation (Pereira et al. 2008). The field observation suggested that CIFC 4106 is the original plant of Híbrido de Timor (Pereira et al. 2005). Additional field observation showed that during artificial hybridization, when CIFC 4106 used as female parent did not obtained any success. In other hand, when used as male parent (donor of pollen)

obtained some success in crossing with low percentage of fertilization (Pereira et al. 2008). These characters are typical for an interspecific hybrid plant. The grouping pattern formed by Structure program followed the origin of the accessions. Most of the accessions derived from the same plant by seed grouped together with shared ancestral probability >0.80.

The existence of the high genetic diversity within Híbrido de Timor proved the importance of this plant for the future breeding program of arabica coffee as a valuable source of genes for economically important diseases and pest. In addition, the collections of Híbrido de Timor found in the germplasm bank of UFV/Epamig have considerable genetic diversity and can be exploited in the breeding program of coffee. It also helps in developing cultivars of *C. arabica* having resistance for disease and pest with high cup quality (Periera et al. 2005, Periera et al. 2008).

The RAPD molecular marker produced a well defined clusters and detected high genetic variability among accessions of Híbrido de Timor which demonstrate its efficiency to study the genetic diversity among accessions of Híbrido de Timor. The efficiency of RAPD molecular marker for the genetic diversity study of was reported in Arabica coffee (Maluf et al. 2005 and Silvestrini et al. 2008). The efficiency of RAPD and SSR molecular markers for the study of genetic diversity also reported by other authors (López-Gartner et al. 2009, Silverstrini et al. 2007, and Silvestrini et al. 2008).

This study demonstrated the importance of genetic diversity study to identify the accession duplicated and establishing the core collection without losing the genetic variability existed within the populations of Híbrido de Timor. Especially in coffee the cost to maintain the germplasm bank is very high due to its perennial nature and need a large area per plant. Besides this, understanding the level of the existed genetic variability among the base population will help in designing appropriate breeding strategy of *C. arabica* in the future.

The combined molecular data and the invidual data showed the existence of high genetic diviersity among accessions of Híbrido de Timor. This is indicated the potential of Híbrido de Timor in the future arabica coffee breeding in Brazil and in the world. Even if accessions of Híbrido de Timor derived from small number of plants it showed high genetic variability which is essencial for the breeding program. The Bayesian clustering analysis and other distance based genetic diversity analysis techniques (Principal coordinate analysis, and UPGMA clustering) methods showed efficient to study the genetic diversity of Híbrido de Timor. The model based clustering analysis

was efficient in assigning genotypes in its respective groups without any ambiguity also proved efficient in this study as reported by López-Gartner et al. 2009.

5. CONCLUSIONS

High genetic diversity was observed among accessions of Híbrido de Timor found in germplasm bank of UFV/Epamig, which is an important for the breeding program of *C. arabica* since it used as the source of genes resistance for diseases and pests.

The clustering analysis using distance based and model based Bayesian analysis were grouped genotypes according to their origin. The model based Bayesian clustering analysis is more efficient in assigning the genotypes in its respective group and assigned genotypes with out ambiguity which is not happen in other methods.

Molecular markers AFLP, RAPD and SSR were efficient in discriminating accessions of Híbrido de Timor and to study the genetic diversity among accessions.

The existence of high genetic diversity among accessions of Híbrido de Timor is fundamental for the future *C. arabica* breeding in Brazil.

6. REFERENCES

Beertrand B, Guyot B, Anthony F, Lashermes P (2003) Impact of the Coffea canephora gene introgression on beverage quality of *C. arabica.* Theor Appl Genet.107 : 387-394.

Bettencourt A (1973) Considerações gerais sobre o 'Híbrido de Timor'. Instituto Agronomico de Campinas, Campinas, Brasil. Circular No. 23 p.20

Brito GG, Caixeta ET, Gallina AP, Maciel-Zambolim E, Zambolim L, Diola V, Loureiro M. E. Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. Euphytica, 2010 (in press)

Bussell JD (1999). The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of Isotoma petraea (Lobeliaceae). Molecular Ecology, 8, 775–789.

Carvalho A, Fazuoli LC, Costa WM (1989) Produtividade do Híbrido Timor, de seus derivados e outras fontes de resistência a Hemileia vastatrix. Bragantia, v. 48, n. 1, p. 73-86.

Charrier A, Eskes A (1997) Coffee. *In* : Tropical Plant Breeding. Eds. Charrier A, Jacquot S.,Hamon, D. Nicolas Coll. Repères, Cirad, France, pp. 128-152.

Chaves GM (1976) Melhoramento do cafeeiro visando à obtenção de cultivares resistentes à Hemileia vastatrix Berk et Br. Revista Ceres, Viçosa, v.23, n.128, p.321-332.

Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P(2000) Characterization of microsatellites loci in Coffea arabica and related coffee species. Mol Ecol 9:1171–1193

Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB, Zambolim EM, Loureiro ME, Pereira AA, Zambolim L (2005) Analysis of AFLP marker associated to the *Mex-1* resistance locus in Icatu progenies. Crop Breeding and Applied Biotechnology 5**:** 387-393.

Excoffier L, Laval G, Schneider S (2006) Arlequin ver. 3.1: An integrated software package for population genetics data analysis. Available at http://cmpg.unibe.ch/software/arlequin3/arlequin31.pdf (verified 19 Apr. 2009).

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: Dominant markers and null alleles. Mol. Ecol. Notes 7:574–578.

Gonçalves W, Pereira AA (1998) Resistência de cafeeiros a nematóides IV-reação de cafeeiros derivados de Híbridos Timor a Meloidogyne exigua. Nematologia Brasileira 22:39-50.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bulletin de la Societé Vanddoise des Sciences Natureles, v. 44, p. 223-270.

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

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

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

López-Gartner G, Cortina H, McCouch SR, Moncada MDP (2009) Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. Tree Genetics and Genomes 5:435-446.

Maluf MP, Silvestrini M, Ruggiero LM, Guerreiro Filho O, Colombo CA (2005) Genetic diversity of cultivated *Coffea arábica* inbred lines assessed by RAPD, AFLP and SSR marker systems. Scientia Agricola 62: 366-373.

Miller MP (1997) Tools for Population Genetic Analyses (TFPGA): a windows program for the analysis of allozyme and molecular population genetic data, Computer software distributed by author.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, USA 70, 3321–3.

Nei M (1978) Estimation of average hetrozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

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

Peakall R, Smouse P (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research *Molecular Ecology Notes*, Vol. 6, No. 1: 288-295.

Periera AA, Moura WM, Zambolim L, Sakiyama NS, Geraldo MC (2002) Melhoramento Genético do Cafeeiro no Estado do Minas Gerais – Cultivares Lançados e em Fase de obtenção. In: O Estado da Arte de Tecnologias na Produção de Café. Ed .ZambolimL. , Viçosa, UFV. pp. 253-296.

Pereira AA, Sakiyama NS, Zambolim L, Moura WM, Zambolim EM, Caixeta ET (2005) Identification and use of sources of durable resistance to coffee leaf rust in the UFV/EPAMIG breeding program. In: Durable resistance to coffee leaf rust. Eds. ZAMBOLIM L., ZAMBOLIM E. M., VÁRZEA V. M. P., Viçosa, UFV. pp. 215-232.

Pereira AA, Baião AC, Sakiyama SN (2008) Híbrido de Timor como fonte de resistência a doenças e de qualidade da bebida do cafeeiro. In: Núcleo de Estudos em Fitopatologia. (Org.). Manejo fitossanitário da cultura do cafeeiro. Brasília: Sociedade Brasileira de Fitopatologia, v. , p. 13-24.

Prakash NS, Marques DV, Varzea VMP, Silva MC, Combes MC, Lashermes P (2004) Introgression molecular analysis of a leaf rust resistance gene from Coffea liberica into C. Arabica L. Theor Appl Genet 109: 1311-1317.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Rovelli P, Mettulio R, Anthony F, Anzueto F, Lashermes P, Graziosi G (2000) Microsatellites in Coffea arabica L. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) Coffee biotechnology and quality. Proceedings of the 3rd International Seminar on Biotechnology in the Coffee. Agroindustry, Londrina, Brazil. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 123–133.

Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O,Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian Coffea arabica L. accessions using microsatellites markers. Genet Resour Crop Evol 54:1367–1379.

Silvestrini M, Maluf M, Silvarolla MB, Guerreiro-Filho O, Medina-Filho HP, Vanini MMT, Oliveira AS, Gaspari-Pezzopane C, Fazuol LC (2008) Genetic diversity of a Coffea Germplasm Collection assessed by RAPD markers. Genet Resour Crop Evol (2008) 55:901–910.

Vekemans X, Beauwens T, Lemaire M, Roldan–Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Mol. Ecol. 11, 139–151.

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research, London, v.23, n.21, p.4407-4414.

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belgian J. Bot. 129:157.

Zambolim L, Vale FXR, Zambolim EM (2005) Doenças do cafeeiro. In: Kimati, H.; Amorim, L.; Rezende, J. A. M.; Bergamim Filho, A.; Camargo, L. E. A. (Eds.). Manual de fitopatologia. 4. ed. São Paulo: Ceres, v. 2, 663 p.