

FLÁVIO DE FRANÇA SOUZA

DIVERSIDADE GENÉTICA, ESTRUTURA POPULACIONAL E
MAPEAMENTO ASSOCIATIVO EM *Coffea canephora*

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
2011

FLÁVIO DE FRANÇA SOUZA

DIVERSIDADE GENÉTICA, ESTRUTURA POPULACIONAL E
MAPEAMENTO ASSOCIATIVO EM *Coffea canephora*

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 em 19 de julho de 2011.

Dra. Eveline Teixeira Caixeta
(Co-orientadora)

Prof. Dr. Cosme Damião Cruz
(Co-orientador)

Dr. Antonio Carlos Baião de Oliveira

Dr. Antônio Alves Pereira

Prof. Dr. Ney Sussumu Sakiyama
(Orientador)

Ofereço...

Aos meus pais, Edvaldo José de Souza e Maria de Lourdes de França Souza, em reconhecimento pela sua dedicação à causa da criação e educação dos seus três filhos. Por ensinar-nos o valor de princípios como honestidade, dignidade, respeito ao próximo, humildade e trabalho...

Às minhas irmãs, Sílvia Leticia de França Souza e Lídia Patrícia de França Souza pela companhia e amizade desde tão nova infância. Aos meus sobrinhos Anderson, Suzana, Lucas, Milena e Vítor, pela alegria que trouxeram ao “clã dos França Souza”.

À memória dos meus avós, José Manoel de Souza, Alzira Mônica de Souza, João Adão de França e Josefa Maria de Jesus, pessoas maravilhosas, com as quais tive a felicidade de conviver e aprender, entre outras coisas, a importância da agricultura.

Dedico...

À minha esposa Farah de Castro Gama e aos meus filhos

Gustavo Castro de França Souza e Giovanna Castro de França Souza.

AGRADECIMENTOS

À Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e à Universidade Federal de Viçosa (UFV), pela oportunidade de realizar esse curso.

Ao professor Ney Sussumu Sakiyama, pela orientação segura e paciente, pelos muitos ensinamentos acadêmicos e pelas inúmeras lições de vida.

À Dra. Eveline T. Caixeta, pela orientação segura, pela disposição de compartilhar o seu conhecimento e pelo apoio em todos os momentos.

À Dra. Eunize Maciel Zambolim, pelos ensinamentos e pelo espírito de colaboração.

Aos amigos e colegas de Laboratório: Luís Felipe V. Ferrão, Rejane Freitas, Kátia Pestana, Samuel Mazzinghy, Tesfahun Alemu, Guilherme Pena, Robson Almeida, Robson Missio, Patrícia Gonçalves, Alexandre Capucho e Telma Faglieri, pela amizade, pelo apoio e pelos muitos ensinamentos.

Aos pesquisadores Antônio Alves Pereira (Epamig), Antônio Carlos Baião de Oliveira (Embrapa Café) e Felipe Lopes da Silva (Epamig), Romário Gava Ferrão (Incaper), Herculano P. Medina Filho (IAC) e Rita Bordignon (IAC) pela amizade, pela ótima parceria que estabelecemos e pelo exemplo de profissionalismo.

Aos professores Pedro Crescêncio Carneiro, Cosme Damião Cruz e Márcio Henrique Barbosa pelos valiosos ensinamentos repassados durante as disciplinas.

Aos bons amigos que fizemos em Viçosa, Magnus e Poty, Bernardino e Gisele, Gustavo e Janaína, Fernando e Lídice pela acolhida calorosa, pelos muitos momentos que compartilhamos e pela amizade sincera.

Aos colegas e amigos da Embrapa Rondônia, no campo experimental de Ouro Preto do Oeste: João Maria Diocleciano, Gilvan de Oliveira Ferro e Milton Messias dos Santos, pelo apoio técnico e logístico, na coleta das amostras e nas avaliações em campo.

À Leida Etelvina, valiosa amiga, por estar sempre cuidando da nossa família e nos apoiando. Aos amigos David e Kárta pela agradável companhia e pelas valiosas lições de vida.

Ao Prof. Manoel Abilio de Queiróz e à Dra. Rita de Cássia Souza Dias, pesquisadores da Embrapa Semiárido, pelo incentivo, pela parceria, pela amizade e por despertarem em mim o interesse pelo melhoramento de plantas.

Aos meus pais, minhas irmãs, meus sobrinhos, meus tios, à minha sogra e aos meus cunhados pela torcida e pelas muitas palavras de incentivo.

À minha esposa Farah, pelo amor, pela presença constante e pelo incentivo; e aos nossos filhos Gustavo e Giovanna por nos trazerem tanta alegria.

A Deus, pelo dom da vida, pela saúde, pelas oportunidades e por colocar tanta gente boa no meu caminho!

BIOGRAFIA

FLÁVIO DE FRANÇA SOUZA, filho de Edvaldo José de Souza e Maria de Lourdes de França Souza, nasceu em Petrolina, Pernambuco, no dia 22 de dezembro de 1973. Iniciou o ensino fundamental, em 1981, nas Escolas Reunidas Ouro Preto, em Petrolina. Em 1985, iniciou a fase complementar do ensino fundamental na Escola Estadual Marechal Antônio Alves Filho – EMAAF. Em 1989, sob influência dos avôs maternos, os agricultores João Adão de França e Josefa Maria de Jesus, ingressou na turma pioneira da Escola Agrotécnica Federal Dom Avelar Vilela, formando-se Técnico em Agropecuária, em 1991. Em 1992, iniciou a graduação em Engenharia Agrônômica pela Faculdade de Agronomia do Médio São Francisco – FAMESF/UNEB, em Juazeiro, Bahia. Durante o curso, despertou interesse pelo estudo da Genética e do Melhoramento de Plantas, no que foi muito incentivado pelo Professor e Pesquisador da Embrapa, Dr. Manoel Abilio de Queiróz. Em 1996, graduou-se Engenheiro Agrônomo e foi convidado a atuar, como bolsista, nos projetos de Recursos Genéticos e de Melhoramento Vegetal de cucurbitáceas na Embrapa Semiárido, em Petrolina. Em 1998, ingressou no programa de pós-graduação em Botânica da Universidade Federal Rural de Pernambuco, tendo obtido o título de mestre, em abril de 2000. Durante os anos de 1999, 2000 e 2001 atuou como professor convidado na Pós-Graduação *Lato Sensu* das Faculdades Integradas da Vitória de Santo Antão - FAINTIVISA, onde ministrou a disciplina Fundamentos de Genética. Em julho 2001, mediante concurso público, foi contratado como Engenheiro Agrônomo, pela Secretaria de Agricultura do Estado da Bahia. Em novembro do mesmo ano, por meio de novo concurso, foi contratado pela Embrapa para ocupar o cargo de Pesquisador na área de Melhoramento Vegetal. Ficou lotado na cidade de Porto Velho, em Rondônia, onde atuou em diversos projetos de pesquisa, desenvolvendo atividades relacionadas ao melhoramento genético de café (canéfora e arábica), melancia, feijão-caupi e feijão comum. Em 2007, ingressou no Programa de Pós-graduação em Genética e Melhoramento da Universidade Federal de Viçosa, em Viçosa, Minas Gerais, onde no dia 19 de julho de 2011, obteve o título de *Doctor Scientiae* em Genética e Melhoramento.

SUMÁRIO

RESUMO.....	vii
ABSTRACT.....	ix
INTRODUÇÃO GERAL.....	1
REVISÃO DE LITERATURA.....	4
1. Taxonomia do café.....	4
2. Morfologia do gênero <i>Coffea</i>	5
3. Espécies comerciais de <i>Coffea</i>	6
3.1. <i>Coffea liberica</i> Bull. ex Hiern.....	9
3.2. <i>Coffea arabica</i> L.....	9
3.3. <i>Coffea canephora</i> Pierre ex A. Froehner.....	10
3.3.1. Auto-incompatibilidade de <i>C. canephora</i>	10
3.3.2. Distribuição natural e estrutura populacional de <i>C. canephora</i>	11
3.3.3. Recursos genéticos de <i>C. canephora</i>	12
3.3.4. Melhoramento genético de <i>C. canephora</i>	14
4. Marcadores moleculares em café.....	22
5. Diversidade genética.....	23
6. Estrutura da população.....	24
7. Desequilíbrio de ligação e mapeamento associativo.....	25
REFERÊNCIAS BIBLIOGRÁFICAS.....	32
CAPÍTULO 1. Análise da diversidade no germoplasma brasileiro de <i>Coffea canephora</i>	41
CAPÍTULO 2. Estudo da estrutura populacional do germoplasma brasileiro de <i>Coffea canephora</i>	62
CAPÍTULO 3. Análise do desequilíbrio de ligação e estudo preliminar do mapeamento associativo em <i>Coffea canephora</i>	89
CONCLUSÃO GERAL.....	127

RESUMO

SOUZA, Flávio de França, D.Sc., Universidade Federal de Viçosa, julho de 2011. **Diversidade genética, estrutura populacional e mapeamento associativo em *Coffea canephora***. Orientador: Ney Sussumu Sakiyama. Co-orientadores: Eveline Teixeira Caixeta e Cosme Damião Cruz.

A seleção assistida por marcadores moleculares apresenta-se como uma técnica potencialmente capaz de auxiliar no melhoramento genético do café canéfora. No entanto, para aplicação dessa ferramenta é preciso que sejam identificados, com precisão, marcadores ligados aos fenótipos de interesse. Até o momento, a busca por esses marcadores tem sido feita por meio do mapeamento baseado na ligação genética, que inclui o desenvolvimento de populações controladas e apresenta baixa resolução. Recentemente, a metodologia de mapeamento associativo, baseada no desequilíbrio de ligação, surgiu como uma alternativa para obtenção de marcadores ligados, por meio do mapeamento de alta resolução em populações naturais. O presente trabalho teve como objetivo estudar a diversidade genética, a estrutura populacional e o desequilíbrio de ligação no germoplasma brasileiro de *C. canephora*, com vistas à implementação da técnica de mapeamento associativo. Foram utilizadas coleções de acessos representativos do germoplasma cultivado e conservado no Brasil, oriundos de quatro instituições de pesquisa, localizadas nos Estados de São Paulo, Espírito Santo, Minas Gerais e Rondônia. As avaliações fenotípicas foram realizadas na Estação Experimental da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), em Ouro Preto do Oeste (RO) e a genotipagem foi realizada no Laboratório de Biotecnologia do Cafeeiro (Biocafê), da Universidade Federal de Viçosa (MG). As análises preliminares de diversidade genética e estrutura populacional foram realizadas utilizando-se um conjunto de 20 microssatélites. Posteriormente, a coleção de acessos com maior diversidade, menor estruturação e menor quantidade de indivíduos aparentados foram genotipadas com 77 microssatélites. Avaliaram-se o desequilíbrio de ligação e a existência de associações entre os marcadores moleculares e algumas características fenotípicas de interesse para o melhoramento da espécie. O germoplasma estudado apresentou diversidade genética expressiva. Observou-se forte estruturação, principalmente, devido a processos evolutivos anteriores a domesticação da espécie. Indícios de sub-

estruturação promovidos por eventos recentes também foram constatados. Os grupos varietais Robusta e Conilon apresentaram-se bastante diferenciados. No grupo Conilon, significativa diferenciação foi observada entre as amostras de Rondônia e do Espírito Santo. Na maioria das populações estudadas, o desequilíbrio de ligação foi baixo. Evidências de associação entre alguns microssatélites e caracteres fenotípicos foram obtidas. Desse modo, este estudo possibilitou um maior conhecimento sobre o germoplasma brasileiro de café canéfora, fornecendo, a melhoristas e curadores de bancos de germoplasma, subsídios para aumento da eficiência na condução dos programas de melhoramento e no manejo dos recursos genéticos da espécie, bem como apresenta considerações importantes a respeito do potencial de uso do mapeamento associativo nesse germoplasma.

ABSTRACT

SOUZA, Flávio de França, D.Sc., Universidade Federal de Viçosa, July, 2011. **Genetic diversity, population structure and association mapping in *Coffea canephora***. Advisor: Ney Sussumu Sakiyama. Co-advisors: Eveline Teixeira Caixeta and Cosme Damião Cruz.

Marker-assisted selection (MAS) is a technique potentially able to aid in *Coffea canephora* breeding. However, its usefulness depends on the precise identification of molecular markers linked to agronomic traits. Nowadays, the search for these markers has been done through linkage mapping, which includes the development of controlled populations and, generally, results on low-resolution frames. Recently, association mapping methodology, based on linkage disequilibrium, emerged as an alternative to obtain linked markers through high-resolution mapping in natural populations. This work aimed to study the genetic diversity, population structure and linkage disequilibrium in Brazilian germplasm of *C. canephora*, in order to implement association mapping technique. We used accession collections representative of the germplasm either grown on commercial plantations or conserved in research institutions in São Paulo, Espírito Santo, Minas Gerais and Rondônia. Phenotypic evaluations were performed at the Experimental Station of Empresa Brasileira de Pesquisa Agropecuária (Embrapa), at Ouro Preto do Oeste, Rondônia. The accessions were genotyped at the Laboratório de Biotecnologia do Cafeeiro (Biocafê), in the Universidade Federal de Viçosa, at Viçosa, Minas Gerais. First genetic diversity analysis and population structure were performed using a set of 20 microsatellites. Subsequently, the collection with greater diversity, less structured and with smaller level of relatedness among individuals was genotyped with a set of 77 microsatellites. We evaluated the linkage disequilibrium and the occurrence of associations among molecular markers and phenotypic traits of interest for breeding. The germplasm showed a considerable genetic diversity. There was a strong structure, mainly due to evolutionary processes prior to domestication of the species. Evidence of sub-structure promoted by recent events was also observed. The varietal groups Robusta and Conilon presented quite divergent. Into Conilon cluster, remarkable differentiation was observed between the samples of Rondônia and Espírito Santo. In most populations, linkage disequilibrium was

low. Evidences of association among some microsatellite and phenotypic traits were obtained. This study increased the knowledge about Brazilian *C. canephora* germplasm, providing to breeders and gene bank curators, subsidies for augmenting the efficiency of breeding programs and to a better management of genetic resources of the species. We also presented important considerations regarding the potential use of mapping association in this germplasm.

INTRODUÇÃO GERAL

A bebida amplamente conhecida como café é produzida a partir dos grãos de duas espécies: café arábica (*Coffea arabica* L.) e café canéfora (*Coffea canephora* Pierre ex A. Froehner). Nos *blends* comerciais (produto obtido pela mistura de grãos das duas espécies), *C. arabica* contribui com o aroma e sabor adocicados, característicos do café, enquanto o *C. canephora* dá solubilidade e corpo à bebida. Outro aspecto de fundamental importância é o fato de que o uso do café canéfora nos *blends* aumenta a lucratividade do produto, sem causar perdas excessivas da qualidade, uma vez que o custo de produção do café canéfora é menor que do café arábica. Além disso, os grãos dessa espécie são o principal componente dos cafés solúveis. Devido ao seu alto teor de cafeína (2%), quando utilizado na produção de cafés descafeinados, gera como subproduto a cafeína, que é matéria prima essencial das indústrias de fármacos, refrigerantes, bebidas isotônicas entre outros.

A produção mundial de café canéfora é da ordem de 2,8 milhões de toneladas de grãos beneficiados ao ano, o que corresponde a cerca de 35 a 40% da produção total de café. O Brasil, o maior produtor e o segundo maior consumidor mundial de café, ocupa a segunda colocação no ranking de produção do café canéfora, sendo superado apenas pelo Vietnã. Na safra 2009/2010, foram produzidas mais de 39 milhões de sacas de café no Brasil, das quais, 10,6 milhões (27%), foram de café canéfora (CONAB, 2010). O café é uma das principais culturas da pauta de exportação do agronegócio brasileiro. Em 2007, as exportações dessa *commodity* totalizaram mais de 24,8 milhões de sacas, rendendo divisas para o Brasil, da ordem de US\$ 3,4 bilhões (ABIC, 2010). Além disso, deve-se ter em conta o papel sócio-econômico desta cultura, uma vez que a cadeia produtiva do café é uma importante fonte geradora de emprego e renda no país.

Introduzido no Brasil no início do século passado, o cultivo comercial do café canéfora foi impulsionado a partir dos anos 1950, com o advento dos cafés solúveis e posteriormente, com o seu uso no café torrado e moído, em misturas com o café arábica. Esse aumento na demanda pelos grãos de *C. canephora* estimulou a produção brasileira e, conseqüentemente, a expansão do parque cafeeiro nacional. Essas lavouras, juntamente com as coleções de trabalho dos programas de melhoramento genético e com os acessos introduzidos no Brasil, compõem o que se pode chamar de germoplasma brasileiro de café canéfora. Apesar da sua magnitude e

importância, são ainda poucos os estudos que tratam da sua caracterização. Além disso, na maioria dos casos, as amostras são setorizadas, o que impede a formação de um quadro panorâmico sobre a diversidade e estrutura populacional desse germoplasma.

A seleção assistida é uma promessa de ferramenta facilitadora dos programas de melhoramento, capaz de aumentar a eficiência das rotinas de seleção pela racionalização dos tamanhos das populações segregantes, em consequência da eliminação precoce dos genótipos não-portadores das características morfo-agronômicas de desejáveis. Por muito tempo a principal metodologia para obtenção de marcadores moleculares ligados às características fenotípicas foi o mapeamento por ligação genética. Todavia, essa técnica apresenta algumas limitações, como a necessidade de formação de populações próprias, tais como F_1 S, F_2 S, retrocruzamentos, duplo-haplóides, entre outras. Recentemente, a estratégia de mapeamento de QTLs por associação genética, baseada no desequilíbrio de ligação, surge como uma alternativa promissora ao mapeamento por ligação, inclusive superando suas principais limitações e possibilitando a identificação de marcadores para a seleção assistida, com alta resolução.

Contudo, o emprego da estratégia de mapeamento associativo requer o conhecimento prévio das características das populações a serem utilizadas, sobretudo no que tange à diversidade genética, estrutura populacional e desequilíbrio de ligação. Desse modo, na concepção do presente trabalho, propôs-se estudar a diversidade e a estrutura populacional do germoplasma brasileiro de *Coffea canephora*, utilizando marcadores microssatélites, com vistas à obtenção de elementos para auxiliar no manejo dos bancos de germoplasma; no direcionamento dos programas nacionais de melhoramento genético da espécie e, sobretudo, na construção de uma estratégia de mapeamento de locos de caracteres quantitativos (QTLs), por meio da técnica de mapeamento associativo, baseada no desequilíbrio de ligação.

Para tanto, foram propostos os seguintes objetivos específicos:

- 1) caracterização molecular de acessos representativos do germoplasma de *Coffea canephora* cultivado e conservado nas plantações comerciais e nas coleções das instituições brasileiras de pesquisa, por meio de marcadores microssatélites;
- 2) Estudo da estrutura genética das populações de *Coffea canephora* existentes no Brasil visando à identificação de amostras adequadas para o estabelecimento de uma

estratégia de mapeamento associativo; 3) Análise do desequilíbrio de ligação entre pares de locos microssatélites em uma população caracterizada fenotipicamente a fim de realizar um estudo preliminar de associação entre marcadores moleculares e características fenotípicas.

REVISÃO DE LITERATURA

1. Taxonomia do café

Os primeiros registros a cerca do café foram encontrados em um manuscrito no Iêmen, em 575. No entanto, as primeiras descrições científicas a respeito do cafeeiro foram apresentadas em 1591 e 1592, pelo botânico veneziano Prospero Alpino, em suas obras *De Medicina Aegyptiorum* e *De Plantis Aegyptii Liber*. Coube a Antoine Jussieu, em sua obra *Histoire Du Café*, publicada em 1716, a primeira classificação botânica do cafeeiro como *Jasminum arabicum*. Posteriormente, em 1737, Carl Von Lineé (Lineu) reclassificou a espécie dando-lhe o nome de *Coffea arabica* (MARTINS, 2008).

A nomenclatura da tribo Coffeae foi originalmente proposta por De Candolle, em 1807. Em sua classificação, esta tribo apresentava grande abrangência e incluía um grande número de gêneros, muitos dos quais foram posteriormente transferidos para outras tribos e subfamílias. Uma das compilações taxonômicas mais impressionantes do gênero *Coffea* foi elaborada por August Chevalier e chama-se *Les Caféiers du Globe*. A obra dividida em três volumes, publicados nos anos de 1929, 1942 e 1947, também traz um conceito do gênero *Coffea* muito mais amplo do que o atualmente aceito. Chevalier dividiu o gênero *Coffea* em quatro secções: *Paracoffea*, *Argocoffea*, *Mascarocoffea* e *Eucoffea*. Esta última agrupava as principais espécies produtoras de cafés e dividia-se em cinco subsecções: *Erythrocoffea* (que inclui, por exemplo, as espécies *C. arabica*, *C. canephora* e *C. congensis*), *Nanocoffea* (p.ex.: *C. humilis*, *C. brevipes*), *Pachycoffea* (p.ex.: *C. liberica*), *Melanocoffea* (p.ex.: *C. stenophylla*) e *Mozambicoffea* (p.ex.: *C. zanguebarie*, *C. racemosa*, *C. salvatrix*, *C. eugenioides*) (BERTHAUD e CHARRIER, 1985).

Trabalhos mais recentes (DAVIS et al., 2006, 2005; MAURIN et al., 2007), envolvendo marcadores morfológicos e moleculares, possibilitaram a reformulação do arcabouço taxonômico do cafeeiro de modo que, atualmente pode-se enunciar o seguinte: O café pertence à família Rubiaceae, subfamília Ixoroideae, tribo Coffeae DC., gênero *Coffea* L. Este gênero encontra-se dividido em dois subgêneros, *Coffea* (95 spp.) e *Baracoffea* (oito spp.), e agrupa 103 espécies (Tabela 1), que ocorrem naturalmente na África tropical e nas ilhas de Madagascar, Comores e Mascarenhas.

O subgênero *Coffea* apresenta ocorrência ampla, cobrindo toda a área de distribuição do gênero, enquanto o subgênero *Baracoffea* está restrito à ilha de Madagascar, nordeste do Quênia e sudeste da Somália.

2. Morfologia do gênero *Coffea*

As espécies do gênero *Coffea* são árvores e arbustos perenes, de madeira dura e densa, com ramificação plagiotrópica; inflorescências axilares pareadas; cálculo presente e geralmente conspicuo; cálice truncado a ondulado, ou levemente lobado e não acrescentado; flores hermafroditas, corolas brancas ou raramente róseas; botões florais com pétalas sobrepostas e contorcidas para a esquerda; anteras protraídas (que se projetam para fora da corola) e semi-protraídas no subgênero *Baracoffea*; frutos contendo duas sementes; semente com um sulco (invaginação) longitudinal na superfície plana (ventral) (DAVIS et al., 2006; MAURIN et al., 2007).

Na ausência de frutos, plantas de algumas espécies do gênero *Coffea* se assemelham a indivíduos de outras taxa da família Rubiaceae, sobretudo com os gêneros *Tricalysia*, *Calycosiphonia*, *Argocoffeopsis* e *Belanophora*, da tribo Coffeeae; *Cremaspora* e *Polysphaeria*, das tribos Cremasporeae e Octotropideae, respectivamente. Uma característica simples que pode ser usada para diferenciar *Coffea* dos demais gêneros é a presença de um cálice reduzido, geralmente em formato de aro, que raramente excede o disco floral e não apresenta lobos (com exceção de *C. kapakata*). Nos outros gêneros, geralmente, os cálices são tubulares, bem desenvolvidos e apresentam lobos na parte superior (DAVIS et al., 2006).

Embora *Psilanthus* seja o gênero mais relacionado com *Coffea*, geralmente, não há dificuldades na distinção entre ambos. As diferenças entre *Coffea* e *Psilanthus* se restringem, basicamente, à morfologia da flor e do pólen. A principal delas refere-se ao fato de que *Psilanthus* apresenta corolas tubulares muito maiores e a maioria das espécies têm inflorescências que se originam inicialmente nas folhas axilares e depois se tornam terminais nas gemas mais curtas. Esta característica também pode ser observada em *Baracoffea*, mas este subgênero tem distribuição restrita e não ocorre em áreas comuns com o gênero *Psilanthus* (DAVIS et al., 2005).

Os grãos típicos de “café”, caracterizados por sementes plano-convexas que apresentam um sulco longitudinal na superfície plana, são produzidos pelas espécies pertencentes aos gêneros *Coffea* e *Psilanthus* (DAVIS et al., 2006; MAURIN et al., 2007).

3. Espécies comerciais de Café

Segundo Berthaud e Charrier (1985), embora muitas espécies tenham sido testadas para exploração comercial, apenas três apresentaram características favoráveis ao cultivo: *Coffea arabica* L., *Coffea canephora* Pierre ex Froehner e *Coffea liberica* Bull. ex Hiern. Atualmente, apenas as duas primeiras têm importância econômica em escala mundial, pois as plantações de *C. liberica* foram dizimadas por uma epidemia de traqueomicose, causada pelo fungo *Fusarium xylarioides* entre as décadas de 1940 e 1950.

As demais espécies do gênero *Coffea* compõem um magnífico acervo de genes e alelos úteis ao melhoramento genético das espécies cultivadas. No entanto, o risco de extinção das populações naturais e vulnerabilidade das coleções de germoplasma *ex situ* são ameaças consideráveis à integridade desse patrimônio. Na Lista Vermelha da União Internacional para a Conservação da Natureza e dos Recursos Naturais ([IUCN](#)) das espécies ameaçadas, observa-se que 72, das 103 espécies de café encontram-se sob algum nível de ameaça de extinção (Tabela 2).

Tabela 1. Taxonomia do gênero *Coffea*

Família	Subfamílias da Família Rubiaceae (3)	Tribos da Subfamília Ixoroideae (8)	Gêneros da Tribo Coffeae (11)	Subgênero do Gênero <i>Coffea</i> (2)	Espécies do Subgênero <i>Coffea</i> (95)			Espécies do Subgênero <i>Baracoffea</i> (8)
Rubiaceae	Ixoroideae Rubioideae Cinchonoideae	Coffeae DC. Alberteae Gardenieae Ixoreae Octotropidae Pavetteae Vanguerieae Incertae Sedis	<i>Coffea</i> <i>Psilanthus</i> <i>Argocoffeopsis</i> <i>Belonophora</i> <i>Calycosiphonia</i> <i>Diplospora</i> <i>Discospermum</i> <i>Nostolachma</i> <i>Tricalysia</i> <i>Sericanthe</i> <i>Xantonnea</i>	<i>Coffea</i> <i>Baracoffea</i>	<i>C. abbayesii</i> <i>C. affinis</i> <i>C. alleizettii</i> <i>C. ambanjensis</i> <i>C. andrambovatensis</i> <i>C. ankaranensis</i> <i>C. anthonyi</i> <i>C. arabica</i> <i>C. arenesiana</i> <i>C. augagneuri</i> <i>C. bakossii</i> <i>C. bertrandii</i> <i>C. betamponensis</i> <i>C. boiviniana</i> <i>C. bonnieri</i> <i>C. brevipes</i> <i>C. bridsoniae</i> <i>C. buxifolia</i> <i>C. canephora</i> <i>C. carrissoi</i> <i>C. charrieriana</i> <i>C. commersoniana</i> <i>C. congensis</i> <i>C. costatifructa</i> <i>C. coursiana</i> <i>C. dactylifera</i> <i>C. dubardii</i> <i>C. eugenioides</i> <i>C. fadenii</i> <i>C. farafanganensis</i> <i>C. fotoana</i> <i>C. fragilis</i> <i>C. gallienii</i> <i>C. heimii</i> <i>C. heterocalyx</i>	<i>C. homollei</i> <i>C. humblotiana</i> <i>C. humilis</i> <i>C. jumellei</i> <i>C. kapakata</i> <i>C. kianjavatensis</i> <i>C. kihansiensis</i> <i>C. kimbozensis</i> <i>C. kivuensis</i> <i>C. lancifolia</i> <i>C. leonimontana</i> <i>C. leroyi</i> <i>C. liaudii</i> <i>C. liberica</i> <i>C. ligustroides</i> <i>C. littoralis</i> <i>C. lulandoensis</i> <i>C. macrocarpa</i> <i>C. magnistipula</i> <i>C. mangoroensis</i> <i>C. manombensis</i> <i>C. mapiana</i> <i>C. mauritiana</i> <i>C. mayombensis</i> <i>C. mcphersonii</i> <i>C. millotii</i> <i>C. minutiflora</i> <i>C. mogenetii</i> <i>C. mongensis</i> <i>C. montekupensis</i> <i>C. montis-sacri</i> <i>C. moratii</i> <i>C. mufindiensis</i> <i>C. myrtifolia</i> <i>C. perrieri</i>	<i>C. pervilleana</i> <i>C. pocsii</i> <i>C. pseudozanguebariae</i> <i>C. racemosa</i> <i>C. rakotonasoloi</i> <i>C. ratsimamangae</i> <i>C. resinosa</i> <i>C. rhamnifolia</i> <i>C. richardii</i> <i>C. sahafaryensis</i> <i>C. sakarahae</i> <i>C. salvatrix</i> <i>C. sambavensis</i> <i>C. schliebenii</i> <i>C. sessiliflora</i> <i>C. stenophylla</i> <i>C. tetragona</i> <i>C. togoensis</i> <i>C. tricalysioides</i> <i>C. tsirananae</i> <i>C. vatovavyensis</i> <i>C. vavateninensis</i> <i>C. vianneyi</i> <i>C. vohemarensis</i> <i>C. zanguebariae</i>	<i>C. ambongensis</i> <i>C. bissetiae</i> <i>C. boinensis</i> <i>C. decaryana</i> <i>C. grevei</i> <i>C. humbertii</i> <i>C. labatii</i> <i>C. pterocarpa</i>

Adaptado de Davis et al., 2006

Tabela 2. Espécies de *Coffea* e sua classificação na lista vermelha da “World Conservation Union (IUCN)”. Adaptado de Davis et al., 2006.

Perigo Crítico de Extinção	Vulnerável	
Africa	Africa	Madagascar
<i>C. anthonyi</i> Stoff. & F.Anthony, ined.	<i>C. arabica</i> L.	<i>C. dubardii</i> Jum.
<i>C. charrieriana</i> Stoff. & F.Anthony, ined.	<i>C. costatifructa</i> Bridson	<i>C. grevei</i> Drake ex A.Chev.
<i>C. fotsoana</i> Stoff. & Sonké	<i>C. dactylifera</i> Robbr. & Stoff.	<i>C. homollei</i> J.-F.Leroy
<i>C. heterocalyx</i> Stoff.	<i>C. fadenii</i> Bridson	<i>C. millotii</i> J.-F.Leroy
<i>C. kihansiensis</i> A.P.Davis & Mvungi	<i>C. kapakata</i> (A.Chev.) Bridson	<i>C. perrieri</i> Drake ex Jum. & H.Perrier
<i>C. kimbozensis</i> Bridson	<i>C. kivuensis</i> Lebrun	<i>C. tricalysioides</i> J.-F.Leroy
<i>C. lulandoensis</i> Bridson	<i>C. ligustroides</i> S.Moore	Dados insuficientes
Madagascar	<i>C. mongensis</i> Bridson	Madagascar
<i>C. andrambovatensis</i> J.-F.Leroy	<i>C. montekupensis</i> Stoff.	<i>C. bissetiae</i> A.P.Davis & Rakotonas., ined.
<i>C. boinensis</i> A.P.Davis & Rakotonas., ined.	<i>C. pseudozanguebariae</i> Bridson	<i>C. minutiflora</i> A.P.Davis & Rakotonas.
<i>C. gallienii</i> Dubard	<i>C. schliebenii</i> Bridson	Risco não avaliado
<i>C. littoralis</i> A.P.Davis & Rakotonas.	<i>C. togoensis</i> A.Chev.	Africa
<i>C. montis-sacri</i> A.P.Davis	<i>C. zanguebariae</i> Lour.	<i>C. affinis</i> De Wild.
<i>C. pterocarpa</i> A.P.Davis & Rakotonas., ined.	Madagascar	Madagascar
<i>C. rakotonasoloi</i> A.P.Davis	<i>C. bertrandii</i> A.Chev.	<i>C. fragilis</i> J.-F.Leroy
Ameaçado de Extinção	<i>C. coursiana</i> J.-F.Leroy	
Africa	<i>C. farafanganensis</i> J.-F.Leroy	
<i>C. bakossii</i> Cheek & Bridson	<i>C. heimii</i> J.-F.Leroy	
<i>C. bridsoniae</i> A.P.Davis & Mvungi	<i>C. mangoroensis</i> Portères	
<i>C. carrissoi</i> A.Chev.	<i>C. pervilleana</i> (Baill.) Drake	
<i>C. leonimontana</i> Stoff.	<i>C. sakarahae</i> J.-F.Leroy	
<i>C. mapiana</i> Sonké, Nguembou & A.P.Davis	<i>C. tetragona</i> Jum. & H.Perrier	
<i>C. pocsii</i> Bridson	Mascarenes	
Madagascar	<i>C. macrocarpa</i> A.Rich.	
<i>C. abbayesii</i> J.-F.Leroy	<i>C. mauritiana</i> Lam.	
<i>C. alleizettii</i> Dubard	Baixo risco de ameaça	
<i>C. ambanjensis</i> J.-F.Leroy	Africa	
<i>C. ambongensis</i> J.-F.Leroy ex A.P.Davis & Rakotonas., ined.	<i>C. humilis</i> A.Chev.	
<i>C. ankaranensis</i> J.-F.Leroy ex A.P.Davis & Rakotonas.	<i>C. magnistipula</i> Stoff. & Robbr.	
<i>C. augagneurii</i> Dubard	<i>C. racemosa</i> Lour.	
<i>C. betamponensis</i> Portères & J.-F.Leroy	<i>C. rhamnifolia</i> (Chiov.) Bridson	
<i>C. bonnieri</i> Dubard	<i>C. salvatrix</i> Swynn. & Philipson	
<i>C. commersoniana</i> (Baill.) A.Chev.	<i>C. sessiliflora</i> Bridson	
<i>C. decaryana</i> J.-F.Leroy	Madagascar	
<i>C. humbertii</i> J.-F.Leroy	<i>C. arenesiana</i> J.-F.Leroy	
<i>C. humblotiana</i> Baill.	<i>C. boiviniana</i> (Baill.) Drake	
<i>C. jumellei</i> J.-F.Leroy	<i>C. buxifolia</i> A.Chev.	
<i>C. kianjavatensis</i> J.-F.Leroy	<i>C. lancifolia</i> A.Chev.	
<i>C. labatii</i> A.P.Davis & Rakotonas., ined.	<i>C. leroyi</i> A.P.Davis	
<i>C. liaudii</i> J.-F.Leroy ex A.P.Davis	<i>C. resinosa</i> (Hook.f.) Radlk.	
<i>C. manombensis</i> A.P.Davis	<i>C. richardii</i> J.-F.Leroy	
<i>C. mcphersonii</i> A.P.Davis & Rakotonas.	Pouco preocupante	
<i>C. mogenetii</i> Dubard	Africa	
<i>C. moratii</i> J.-F.Leroy ex A.P.Davis & Rakotonas.	<i>C. brevipes</i> Hiern	
<i>C. ratsimamangae</i> J.-F.Leroy ex A.P.Davis & Rakotonas.	<i>C. canephora</i> Pierre ex A.Froehner	
<i>C. sahafaryensis</i> J.-F.Leroy	<i>C. congensis</i> A.Froehner	
<i>C. sambavensis</i> J.-F.Leroy ex A.P.Davis & Rakotonas.	<i>C. eugenioides</i> S.Moore	
<i>C. tsirananae</i> J.-F.Leroy	<i>C. liberica</i> Bull. ex Hiern	
<i>C. vatovavyensis</i> J.-F.Leroy	<i>C. mayombensis</i> A.Chev.	
<i>C. vavateninensis</i> J.-F.Leroy	<i>C. mufindiensis</i> Hutch. ex Bridson	
<i>C. vianneyi</i> J.-F.Leroy	<i>C. stenophylla</i> G.Don	
<i>C. vohemarensis</i> A.P.Davis & Rakotonas.		
Mascarenes		
<i>Coffea myrtifolia</i> (A.Rich. ex DC.) J.-F.Leroy		

Adaptado de Davis et al., 2006

3.1. *Coffea liberica* Bull. ex Hiern.

Essa espécie foi inicialmente encontrada na África Ocidental, em 1792, no entanto, seus ecótipos centro-africanos só foram descobertos no início do século XX (CHEVALIER, 1929). O seu cultivo iniciou-se na África antes da chegada dos colonizadores europeus. O vigor do café libérica e sua aparente resistência à ferrugem promoveram sua expansão na Indonésia entre 1880 e 1905. Essa espécie teve grande importância econômica no período de 1930 a 1950, posteriormente foi substituído por *C. canephora*. Atualmente, responde por menos de 1% do mercado mundial de cafés e é cultivado em pequena escala na Malásia e no ocidente da África. Juntamente com *C. canephora*, apresenta uma das distribuições naturais mais amplas do gênero, ocorrendo praticamente em toda extensão da floresta tropical africana. Além disso, apresenta grande variabilidade em nível molecular, morfológico e agrônômico (LEBRUN, 1941; BERTHAUD e CHARRIER, 1985; N'DIAYE et al., 2005; DAVIS et al., 2006). A espécie é geneticamente estruturada de modo que ocorrem duas populações altamente diferenciadas no oeste e na região central do continente africano (BERTHAUD e CHARRIER, 1985). O notável polimorfismo levou a sugestão de um grande número de espécies, subespécies e variedades, que foram organizadas sob a sinonímia de *Coffea liberica*, com duas variedades: *C. liberica* var. *liberica* e *C. liberica* var. *dewevrei* (LEBRUN, 1941). Assim, cabe esclarecer que os termos: “Café libérica” ou “liberiano” referem-se a *C. liberica* var. *liberica*, enquanto o termo “Café excelsa”, refere-se a *C. liberica* var. *dewevrei* (DAVIS et al., 2006).

3.2. *Coffea arabica* L.

Devido à qualidade de bebida superior, *C. arabica* responde por 60-70 % do café consumido no mundo. As plantas da espécie *C. arabica* são autógamas, alotetraplóides verdadeiras ($2n=4X=44$) e seus cultivares comerciais apresentam pequena variabilidade genética. O centro primário de diversidade genética da espécie são as terras altas do sudoeste da Etiópia e o Sudão. Durante o século XVII, o consumo de café se espalhou rapidamente pelo continente europeu, produzindo grandes lucros para o Iêmen, único país produtor à época. Algumas sementes foram coletadas ilegalmente naquele país, por exploradores holandeses e foram plantadas na ilha de Java. Suas progênies foram levadas ao Suriname, que assim como Java, estava sob domínio da Holanda. Daí, uma sucessão de eventos ilícitos, envolvendo

roubo de sementes e mudas, possibilitou a introdução do café na Guiana Francesa e de lá para o Brasil (BERTHAUD e CHARRIER, 1985). O café arábica foi introduzido no Brasil em 1727, no Estado do Pará, por meio de sementes trazidas da Guiana Francesa, pelo sargento mor Francisco de Melo Palheta (MARTINS, 2008). No entanto, as primeiras plantações comerciais foram estabelecidas no Vale do Paraíba por volta de 1761. Nos anos de 1852 e 1896, ocorreram mais dois ciclos de introdução de genótipos provenientes da Ilha de Reunião e da Ilha de Sumatra, respectivamente (CARVALHO, 1993). Nessas ocasiões, foram introduzidas as cultivares Bourbon e Sumatra que tiveram grande importância para a cafeicultura brasileira, por serem as genitoras da cultivar Mundo Novo, que por sua vez, foi a genitora masculina na síntese das cultivares Catuaí Vermelho e Catuaí Amarelo, obtidas pelo Instituto Agrônomo de Campinas, em 1949 e que atualmente, ocupam cerca de 80% da área plantada com café no Brasil.

3.3. *Coffea canephora* Pierre ex A. Froehner

O café canéfora, ou café robusta, é uma espécie diplóide ($2n=2x=22$), estritamente alógama, que apresenta sistema de auto-incompatibilidade, como a maioria das outras espécies diplóides do gênero *Coffea*. Até o presente, as únicas espécies autoférteis conhecidas no gênero são: *C. arabica* e as espécies diplóides *C. anthonyi* e *C. heterocalix* (MAURIN et al., 2007).

3.3.1. Auto-incompatibilidade de *C. canephora*

A auto-incompatibilidade consiste na impossibilidade das flores de uma planta ser fecundadas pelo seu próprio pólen. Esse fenômeno é relativamente comum entre as angiospermas e funciona como um sistema eficaz de manutenção da variabilidade genética da espécie. Esta incompatibilidade pode ocorrer entre plantas diferentes, quando elas possuem alelos em comum do sistema de incompatibilidade. Nas angiospermas, dois tipos principais de incompatibilidade são conhecidos: a incompatibilidade esporofítica e a gametofítica. No sistema esporofítico, os grãos de pólen não germinam nos estigmas das plantas que possuem os mesmos alelos de incompatibilidade. Todos os grãos de pólen apresentam a mesma reação, determinada pelo genótipo da planta fornecedora do pólen. No sistema gametofítico, a incompatibilidade depende do genótipo haplóide do grão de pólen. Nas espécies

diplóides do gênero *Coffea* a auto-incompatibilidade é do tipo gametofítica, monogênica e está associada a um loco gênico “S” com múltiplos alelos (BERTHAUD, 1980).

3.3.2. Distribuição natural e estrutura populacional de *C. canephora*

Em condições naturais, as subpopulações de *C. canephora* geralmente são formadas por um pequeno grupo de plantas matrizes, com poucas progênes espalhadas por áreas de tamanho limitado (cerca de 1 ha). O fluxo gênico interpopulacional é baixo, uma vez que a dispersão de pólen, embora possa alcançar um raio de alguns quilômetros, geralmente ocorre dentro dos limites das subpopulações. Por outro lado, a disseminação das sementes, que é realizada pelas aves e mamíferos, pode atingir maiores distâncias (BERTHAUD, 1986).

Avaliações fenotípicas, bioquímicas e moleculares têm sido empregadas para estudo da diversidade genética e da estrutura populacional de *C. canephora*, em populações naturais e nas coleções de germoplasma (BERTHAUD, 1986; MONTAGNON et al., 1998a; DUSSERT et al., 1999; CUBRY, 2009). Esses estudos convergem para a existência de dois grupos: o Guineano, que compreende os genótipos do oeste africano (Guiné e Costa do Marfim), de folhas menores, menor vigor, menor porte, frutos pequenos, bebida de qualidade inferior, tolerantes à seca e suscetíveis à ferrugem (*Hemileia vastatrix* Berk. et Br.); e o Congolês, composto por genótipos da região central da África, divididos em quatro subgrupos: SG1, SG2, B e C. Recentemente, um novo subgrupo, composto por acessos selvagens de Uganda foi proposto (MUSOLI et al., 2010)

O SG1 reúne os genótipos chamados de Kouillous (Conilons, no Brasil) que ocorrem do Benin ao Gabão e apresentam características adaptativas semelhantes, em parte, àsquelas do grupo Guineano. Os subgrupos SG2, B e C compreendem os genótipos do tipo Robusta, que consistem de plantas mais altas, vigorosas, de folhas e frutos maiores, com melhor qualidade de bebida, maior resistência à ferrugem e maior sensibilidade à seca (CUBRY, 2009). No processo de melhoramento genético da espécie, o grupo guineano ficou praticamente ausente, mantendo-se restrito à sua área de origem. De fato, até recentemente, Costa do Marfim e Guiné eram os únicos países que possuíam populações cultivadas e selvagens do grupo Guineano (MONTAGNON et al., 1998b).

A estrutura populacional descrita anteriormente está fortemente relacionada com o isolamento geográfico e aos eventos históricos que remetem às últimas glaciações, ocorridas há 18 mil anos. Na natureza, os grupos Congolês e Guineano encontram-se separados pelo intervalo Dahomey, que compreende uma estreita faixa de terras áridas (cerca de 300 km de largura), situada entre os blocos de floresta do centro e do oeste africano (MAURIN et al., 2007). Um padrão semelhante de diferenciação é observado na espécie *C. liberica*, na qual se verifica, inclusive, redução da fertilidade dos híbridos obtidos entre os genótipos do oeste (*C. liberica* var. *liberica*) e do centro da África (*liberica* var. *dewevrei*). Outras espécies de plantas e animais também apresentam um curso evolucionário similar (GOMEZ et al., 2009). Essa coincidência levou a formulação da “Teoria do Refúgio”, na qual, postula-se que em determinados períodos geológicos, a distribuição da floresta africana não foi estável, ocorrendo sucessivos eventos de expansão e retração da sua área. Durante a fase de expansão, a recolonização ocorreria a partir de áreas propícias, chamadas de ‘refúgios’ onde a floresta conseguiu sobreviver durante as fases desfavoráveis. Uma vez isoladas em seus refúgios, essas populações deixaram de compartilhar novos eventos de mutação e recombinação (BERTHAUD e CHARRIER, 1985). A influência de forças evolutivas que atuam sob condições ambientais distintas promove a fixação de combinações alélicas particulares, que conferem vantagens adaptativas aos indivíduos de cada população levando-as à diferenciação (GOMEZ et al., 2009).

3.3.3. Recursos genéticos de *C. canephora*

A grande variabilidade observada entre e dentro das populações naturais de *C. canephora* faz desta espécie uma magnífica fonte de alelos para o desenvolvimento de novas cultivares, bem como, para o melhoramento de *C. arabica*. Considerando seu forte caráter alogâmico, acrescente-se, o importante papel das lavouras primitivas realizadas no continente africano, haja vista que, inicialmente, muitas populações selvagens de *C. canephora* foram submetidas diretamente ao cultivo e o intercâmbio de material genético entre as regiões produtoras era bastante intenso (BERTHAUD e CHARRIER, 1985). Desse modo, essas lavouras tornaram-se locais favoráveis ao inter cruzamento de diferentes tipos varietais, propiciando a recombinação e o surgimento de genótipos que não ocorreriam na natureza.

Todavia, o potencial genético de *C. canephora* ainda é subutilizado. Por exemplo, o grupo guineano, que possui alelos de tolerância à seca, ficou restrito à sua área de origem (Guiné e Costa do Marfim), não havendo registro de sua introdução nos programas de melhoramento de outros países (MONTAGNON et al., 1998b). Além disso, até recentemente, acessos representativos desse grupo encontravam-se preservados apenas na Costa do Marfim. Por sua vez, o grupo congolês, apresenta maior variabilidade e foi bastante difundido entre os países produtores. No entanto, a maioria das introduções, provavelmente, foi feita a partir de sementes e mudas oriundas de um número limitado de plantas. Acredita-se que o estoque de Robusta introduzido em Java em 1901, proveio de plantas cultivadas no Zaire. O acervo de Java foi, posteriormente, incrementado com acessos do Gabão (kouillou) e Uganda (CHARRIER e BERTHAUD, 1985). O material selecionado em Java foi então, reintroduzido na África de modo que a base genética da cafeicultura de muitos países africanos era muito similar (DUSSERT et al., 1999).

No caso do Brasil, onde a espécie foi introduzida no início do século passado (FERRÃO et al., 2007a; FAZUOLI et al., 2009), houve maior expansão do SG1, representado pelo tipo varietal ‘Conilon’ (adaptação do termo ‘Kouillou’). O SG2, que reúne as progênies do tipo varietal ‘Robusta’, apresenta-se precariamente representado nas lavouras brasileiras, embora seja portador de inúmeros alelos de interesse, sobretudo com relação à resistência à ferrugem, fusariose (*Fusarium xylarioides*), CBD (*Colletotrichum kahawae*), nematóides e bicho mineiro; maior tamanho de grãos e menor teor de cafeína. Ademais, os acessos mantidos nas coleções nacionais de germoplasma são progênies de um número relativamente pequeno de indivíduos e advêm de outras coleções estabelecidas, sobretudo, na Indonésia e na Costa Rica (FAZUOLI et al., 2009).

A espécie *C. canephora* apresenta ampla variabilidade e tem uma das distribuições geográficas mais amplas dentro do subgênero *Coffea*, podendo ser encontrada em populações naturais em toda a faixa centro-ocidental do continente africano, desde a Guiné até Angola (com exceção do intervalo Dahomey, que é uma faixa de savana entre o Togo e Benin) (MAURIN et al., 2007).

O germoplasma de *C. canephora* encontra-se conservado em coleções *ex situ* em diferentes países, como: Costa do Marfim, Camarões, Uganda, Índia, Indonésia e Brasil (BERTHAUD e CHARRIER, 1985).

3.3.4. Melhoramento genético de *C. canephora*

Segundo Sakiyama et al. (2005), os programas de melhoramento genético do café arábica têm sido direcionados principalmente para obtenção de cultivares resistentes a doenças e pragas; cultivares com produção elevada e estável e com precocidade da primeira colheita; cultivares com adaptação específica a diferentes regiões cafeeiras; cultivares adaptadas ao cultivo sob irrigação; cultivares tolerantes à geadas, à seca e a alumínio tóxico; cultivares de porte baixo e formato de copa adequado para a colheita mecanizada; cultivares com maturações uniformes, precoces, médias e tardias; e cultivares de boa qualidade de bebida. Os principais métodos utilizados no melhoramento genético do café arábica no Brasil têm sido: introdução, seleção de plantas individuais com teste de progênie, método genealógico, retrocruzamentos e, em menor escala, os métodos de seleção recorrente e SSD (*Single Seed Descent*). As cultivares de arábica compõem-se basicamente de linhagens endogâmicas, geralmente lançadas para cultivo comercial a partir da geração F₄, ou de multilinhas. O avanço no desenvolvimento de protocolos eficientes para propagação de mudas clonais de *C. arabica* a um baixo custo, também poderá viabilizar o uso de híbridos em geração F₁.

Diferenças biológicas relevantes entre *C. canephora* e *C. arabica* pressupõem o estabelecimento de estratégias de melhoramento bastante distintas para cada uma dessas espécies. No caso do café canéfora, o melhoramento genético tem sido focado principalmente no aumento da produtividade, na resistência à seca, na uniformidade de maturação e na resistência a doenças (FERRÃO et al. 2007a). Outras características desejáveis como o aumento do tamanho de grãos, a estabilidade de produção, a melhoria da qualidade de bebida, redução do teor de cafeína, também têm sido incorporadas como objetivo dos programas de melhoramento (CHARRIER e BERTHAUD, 1985; SOUZA, 2003; FERRÃO et al. 2007a). Uma mutação que determinasse a redução do porte em *C. canephora*, a exemplo daquela que deu origem às cultivares Caturra, San Ramón, San Bernardo e Villa Sachi em *C. arabica* seria de grande interesse para o melhoramento da espécie. Relatos da ocorrência de plantas anãs de *C. canephora* no germoplasma indiano foram apresentados (KUMAR et al. 1994). Além disso, também seriam interessantes polimorfismos que

conferissem resistência à broca do café e às cochonilhas; e que facilitasse o desprendimento dos frutos maduros, viabilizando o emprego da colheita mecanizada.

O melhoramento genético de *C. canephora* teve início ainda por volta de 1900, quando as primeiras plantas foram levadas da África para ilha asiática de Java, que à época encontrava-se sob domínio holandês. Estas plantas pertenciam ao tipo Robusta e foram coletadas a partir de populações selvagens sob cultivo no Zaire, em 1895. Posteriormente, a coleção javanesa foi enriquecida com introduções do Gabão (Kouillou) e Uganda. Portanto, as bases biológicas e os fundamentos para o melhoramento de *C. canephora* foram estabelecidos pelos pesquisadores holandeses que trabalharam em Java, até 1940 (CHARRIER e BERTHAUD, 1985). Todavia, o cultivo rudimentar de *C. canephora* já era praticado na África, antes da chegada dos colonizadores europeus, sobretudo em Uganda e na parte oriental da bacia do rio Congo (MONTAGNON et al. 1998a). O declínio das lavouras de café arábica, devastadas pelo ataque da ferrugem, provocada pelo fungo *Hemileia vastatrix*, promoveu o interesse por uma dezena de espécies diplóides do gênero *Coffea*, entre as quais se destacaram *C. liberica* e *C. canephora*. Em especial, uma subpopulação de *C. canephora*, denominada Robusta, prontamente se adaptou às condições de cultivo nas florestas tropicais úmidas (CHARRIER e BERTHAUD, 1985).

Sendo uma espécie alógama, de fácil propagação vegetativa, que apresenta germoplasma geneticamente estruturado em populações polimórficas, compondo grupos heteróticos bem definidos, com indivíduos altamente heterozigotos, *C. canephora* se adéqua perfeitamente à estratégia de seleção recorrente recíproca interpopulacional (RESENDE e BARBOSA, 2005). Na prática, mesmo fora desse contexto, diferentes técnicas de melhoramento vêm sendo utilizadas com sucesso. Dentre elas, podem ser destacadas as seguintes (CHARRIER e BERTHAUD, 1985; MONTAGNON et al. 1998a; FERRÃO, et al., 1999):

- a) Introdução de germoplasma e seleção adaptativa;
- b) Seleção massal;
- c) Seleção entre e dentro de progênies de meio-irmãos;
- d) Seleção clonal;
- e) Hibridação:
 - Capacidade geral de combinação
 - Capacidade específica de combinação
 - Híbridos intra-específicos entre os grupos Guineano e Congolês e entre os tipos varietais Conilon ou Robusta, dentro do grupo Congolês

- Híbridos intra-específicos entre cafeeiros dos tipos varietais Conilon ou Robusta
- Híbridos inter-específicos entre cafeeiros de *C. canephora* e as espécies *C. congensis*, *C. liberica* e *C. eugenioides* e *C. arabica*
- f) Haplodiploidização
- g) Método do retrocruzamento
- h) Seleção recorrente
- i) Seleção recorrente recíproca

Após a introdução de *C. canephora* em Java, rapidamente, a cultura se espalhou pela África, Ásia e América, onde surgiram vários programas de melhoramento, em diversos países. Na fase inicial, quando da introdução do germoplasma, os pesquisadores, lançaram mão do melhoramento e seleção dentro de populações. Em seguida, passou-se ao melhoramento e seleção em populações sintéticas, com a obtenção de progênies para exploração de clones superiores (MONTAGNON et al. 1998a). Em alguns países, como a Costa do Marfim, e mais recentemente no Brasil, tem-se buscado obter híbridos superiores dentro de programas de seleção recorrente recíproca, explorando a diversidade existente entre os grupos heteróticos já identificados na espécie (LEROY et al. 1993, 1994, 1997; SOUZA e SANTOS, 2009; FAZUOLI et al., 2009).

As cultivares de café canéfora compõem-se basicamente de variedades clonais e variedades sintéticas. Além disso, o uso de híbridos biclonais em cultivos comerciais tem ganhado força, sobretudo no continente africano e pode vir a ser uma boa alternativa para as regiões brasileiras, onde a adoção de variedades clonais apresenta baixa viabilidade técnica (MONTAGNON et al., 1998b; SOUZA e SANTOS, 2009). Por guardarem grande semelhança com *C. canephora*, os híbridos interespecíficos entre esta espécie e *C. congensis* também podem ser considerados como uma modalidade de cultivar de canéfora. Esses híbridos são conhecidos como ‘Congustas’, ‘Conuga’ (Indonésia) (CHARRIER e BERTHAUD, 1985) e híbrido ‘CxR’, (Índia) (BATH et al., 2005)

No continente asiático, merecem destaque os programas de melhoramento desenvolvidos em Java, na Índia, Indonésia e Malásia. As primeiras ações de melhoramento de *C. canephora* foram registradas em Java (CHARRIER e BERTHAUD, 1985). Inicialmente, introduziu-se germoplasma do tipo Robusta a partir do Zaire (atualmente República Democrática do Congo), Uganda e Gabão.

Após alguns ciclos de seleção massal, algumas plantas matrizes foram selecionadas para produção de sementes e cerca de 15 clones foram obtidos e disponibilizados aos produtores para cultivo comercial. Híbridos interespecíficos naturais entre *C. congensis* e *C. canephora* foram identificados, selecionados e dirigidos ao cultivo comercial sob a denominação de ‘Congusta’. Assim como os clones de Robusta, esse material teve boa aceitação entre os cafeicultores javaneses (MONTAGNON et al., 1998b).

Na Índia, *C. canephora* foi introduzido entre os anos de 1900 e 1905, a partir de sementes advindas de Java. Posteriormente, ocorreram novas introduções (1950 a 1970) de acessos de Robusta, provenientes da Costa do Marfim, Vietnã, Uganda, Madagascar, Indonésia e Costa Rica. O emprego da seleção massal sobre esse germoplasma permitiu a obtenção de clones que se tornaram muito populares, entre os agricultores, como o ‘Robusta Balehonnur’. O cultivo comercial do híbrido interespecífico ‘CxR’ (Congensis x Robusta) também foi muito bem sucedido na Índia (MONTAGNON et al. 1998b).

Na Indonésia, as primeiras seleções para produtividade em *C. canephora* foram realizadas em progênies de polinização aberta resultantes do intercruzamento de clones selecionados. Verificou-se que os clones BP39 e BP42 originaram progênies mais produtivas e de adaptação ampla. Sementes híbridas produzidas em jardins biclonais destes dois genótipos foram recomendadas no final da década de 30. A seleção massal com testes clonais também foi realizada, resultando na recomendação de variedades policlonais. O melhoramento de canéfora foi interrompido naquele país durante a segunda guerra mundial, no entanto, os materiais recomendados antes da guerra ainda continuam em uso pelos agricultores (MONTAGNON et al., 1998b).

No Vietnã, o tipo varietal Robusta foi introduzido pelos colonizadores franceses no início do século passado e se adaptou muito bem às condições ambientais locais. Com a guerra e o estabelecimento de sanções ao governo comunista houve uma queda brusca das exportações de café, desestimulando o seu cultivo. A partir dos anos 80, com a reabertura do mercado, a produção cresceu rapidamente, de modo que o Vietnã é, atualmente, o maior produtor mundial de café canéfora. Cerca de 3% da área vietnamita de café é dedicada a outras variedades, como Catimor (*C. arabica*) e ‘Chari’ (*C. liberica* var. dewevrei) (Vietnamese Coffee Online. Disponível no <http://www.vietnamese-coffee.com>).

Na década de 60, do século passado, a França promoveu a criação de um grande programa de melhoramento genético em suas colônias na África, sobretudo em Madagascar, Costa do Marfim e República Centro-Africana. Esse programa compreendia duas frentes principais baseadas na “seleção vegetativa” e na “seleção generativa”, com vistas à obtenção de clones e híbridos superiores, respectivamente. Os clones obtidos por meio da primeira estratégia foram bem sucedidos nos três países, no entanto, os híbridos foram bem aceitos apenas na Costa do Marfim (MONTAGNON et al., 1998b; CAPOT, 1977).

O programa de melhoramento genético de café canéfora na Costa do Marfim foi um dos mais dinâmicos e arrojados que se tem notícia. O germoplasma conservado nas coleções marfinenses ainda é um dos mais diversificados e inclui acessos coletados em plantações locais e em populações naturais de vários países localizados sobre a área de diversidade do gênero *Coffea*, como República Democrática do Congo, Madagascar, Camarões e Uganda. A seleção clonal possibilitou a identificação de sete clones elites e a seleção generativa promoveu a obtenção de inúmeros híbridos superiores. No entanto, esses híbridos produziam cerca de 35% menos que as variedades clonais e foram utilizados comercialmente apenas em áreas marginais. O uso da seleção recorrente na Costa do Marfim, possibilitou a obtenção de 200 clones e 40 progênies híbridas inter-grupos que apresentaram desempenho superior aos clones mais produtivos em cultivo naquele país (MONTAGNON et al., 1998b; LEROY et al., 1997, 1994, 1993; CAPOT, 1977).

Em Madagascar, o germoplasma inicialmente utilizado no melhoramento foi composto por sementes advindas de Java e consistiam de híbridos interespecíficos entre *C. canephora* e *C. congensis*, além de vários acessos dos tipos varietais Robusta e Kouillou de origem incerta. Em 1962, 18 clones foram disponibilizados para os produtores e no final da década de 70, seis híbridos também foram entregues para cultivo comercial. A técnica de seleção clonal também foi empregada nos híbridos ‘Congustas’, que passaram a ser propagados vegetativamente (MONTAGNON et al., 1998b).

Na República Centro-Africana, as sementes utilizadas nas primeiras seleções foram trazidas do Zaire e compunham-se basicamente do tipo varietal Robusta. Posteriormente, outros materiais foram introduzidos, mas não foram utilizados no programa de melhoramento daquele país, pois o seu desempenho ficou muito aquém

dos Robustas. No final da década de 60, 30 clones elites estavam prontos para serem disponibilizados aos produtores (MONTAGNON et al., 1998b).

O programa de melhoramento do café canéfora em Camarões teve início com a introdução de clones elites, oriundos dos programas de melhoramento de outros países. Assim, foram selecionados para cultivo local cinco clones da República Centro-africana, três de Java, dois da Costa do Marfim, um de Madagascar e um do Zaire. Na sequência, iniciou-se um programa baseado na obtenção de híbridos entre clones, cujos parentais eram escolhidos com base no resultado de cruzamentos dialélicos. Esta estratégia levou a obtenção de híbridos que atingiam até 75% da produtividade dos melhores clones (BOUHARMONT et al., 1979; 1986).

Em Uganda, *C. canephora* já era cultivado antes da chegada dos colonizadores europeus, portanto, não devem ser menosprezadas as iniciativas dos agricultores locais quando da escolha das plantas mais produtivas para composição de novas lavouras. Todavia, o melhoramento da espécie naquele país iniciou-se oficialmente em 1916, a partir de sementes vindas de Java, de clones de várias origens e de sementes colhidas em populações selvagens locais. O germoplasma era dividido em dois grupos fenotípicos: ereto e pendente. Utilizou-se a seleção recorrente intrapopulacional, promovendo-se a clonagem e recombinação das plantas superiores em cada grupo. Essa estratégia possibilitou a obtenção de cinco clones pendentes e três eretos (MONTAGNON et al., 1998b).

Nas Américas, predomina o cultivo do café arábica, sendo que, a produção de café canéfora em larga escala limita-se praticamente ao Brasil. No México, cinco seleções clonais foram feitas em progênies oriundas da Costa Rica e foram disponibilizadas aos produtores na forma de uma variedade policlonal denominada 'Romex' (MONTAGNON et al., 1998b). Um fato interessante sobre a Costa Rica é que embora possua uma das principais coleções de *C. canephora* das Américas, esta espécie não é cultivada naquele país.

No Brasil, o primeiro método utilizado foi a introdução de germoplasma, que ocorreu por volta de 1908, pouco tempo depois do início do cultivo comercial do café canéfora em Java. As primeiras sementes eram do grupo Kouillou (que no Brasil foi chamado de Conilon) e foram cultivadas no sudeste brasileiro, principalmente no Estado do Espírito Santo (FERRÃO et al., 2007a). A introdução de genótipos do tipo varietal Robusta aconteceu posteriormente, ainda na segunda década do século passado, por meio de acessos coletados na Indonésia e incorporados ao acervo do

Horto Florestal da Companhia Paulista de Estradas de Ferro. Nos anos de 1971 e 1974 ocorreram novos eventos de introdução, dessa vez, a partir de seleções realizadas na coleção de germoplasma da Costa Rica (FAZUOLI et al, 2009). A princípio, os acessos ficaram restritos às instituições de pesquisa. Na década de 80, o IAC disponibilizou algumas populações de Robusta para cultivo comercial, entre elas as progênies obtidas por meio de seleção massal, para maior tamanho de fruto, denominada de ‘Guarini IAC 1955’ e para resistência a nematóides, chamada de ‘Apoatã IAC 2258’. Esta última foi amplamente utilizada no Estado do Paraná como porta-enxertos (FAZUOLI, 1986). Embora tenham características importantíssimas, como alta produção, resistência a doenças e melhor qualidade de grãos, essas cultivares têm participação inexpressiva na cafeicultura nacional.

Nos últimos anos, a seleção massal com testes clonais possibilitou ao Incaper disponibilizar uma série de cinco cultivares policlonais do tipo varietal Conilon e uma variedade sintética, denominada Robustão Capixaba, tolerante à seca. As variedades policlonais apresentam alto rendimento, têm épocas de maturação diferenciadas, permitindo o escalonamento da colheita (FERRÃO et al., 2007a). Esses genótipos tiveram ampla aceitação no Espírito Santo e também têm sido plantados em outros Estados.

O sucesso das cultivares de Conilon, em detrimento das cultivares de Robusta, se deve basicamente à sua tolerância à seca e à facilidade de manejo, aliadas ao contexto histórico-cultural que foi criado pela ausência de plantas do tipo Robusta na fase de implantação das primeiras lavouras de canéfora no Brasil. A produtividade não é fator determinante, pois os ensaios de competição que têm sido realizados apontam para uma equivalência na produtividade de ambos os tipos varietais (VENEZIANO, 2003).

Como se vê, a seleção massal com testes clonais ainda é a técnica prevalecente nos programas de melhoramento de canéfora na maioria dos países produtores, pois a propagação vegetativa possibilita o aproveitamento integral da variância genética e possibilita a fixação de caracteres de interesse. Essas variedades também têm sido as preferidas pelos agricultores. No entanto, esse tipo de variedade requer um cuidado especial dos melhoristas, que devem se preocupar com a incorporação frequente e contínua de novos clones elites no sistema produtivo, a fim de minimizar os riscos de estreitamento da base genética da cultura.

Os híbridos biconais, policlonais e as variedades sintéticas, embora, possam apresentar-se menos produtivos e mais heterogêneos, também têm o seu espaço assegurado, pois outro aspecto que deve ser levado em consideração é o fato de que a propagação de variedades por sementes é mais fácil, prática e econômica do que aquela realizada por meio de clones. Calcula-se que um campo de sementes de 1,0 ha produza mais de 1000 kg de sementes viáveis, o que seria suficiente para implantação de 2000 ha. Por outro lado, um jardim clonal de área equivalente produz cerca de dois milhões de clones, o que daria para o plantio de 1000 ha (MONTAGNON et al., 1998a). Além disso, em comparação com as cultivares propagadas por sementes, as clonais exigem um melhor manejo da lavoura, sobretudo, quanto ao controle fitossanitário e à disponibilidade de água e de nutrientes no solo, haja vista que o seu sistema radicular tende a ser menos desenvolvido. A melhoria das condições de manejo pode ser uma condição que não está imediatamente ao alcance de considerável parcela dos cafeicultores brasileiros. Portanto, para esses, o uso de variedades propagadas por sementes, como é o caso dos híbridos de clones e das variedades sintéticas, parece ser ainda a melhor alternativa (SOUZA e SANTOS, 2009).

Especialmente, os híbridos biconais têm grandes chances de ser mais bem aproveitados, mesmo nos sistemas produtivos mais tecnificados, uma vez que apresentam relativa uniformidade e o seu potencial produtivo tende a crescer rapidamente com o emprego de cruzamentos mais heteróticos, que aproveitem a diversidade entre os grupos (Congolês x Guineano) ou subgrupos (Kouillou x Robusta), nos quais a espécie encontra-se estruturada. Em condições experimentais, alguns desses híbridos têm superado os clones mais produtivos (SOUZA e SANTOS, 2009; FAZUOLI et al., 2009).

Os principais programas brasileiros de melhoramento genético de café canéfora são conduzidos, pelo IAC, em São Paulo, e pelo INCAPER, no Espírito Santo. Além desses, merece destaque o programa de melhoramento da Embrapa, em Rondônia. Este programa teve início na década de 70, com a introdução e avaliação de diversas progênies de Conilon e Robusta, em Rondônia (SOUZA et al., 2003). Entre os anos de 1983 e 1990, foram avaliadas 18 progênies de *Coffea canephora*, das variedades Conilon, Robusta, Apoatã e Laurenti. As progênies foram avaliadas quanto à produção, altura de planta, diâmetro de copa, percentagem de frutos com lojas vazias, tipos de sementes (chato, moca e concha), peneira média, peso de sementes

tipo chato, rendimento (peso de café beneficiado/café em coco), qualidade de bebida, teor de sólidos solúveis e teor de cafeína. A avaliação desse germoplasma resultou na recomendação das seguintes progênies: ‘Kouillou IAC 69-5’, ‘Kouillou IAC 66-3’, ‘Robusta IAC 2259’, ‘Robusta IAC 2258-1’, ‘Robusta IAC 1647’ (VENEZIANO, 2003).

Nos últimos anos, a ampla variabilidade observada nas lavouras de café Conilon em Rondônia motivou a realização de coletas anuais de clones de plantas fenotipicamente superiores, identificadas nos plantios comerciais do Estado e no Campo Experimental da Embrapa em Ouro Preto do Oeste. Cerca de 700 clones foram coletados e estão sendo avaliados quanto à produtividade, qualidade de produção, resistência à ferrugem, estabilidade de produção e época de maturação, entre outros (VENEZIANO et al., 2003). Recentemente, técnicas de análise multivariada foram utilizadas para possibilitar o agrupamento dos clones, auxiliando na identificação de materiais elites que possam compor variedades clonais ou serem recombinados para a obtenção de populações superiores, com grande variabilidade genética, que servirão de fonte de germoplasma para o desenvolvimento de futuras variedades (SOUZA, 2003).

Em 1994, a Embrapa iniciou um trabalho em parceria com a Empresa Capixaba de Pesquisa Agropecuária (na época denominada ENCAPA e atualmente INCAPER) com o objetivo de avaliar e selecionar clones de plantas superiores para composição de variedades clonais adaptadas às condições do Estado de Rondônia. Trinta e seis clones da Embrapa Rondônia e 29 do INCAPER foram avaliados quanto à altura de planta, diâmetro da copa, número de ramos plagiotrópicos, época de maturação, percentagem de frutos com lojas vazias, produção e peneira média (VENEZIANO et al., 2003).

4. Marcadores moleculares em café

O desenvolvimento de marcadores moleculares tem possibilitado novas oportunidades no melhoramento genético do cafeeiro. As principais técnicas empregadas têm sido o RFLP (*Restriction Fragment Length Polymorphism*), RAPD (*Random Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphism*) e os microssatélites ou SRRs (*Simple Sequence Repeats*).

A seleção convencional em populações melhoradas de café é trabalhosa e demanda muito tempo, além disso, os efeitos da interação com o ambiente podem levar a resultados insatisfatórios. O emprego de marcadores moleculares na seleção poderia superar a maioria das limitações correntes.

Grande parte dos trabalhos com marcadores moleculares em café estão relacionados com o estudo da diversidade genética do gênero *Coffea* (BERTHAUD, 1986; LASHERMES et al., 1993; DUSSERT et al., 1999; COMBES et al., 2000; ANTHONY et al., 2002; PRAKASH et al., 2005; MALUF et al., 2005; SILVESTRINI, 2008). A identificação de marcadores para seleção assistida também tem sido alvo de pesquisas (NOIR et al., 2003; PRAKASH et al., 2004; MAHÉ et al., 2008, DINIZ et al., 2005; BRITO et al., 2010; DIOLA et al., 2011). A construção de mapas genéticos é outra ferramenta que tem sido proposta para auxiliar no melhoramento de características de interesse nas principais espécies do gênero *Coffea* (KY et al. 1999, 2000; OLIVEIRA et al., 2007; CUBRY, 2009).

5. Diversidade genética

O conhecimento sobre a diversidade genética e o nível de diferenciação genética entre populações das espécies auxilia na definição dos estoques genéticos e no estabelecimento de estratégias de uso e manejo desses recursos (CRUZ, 2005). No caso de *C. canephora*, estudos de diversidade podem facilitar a orientação dos programas de melhoramento na escolha de genitores para cruzamentos ou de genótipos para composição de variedades clonais (SOUZA et al., 2003). Ademais, esses estudos podem minimizar o risco de erosão genética causados pelo desenvolvimento de cultivares superiores, de base genética estreita como é o caso dos clones. Além disso, o conhecimento da diversidade genética poderá auxiliar no manejo dos bancos de germoplasma, oferecendo parâmetros para estabelecimento de tamanho de populações a serem conservadas e eliminação de duplicatas, entre outras aplicações (CRUZ, 2005).

Diferentes técnicas moleculares têm sido utilizadas para avaliar a diversidade genética em *Coffea canephora*. Inicialmente, os estudos com germoplasma selvagem e cultivado na África, utilizando isoenzimas revelaram a existência dos grupos Guineano e Congolês, com uma alta diversidade genética e forte estruturação geográfica (BERTHAUD, 1986). A existência de cinco grupos de diversidade (que

receberam denominações de “A” a “E”) foi revelada mais tarde por meio de marcadores RFLP (DUSSERT et al., 1999). Com advento dos marcadores obtidos por amplificação via PCR (*Polymerase Chain Reaction*), RAPD, AFLP e SSR, verificou-se um número crescente de estudos genéticos com essa espécie. Alguns trabalhos relatam o uso desses marcadores na análise da diversidade e a estrutura genética das populações de *C. canephora* na África (GOMEZ et al., 2009) e na Índia (PRAKASH et al., 2005). Todavia, esta questão ainda é pouco abordada no germoplasma Sul-americano, onde maiores esforços têm sido concentrados na espécie *C. arabica*. O uso de marcadores RAPD demonstrou uma elevada diversidade e estrutura genética na coleção de germoplasma do IAC (SILVESTRINI et al., 2008). Por outro lado, estudos envolvendo marcadores RAPD e SSR nas populações que originaram as variedades clonais do programa de melhoramento genético conduzido no Estado do Espírito Santo apontaram para uma menor estruturação e redução na diversidade (FERRÃO et al., 2007b).

6. Estrutura da população

Em genética, o termo população é utilizado para designar o grupo de indivíduos genotipicamente distintos, com sistema de acasalamento definido, potencialmente aptos à produção de descendentes férteis, em frequência proporcional à contribuição gamética de seus genitores. O conhecimento da estrutura genética de uma população possibilita realizar ou prever mudanças em magnitude e sentido desejados, sendo, portanto útil na condução de programas de melhoramento genético e no gerenciamento de bancos de germoplasma. Além disso, poderá fornecer as informações necessárias para entender como se processa a evolução na população estudada (CRUZ, 2005).

A estrutura populacional de uma espécie é o conjunto de suas características genéticas e demográficas, é resultado da ação e das interações de uma série de mecanismos evolutivos e ecológicos (MARTINS, 1987). Em nível genético, a estrutura de uma população pode ser definida pela frequência dos alelos que compõem os diferentes genótipos dos diferentes indivíduos integrantes da população (CRUZ, 2005).

Um modelo baseado em estatística bayesiana e implementada no *software* STRUCTURE (PRITCHARD et al., 2000) foi utilizado para determinar a estrutura genética de populações de *Coffea arabica* (LOPES-GARTNER et al., 2009) e *Coffea canephora* (GOMÉZ et al., 2009; CUBRY, 2009).

7. Desequilíbrio de ligação e mapeamento associativo

O mapeamento de QTLs pode ser realizado por meio de duas abordagens distintas: com base na ligação genética ou com base no desequilíbrio de ligação (LD). No mapeamento baseado na ligação genética, utilizam-se populações derivadas de cruzamentos controlados, sendo que as mais informativas são aquelas obtidas a partir do cruzamento entre dois indivíduos homocigotos e contrastantes. Nas espécies vegetais autógamas, os tipos mais comuns de populações de mapeamento são as F_2 , os retrocruzamentos, as linhagens endogâmicas recombinantes (RILs) e os duplo-haplóides. No caso das espécies alógamas, utilizam-se cruzamentos exogâmicos como famílias de irmãos completos, famílias de meio-irmãos, ou outras genealogias mais complexas (CRUZ et al., 2009). O mapeamento utilizando famílias pode ser considerado um caso especial de mapeamento baseado no LD, no qual o desequilíbrio é gerado pelo estabelecimento de uma população a partir de um reduzido número de fundadores em um passado recente. Uma população F_2 , por exemplo, é derivada de uma planta F_1 . O processo meiótico e um delineamento experimental apropriado asseguram que a força da correlação entre o marcador e a característica fenotípica de interesse é proporcional à distância genética entre o marcador e o QTL, sendo nula a correlação entre locos não ligados (MACKAY e POWELL, 2007).

A análise de associação, ou mapeamento baseado no desequilíbrio de ligação (LD), tem sido muito utilizada no estudo genético de doenças humanas, como mal de Alzheimer e fibrose cística (FLINT-GARCIA et al., 2003). Esta metodologia tem o potencial de identificar e mapear QTLs; identificar o polimorfismo gênico causador da diferença entre dois fenótipos alternativos, bem como permite a identificação de blocos haplótipos e haplótipos representando diferentes alelos de um gene (GUPTA et al., 2005).

O mapeamento baseado no LD explora a história de recombinações da população e provê oportunidade de mapeamento fino de alta resolução. Por outro

lado, o mapeamento baseado em famílias tem maior poder na detecção de QTLs, de modo, que essas duas abordagens são complementares (MACKAY e POWELL, 2007).

Como já demonstrado, o mapeamento convencional apresenta algumas limitações, como a necessidade de formação de gerações avançadas de populações biparentais (F_2 , RC, RILs), o que requer tempo, com alto custo e menor resolução na detecção de QTLs, uma vez que apenas dois alelos são amostrados simultaneamente em populações biparentais (FLINT-GARCIA et al., 2003). Além disso, a estratégia de mapeamento de QTLs em estudos de ligação, utilizando famílias e *pedigrees*, tem sido limitada aos locos que exercem maior efeito na variação fenotípica da característica quantitativa (RISCH, 2000). O mapeamento associativo pode superar essas limitações. A estratégia pode ser aplicada inclusive a espécies nas quais o mapeamento baseado em famílias e *pedigrees* não é factível devido à impossibilidade de realização de cruzamentos (FERREIRA e GRATAPAGLIA, 2006). Nessa abordagem, podem ser utilizadas as próprias populações de melhoramento ou mesmo acessos de coleções de germoplasma. No entanto, diferentemente das populações humanas e animais, em espécies vegetais sob melhoramento, o acasalamento ao acaso é uma exceção havendo prevalência de mecanismos que levam à endogamia. Além disso, o tamanho efetivo populacional é pequeno e uma forte pressão de seleção é aplicada (STICH et al., 2005).

A ligação genética (*linkage*) é a co-segregação de alelos de locos diferentes pelo fato destes estarem localizados fisicamente próximos, no mesmo cromossomo. O Desequilíbrio de Ligação (DL), também conhecido como desequilíbrio de fase gamética, desequilíbrio gamético e associação alélica, consiste na vinculação não aleatória entre alelos de diferentes locos. Esse conceito pode ser estendido às associações não aleatórias que ocorrem entre dois marcadores moleculares, entre dois genes, entre dois QTLs ou entre um gene (ou QTL) e um loco marcador (GUPTA et al., 2005). Se em uma população, dois alelos de locos distintos são encontrados juntos com frequência maior do que a esperada (nesse caso, a frequência esperada equivale ao produto da frequência dos alelos), diz-se que eles estão em desequilíbrio de ligação (FERREIRA e GRATTAPAGLIA, 2006).

Embora, às vezes, seja confundida com o DL, a ligação genética é apenas uma das causas do desequilíbrio. Associações significativas podem ocorrer entre alelos de locos distantes, ou mesmo, situados em cromossomos diferentes

(ABDURAKHMONOV e ABDUKARIMOV, 2008). Estas associações podem ser devidas a outras causas, como a presença de estrutura na população, seleção, efeito fundador, epistasia, entre outros, e são considerados fatores perturbadores da análise de associação.

Do ponto de vista estatístico, o DL corresponde à covariância dos polimorfismos exibidos entre dois genes/marcadores. Algumas estatísticas foram desenvolvidas para estimá-lo, sendo que as mais utilizadas são o D' e o r^2 . No entanto, o princípio básico por trás dessas estatísticas é a diferença entre a frequência de haplótipos esperados e observados (FLINT-GARCIA et al. 2003). Assim, considerando-se par de locos, com dois alelos em cada (A e a , B e b), com frequências alélicas π_A , π_a , π_B e π_b , o DL entre esses locos pode ser quantificado usando a seguinte expressão:

$$D_{ab} = (\pi_{AB} - \pi_A \pi_B),$$

onde: D_{ab} corresponde ao desequilíbrio de ligação entre os locos A e B e π_{AB} é a frequência do haplótipo AB .

A estatística D' é útil para determinar a ocorrência de recombinação entre pares de alelos e é estimada pela equação:

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_b, \pi_a \pi_B)} \quad \text{se, } D_{ab} < 0$$

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_B, \pi_a \pi_b)} \quad \text{se, } D_{ab} > 0$$

A estatística r^2 é informativa para avaliar a resolução das técnicas de mapeamento associativo e é estimada por meio da equação:

$$r^2 = \frac{(D_{ab})^2}{\pi_A \pi_a \pi_B \pi_b}$$

De modo geral, D' mede apenas as diferenças de recombinação, enquanto r^2 resume o histórico de mutações e recombinações. Além disso, r^2 indica como os marcadores podem estar correlacionados com o QTL de interesse, sendo geralmente, preferido nos estudos de mapeamento associativo. Tanto D' quanto r^2 variam de 0 a 1. Valores próximos de zero evidenciam independência entre os alelos, ou seja,

equilíbrio. Por outro lado, valores que tendem à unidade, sugerem associação, isto é: desequilíbrio (FERREIRA e GRATTAPAGLIA, 2006).

A significância estatística do DL (*P*-valor) entre dois locos pode ser testada usando-se tabelas de contingência 2 x 2 e aplicando-se o teste de χ^2 ou o teste exato de Fisher. Probabilidades menores que 5% indicariam falta de independência entre os locos, sugerindo desse modo que os dois estariam associados (GUPTA et al., 2005). Se mais de dois alelos estão presentes procede-se uma análise de permutação multifatorial, que consiste na realização de permutações para calcular a proporção das distribuições de permutação gamética que são menos prováveis do que a distribuição gamética sob hipótese nula de independência (WEIR, 1996).

O poder estatístico para detectar uma associação significativa, ou seja, a probabilidade de rejeitar os falso-positivos depende da magnitude da associação e da frequência do alelo de interesse na população. Além disso, o tamanho populacional e os padrões de desequilíbrio de ligação no genoma também afetam o poder dos testes estatísticos.

A recombinação é o fenômeno capaz de reduzir a extensão do desequilíbrio de ligação. Diferentemente do genoma humano e de outros mamíferos, o genoma vegetal é heterogêneo quanto à ocorrência de recombinação. Ou seja, ocorrem regiões com altas taxas de recombinação e longas sequências com baixa taxa de recombinação. Geralmente as primeiras são ricas em genes, enquanto as últimas apresentam baixa densidade gênica (FERREIRA e GRATTAPAGLIA, 2006).

O conhecimento a cerca da extensão do desequilíbrio de ligação e sua distribuição no genoma é fundamental no mapeamento associativo. O comprimento das regiões cromossômicas em DL é uma informação crucial, pois, regiões em desequilíbrio devem estar presentes para que se possa detectar associações do tipo marcador-fenótipo e a extensão dessas regiões é que vai estabelecer a resolução do mapeamento.

Dois métodos têm sido usados para visualizar o declínio do desequilíbrio de ligação ao longo de uma região genômica.

- i) Gráficos de declínio do DL: baseiam-se na plotagem das estimativas r^2 ou D' , tomadas entre pares de locos, em função da distância entre os locos.
- ii) Matriz de DL: consiste num diagrama de dupla entrada contendo uma sequência de locos na parte superior e repetida na lateral do quadro, de

modo que as estimativas de desequilíbrio de ligação entre os pares de locos (r^2 ou D') são representados, na diagonal superior da matriz, por cores que indicam a magnitude do desequilíbrio na intersecção dos respectivos locos. De forma semelhante, representam-se os valores de significância da estimativa na diagonal inferior da matriz.

Informações a respeito das causas geradoras de DL na população em estudo são essenciais para o alcance de conclusões a cerca dos prospectos do mapeamento associativo. O DL causado pela ligação genética é útil no mapeamento, mas aquele gerado por parentesco, estratificação da população e deriva genética podem produzir associações espúrias (STICH et al., 2005) e levar o pesquisador a conclusões errôneas.

O uso de coleções de germoplasma em estudos de genética de associação tem forte apelo científico, pois tais coleções, que visam à conservação de espécies de importância agrícola, abrigam genes importantes para o melhoramento genético, os quais podem ser identificados, por meio do mapeamento associativo, a depender da amostragem adequada dos acessos (FERREIRA e GRATAPAGLIA, 2006).

Existem diversos tipos de metodologias, as quais foram inicialmente desenvolvidas para estudos de associação em humanos e adaptados para outros organismos. Ultimamente, consideráveis avanços têm possibilitado o desenvolvimento de metodologias mais poderosas, mais precisas e menos viesadas para o mapeamento associativo em plantas (ABDURAKHMONOV e ABDUKARIMOV, 2008). As estatísticas para a análise de associação, sob condições ideais, seriam regressão linear, análise de variância, teste t ou teste χ^2 (ZHU et al., 2008). Porém, a presença de estrutura na população estudada pode levar a obtenção de falsas associações. O problema ocorre porque qualquer característica fenotípica que também estiver correlacionada com a estrutura da população em locos neutros apresentará um número inflacionado de associações positivas (HALL et al., 2010).

Os principais métodos de mapeamento associativo desenvolvidos para contornar esse problema são o ‘Controle Genômico’, ‘Associação Estruturada’ e o emprego de modelos mistos. Além desses, outros de menor relevância para espécies vegetais são sucintamente descritos a seguir.

O método Caso-controle (*Control-case* – CC) possibilita a identificação dos genes causais de um fenótipo (seja, por exemplo, a suscetibilidade a uma

determinada doença) por meio da comparação das frequências alélicas entre dois grupos de indivíduos não relacionados, sendo um grupo composto por indivíduos portadores do fenótipo, chamado de “caso”, e outro grupo de indivíduos não portadores (sadios), chamado de “controle”. O teste χ^2 , teste exato de Fisher ou a correção de continuidade de Yates podem ser usados para a comparação das frequências dos alelos e detecção das associações entre marcador e fenótipo. Os resultados deste método são muito influenciados pela ocorrência de subestruturação nas amostras das populações avaliadas (ABDURAKHMONOV E ABDUKARIMOV, 2008).

O método do Risco Relativo haplotípico (*haplotype relative risk – HRR*), idealizado por Falk e Rubinstein (1987), reduz o efeito da estrutura. Neste método, um pseudo-grupo controle, contendo combinação de dois alelos que não são transmitidos para progênie afetada, é criado. Em seguida, as frequências dos alelos marcadores no grupo caso e no grupo pseudo-controle são correlacionadas (ABDURAKHMONOV e ABDUKARIMOV, 2008).

O Teste de Transmissão de desequilíbrio (*Transmission disequilibrium test – TDT*), desenvolvido por Spielman et al., (1993), é um método de mapeamento associativo baseado em famílias, que elimina o efeito do confundimento causado pela estruturação, utilizando a comparação entre a transmissão ou não-transmissão de alelos marcadores para progênies afetadas, por meio de um teste χ^2 , assumindo uma ligação entre marcador e fenótipo. O método requer a genotipagem de um genitor heterozigoto, um genitor homozigoto e uma progênie afetada. Muitas derivações desse método foram produzidas (ABDURAKHMONOV e ABDUKARIMOV, 2008). Por exemplo, o QTDT (*Quantitative TDT*) é empregado na análise de caracteres quantitativos (ZHU et al., 2008).

O Controle Genômico (*Genomic Control – GC*) consiste no emprego de um grande número de marcadores neutros ou que não estejam envolvidos no controle da característica fenotípica de interesse. As associações são estimadas e o teste estatístico é aplicado para estabelecimento do valor crítico correspondente a taxa desejada de erro do tipo I. Embora seja computacionalmente simples, esse método requer o uso de muitos marcadores-controle para capturar adequadamente a extensão da variação da estrutura populacional ao longo do genoma de um organismo. Além disso, em algumas situações, é possível superestimar os efeitos da estrutura, o que

resulta na perda de poder estatístico para detectar as associações verdadeiras (HALL et al., 2010).

O método de associação estruturada (*Structured association – SA*) foi desenvolvido por Pritchard et al. (2000) e consiste na inferência de detalhes sobre a estrutura da população e da ancestralidade de uma amostra de indivíduos, por meio de um conjunto de marcadores não ligados, analisados em um modelo bayesiano. Estas informações são utilizadas para identificar populações dentro das quais ocorre acasalamento ao acaso (HALL et al., 2010). Este método encontra-se implementado no software STRUCTURE. O uso desse método no estudo da associação do gene *dwarf8* com o tempo de florescimento, em milho reduziu a incidência de falso-positivos devido à estrutura populacional em 8% (SORKHEH et al., 2008).

A metodologia de modelos mistos (*Mixed Linear Model – MLM*) é uma das mais recentes e promissoras técnicas para corrigir os falsos efeitos causados pela estrutura da população. Neste método, preconizado por Yu et al. (2006), marcadores aleatórios são utilizados para estimar as matrizes que representam a estrutura populacional (Q) e os coeficientes de parentesco (K) da população. Essas matrizes são incorporadas ao escopo da metodologia de modelos mistos para estimar as associações SSR-carater fenotípico. (ZHU et al., 2008). Q é uma matriz de co-ancestralidade do tipo $n \times p$, (onde n é o número de indivíduos e p é o número de subpopulações definidas) e K consiste em uma matriz de parentesco do $(n \times n)$, na qual, os elementos fora da diagonal são uma estimativa da probabilidade de identidade por descendência, baseada em marcadores moleculares. Os elementos da diagonal de K equivalem a 1, para endogâmicos e $[0,5 \times (1 + F_x)]$ para não-endogâmicos, onde F_x é o coeficiente de endogamia (ZHU et al., 2008).

REFERÊNCIAS BIBLIOGRÁFICAS

ABDURAKHMONOV, I.Y.; ABDUKARIMOV, A. Application of association mapping to understanding genetic diversity of plant germplasm resources. **International Journal of Plant Genomics**. 2008.

ABIC - Exportações brasileiras de café em grãos. Disponível no site: http://www.abic.com.br/estat_exportacoes.html. Consultado em 24jul2010

ANTHONY, F.; COMBES, M.C.; ASTORGA, C.; BERTRAND, B.; GRAZIOSI, G.; LASHERMES, P. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. **Theoretical and Applied Genetics**, 104:894-900, 2002.

BHAT, P.R.; KRISHNAKUMAR, V.; HENDRE, P.S.; RAJENDRAKUMAR, P.; VARSHNEY, R.K.; AGGARWAL, R.K. Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* x *Coffea congensis*). **Molecular Ecology Notes**, 5:80-83. 2005.

BERTHAUD, J.; CHARRIER, A. Genetic resources of *Coffea*. In: CLARCK, R.J. AND MACRAE, R. (eds). **Coffee – Agronomy**. London, Elsevier Applied Science, p.1-40. 1985.

BERTHAUD, J. L'incompatibilité chez *Coffea canephora*: méthode de test et déterminisme génétique. **Café Cacao Thé**, 24: 167-174. 1980

BERTHAUD, J. **Les ressources génétique pour l'amélioration des caféiers africains diploïdes. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application**. 1986. 379f. Tese (Doutorado em Agronomia) – ORSTOM, Paris.

BOUHARMONT, P.; AWEMO, J. La sélection végétative du caféier Robusta au Cameroun. 1ère Partie : Programme de sélection. **Café Cacao Thé**, 23: 227–254. 1979.

- BOUHARMONT, P.; LOTODÉ, R.; AWEMO, J.; CASTAING, X. La sélection générative du caféier Robusta au Cameroun. Analyse des résultats d'un essai d'hybrides dialléle partiel implanté en 1973. **Café Cacao Thé**, 30: 93–112. 1986
- BRADBURY, P.J.; ZHANG, Z.; KROON, D.E.; CASSTEVENS, T.M.; RAMDOSS, Y.; BUCKLER, E.S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, 23: 2633–2635. 2007.
- BRITO, G. G.; Caixeta, E.T.; GALLINA, A.P.; 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** 173:255-264, 2010.
- CAPOT, J. L' amélioration du caféier robusta en côte D'Ivoire. **Café Cacao Thé**, 21:233-244. 1977
- CARVALHO, A. **Histórico do desenvolvimento do cultivo do café no Brasil**. Campinas, Instituto Agronômico de Campinas. 1993. 7p. (Documentos IAC, 34).
- CHARRIER, A.; BERTHAUD, J. Principles and methods of coffee plant breeding: *Coffea canephora*. In: CLARCK, R.J.; MACRAE, R. (eds). **Coffee – Agronomy**. London, Elsevier Applied Science. P. 167-198. 1985.
- CHEVALIER, A. **Les caféiers du globe. I. Généralités sur les caféiers**. Encyclopédie biologique n°5, Paul Lechevalier, Paris, 1929.
- COMBES, M.C.; ANDRZEJEWSKI, S.; ANTHONY, F.; BERTRAND, B.; ROVELLI, P.; GRAZIOSI, G.; LASHERMES, P. Characterization of microsatellites loci in *Coffea arabica* and related coffee species. **Molecular Ecology**, 9:1171-1193, 2000.
- CONAB. **Acompanhamento da Safra Brasileira Café Safra 2010**, segunda estimativa, maio/2010. Companhia Nacional de Abastecimento. Brasília: Conab, 2010.
- CRUZ C.D. **Introdução à genética quantitativa**. Viçosa:UFV. 394p.: il 2005.

CRUZ, C.D.; GOOD GOD; P.I.V; BHERING, L.L. Mapeamento de QTLs em populações exogâmicas. In: BORÉM, A.; CAIXETA, E.T. **Marcadores moleculares** 2ed. Viçosa:UFV. 443-485. 2009.

CUBRY, P. **Structuration de la diversité génétique et analyse des patrons de déséquilibre de liaison de l'espèce *Coffea canephora* Pierre ex. Froehner.** 2009. 249f. Tese (Doutorado em Agronomia) - Université Montpellier II, Montpellier. 2009.

DAVIS, A.P.; GOVAERTS, R.; BRIDSON, D.M.; STOFFELEN, P. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). **Botanical Journal of the Linnean Society**, 152:465-512, 2006.

DAVIS, A.P.; BRIDSON, D.M.; RAKOTONASOLO, F. A reexamination of *Coffea* subgenus *Baracoffea* and comments on the morphology and classification of *Coffea* and *Psilanthus* (Rubiaceae-Coffeae). In: Keating RC, Hollowell VC, Croat T, eds. Festschrift for William G. D'Arcy: the legacy of a taxonomist (Monograph in Systematic Botany 104). Missouri: MBG Press, 398–420. 2005.

DIOLA, V. BRITO, G.G.; CAIXETA, E.T.; ZAMBOLIM, E.M.; SAKIYAMA, N.S.; LOUREIRO, M.E. High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics & Genomes**, 7:1-10, 2011.

DINIZ, L.E.C.; SAKIYAMA, N.S.; LASHERMES, P.; CAIXETA, E.T.; OLIVEIRA, A.C.B.; MACIEL-ZAMBOLIM, E.; LOUREIRO, M.E.; PEREIRA, A.A.; ZAMBOLIM, L. Analysis of AFLP markers associated to the *Mex-1* resistance locus in Icatu progenies. **Crop Breeding and Applied Biotechnology**, 5:387-393, 2005.

DUSSERT, S.; LASHERMES, P.; ANTHONY, F.; MONTAGNON, C.; TROUSLOT, P.; COMBES, M.C.; BERTHAUD, J.; NOIROT, M.; HAMON, S. Le caféier, *Coffea canephora*. In: Hamon P., Seguin M., Perrier X., Glaszmann J.-C. (eds.) **Diversité génétique des plantes tropicales cultivées.** CIRAD, Montpellier, France, pp. 175–794. 1999.

FALK C.T.; RUBINSTEIN, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. **Annals of Human Genetics**, 51:227–233, 1987.

FAZUOLI, L.C.; MISTRO, J.C.; BRAGHINI, M.T. Melhoramento do café robusta no Instituto Agronomico de Campinas. In: Zambolim L (Ed). **Tecnologias para produção do café conilon**. Editora UFV, Viçosa, p. 201-247. 2009.

FAZUOLI, L.C. Genética e Melhoramento do cafeeiro. In: RENA, A.B.; MALAVOLTA, E.; ROCHA, M. IAMADA, T. ed.. **Cultura do cafeeiro**: fatores que afetam a produtividade. Piracicaba: POTAFOS. p.88-113. 1986.

FERRÃO, R.G.; FONSECA, A.F.A.; FERRÃO, M.A.G. Programa de melhoramento genético de café Robusta no Brasil. In: NURMBERG et al. (Eds.). **Simpósio de atualização em genética e melhoramento de plantas**. Lavras, MG: UFLA, p.50-65,1999.

FERRÃO, R.G.; FONSECA, A.F.A.; FERRÃO, M.A.G.; BRAGANÇA, S.M.; VERDIN-FILHO, A.C.; VOLPI, P.S. Cultivares de café conilon. In: FERRÃO RG, FONSECA AFA, BRAGANÇA SM, FERRÃO MAG, DE MUNER LH (eds) **Café conilon**. Vitória, pp. 203-225. 2007a

FERRÃO, M.A.G.; FONSECA, A.F.A.; FERRÃO, R.G.; OLIVEIRA, M.A.; BARBOSA, W.M.; D'ISEP, M.S.P.; BARBOSA, R.P. Técnicas moleculares e biotecnológicas aplicadas ao café. In: FERRÃO, R.G.; FONSECA, A.F.A.; BRAGANÇA, S.M.; FERRÃO, M.A.G.; DE MUNER, L.H. (eds) **Café conilon**. Vitória, pp. 175-201. 2007b

FERREIRA M.E.; GRATTAPAGLIA, D. Genética de associação em plantas. In: Aluizio Borém; Eveline T. Caixeta. (eds). **Marcadores Moleculares**. 1 ed. Viçosa: UFV, v. 1, p. 273-306. 2006.

FLINT-GARCIA, S.A.; THUILLET, A.; YU, J.; PRESSOIR, G.; ROMERO, S.M.; MITCHELL, S.E.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M.M.; BUCKLER, E.S. Maize association population: A high-resolution platform for quantitative trait locus dissection. **Plant Journal**, 44:1054–1064. 2005.

GOMÉZ, C; DUSSERT, S; HAMON, P; HAMON, S; KOCHKO, A; PONCET, V. Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. **Evolutionary Biology**, 9:1-19. 2009.

GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. **Plant Molecular Biology**, 57:461-485, 2005.

HALL, D.; TEGSTRÖM, C.; INGVARSSON, P.K. Using association mapping to dissect the genetic basis of complex traits in plants. **Briefings in Functional Genomics and Proteomics**, 1- 9, 2010

KY, C-L.; BARRE, P.; LORIEUX, M.; TROUSLOT, P.; AKAFFOU, S.; LOUARN, J.; CHARRIER, A.; HAMON, S.; NOIROT, M. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetics**, 101:669–676, 2000.

KY, C-L.; LOUARN, J.; GUYOT, B.; CHARRIER, A.; HAMON, S.; NOIROT, M. Relations between and inheritance of chlorogenic acid contents in an interspecific cross between *Coffea pseudozanguebariae* and *C. liberica* var. Dewevrei. **Theoretical and Applied Genetics**, 98:628–637, 1999.

KUMAR A.; SRINIVASAN, C.S.; NATARAJ, T. A preliminary note on the occurrence of dwarf mutants in Robusta coffee (*Coffea canephora* Pierre ex. Froehner). **Journal of Coffee Research**, 24:41-45, 1994

LASHERMES, P.; CROS, J.; MARMEY, P.; CHARRIER, A. Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. **Genetic Resources and Crop Evolution**, 40:91-93, 1993.

LEBRUN J. **Recherches morphologiques et systématiques sur les caféiers du Congo**. Publ. INEAC Hors Série 184p. 1941.

LEROY, T.; MONTAGNON, C.; CHARRIER, A.; ESKES, A.B. Reciprocal recurrent selection applied to *Coffea canephora* Pierre. I. Characterization and evaluation of breeding populations and values of intergroup hybrids. **Euphytica**, 67: 113-125, 1993.

LEROY, T.; MONTAGNON, C.; CILAS, C.; CHARRIER, A.; ESKES, A.B. Reciprocal recurrent selection applied to *Coffea canephora* Pierre. II. Estimation of genetic parameters. **Euphytica**, 74: 121-128, 1994.

LEROY, T.; MONTAGNON, C.; CILAS, C.; YAPO, A.; CHARMETANT, P.; ESKES, A.B. Reciprocal recurrent selection applied to *Coffea canephora* Pierre. III. Genetic gains and results of first intergroup crosses. **Euphytica**, 95: 347-354, 1997.

LOPES-GARTNER, G.; CORTINA, H.; McCOUCH; MONCADA, MP. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. **Tree Genetics & Genomes**, 5:435-446, 2009.

MACKAY, I.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, 12:57-63, 2007.

MAHÉ, L.; COMBES, M.C.; VÁRZEA, V.M.P.; GUILHAUMON, C.; LASHERMES, P. Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). **Molecular Breeding**, 21:105-113, 2008.

MALUF, M.P.; SILVESTRINI, M.; RUGGIERO, L.M.C.; GUERREIRO-FILHO, O.; COLOMBO, C.A. Genetic diversity of cultivated *Coffea* inbred lines assessed by RAPD, AFLP and SSR marker systems. **Scientia Agricola**, 62:366-373, 2005.

MARTINS, A.L. **História do café**. São Paulo: Contexto. 2008. 316p.

MARTINS, P.S. Estrutura populacional, fluxo gênico e conservação "in situ". **IPEF**, 35:71-78, 1987.

MAURIN, O.; DAVIS, P.A.; CHESTER, M.; MVUNG, E.F.; JAUFEEERALLY-FAKIM, Y.; FAY, M.F. Towards a phylogeny for *Coffea* (Rubiaceae): Identifying well-supported lineages based on nuclear and plastid DNA sequences. **Annals of Botany**, 1-19, 2007.

MONTAGNON, C.; LEROY, T.; ESKES, A.B. Amélioration variétale de *Coffea canephora*. I. Critères et méthodes de sélection. **Plantations, recherche, développement**, p.18-33, 1998a.

MONTAGNON, C.; LEROY, T.; ESKES, A.B. Amélioration variétale de *Coffea canephora*. II. Les programmes de sélection et leurs résultats. **Plantations, recherche, développement**, p.89-98, 1998b.

MUSOLI CP, CUBRY P, ALUKA P, BILLOT C, DUFOUR M, DE BELLIS F, POT D, BIEYSSE D, CHARRIER A, LEROY T (2009) Genetic differentiation of wild and cultivated populations: Diversity of *Coffea canephora* Pierre in Uganda. **Genome**, 52:634-646, 2010.

N'DIAYE A, PONCET V, LOUARN J, HAMON S, NOIROT M. Genetic differentiation between *Coffea liberica* var. *liberica* and var. *Dewevrei* and comparison with *C. canephora*. **Plant Syst. Evolution**, 253:95-104, 2005.

NOIR, S; ANTHONY, F; BERTRAND, B; COMBES, M.-C.; LASHERMES, P. Identification of a major gene (Mex-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica*. **Plant Pathology**, 52:97–103, 2003.

OLIVEIRA, A.C.B.; SAKIYAMA, N.S.; CAIXETA, E.T.; MACIEL-ZAMBOLIM, E.; RUFINO, R.J.N.; ZAMBOLIM, L. Partial map of *Coffea arabica* L. and recovery of the recurrent parent in backcross progenies. **Crop Breeding and Applied Biotechnology** 7:196-203, 2007.

PRAKASH, N.S.; MARQUES, D.V.; VARZEA, V.M.P.; SILVA, M.C.; COMBES, M.C.; LASHERMES, P. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *Coffea arabica* L. **Theoretical and Applied Genetics**, 109:1311–1317, 2004.

PRAKASH, N. S.; COMBES, M.; DUSSERT, S.; NAVEEN, S.; PHILIPPE LASHERMES, L. Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. **Genetic Resources and Crop Evolution**, 52:333–343, 2005.

PRITCHARD, J.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, 155:945–959, 2000.

RESENDE, M. D. V. de; BARBOSA, M. H. P. **Melhoramento genético de plantas de propagação assexuada**. Colombo: Embrapa Florestas, 2005. 130 p.

RISCH, N.J. Searching for genetic determinants in the new millennium. **Nature** 405:847–856. 2000.

SAKIYAMA, N.S.; PEREIRA, A.A.; ZAMBOLIM, L. Melhoramento de Café Arábica. In: **Melhoramento de espécies cultivadas**. (editor) Aluizio Borém. Viçosa:UFV, 2005. 817p.:il.

SCHUSTER, I.; CRUZ, C.D. **Estatística Genômica Aplicada a Populações Derivadas de Cruzamentos Controlados**. 1. ed. Viçosa: UFV, 2004. v. 1. 568 p.

SILVESTRINI, M.; MALUF, M.P.; SILVAROLLA, M.B.; GUERREIRO-FILHO, O.; MEDINA-FILHO, H.P.; VANINI, M.M.T.; OLIVEIRA, A.S.; GASPARIPEZZOPANE, C.; FAZUOLI, L.C. Genetic diversity of a *Coffea* germplasm collection assessed by RAPD markers. **Genetic Resources and Crop Evolution**, 55:901-910, 2008.

SORKHEH, K; MALYSHEVA-OTTO, L.V.; WIRTHENSOHN, M.G.; TARKESH-ESFAHANI, S.; MARTÍNEZ-GÓMEZ, P. Linkage disequilibrium, genetic association mapping and gene localization in crop plants. **Genetics and Molecular Biology**, 31:805-814, 2008.

SOUZA, F.F. Análise de correlações entre caracteres morfo-agronômicos em clones de café Conilon da coleção de germoplasma da Embrapa Rondônia. In: Congresso Brasileiro de Melhoramento de Plantas, 2, 2003, Porto Seguro, **Anais...** Cruz das Almas: Embrapa Mandioca e Fruticultura, 2003 (CD-ROM).

SOUZA, F.F.; SANTOS, M.M.; VENEZIANO, W. Manejo de recursos genéticos de café em Rondônia. In: Simpósio de Pesquisa dos cafés do Brasil, 3, 2003, Porto Seguro. **Anais...** Brasília: Embrapa Café, 2003, p. 238.

SOUZA, F.F.; SANTOS, M.M. Melhoramento genético do café canéfora em Rondônia In: Zambolim L (Ed). **Tecnologias para produção do café conilon**. Editora UFV, Viçosa, p. 175-200. 2009.

SPIELMAN, R.S.; MCGINNIS, R.E.; J. EWENS, J. Transmission Test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). **American Journal of Human Genetics**, 52:506–516, 1993.

STICH, B.; MELCHINGER, A.E.; FRISCH, M.; MAURER, H.P.; HECKENBERGER, M.; REIF, J.C. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. **Theoretical and Applied Genetics**, 111:723-730, 2005.

VENEZIANO, W. **Recursos Genéticos de café: características, uso e manejo**. In: Seminário Internacional do Agronegócio do Café na Amazônia, 1, Porto Velho: Embrapa Rondônia, 2003. 124p. (Embrapa Rondônia. Documentos, 78)

VENEZIANO, W.; FONSECA, A.F.A.; FAZUOLI, L.C. Avaliação de clones de café conilon em Rondônia. In: Simpósio de Pesquisa dos cafés do Brasil, 3, 2003, Porto Seguro. **Anais...** Brasília: Embrapa Café, 2003, p. 219.

WEIR, B.S. **Genetic Data Analysis II - Methods for Discrete Population Genetic Data**. Sinauer Associates, Sunderland. 1996.

YU, J.; PRESSOIR, G.; BRIGGS, W.H.; BI, I.V.; YAMASAKI, M.; DOEBLEY, J.F.; McMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.N.; HOLLAND, J.B.; RESOVICH, S.; BUCKLER, E.S.; A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, 2:203-208, 2006.

ZHU, C.; GORE, M.; BUCKLER, E.S.; YU, J. Status and prospects of association mapping in plants. **Plant Genome**, 1:5–20, 2008.

CAPÍTULO 1

Análise da diversidade no germoplasma brasileiro de *Coffea canephora*

Molecular overview about *Coffea canephora* germplasm in Brazil

ABSTRACT

Coffea canephora germplasm, commercially grown or maintained in Brazilian institutions of agricultural research, shows wide phenotypic variability, however its molecular polymorphism remain precariously known. This work aimed to characterize *C. canephora* accessions representative of commercial crops and germplasm collections, in Brazil. A total of 130 accessions from germplasm banks from São Paulo, Minas Gerais, and also collected in commercial plantations at Espírito Santo and Rondônia were evaluated with a set of 20 microsatellite primers. Multivariate methods were used to estimate the relationship among accessions. Microsatellites detected high level of polymorphism and two major genetic groups were identified. First was composed by Kouillou accessions from Espírito Santo and Rondônia and the second, by Robusta accessions from São Paulo and Minas Gerais. Both group presented expressive but limited variability. Thus, new efforts are necessary to increase genetic diversity of Brazilian gene banks. Moreover, breeding programs should enlarge Robusta alleles in future cultivars.

Keywords: SSR, Robusta coffee, genetic diversity

INTRODUCTION

Coffea canephora Pierre ex Froehner presents a wide genetic variability, with one of the widest geographic natural distribution within the subgenus *Coffea*, which extend west-east, from Guinea to Uganda, and north-south, from Cameroon to Angola (Maurin et al. 2007). Likewise most diploid species in genus *Coffea*, *C. canephora* is allogamous and presents a self-incompatibility system.

Phenotypic, biochemical and molecular studies carried out in order to assess the genetic diversity of *C. canephora* in Africa indicated the existence of two major groups: 'Guinean' and 'Congolese' (Berthaud 1986, Montagnon et al. 1992, Dussert et al. 1999). The Guinean is composed by populations from Guinea and Ivory Coast, that present small growth habit, small elongated leaves, small internodes length, secondary branching, low bean weight, high caffeine content, early harvest, drought tolerance and susceptibility to leaf rust, caused by the fungus *Hemileia vastatrix* (Leroy et al. 1993). The genotypes of Guinean group have been restricted to the West Africa and thus, they probably kept out from the *C. canephora* breeding programs around the world (Montagnon et al. 1998a). On the other hand, Congolese group, composed by populations from Central African Republic and Cameroon, is divided in four subgroups, *i.e.*: SG1, C, B e SG2 (Dussert et al. 1999). Genotypes from SG1 present medium-sized narrow leaves, medium large beans with higher caffeine content; very late bearing plants, generally well branched, moderately resistant to leaf rust and highly resistant to drought. Genotypes from SG2, B and C present large broad leaves, large beans with lower caffeine content. SG2 plants are slightly branched, moderately late-bearing plants, highly resistant to leaf rust and very sensitive to drought (Montagnon et al. 1998b). In Brazil, the varietal types called Conilon (deformation of Kouillou) comprise SG1. Nevertheless, due the resemblance of some adaptive traits between Guinean Group and SG1, the assumption that

Conilon would become to Guinean group had prevailed for so long. Kouillou genotypes arrived in Brazil, by Rio de Janeiro and Espírito Santo, about 1912 (Ferrão et al. 2007a). Robusta genotypes were introduced in São Paulo, a little bit later in the same decade. Additionally, following introductions of Robusta occurred in São Paulo and Minas Gerais during the 70's (Fazuoli et al. 2009).

In Brazil, the main coffee germplasm collections are placed in governmental institutions, in São Paulo, at Instituto Agrônomo de Campinas (IAC); in Espírito Santo, at the Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper); in Minas Gerais, at the Universidade Federal de Viçosa (UFV), in a partnership with Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig); and in Rondônia, at Embrapa (Embrapa Rondônia). Furthermore, Espírito Santo and Rondônia response for 75% of *C. canephora* produced in Brazil. For this reason, collections of the Incaper and Embrapa Rondônia harbor a great amount of accessions representative from commercial plantations. Embrapa Rondônia has the particularity of containing a significant number of accessions resulting from exchanges with the institutions mentioned before (*i.e.*: IAC, UFV and Incaper), besides an expressive sample of local genotypes (Souza and Santos 2009). Consequently, its variability is representative of the germplasm commercially grown and conserved in Brazil.

Coffea canephora accessions have been studied using phenotypic traits (Fonseca et al. 2006, Ivoglo et al. 2008, Souza and Santos 2009) and RAPD markers (Ferrão et al. 2007b, Silvestrini et al. 2008). These studies have confirmed there is a wide variability within the germplasm maintained in the Brazilian collections. However, despite the advantages of microsatellites - *e.g.*: high reproducibility, multi-allelic locus, co-dominant inheritance, high degree of polymorphism, relative

abundance and good coverage of the genome (Powell et al. 1996) - there are a few works using these markers. Furthermore, there is no report of comparisons about the diversity among accessions from different institutions, providing a well representative coverage of this germplasm.

Thus, this study aimed to characterize accessions representing the germplasm cultivated in Brazilian plantations and conserved in research institutions, in order to propose guide lines to management of gene banks and breeding strategies.

MATERIAL AND METHODS

Plant material

A total of 130 accessions of *C. canephora* (Table 1), maintained at the Coffee Germplasm Collection of Embrapa, in Rondônia State, were genotyped. These accessions comprise a good sample of the germplasm used in the Brazilian breeding programs. They were introduced in Embrapa gene bank during the last three decades, by means of interchange with other public institutions, *i.e.*: IAC, in São Paulo (43 accessions); UFV, in Minas Gerais (11 accessions) and Incaper, in Espírito Santo (40 accessions). The other accessions (36) were collected in traditional coffee producing areas at Rondônia State. One *C. arabica* and one Híbrido de Timor (*C. arabica* x *C. canephora*) accessions were included in the analysis as out group species.

DNA extraction

Young and completely extended leaves were collected from each accession, frozen at $-80\text{ }^{\circ}\text{C}$, lyophilized, ground to become a fine powder and kept at $-20\text{ }^{\circ}\text{C}$

until used. Genomic DNA was extracted using the method described by Diniz et al (2005) and all DNA samples were prepared to a final concentration of $25\text{ng}\cdot\mu\text{L}^{-1}$.

Table 1 List of accessions genotyped in this study.

Accession	Code	AS ¹	Accession	Code	AS ¹	Accession	Code	AS ¹
Kouillou IAC66-1.1	K66-11	1	Robusta UFV 3587.3	R35873	2	Encapa 07	ES07	4
Kouillou IAC 66-1.2	K66-12	1	Robusta UFV 3751.1	R37511	2	Encapa 14	ES14	4
Kouillou IAC 66-1.3	K66-13	1	Robusta UFV 3751.2	R37512	2	Encapa 16	ES16	4
Kouillou IAC 66-3..1	K66-31	1	Robusta UFV 3754.1	R37541	2	Encapa 19	ES19	4
Kouillou IAC 66-3.2	K66-32	1	Robusta UFV 3754.2	R37542	2	Encapa 28	ES28	4
Kouillou IAC 68-7.1	K68-71	1	Robusta UFV 3755.1	R37551	2	Encapa 104A	ES104A	4
Kouillou IAC 68-7.2	K68-72	1	Robusta UFV 3755.2	R37552	2	Encapa 104B	ES104B	4
Kouillou IAC 68-7.3	K68-73	1	Robusta UFV 3755.3	R37553	2	Encapa 106	ES106	4
Kouillou IAC 69-15	K69-15	1	Cpafro 010	RO010	3	Encapa 110A	ES110A	4
Kouillou IAC 69-5.1	K69-51	1	Cpafro 016	RO016	3	Encapa 110B	ES110B	4
Kouillou IAC 69-5.2	K69-52	1	Cpafro 022	RO022	3	Encapa 112	ES112	4
Kouillou IAC 69-5.3	K69-53	1	Cpafro 024	RO024	3	Encapa 116	ES116	4
Kouillou IAC 70-1.1	K70-11	1	Cpafro 036	RO036	3	Encapa 120	ES120	4
Kouillou IAC 70-1.2	K70-12	1	Cpafro 044	RO044	3	Encapa 132	ES132	4
Kouillou IAC 70-1.3	K70-13	1	Cpafro 045	RO045	3	Encapa 139	ES139	4
Kouillou IAC 70-14.1	K70-141	1	Cpafro 056	RO056	3	Encapa 143	ES143	4
Kouillou IAC 70-14.2	K70-142	1	Cpafro 063	RO063	3	Encapa 148	ES148	4
Kouillou IAC 70-14.3	K70-143	1	Cpafro 077	RO077	3	Encapa 149	ES149	4
Laurenti.1	Laur1	1	Cpafro 085	RO085	3	Encapa 154	ES154	4
Laurenti.2	Laur2	1	Cpafro 086	RO086	3	Encapa 201	ES201	4
Apoatã IAC 2258.1	Apo1	1	Cpafro 089	RO089	3	Encapa 26	ES26	4
Apoatã IAC 2258.2	Apo2	1	Cpafro 098	RO098	3	Encapa 29	ES29	4
Apoatã IAC 2258.3	Apo3	1	Cpafro 100	RO100	3	Encapa 36	ES36	4
Robusta IAC 640.1	R6401	1	Cpafro 103	RO103	3	Encapa 45	ES45	4
Robusta IAC 640.2	R6402	1	Cpafro 119	RO119	3	Encapa 49	ES49	4
Robusta IAC 640.3	R6403	1	Cpafro 127	RO127	3	Encapa 99	ES99	4
Robusta IAC 1641.1	R16411	1	Cpafro 138	RO138	3	Encapa V.1	ESV1	4
Robusta IAC 1641.2	R16412	1	Cpafro 140	RO140	3	Encapa V.2	ESV2	4
Robusta IAC 1655.1	R16551	1	Cpafro 142	RO142	3	Encapa V.3	ESV3	4
Robusta IAC 1655.2	R16552	1	Cpafro 143	RO143	3	Encapa V.4	ESV4	4
Robusta IAC 1675.1	R16751	1	Cpafro 147	RO147	3	Encapa V.5	ESV5	4
Robusta IAC 1675.2	R16752	1	Cpafro 155	RO155	3	Encapa V.6	ESV6	4
Robusta IAC 1675.3	R16753	1	Cpafro 156	RO156	3	Encapa V.7	ESV7	4
Robusta IAC 2257.1	R22571	1	Cpafro 160	RO160	3	Encapa V.9	ESV9	4
Robusta IAC 2257.2	R22572	1	Cpafro 164	RO164	3	Encapa V.10	ESV10	4
Robusta IAC 2259.1	R2259	1	Cpafro 183	RO183	3	Encapa V.11	ESV11	4
Robusta IAC 2286.1	R22861	1	Cpafro 184	RO184	3	Encapa V.12	ESV12	4
Robusta IAC 2286.2	R22862	1	Cpafro 189	RO189	3	Encapa V.13	ESV13	4
Robusta Col - 10.1	RCol-101	1	Cpafro 190	RO190	3	<i>C. arabica</i> var.		
Robusta Col - 10.2	RCol-102	1	Cpafro 193	RO193	3	Typica UFV	<i>Carabica</i>	2
Robusta Col - 10.3	RCol-103	1	Cpafro 194	RO194	3	2945		
Robusta Col - 5.1	RCol-51	1	Cpafro 196	RO196	3	Híbrido de		
Robusta Col - 5.2	RCol-52	1	Cpafro 199	RO199	3	Timor CIFC	HibTimor	2
Robusta UFV 3580	R3580	2	Cpafro 203	RO203	3	1343/269		
Robusta UFV 3587.1	R35871	2	Encapa 02	ES02	4			
Robusta UFV 3587.2	R35872	2	Encapa 03	ES03	4			

¹Accession Source: (1) Coffee Germplasm Collection of Instituto Agronômico de Campinas (IAC), São Paulo; (2) Coffee Germplasm Collection of Universidade Federal de Viçosa (UFV), Minas Gerais; (3) Accessions collected in commercial coffee fields in Rondônia State, by Embrapa, and (4) in Espírito Santo State and selected in Breeding Program of the Instituto Capixaba de Assistência Técnica, Pesquisa e Extensão Rural (INCAPER).

Microsatellite markers

Twenty microsatellites were used in this study. These DNA markers were developed from non-redundant Express Sequence Tags (EST) of the Brazilian Coffee Genome Project.

Each reaction was set to a final volume of 20 μ L, containing 50 ng of genomic DNA, 0.6 unit of *Taq* DNA polymerase, buffer 1x, 1mM of $MgCl_2$, 150 μ M of each dNTP and 0.1 μ M of each primer. PCR amplifications were carried out using touchdown proceeding, which comprises initial denaturation at 94 °C for 2 min, followed by 10 cycles of denaturation at 94 °C for 0.5 min, annealing at 67°C for 0.5 min, decreasing 1 °C after each cycle, and extension at 72 °C for 0.5 min. After that, another set of 30 cycles, comprising denaturation at 94 °C for 0.5 min, annealing at 55 °C for 0.5 min and extension at 72 °C for 0.5 min, was accomplished followed by a final 8 min extension time at 72 °C. Before running, PCR products were denatured in 8 μ L of denaturing dye (95% formamide) at 94 °C for 5 min and 7 μ L of sample were loaded on a standard 6% polyacrylamide gel at 50 °C and run at a constant power of 90 W for about 2 h. Post-PCR multiplex (Figure 1), which involved the multi-loading of individual PCR assays (two to four SSRs per running), was performed spacing successive loads by 10 to 30 min during electrophoresis, depending on the prior information about fragment size. At last, the gel was treated with ethanol (10%) + acetic acid (1%), followed by nitrate acid (1,5%); stained with silver nitrate (4%), developed with sodium carbonate (3%) and formaldehyde (0.03%); and fixed with acetic acid (5%) and dried for posterior analysis at a transilluminator apparatus.

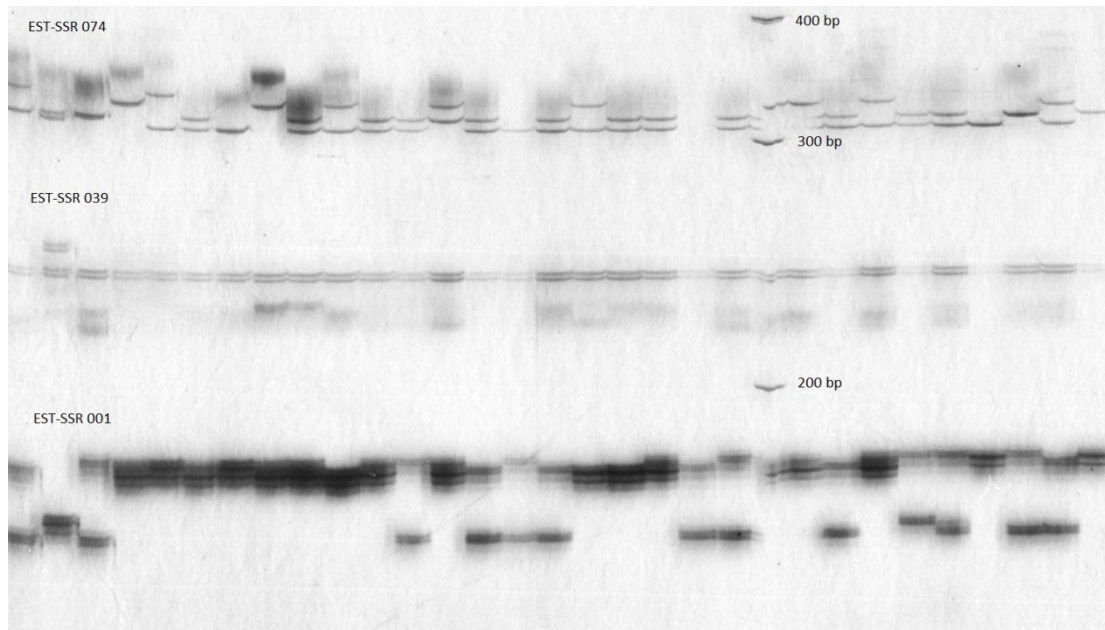


Figure 1. Polyacrylamide gel revealing a Post-PCR multiplex, which involved the loading of three individual PCR assays (EST-SSR 001, EST-SSR 039 and EST-SSR 074).

Data Analysis

The evaluation of each locus was performed considering homozygotes and heterozygotes, individuals which showed one or two alleles, respectively. The dissimilarity between the accessions was estimated based on the complement ($1 - S_{ii'}$) of the weighted coincidence index, with the equation: $S_{ii'} = \frac{1}{2} \sum_{j=1}^L p_j c_j$, where $S_{ii'}$ is the similarity between the accessions i and i' ; L is the total number of loci; c_j is the number of common alleles between i and i' ; and p_j is the weight associated to j locus, obtained by a_j/A , being a_j , the number of alleles in locus j and A total number of alleles. Principal Coordinate Analysis (PCoA) was performed to view the overall diversity. The dissimilarity matrix was also represented in a dendrogram based on the unweighted pair-group method using arithmetic averages (UPGMA). Goodness-of-fit of the tree was tested comparing cophenetic value matrix, with the original dissimilarity matrix. Statistical procedures were accomplished using the softwares: Genes (Cruz, 2001), Darwin 5.0 (Perrier et al. 2003) and NTSYS-pc (Rohlf, 1998).

RESULTS AND DISCUSSION

The overall diversity among the 130 accessions of *C. canephora* is represented in the bi-dimensional graphics based on the principal coordinate analysis (Figure 2). The first, second and third axis exhibited, respectively, 19.1%, 5.8% and 3.3% of the total variability. Accessions were clearly separated in two major groups, in accordance with their phenotypic characteristics, i.e. Kouillous and Robustas. Kouillou accessions were plotted at the left side of the plan composed by the 1st and 2nd coordinates. A division by location was also observed into Kouillou cluster. Populations from Rondônia and Espírito Santo occupied, respectively, upper and lower halves of the plan. Robusta accessions were positioned in the right side, without presenting notorious sub-clustering. On the plan composed by 1st and 3rd coordinates, Robustas and Kouillous continued taken apart, reinforcing their genetic distance. Additionally, accessions assumed a more spread distribution, such that Kouillous from Rondônia and Espírito Santo appeared a little merged.

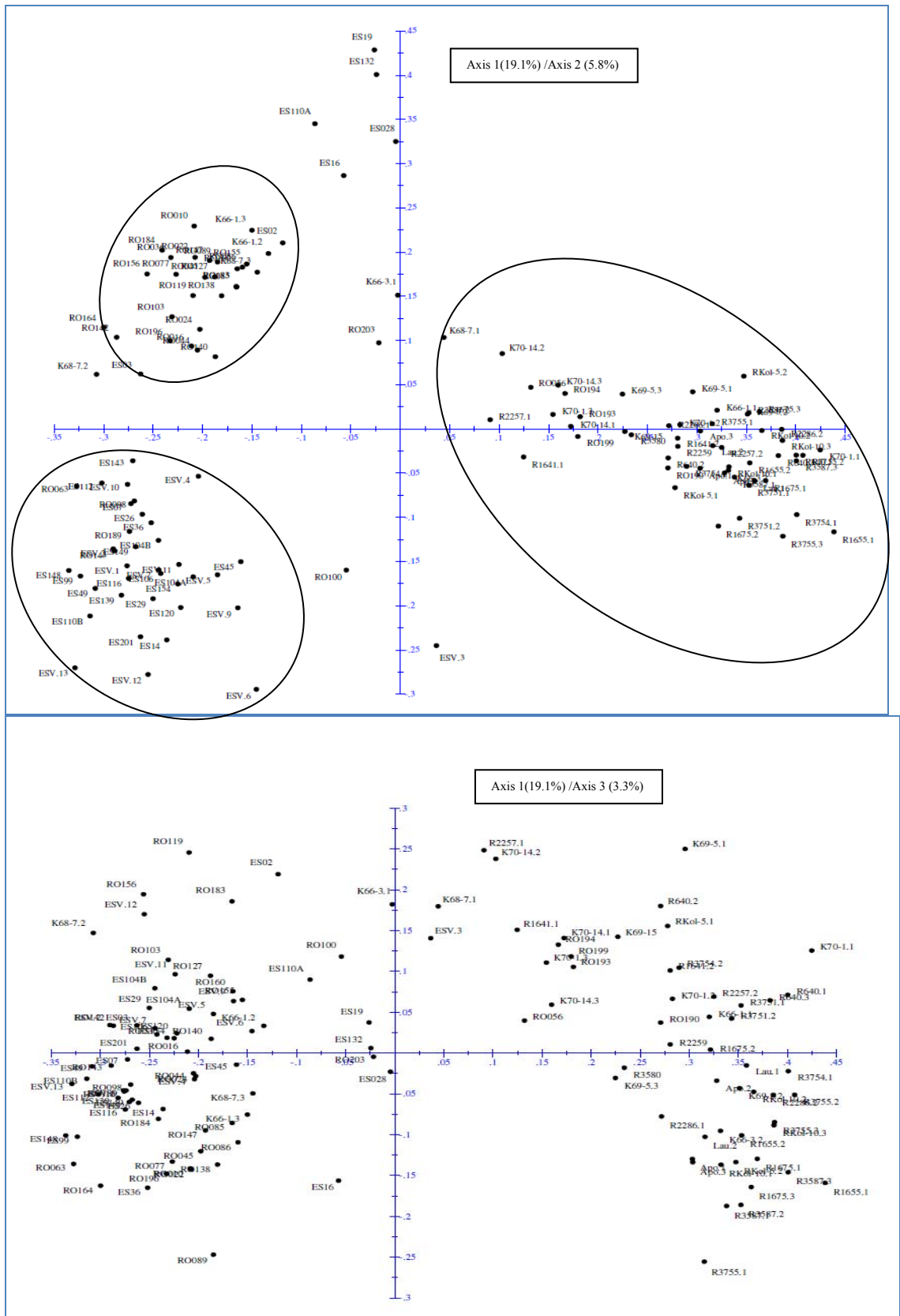


Figure 2. Principal coordinate analysis based on SSR data of 130 *C. canephora* accessions.

Relationships among *C. canephora* accessions and also with others species of the genus *Coffea* were evaluated with UPGMA clustering technique (Figure 3). The cophenetic correlation was high (81.5%), and the levels of stress and distortion were low (1.4% and 11.8%, respectively), demonstrating that the dendrogram satisfactorily represents the original matrix of dissimilarities. Fourteen nodes on the dendrogram presented high bootstrap values (>50%), indicating a strong consistency of respective subclusters. General distribution of the accessions, basically, corroborated the previous configuration observed on PCoA. *Coffea canephora* accessions were clearly clustered by their phenotype, and Kouillous were subclustered by location. Some misclassifications were observed: five accessions of Rondônia (RO098, RO189, RO143, RO100 and RO063) were positioned in the Espírito Santo cluster and two accessions of Espírito were clustered into Rondonian group (ES02 and ES03). Accessions of Rondônia were slightly more dissimilar (0.18 to 0.62) than Espírito Santo (0.10 to 0.53). Although good amount of variability was found, it has known that some genetic drift occurred early, during the Kouillou introduction, in the beginning of the last century, when a big amount of seeds from a few plants were imported to compose the first Brazilian plantations, on Southeast region (Ferrão et al. 2007a, Fazuoli et al. 2009). So, the diversity observed is a minimal sample of the natural species diversity. Moreover, this variability may be in risk due to the intense selection practiced by farmers and breeders across the decades. The use of vegetative propagation focused on a small set of genotypes (clones) may also contribute to aggravate this scenario.

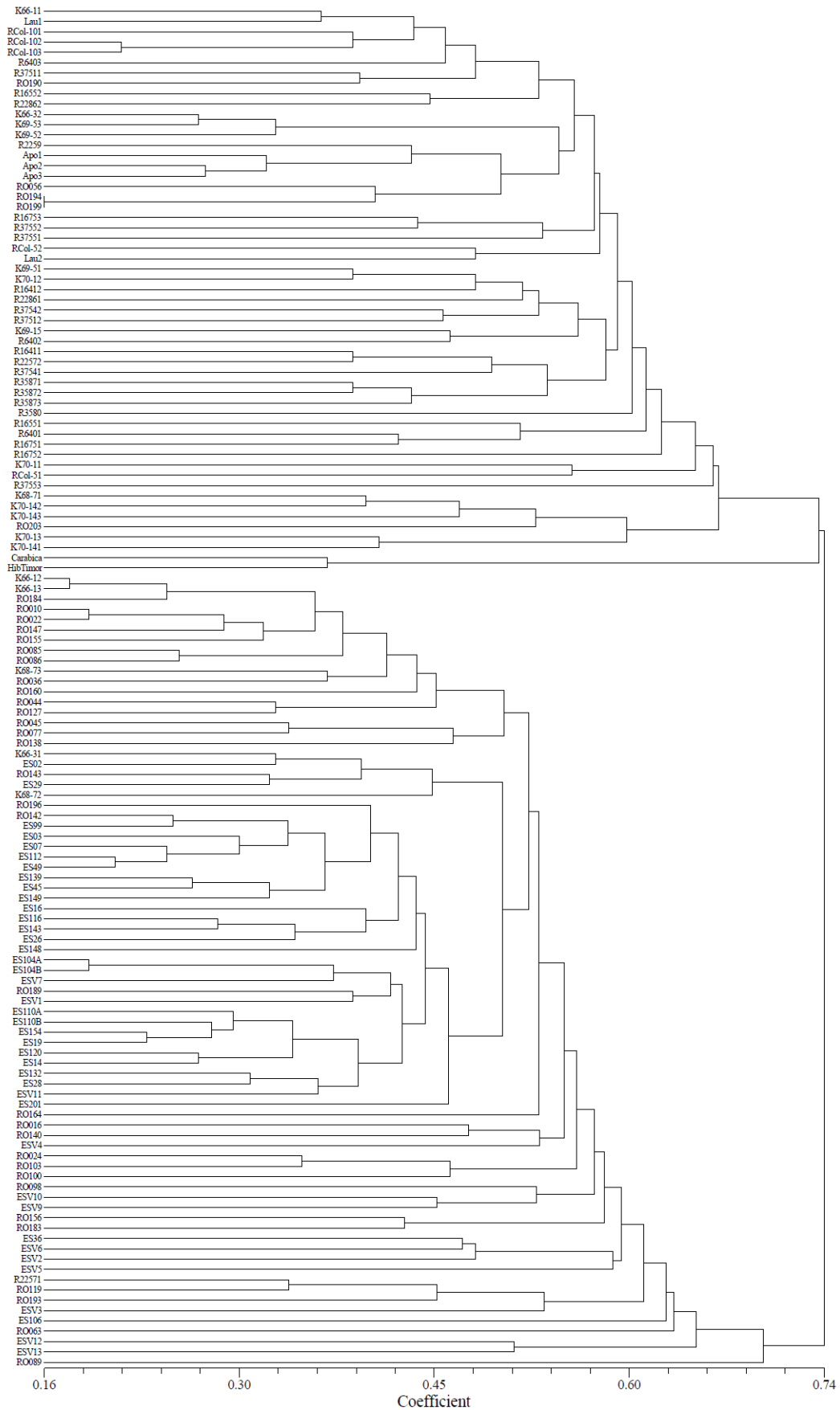


Figure 3. Dendrogram representing the dissimilarity among 130 *C. canephora* accessions, obtained by UPGMA method, based on the weighted coincidence index estimated over the polymorphism of 20 microsatellites.

In Robusta cluster, accessions presented a little more dissimilar (0.22 to 0.66) and no clear distinction was verified among those from IAC and UFV collections. This result expresses a wider genetic diversity of Robusta accessions. Prakash et al. (2005) also reported a high amount of diversity in a sample of the Indian Robusta coffee. Nevertheless, only one diversity group, from the five previously describe by Dussert et al. (1999), was represented in that genepool. It is possible that a similar situation takes place in Brazil, where the prior cycle of Robusta introductions occurred by collecting cultivated accessions from Indonesia (Fazuoli et al. 2009), likewise Indian case. Other part of the *C. canephora* accessions from IAC and UFV were introduced later, during 70's and 80's of last century, from the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) at Turrialba, Costa Rica, which received germplasm from FAO missions in African Continent (Fazuoli et al. 2009). During these expeditions, germplasm was collected in a wide geographical range and comprised wild and cultivated accessions (Montagnon 1998a). Even though, all of these accessions represent a minimal part of the natural variability of Robusta germplasm.

Notably, the accessions RO056, RO190, RO193, RO194 and RO199, from Rondônia; and ESV.3, from Espírito Santo, which are part of Kouillou collections, were grouped among Robusta accessions. Although this result seemed controversy, phenotypic data from field evaluations showed that those plants present intermediate characteristics between both Kouillou and Robusta groups, and for this reason, they had caused frequent doubts of classification. In this work, microsatellites SSRs were able to confirm the hybrid origin of those genotypes.

Considering that only six, in a total of 73 accessions collected in traditional cultivation areas, share alleles with Robusta genotypes, it is possible to suppose that

the presence of this germplasm is considerably reduced on the most commercial plantations. Thus, increasing the participation of Robusta alleles in Brazilian breeding programs would imply a lot of benefits, like the enhancement of resistance against leaf rust and nematode; raising productivity and improving beverage quality. Moreover, the exploitation of heterosis resulting from intergroup crosses can be very advantageous. This strategy has been used for so long in some breeding programs around the world, with remarkable success (Bouharmont et al. 1986, Leroy et al. 1997). Furthermore, 'Robusta' and 'Kouillou' are divergent heterotic groups with complementary characteristics. Robusta plants present high resistance to rust and nematodes, and give good beverage. On the other hand, Kouillou plants are tolerant to drought and they are easier to cultivate due to smaller size. So, these populations compose a perfect combination to use in a reciprocal recurrent selection program, as it has been already performed in Ivory Coast, since 1984 (Leroy et al. 1993, 1994 and 1997). Although higher genetic gains could be obtained using Guinean *versus* Congolese strategy, considerably progress has been also achieved between Congolese subgroups (Leroy et al. 1993).

Twelve accessions from the IAC collections labeled as "Kouillou" were grouped among the Robusta. This may be a consequence of spontaneous intercrosses between groups occurred on field at germplasm collection, once they are kept *ex situ*. Alternatively, it may be due to the own great variability of these populations. In a similar study, using RAPD markers, Silvestrini et al. (2008) also found some cases, in which accessions from Kouillou populations (i.e.: K68-02 and K69-13), conserved at IAC germplasm collection, were grouped together with Robusta genotypes and even with accessions of *C. congensis*.

Coffea arabica and Híbrido de Timor composed an internal subcluster, among *C. canephora* accessions. These genotypes were closer to Robusta cluster, suggesting that *C. arabica* shares more alleles with Robusta than Kouillou. This result points to a putative origin of *C. arabica*, from a spontaneous cross between a Robusta ancestor and another species. At the same way, Híbrido de Timor should have arisen from a *C. arabica* plant that received unreduced gametes from a Robusta plant. A Robusta genitor, instead Kouillou, may explain the remarkable resistance to leaf rust and nematodes observed on derivatives of Híbrido de Timor (Chaves 1976, Gonçalves and Pereira 1998). Missio et al. (2009) also observed that *C. arabica* accessions were more similar to Robusta than Kouillou. These results implies that further studies, which aim to obtain a better comprehension about ancestry origin of *C. arabica* and Híbrido de Timor, have to take in to account the diversity of *C. canephora* and other putative genitors, in order to avoid spurious conclusions. Moncada and McCouch (2004) studying *C. arabica* and Híbrido de Timor from different origins and their genetic relation to other species of *Coffea*, reported that H. de Timor shared more alleles with *C. arabica* than with the *C. canephora* sampled in that work. The authors enumerated some hypothesis, including the possibility of that *C. canephora* accession was not representative of the H. de Timor ancestral genitors. Tesfaye et al. (2007) suggest that the question of *C. arabica* origins remains unsolved. Using chloroplast microsatellites, they concluded that *C. arabica* has a recent origin in a unique allopolyploidization event or by severe bottleneck effects during its evolutionary history. Furthermore, *C. eugenioides* is not the maternal specie of *C. arabica*, in contrast with previous reports (Raina et al. 1998, Lashermes et al. 1999). With regard to the paternal parent, they recommend the analysis with nuclear

markers based on a representative sampling of putative *C. arabica* relatives, including their geographically different populations.

Molecular characterization allowed to assess the diversity of *C. canephora* gene pool in Brazil and added a small clue about the origin of *C. arabica* and Híbrido de Timor. For the first time a well-representative sample of accessions from cultivated areas and also conserved at the main coffee germplasm collections were taken together to analysis with microsatellites. These markers revealed a high degree of polymorphism, which allowed a satisfactory understanding of the genetic diversity among the *Coffea canephora* accessions. Moreover they allowed the proper identification of the main current phenotypic varietal groups and showed to be able to resolve doubts about the accession classification. This ability represents a great advantage, because the high intra-specific variability and the environmental effects produce a great amount of types, which can hinder the identification of Kouillous and Robustas based only on the phenotypic evaluation.

Despite the polymorphism found on Brazilian accessions, the diversity showed to be in risk, mainly in Kouillou group. Even Robusta sample presented a modest diversity in comparison with natural diversity of this group in Africa. Nowadays, Brazil plays a fundamental role in the *C. canephora* world production. So, a hurry action should be performed in order to enlarge the representation of natural diversity of this species on Brazilian gene banks. The establishment of new ways of interchanging with other collections around the world should be an interesting initiative to promote introduction of new accessions. For while, the present diversity has been enough to support the advances of Brazilian breeding programs and a new threshold is expected with the intensification of the use of Robusta germplasm on breeding programs.

ACKNOWLEDGEMENTS

Authors thank to Brazilian Consortium for Coffee Research and Development (Consortio Pesquisa Café) and Agrofuturo/Embrapa for the financial support.

RESUMO

O germoplasma de *Coffea canephora*, cultivado ou conservado nas instituições brasileiras de pesquisa agropecuária, apresenta grande variabilidade fenotípica, porém seu polimorfismo molecular continua precariamente estudado. Este trabalho objetivou caracterizar acessos de *C. canephora* oriundos de cultivos comerciais e bancos de germoplasma brasileiros. Um total de 130 acessos das coleções de São Paulo, Minas Gerais, e coletados em plantios comerciais no Espírito Santo e Rondônia foram genotipados com 20 *primers* microssatélites. Métodos multivariados foram utilizados para estimar a relação entre os acessos. Os microssatélites detectaram alto polimorfismo e dois grupos foram identificados. O primeiro foi constituído pelos genótipos Kouillou do Espírito Santo e Rondônia e o segundo pelos genótipos Robusta de São Paulo e Minas Gerais. Ambos apresentaram variabilidade expressiva, mas limitada. Assim, novos esforços devem ser realizados para aumentar a diversidade genética das coleções brasileiras. Ademais, os programas de melhoramento devem ampliar a participação dos alelos de Robusta nas futuras cultivares.

Palavras-chave: SSR, Café robusta, diversidade genética

REFERENCES

- Berthaud J (1986) Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application. Paris, ORSTOM, 379p.
- Bouharmont P, Lotodé R, Awemo J and Castaing X (1986) La sélection générative du caféier Robusta au Cameroun – Analyse des résultats d'un essai d'hybrides diallele partiel implanté en 1973. **Café Cacao Thé**, **30**: 93-112.
- Chaves GM (1976) Melhoramento do cafeeiro visando à obtenção de cultivares resistentes à *Hemileia vastatrix* Berk et Br. **Revista Ceres** **23**: 321-332.
- Cruz CD (2001) **Programa Genes** (versão Windows): aplicativo computacional em genética e estatística. Editora UFV, Viçosa, 648p.
- Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB, Zambolim E, 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.
- Dussert S, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes MC, Berthaud J, Noirot M, and Hamon S. (1999) Le caféier, *Coffea canephora*. In: Hamon P, Seguin M, Perrier X, Glaszmann JC. **Diversité génétique des plantes tropicales cultivées**. Montpellier: CIRAD, p.175-194.
- Fazuoli LC, Mistro JC and Braghini MT (2009) Melhoramento do café robusta no Instituto Agronomico de Campinas. In: Zambolim L (Ed). **Tecnologias para produção do café conilon**. Editora UFV, Viçosa, p. 201-247.
- Ferrão MAG, Ferrão RG, Fonseca AFA, Verdim Filho AC and Volpi PS (2007a) Origem, dispersão, taxonomia e diversidade genética de *Coffea canephora* In:

Ferrão RG, Fonseca AFA, Bragança SM, Ferrão MAG and De Muner LH (eds.)

Café conilon. Incaper, Vitória, p. 67-91.

Ferrão MAG, Fonseca AFA, Ferrão RG, Oliveira MA, Barbosa WM, D'isep MSP and Barbosa RP (2007b) Técnicas moleculares e biotecnológicas aplicadas ao café. In: Ferrão RG, Fonseca AFA, Bragança SM, Ferrão MAG and De Muner LH (eds.) **Café conilon.** Incaper, Vitória, p. 175-201.

Fonseca AFA, Sedyama T, Cruz CD, Sakiyama NS, Ferrão MAG, Ferrão RG and Bragança SM (2006) Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira** **41**: 599–605

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

Ivoglo MG, Fazuoli LC, Oliveira ACB, Gallo PB, Mistro JC, Silvarolla MB, Tomabraghini M (2008) Divergência genética entre progênes de café robusta. **Bragantia** **67**:823–831

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

Leroy T, Montagnon C, Charrier A and Eskes AB (1993) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. I. Characterization and evaluation of breeding populations and values of intergroup hybrids. **Euphytica** **67**: 113-125.

- Leroy T, Montagnon C, Cilas C, Charrier A and Eskes AB (1994) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. II. Estimation of genetic parameters. **Euphytica** **74**: 121-128.
- Leroy T, Montagnon C, Cilas C, Yapo A, Charmetant P and Eskes (1997) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. III. Genetic gains and results of first intergroup crosses. **Euphytica** **95**: 347-354.
- Maurin O, Davis PA, Chester M, Mvung EF, Jaufeerally-Fakim Y and Fay MF (2007) Towards a phylogeny for *Coffea* (Rubiaceae): Identifying well-supported lineages based on nuclear and plastid DNA sequences. **Annals of Botany**. 1-19.
- Missio RF, Caixeta ET, Zambolim EM, Pena GF, Ribeiro AP, Zambolim L, Pereira AA, Sakiyama NS, (2009) Assessment of EST-SSR markers for genetic analysis on coffee. **Bragantia** **68**: 573-581.
- Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome** **47**: 501-509
- Montagnon C, Leroy T and Yapo A (1992) Genotypic and phenotypic diversity of some coffee groups (*Coffea canephora* Pierre) in the collections – consequences on their use in breeding. **Café Cacao and Thé** **36**: 187-198.
- Montagnon C, Leroy T and Eskes AB (1998a) Amélioration variétale de *Coffea canephora*. II. Les programmes de sélection et leurs résultats. **Plantations, Recherche, Développement** **5**: 89-98.
- Montagnon C, Leroy T and Eskes AB (1998b) Amélioration variétale de *Coffea canephora*. I. Critères et méthodes de sélection. **Plantations, Recherche, Développement** **5**: 18-28.

- Perrier X, Flori A and Bonnot F (2003) Data analysis methods. In Genetic diversity of cultivated tropical plants. Science Publishers. Inc., Enfield, NH., p 43-76.
- Powell W, Machray GC and Provan J (1996) Polymorphism revealed by simple sequence repeats. **Trends Plant Science** 1: 215-222.
- Prakash NS, Combes MC, Dussert S, Naveen S, Lashermes P (2005) Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLP. **Genetic Resources and Crop Evolution** 52: 333-343.
- Raina SN, Mukai Y and Yamanoto M (1998) *In situ* hybridization identifies the diploid progenitor species of *C. arabica* (Rubiaceae). **Theoretical and Applied Genetics** 97: 1204-1209.
- Rohlf FJ (1998) **NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1**. New York: Exeter Software. 83p.
- Silvestrini M, Maluf MP, Silvarolla MB, Guerreiro-Filho O, Medina-Filho HP, Vanini MMT, Oliveira AS, Gaspari-Pezzopane C and Fazuoli LC (2008) Genetic diversity of a *Coffea* germplasm collection assessed by RAPD markers. **Genetic Resources and Crop Evolution** 55: 901-910.
- Souza FF and Santos MM (2009) Melhoramento genético do café canéfora em Rondônia In: Zambolim L (Ed). **Tecnologias para produção do café conilon**. Editora UFV, Viçosa, p. 175-200.
- Tesfaye K, Borsch T, Govers K and Bekele E (2007) Characterization of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. **Genome** 50: 1112-1129.

CAPÍTULO 2

Estudo da estrutura populacional do germoplasma brasileiro de *Coffea canephora*

Genetic structure of Brazilian *Coffea canephora* germplasm assessed by microsatellites

Abstract

This work aimed to evaluate the genetic diversity and structure of the Brazilian *Coffea canephora* germplasm, in order to achieve basic information to perform further genetic association studies. The survey was carried out using 20 microsatellites and 130 accessions from the four principal collections in Brazil. Parameters of genetic diversity and structure were estimated using Nei Statistics, Multivariate Analysis and Bayesian Analysis. High allelic and genotypic diversity was found. The germplasm presented strongly structured. Multivariate methods pointed to the presence of three genetic clusters and Bayesian approach basically confirmed this structure. As a consequence, further construction of an association map has to consider this genetic arrangement or may be restricted to a single collection in order to assure its reliability. All collections are able to be used in association studies, but samples from Espírito Santo and Rondônia represent better the cultivated germplasm in Brazil. Besides, it is really recommended the use of association-mapping methods that correct the linkage disequilibrium caused by familial relatedness and population structure.

Keywords: Microsatellites, Robusta Coffee, Nei's Statistics,

Introduction

The genus *Coffea* L. comprises over than 103 species (Davis et al. 2006), from which three are commercially exploited to produce the beverage known as ‘coffee’. *Coffea arabica* represents at least 63% from the world-wide production; *Coffea canephora* provides about 36% and *C. liberica* responses for less than 1%. *C. canephora*, generally known as ‘Robusta Coffee’, originated at West and Central Africa and, nowadays, is mainly grown at the tropical and subtropical countries (Berthaud and Charrier 1988) in Africa, Asia and South America. In industry, grains of Robusta coffee are mainly used in the production of instant coffee, blends with *C. arabica* and like source of caffeine to medicines, soft drinks and stimulants.

Coffea canephora has been cultivated since the last decades of the XIX century. First plantations were established in Africa, at Uganda and Congo. The beginning official efforts of breeding took place early in the last century at Java, where several varieties were developed and spread out in Southeast Asia and Central Africa. These genotypes soon replaced *C. arabica* that had been hardly affected by leaf rust caused by the fungus *Hemileia vastatrix*. Simultaneously, at West Africa, native populations traditionally grown in many locals assumed status of varieties and comprised, jointly with Java’s genotypes, an important germplasm, which was further introduced into many countries, including Brazil, in different occasions (Montagnon et al. 1998a).

For so long, Brazil has been recognized as the major producer of Arabica coffee, but the production of *Coffea canephora* is consistently increasing. This growing is supported by the results of the main genetic national breeding programs, and some cultivars have been developed to attend different demands (Fazuoli 1986;

Ferrão et al. 2007a). First cultivars were improved using phenotypic selection on the Robusta populations introduced in the germplasm collection of the Instituto Agrônomo de Campinas (IAC), in São Paulo (Fazuoli 1986). Lately, the Instituto Capixaba de Assistência Técnica, Pesquisa e Extensão Rural (Incaper), in Espírito Santo state, developed some clonal cultivars by means of an initial prospection and cutting of superior plants in Kouillou plantations. This massal selection was followed by rigorous evaluation assays under controlled conditions at the experimental stations in different places (Ferrão et al. 2007a). Analogous strategy was implemented in the genetic breeding program carried out in the Empresa Brasileira de Pesquisa Agropecuária (Embrapa), in Rondônia state (Souza and Santos 2009). Recently, inter-varietal hybrids (Robusta x Kouillou), obtained by crosses between pair of clones, have been tested and demonstrated high potential of use in commercial cultivation (Fazuoli 2008; Souza and Santos 2009).

Brazilian breeding programs of *C. canephora* are well supported by the genetic resources in germplasm collections. The first introduction cycle in IAC gene bank were initially accomplished from cultivated accessions from Indonesia (Fazuoli et al. 2009). After that, a big amount of accessions were introduced from the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) at Turrialba, Costa Rica in 1971 and 1974 (Fazuoli et al. 2009). The accessions of the Universidade Federal de Viçosa (UFV) were also introduced from Turrialba collection, in the same occasion. Accessions of Incaper were collected at farmers on Espírito Santo and a few others were introduced from IAC. Germplasm collection of Embrapa, at Rondônia, comprises more than 700 accessions of *C. canephora* (Souza and Santos 2009), including many progenies that came from collections of IAC, UFV and Incaper. Each accession is maintained with at least 10 individuals. Remain

accessions are clones and progenies from plants sampled in several plantations at traditional producing coffee areas at Rondônia state. Consequently, Embrapa collection harbors representative samples of the most *Coffea canephora* germplasm in Brazil.

Some accessions in those gene banks were characterized by means of morphologic (Ivoglio et al. 2008; Souza and Santos 2009), agronomic (Gaspari-Pezzopane et al. 2004; Fonseca et al. 2006; Souza and Santos 2009) and molecular (Ferrão et al. 2007b; Silvestrini et al. 2008) traits. Nevertheless, a better knowledge about the genetic diversity and structure of the available germplasm is still necessary and may promote a better conservation and management of these genetic resources and, also increase the efficiency of the breeding programs. Besides, information about genetic diversity and population structure is very important to choose appropriated samples and elect the better technique in order to perform association maps. This approach is based on linkage disequilibrium (DL) and has been successfully employed in genetics to detect QTL coding for simple and complex human diseases (Corder et al. 1994; Kerem et al. 1989). Associate mapping is arising like a promising alternative to map QTLs in plants, overcoming the limitations of conventional linkage mapping (Kraakman et al. 2004). The method is performed using natural populations, accessions of germplasm collections or even breeding lines. It also permits the use of pre-existent data accumulated on phenotypic evaluations. However, the success of this approach depends on the understanding of population structure (López-Gartner et al. 2009). Ignore it may lead one to find spurious marker-trait associations (Pritchard et al. 2000).

Different molecular techniques have been used to assess the genetic diversity in *Coffea canephora*. Initially, studies with wild and cultivated germplasm from

Africa, using isoenzymes revealed the existence of two principal groups, *i.e.*, ‘Guinean’, from West Africa, and ‘Congolese’, from Central Africa, with a high genetic diversity and strong geographic structure (Berthaud 1986). The existence of five diversity groups was later revealed by analysis of Restriction Fragment Length Polymorphism (RFLP) probes (Dussert et al. 1999). The advent of polymerase chain reaction (PCR) marker techniques, *e.g.*, Randomized Amplified Polymorphism of DNA (RAPD), Simple Sequence Repeats (SSR, Microsatellites), Amplified Fragment Length Polymorphism (AFLP), *etc.*, has increased the number of genetic studies with *C. canephora*. There are a few papers reporting the use of these tools in the survey of genetic diversity or population structure in this species, in Africa (Dussert et al. 1999; Cubry 2009, Gomez et al. 2009) and Asia (Prakash et al. 2005). However, this issue is still weakly approached on South American germplasm, where most efforts are focused on the study of *C. arabica*. Analysis of RAPD demonstrated a fairly high and structured genetic diversity in accessions of *C. canephora* from IAC collection (Silvestrini et al. 2008).

This study aimed to evaluate genetic diversity and structure of *Coffea canephora* in Brazil, in order to establish a strategy to construct an association map, exploiting the appropriated germplasm and available data, already achieved by the Brazilian gene banks and/or breeding programs.

Material and methods

Plant material and DNA extraction

A total of 130 accessions of *C. canephora*, maintained at the Embrapa Coffee Germplasm Collection, at Rondônia State, were used for genetic analysis (Table 1).

Table 1 List of accessions genotyped in this study.

Accession	Code	GB ¹	Accession	Code	GB ¹	Accession	Code	GB ¹
Kouillou IAC66-1.1	K66-11	1	Robusta UFV 3587.3	R35873	2	Encapa 07	ES07	4
Kouillou IAC 66-1.2	K66-12	1	Robusta UFV 3751.1	R37511	2	Encapa 14	ES14	4
Kouillou IAC 66-1.3	K66-13	1	Robusta UFV 3751.2	R37512	2	Encapa 16	ES16	4
Kouillou IAC 66-3.1	K66-31	1	Robusta UFV 3754.1	R37541	2	Encapa 19	ES19	4
Kouillou IAC 66-3.2	K66-32	1	Robusta UFV 3754.2	R37542	2	Encapa 28	ES28	4
Kouillou IAC 68-7.1	K68-71	1	Robusta UFV 3755.1	R37551	2	Encapa 104A	ES104A	4
Kouillou IAC 68-7.2	K68-72	1	Robusta UFV 3755.2	R37552	2	Encapa 104B	ES104B	4
Kouillou IAC 68-7.3	K68-73	1	Robusta UFV 3755.3	R37553	2	Encapa 106	ES106	4
Kouillou IAC 69-15	K69-15	1	Cpafro 010	RO010	3	Encapa 110A	ES110A	4
Kouillou IAC 69-5.1	K69-51	1	Cpafro 016	RO016	3	Encapa 110B	ES110B	4
Kouillou IAC 69-5.2	K69-52	1	Cpafro 022	RO022	3	Encapa 112	ES112	4
Kouillou IAC 69-5.3	K69-53	1	Cpafro 024	RO024	3	Encapa 116	ES116	4
Kouillou IAC 70-1.1	K70-11	1	Cpafro 036	RO036	3	Encapa 120	ES120	4
Kouillou IAC 70-1.2	K70-12	1	Cpafro 044	RO044	3	Encapa 132	ES132	4
Kouillou IAC 70-1.3	K70-13	1	Cpafro 045	RO045	3	Encapa 139	ES139	4
Kouillou IAC 70-14.1	K70-141	1	Cpafro 056	RO056	3	Encapa 143	ES143	4
Kouillou IAC 70-14.2	K70-142	1	Cpafro 063	RO063	3	Encapa 148	ES148	4
Kouillou IAC 70-14.3	K70-143	1	Cpafro 077	RO077	3	Encapa 149	ES149	4
Laurenti.1	Laur1	1	Cpafro 085	RO085	3	Encapa 154	ES154	4
Laurenti.2	Laur2	1	Cpafro 086	RO086	3	Encapa 201	ES201	4
Apoatã IAC 2258.1	Apo1	1	Cpafro 089	RO089	3	Encapa 26	ES26	4
Apoatã IAC 2258.2	Apo2	1	Cpafro 098	RO098	3	Encapa 29	ES29	4
Apoatã IAC 2258.3	Apo3	1	Cpafro 100	RO100	3	Encapa 36	ES36	4
Robusta IAC 640.1	R6401	1	Cpafro 103	RO103	3	Encapa 45	ES45	4
Robusta IAC 640.2	R6402	1	Cpafro 119	RO119	3	Encapa 49	ES49	4
Robusta IAC 640.3	R6403	1	Cpafro 127	RO127	3	Encapa 99	ES99	4
Robusta IAC 1641.1	R16411	1	Cpafro 138	RO138	3	Encapa V.1	ESV1	4
Robusta IAC 1641.2	R16412	1	Cpafro 140	RO140	3	Encapa V.2	ESV2	4
Robusta IAC 1655.1	R16551	1	Cpafro 142	RO142	3	Encapa V.3	ESV3	4
Robusta IAC 1655.2	R16552	1	Cpafro 143	RO143	3	Encapa V.4	ESV4	4
Robusta IAC 1675.1	R16751	1	Cpafro 147	RO147	3	Encapa V.5	ESV5	4
Robusta IAC 1675.2	R16752	1	Cpafro 155	RO155	3	Encapa V.6	ESV6	4
Robusta IAC 1675.3	R16753	1	Cpafro 156	RO156	3	Encapa V.7	ESV7	4
Robusta IAC 2257.1	R22571	1	Cpafro 160	RO160	3	Encapa V.9	ESV9	4
Robusta IAC 2257.2	R22572	1	Cpafro 164	RO164	3	Encapa V.10	ESV10	4
Robusta IAC 2259.1	R2259	1	Cpafro 183	RO183	3	Encapa V.11	ESV11	4
Robusta IAC 2286.1	R22861	1	Cpafro 184	RO184	3	Encapa V.12	ESV12	4
Robusta IAC 2286.2	R22862	1	Cpafro 189	RO189	3	Encapa V.13	ESV13	4
Robusta Col - 10.1	RCol-101	1	Cpafro 190	RO190	3			
Robusta Col - 10.2	RCol-102	1	Cpafro 193	RO193	3			
Robusta Col - 10.3	RCol-103	1	Cpafro 194	RO194	3			
Robusta Col - 5.1	RCol-51	1	Cpafro 196	RO196	3			
Robusta Col - 5.2	RCol-52	1	Cpafro 199	RO199	3			
Robusta UFV 3580	R3580	2	Cpafro 203	RO203	3			
Robusta UFV 3587.1	R35871	2	Encapa 02	ES02	4			
Robusta UFV 3587.2	R35872	2	Encapa 03	ES03	4			

¹Gene Bank: (1) Instituto Agronômico de Campinas (IAC), São Paulo; (2) Universidade Federal de Viçosa (UFV), Minas Gerais; (3) Embrapa, Rondônia; and (4) Instituto Capixaba de Assistência Técnica, Pesquisa e Extensão Rural (INCAPER), Espírito Santo.

These accessions represent either the germplasm commercially cultivated, used in breeding programs, as so conserved at the main Brazilian gene banks. We used 43, 11 and 40 accessions from IAC (São Paulo), UFV (Minas Gerais) and Incaper (Espírito Santo) collections, respectively. The 36 remain accessions were collected in traditional coffee producing areas at Rondônia State. Young and completely extended leaves were collected from each accession, frozen at $-80\text{ }^{\circ}\text{C}$, lyophilized, ground until becoming a fine powder and kept at $-20\text{ }^{\circ}\text{C}$ until used. Genomic DNA was extracted from ground tissue of each accession, using the method described by Diniz et al. (2005) and all DNA samples were prepared to a final concentration of $25\text{ ng}\cdot\mu\text{L}^{-1}$.

SSR markers, PCR Reaction and electrophoresis

The twenty microsatellites used in this study were developed from non-redundant Express Sequence Tags (EST) of the Brazilian Coffee Genome Project. Each reaction was set to a final volume of $20\text{ }\mu\text{L}$, containing 50 ng of genomic DNA, 0.6 unit of *Taq* DNA polymerase, buffer 1x, 1 mM of MgCl_2 , $150\text{ }\mu\text{M}$ of each dNTP and $0.1\text{ }\mu\text{M}$ of each primer. PCR amplifications were carried out using touchdown proceeding, which comprises initial denaturation at $94\text{ }^{\circ}\text{C}$ for 2 min , followed by 10 cycles of denaturation at $94\text{ }^{\circ}\text{C}$ for 0.5 min , annealing at $67\text{ }^{\circ}\text{C}$ for 0.5 min , decreasing 1°C after each cycle, and extension at $72\text{ }^{\circ}\text{C}$ for 0.5 min . After that, another set of 30 cycles, comprising denaturation at $94\text{ }^{\circ}\text{C}$ for 0.5 min , annealing at $55\text{ }^{\circ}\text{C}$ for 0.5 min and extension at $72\text{ }^{\circ}\text{C}$ for 0.5 min , was accomplished followed by a final 8 min extension time at $72\text{ }^{\circ}\text{C}$. Before running, PCR products were denatured in $8\text{ }\mu\text{L}$ of denaturing dye (95% formamid) at $94\text{ }^{\circ}\text{C}$ for 5 min and $7\text{ }\mu\text{L}$ of sample were loaded on a standard 6% polyacrylamide gel at $50\text{ }^{\circ}\text{C}$ and run at a constant

power of 90 W for about 2 h. Post-PCR multiplex, which involved the multi-loading of individual PCR assays (two to four SSRs per running), was performed spacing successive loads by 10 to 30 min during electrophoresis, depending on the prior information about fragment size. At last, the gel was treated with ethanol (10%) + acetic acid (1%), followed by nitrate acid (1,5%); stained with silver nitrate (4%), developed with sodium carbonate (3%) + formaldehyde (0.03%); and fixed with acetic acid (5%) and dried for posterior analysis at a transilluminator apparatus.

Data Analysis

Microsatellite alleles were scored in all 130 accessions using a molecular weight standard ladder and compared to the expected size of the microsatellite sequence. The evaluation of each locus was performed considering as homozygotes and heterozygotes, individuals which showed one or two alleles, respectively. CONVERT software (Glaubitz 2004) was used in order to facilitate the manipulation of data set through different statistical programs. The software Genetix 4.03 (Belkhir 2001) was used to estimate the number of alleles per locus and heterozygosity parameters. The polymorphism information content (PIC) of each SSR was calculated conforming Botstein et al. (1980), using the formula: $PIC = 1 - \sum_{i=1}^a p_i^2 - \sum_{i,j=1}^a \sum_{i \neq j} p_i^2 p_j^2$, where p_i and p_j are the frequencies of i^{th} and j^{th} alleles at a given locus and a is the number of alleles.

Gene diversity statistics of Nei (1978) and genetic structure of the groups was further investigated by AMOVA, using ARLEQUIN 3.1 (Excoffier et al. 2005). A Factorial Correspondence Analysis, based on the modified Roger's distance (Goodman and Stuber, 1983) was performed, using the software Genetix 4.03 (Belkhir 2001).

Population structure of *C. canephora* was also assessed by means of a Bayesian clustering analysis performed by the software STRUCTURE (Pritchard et al. 2000). In this software, multilocus genotypes are used to assign fractions of genetic ancestry of each individual into genetic clusters based on a predefined number of K clusters. The allelic frequencies within each cluster are in Hardy-Weinberg equilibrium and the loci are not linked. K was estimated using a Monte Carlo Markov chain (MCMC) approach for each K, after a 50,000 burn-in period and a length of 100,000 iterations. The posterior log-likelihood of K was estimated for k = 2 to 12 genetic clusters with 20 independent runs at each k. It was used the admixture model of ancestry and specified that allele frequencies were correlated, without prior population information. The true number of genetic clusters was estimated by calculating ΔK , as described by Evanno et al. (2005). Individuals were then assigned to a single cluster if their membership coefficient in that cluster was ≥ 0.70 .

Results

Molecular diversity

Molecular diversity analysis of *Coffea canephora* accessions was performed using 20 polymorphic SSR (Table 2). A total of 104 alleles were recorded across the 130 accessions, and the number of alleles per locus ranged from two (EST-SSR096) to 11 (EST-SSR002), with an average of 5.2. Alleles size across all loci ranged from 92 to 284 bases pair (bp) and the higher range was observed at loci EST-SSR022 (74 bp).

Table 2 Genetic diversity parameters estimated in 20 polymorphic microsatellites in four *Coffea canephora* germplasm collections in Brazil.

Locus	K	Size Range (bp)	Ho	He	PIC	H'
EST-SSR001	7	147 – 166	0.546	0.829	0.805	1.792
EST-SSR002	11	132 – 159	0.600	0.829	0.813	2.044
EST-SSR003	9	218 – 250	0.185	0.787	0.761	1.823
EST-SSR010	4	120 – 127	0.385	0.618	0.547	1.089
EST-SSR014	3	157 – 162	0.315	0.532	0.473	0.908
EST-SSR019	5	199 – 218	0.446	0.742	0.701	1.460
EST-SSR022	4	92 – 166	0.285	0.519	0.477	0.980
EST-SSR025	5	106 – 116	0.223	0.581	0.543	1.160
EST-SSR026	4	222 – 228	0.246	0.686	0.623	1.222
EST-SSR027	3	233 – 263	0.146	0.463	0.383	0.746
EST-SSR030	6	193 – 216	0.492	0.722	0.679	1.433
EST-SSR036	5	264 – 272	0.323	0.522	0.490	1.057
EST-SSR039	7	196 – 218	0.215	0.340	0.331	0.814
EST-SSR058	4	200 – 208	0.339	0.613	0.536	1.072
EST-SSR067	4	109 – 116	0.369	0.690	0.642	1.280
EST-SSR069	4	92 – 106	0.231	0.481	0.443	0.917
EST-SSR074	9	250 – 283	0.615	0.706	0.670	1.509
EST-SSR096	2	153 – 157	0.077	0.500	0.375	0.693
EST-SSR097	4	240 – 284	0.300	0.685	0.624	1.232
EST-SSR108	4	120 – 128	0.285	0.492	0.441	0.877
Mean	5.2		0.331	0.617	0.568	1.205

Number of alleles per locus (k); observed heterozygosity (Ho); Genetic diversity (He); polymorphic information content (PIC); Shannon index (H').

The observed heterozygosity ranged from 0.077 (EST-SSR096) to 0.615 (EST-SSR074) and the expected heterozygosity, here assumed as a genetic diversity parameter, ranged from 0.340 (EST-SSR039) to 0.829 (EST-SSR001 and EST-SSR002). Loci informativeness was evaluated using the polymorphic information content (PIC). This parameter represents the value of a marker for detecting within population polymorphism. It depends on the number of detectable alleles and the distribution of their frequency (Guo and Elston 1999). PIC ranged from 0.331 (EST-SSR039) to 0.813 (EST-SSR002). The average of genotypic diversity, inferred by Shannon index (H'), was 1.205. This diversity was higher at locus EST-SSR002, (H'= 2.044), and lower at locus EST-SSR096 (H'=0.693). This last locus presented two alleles and the three possible genotypes were found. Nevertheless, the genotypic frequencies were very unbalanced (56.0%, 35.0% and 9.0%), what contributed to obtain a low H'.

Collection diversity, differentiation and genetic distance

Genetic diversity in main Brazilian *Coffea canephora* gene banks was evaluated (Table 3). On the IAC collection, composed by 43 accessions, it was observed the highest mean number of alleles per locus (4.85); the highest genetic diversity ($H_e=0.589$) and highest genotypic diversity ($H'=1.149$). On the other hand, the Incaper collection, comprised by 40 individuals, showed lower values for all diversity parameters. Collections from UFV and Embrapa presented intermediary values of diversity.

The Nei's distance was also estimated for all pair of collections (Table 3). The highest distance was observed between the samples of UFV and Incaper (0.753) and the smallest between UFV and IAC samples (0.096). Embrapa collection presented intermediary distance values, being closer to Incaper collection.

Table 3 Genetic diversity and genetic distance in four Brazilian germplasm collections of *Coffea canephora*, assessed by the polymorphism evaluated in 20 SSR.

Collection	N	A	Ho	He	H'	Nei's distance			
						IAC	UFV	Embrapa	Incaper
IAC	43	4.85	0.405	0.589	1.149	-	0.096	0.326	0.492
UFV	11	3.80	0.423	0.523	0.986		-	0.581	0.753
Embrapa	36	4.15	0.278	0.508	0.945			-	0.097
Incaper	40	3.75	0.275	0.449	0.824				-
Mean			0.345	0.517	0.976				

Germplasm bank (GB); Number of accessions per collection (N); mean number of alleles per loci (A); observed heterozygosity (Ho); Expected Heterozygosity; Genetic diversity (He); Shannon index (H').

The results of AMOVA (Excoffier et al. 2005) point that genetic differences within groups contributed more to the global diversity, with between-group variance, being highly significant ($P>0.001$). In this analysis, about 36% of the total molecular variance was due to the differences among collections, whereas 64% was located into those groups.

Distribution of all accessions in a bi-dimensional plan, based on a factorial correspondence analysis allowed a better comprehension about the diversity within and among collections (Figure 1).

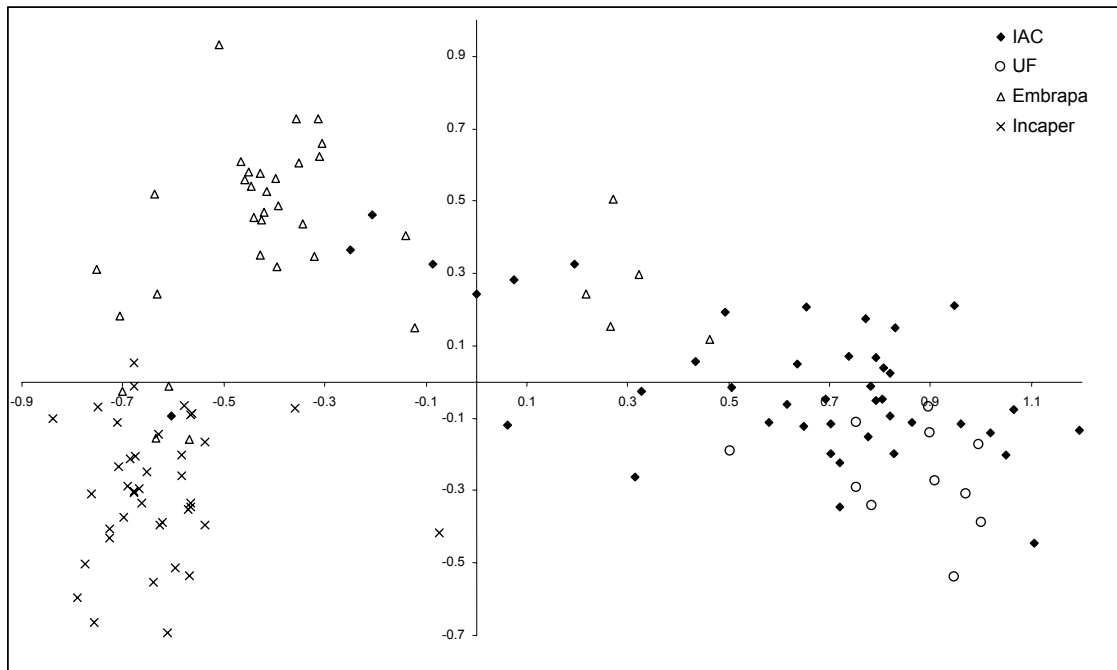


Figure 1

Factorial correspondence analysis representing the diversity of four *C. canephora* Brazilian collections based on the polymorphism observed in a set of 20 microsatellites.

Basically, three clusters of accessions were formed. First was composed mainly by individuals from Incaper collection; second comprised accessions from Embrapa and the third group was more diverse and it included most accessions from IAC and all accessions from UFV. Accessions from IAC were widely dispersed, while accessions from Incaper presented strongly concentrated. Accessions of UFV presented merged with those from IAC, suggesting they belong to the same genetic background.

Genetic structure

Using Evanno's method over STRUCTURE results, four clusters ($k=4$) were inferred, for which the highest likelihood scores were obtained with estimated Ln

probability of -4276.4 ; the mean value of Ln likelihood was -4159.4 and the alpha mean value was 0.04 . This value of alpha, so close to zero, indicates that most individuals essentially belong to one population or another. Alpha values close to one would mean that most individuals were admixed (Falush et al. 2003). Estimated membership coefficients for each collection are presented in Table 4.

Table 4 Gene proportion of *C. canephora* collections in each of the four clusters inferred by Bayesian analysis

Collection	Structure Inferred clusters				Number of Individuals
	I	II	III	IV	
IAC	0.331	0.102	0.531	0.036	43
UFV	0.161	0.008	0.821	0.010	11
Embrapa	0.109	0.642	0.046	0.202	36
Incaper	0.022	0.024	0.007	0.948	40

IAC collection presented 53.1% at cluster III, 33.1% at cluster I and 10.2% at cluster II. UFV accessions also had the most proportion at cluster III (82.1%) and a minor part at cluster I (16.1%), demonstrating a very similar genetic profile with IAC collection, which is in accordance with the results of genetic differentiation and genetic distance presented before. Embrapa collection presented the most diverse ancestry origins: 64.2% at cluster II, 20.2% at cluster IV and 10.9% at cluster I. Incaper collection had the most concentrated membership, with 94.8% at the cluster IV. The number of individuals per cluster, considering a 70% of membership to include the accession in to the cluster, is presented on Table 5.

Table 5 Number of *C. canephora* accessions per Bayesian cluster, considering 70% of membership.

Collection	Inferred clusters (>70% membership)					Number of Individuals
	I	II	III	IV	AD	
IAC	12	3	23	1	4	43
UFV	1	0	9	0	1	11
Embrapa	3	22	0	6	5	36
Incaper	0	0	0	38	2	40
Total	16	25	32	45	12	130

Cluster I was composed by 16 accessions, being 12 from IAC, three from Embrapa and one from UFV. Cluster II was comprised for 22 accessions from Embrapa and three from IAC. Cluster III was composed by 23 accessions from IAC and nine from UFV. Cluster IV, which harbors 45 accessions, was formed basically by individuals from Incaper collection (38) Six accessions from Embrapa were also included. The set of admixtures were comprised by 12 accessions, being Embrapa and IAC the collections with larger number of individuals (5 and 4, respectively). Therefore, the criterion of 0.7 of membership made possible to classify unambiguously about 91% of the accessions into the four structured groups.

The bar graphics (Figure 2) provides an overview about the distribution of ancestry in the accessions surveyed. Colors red and blue (clusters I and III) indicate Robusta ancestry, while green and yellow (clusters II and IV) are associated with Kouillou. So, individuals with mixed ancestry (admixture) were considered as hybrids. IAC and UFV collections are composed basically of Robusta genotypes. A few Kouillou accessions are also found in IAC. On the other hand, Kouillou genotypes are predominant in Embrapa and Incaper collections. Some Robusta and intergroup hybrids are also seen in Embrapa (Figure 2A). The ordering accessions by percent of ancestry (Figure 2B) shows a largest presence of Kouillou genotypes.

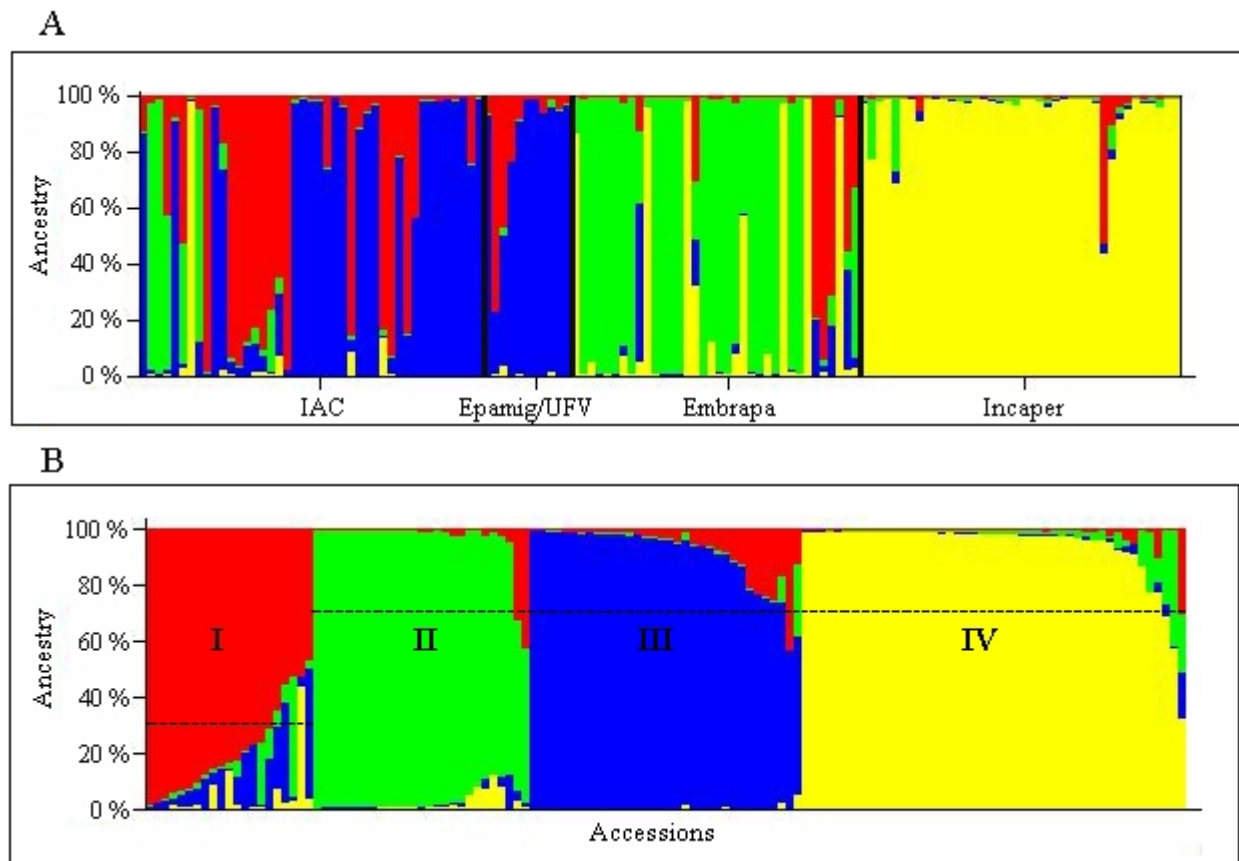


Figure 2

Genetic structure in Brazilian *Coffea canephora* germplasm, inferred by Bayesian analysis of 20 polymorphic microsatellite loci, in 130 accessions. Two overviews of the accession ancestry ordinated by (A) original collection and (B) by percent of ancestry.

Discussion

Prior studies had reported the phenotypic (Fonseca et al. 2006; Ivoglo et al. 2008) and the molecular (Silvestrini et al. 2008) diversity of Brazilian *C. canephora* germplasm. Nevertheless, since each institution has evaluated only its own accessions separately, overall information remained lacking. So, this is the first report where genotypes from the mainly coffee research institutions are assembled and analyzed simultaneously. Therefore, the accessions used in this work represent either the germplasm of *C. canephora* commercially cultivated and conserved in the

Brazilian plantations and gene banks. Here, we also presented the first results of a microsatellite set designed from the non-redundant expressed sequence tags (EST) of Brazilian Coffee Genome Project, in an evaluation of a wide sample of *Coffea canephora* accessions.

Despite microsatellites tested in this study had been originally developed from *C. arabica*, most of them presented highly polymorphic and shown to be very suitable to study the diversity in *C. canephora*. This remarkable polymorphism can be attributed to the outcross reproduction system and the diverse geographical origin of this species. As regard to the high transferability observed in our results, former studies also have demonstrated high rates of microsatellites transferability among different species in the genus *Coffea* (Poncet et al. 2004; 2007).

The number of alleles per locus, which ranged from two to 11, was very similar to other studies that used different sets of microsatellite and diverse germplasm (Poncet et al. 2004; Moncada and McCouch 2004; Prakash et al. 2005). A total of 104 alleles were detected and most of them were presented in the IAC, suggesting that this collection contain most of the variability present in the evaluated germplasm. In fact, IAC coffee gene bank is the first established in Brazil and nowadays, is one of the most representative *Coffea* collections conserved *ex situ* around the world (Silvestrini et al. 2008). It harbors accessions from 14 species of genus *Coffea* and three species of the genus *Psilanthus* (Medina-Filho et al. 2007). Moreover, IAC germplasm collection has supplied with a lot of accessions the more recent collections, and thus contributing to the mainly breeding programs in different Brazilian research institutions. On the other hand, lower diversity estimatives were observed on Incaper collection, suggesting more uniformity within group.

The observed heterozygosity across all loci ($H_o=0.331$), which represents the actual heterozygosity, was analogous to that reported by Poncet et al (2004) and Gomez et al (2009) ($H_o=0.380$, in both reports), in different *C. canephora* populations. Likewise, genetic diversity ($H_e = 0.51$) were similar to those found by Poncet et al. (2007) ($H_e=0.51$) and Gomez et al. (2009) ($H_e=0.65$).

IAC collection presented highest genetic diversity ($H_e=0.589$) and Incaper the smallest ($H_e=0.449$), but all collections presented H_e lower than H_o , that indicates a putative deficit of heterozygotes. However, considering *Coffea canephora*, as the most diploids species of *Coffea*, presents gametophytic self-incompatibility controlled by a single locus with multiple alleles (Berthaud 1980), a high level of heterozygosity would be expected. Poncet el (2004) found similar results in others *C. canephora* populations and attributed the heterozygote deficiency to the population factors. Cocoa (*Theobroma cacao* L.) is also a self-incompatible species and have presented heterozygote deficit in many studies (N’Goran et al. 2000; Sereno et al. 2006; Efombagn et al. 2008). These results have been attributed to the genetic drift, sub-grouping or sampling-related individuals. Small populations favor crosses between relatives increasing the homozygosity in the subpopulations. In the case of Brazilian *C. canephora* germplasm, genetic drift occurred since the first introduction cycles, in the beginning of the last century. A big amount of seeds from a few plants were imported to compose the first Brazilian plantations, in Rio de Janeiro and Espírito Santo (Ferrão et al. 2007a, Fazuoli et al. 2009). Additionally, in a wild status at Africa, *C. canephora* populations usually occur isolated and they are composed by a few mother trees (Berthaud 1986), what naturally would favor the augment of inbreeding. Nevertheless, it is very probable that homozygote excess observed in these collections is mainly due to Wahlund effect, as a consequence of the population

subdivision. In this case, the comparison between observed and expected heterozygosity is better interpreted as an important estimate about differentiation level of the subpopulations (Cruz et al. 2011), instead an inbred parameter.

All approaches converged to indicate the allocation of most genetic diversity within groups instead between groups. Similar result has been reported in other outcross perennial species (Sereno et al. 2006; Ronning and Schnell 1994; N’Goran et al. 2000) and it suggests a presence of strong intra-population structure with small differentiation between populations. These results are in accordance with Hamrick (1992), which reported that about 80% of the total diversity in tropical plant species occurs into the subpopulations.

According to the Bayesian analysis, the accessions were classified in four structured clusters ($k=4$). These groups did not match completely with prior given collections. Accessions from IAC and Embrapa comprised a separated cluster (I) and remain accessions of IAC presented merged with those from UFV, which composed the cluster III. The observed structure was basically correlated with well known phenotypic grouping and with historic of cultivation of the species in Brazil. In the former classification proposed by Berthaud (1986), *Coffea canephora* diversity was divided in two mainly geographical groups, *i.e.* Guinean and Congolese. Guinean group was kept restricted to Guinea and Ivory Coast. On other hand, Congolese group, which consists of, at least, four phenotypic subgroups (Dussert et al. 1999) has gotten out from Africa and spread out into producing countries around the world, including South America (Montagnon et al. 1998b). In Brazil, subgroups 1 and 2 (SG1 and SG2) are represented by the populations known as Conilon (deformation of Kouillou) and Robusta, respectively. However, the most Brazilian plantations of *Coffea canephora* are composed by genotypes from Kouillou. Historically, the

advantages of growing Robusta genotypes have been informed precariously among coffee farmers. Nevertheless, all governmental coffee research institutions maintain germplasm stocks of both groups. In the present study, the accessions comprise as advanced generation offsprings from individuals collected at Africa and Indonesia, as clones from open pollination progenies selected in commercial plantations in Brazil.

The structured clusters I and III correspond to Robusta genotypes (stocked in gene banks, rarely found in commercial plantations) and the clusters II and IV comprise Kouillou genotypes, cultivated at Rondônia and Espírito Santo states, respectively. Kouillou accessions from Espírito Santo (cluster IV) were clearly distinct from Rondônia ones and composed a homogeneous and well defined cluster, where they share more than 94% ancestry. This uniformity may be a consequence of an intensive process of artificial selection allied to a predominant clonal propagation strategy adopted in Espírito Santo. This scenario may represent a risk to coffee business, with some consequences to Brazilian production, once that state responses to 70% of national production. So, efforts of collect and conservation of *C. canephora* accessions must be emphasized on areas where germinative propagation is still the predominant system.

Rondonian germplasm (cluster II) began to be formed at the last four decades by means of the unofficial introduction of seeds and clonal plantlets coming from Espírito Santo and São Paulo. Successive cycles of artificial selection and new introduction events would not enough to explain the remarkable differences from this germplasm to the original source at Espírito Santo. Considering a mean generation time of about 10 years, only three or four generations would have advanced since *C. canephora* was introduced in Rondônia. Thus, probably, this differentiation may be caused by a sampling phenomenon like founder effect.

Understanding the germplasm variability and the genetic relationship between the accessions will facilitate the management of genetic resources and increase the efficiency of the breeding programs. Besides, a better comprehension about the population structure is necessary to choose a suitable germplasm to association mapping. When the research aim to find genes using association mapping, the presence of undetected population structure may support spurious associations and thus nullify standard tests (Ewens and Spielman, 1995).

In conclusion, genetics analysis in Brazilian *Coffea canephora* germplasm presented a clear evidence of structure and substructure, which is highly correlated both with prior phenotypic grouping (Kouillou vs Robusta) and with the cultivation historic of the specie in Brazil (Rondônia vs Espírito Santo). Thus, the construction of an association map, using this germplasm, has to consider this genetic arrangement or must be restricted to a single population in order to assure its reliability and to increase its resolution. All collections are able to be used in association studies, but samples from Incaper and Embrapa are more representative of the cultivated germplasm in Brazil. Besides, it is really recommended the use of methods that correct the linkage disequilibrium caused by familial relatedness and population structure, like QK mixed-model proposed by Yu et al. (2006).

References

- Belkhir K (2001) GENETIX, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier (France).
- Berthaud J. (1980) L'Incompatibilité chez *Coffea canephora* : méthode de test et déterminisme génétique. Café Cacao Thé 24 :267-274.

- Berthaud J (1986) Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application. Thèse de doctorat, Paris, France, ORSTOM. Travaux et documents, 188, 379p.
- Berthaud J, Charrier A (1988) Genetics resources of *Coffea*. In: Clarke RJ, Macrae R (eds) Coffee, vol 4. Agronomy. Elsevier Applied Science, London, pp 1 – 42.
- Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331.
- Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmechel KE, Small GW, Roses AD, Haines JL, Pericak- Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* 7:180–184
- Cruz CD, Ferreira FM, Pessoni LA (2011) Biometria aplicada ao estudo da diversidade genética. Visconde do Rio Branco : Ed Suprema. 620p. :il.
- Cubry P (2009) Structuration de la diversité génétique et analyse des patrons de déséquilibre de liaison de l'espèce *Coffea canephora* Pierre ex. Froehner. Doctoral Thesis. Université Montpellier II. 242p.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated checklist of the genus *Coffea* L. (Rubiaceae). *Botanical Journal of the Linnean Society* 152:465–512.
- Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB, Zambolim E, 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.
- Dussert S, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes MC, Berthaud J, Noirot M, Hamon S. (1999) Le caféier, *Coffea canephora*. In: Hamon P, Seguin M, Perrier X, Glaszmann JC. Diversité génétique des plantes tropicales cultivées. Montpellier: CIRAD, p.175-194.
- Efombagn MIB, Motamayor JC, Sounigo O, Eskes AB, Nyasse S, Cilas C, Schnell R, Manzanares-Dauleux MJ, Kolesnikova-Allen M. (2008) Genetic diversity and structure of farm and GenBank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. *Tree Genet Genome* 4: 821–831.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47–50.
- Evanno G, Regnaut, S, Goudet J. (2005). Detecting the number of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611–2620.
- Ewens, W. J. and R. S. Spielman (1995) The transmission/disequilibrium test: history, subdivision, and admixture. *Am. J. Hum. Genet.* 57: 455–464.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Fazuoli LC, Mistro JC, Braghini MT (2009) Melhoramento do café robusta no Instituto Agronomico de Campinas. In: Zambolim L (Ed). *Tecnologias para produção do café conilon*. Editora UFV, Viçosa, p. 201-247.

- Fazuoli LC (1986) Genética e melhoramento do cafeeiro. In: Rena AB, Malavolta E, Yamada J (eds) Cultura do cafeeiro: fatores que afetam a produtividade do cafeeiro. Piracicaba, pp. 87-113.
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Ferrão RG, Fonseca AFA, Ferrão MAG, Bragança SM, Verdin-Filho AC, Volpi PS (2007a) Cultivares de café conilon. In: Ferrão RG, Fonseca AFA, Bragança SM, Ferrão MAG, De Muner LH (eds) Café conilon. Vitória, pp. 203-225.
- Ferrão MAG, Fonseca AFA, Ferrão RG, Oliveira MA, Barbosa WM, D’Isep MSP, Barbosa RP (2007b) Técnicas moleculares e biotecnológicas aplicadas ao café. In: Ferrão RG, Fonseca AFA, Bragança SM, Ferrão MAG, De Muner LH (eds) Café conilon. Vitória, pp. 175-201.
- Fonseca AFA, Sedyama T, Cruz CD, Sakiyama NS, Ferrão MAG, Ferrão RG, Bragança SM (2006) Divergência genética em café conilon. *Pesquisa Agropecuária Brasileira*, 41(4):599–605
- Gaspari-Pezzopane C, Medina-Filho HP, Bordignon R (2004) Variabilidade genética do rendimento intrínseco de grãos em germoplasma de *Coffea*. *Bragantia* 63(1):39–54
- Glaubitz J (2004) CONVERT :a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes* 4: 309-310.
- Gomez C, Dussert S, Hamon P, Hamon S, Kochko A, Poncet V. (2009) Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in Guineo-

- Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *Evolutionary Biology* 9:167 1-19
- Goodman MM and Stuber CW (1983) Races of maize. VI isozyme variation among races of maize in Bolivia. *Maydica* 28:273-280.
- Guo X and Elston RC (1999) Linkage informative content of polymorphic genetic markers. *Hum Hered* 49:112–118.
- Hamrick JL (1992) Distribution of genetic diversity in tropical tree populations: implications for the conservation of genetic resources. In: Lambeth CC and Dvorak W (ed.) *Resolving tropical forest resource concerns through tree improvement, gene conservation and domestication of new species*. NC State University, Raleigh, pp 74-82.
- Ivoglo MG, Fazuoli LC, Oliveira ACB, Gallo PB, Mistro JC, Silvarolla MB, Tomabraghini M (2008) Divergência genética entre progênies de café robusta. *Bragantia* 67(4):823–831
- Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox DK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science* 245:1073–1080
- Kraakman ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446
- 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

- Moncada P, McCouch SR (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* 47:501–509
- Montagnon C, Leroy T, Eskes AB (1998a) Amélioration variétale de *Coffea canephora*. I. Critères et méthodes de sélection. Plantations, recherche, développement. Jan-Fev.
- Montagnon C, Leroy T, Eskes AB (1998b) Amélioration variétale de *Coffea canephora*. II. Les programmes de sélection et leurs résultats. Plantations, recherche, développement. Mars-Avril.
- Medina-Filho HP, Maluf MP, Bordignon R, Guerreiro-Filho O Fazuoli LC (2007) Traditional breeding and modern genetics: a summary of tools and developments to exploit biodiversity for the benefit of the coffee agroindustrial chain. *Acta Hort* 745:351–368
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.
- N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000) The genetic structure of cocoa populations (*Theobroma cacao* L.) revealed by RFLP analysis *Euphytica*, 115 (2), pp. 83-90.
- Poncet V, Hamon P, Minier J, Carasco C, Hamon S, Noirot M (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* 47:1071–108
- Poncet V, Dufour M, Hamon P, Hamon S, Kochko A, Leroy T (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. *Genome* 50:1156-1161.

- Prakash NS, Combes MC, Dussert S, Naveen S, Lashermes P (2005) Analysis of genetic diversity in Indian robusta coffee genepool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLP. *Genetic Resources and Crop Evolution* 52: 333-343.
- Pritchard JK, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus genotypes. *Genetics* 155: 945–959.
- Ronning CM, Schnell RJ (1994) Allozyme diversity in a germplasm collection of *Theobroma cacao* L. *Journal of Heredity* 85:291-294.
- Sereno ML, Albuquerque PSB, Vencovsky R, Figueira A (2006) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conservation Genetics* 7:13-24.
- Silvestrini M, Maluf MP, Silvarolla MB, Guerreiro-Filho O, Medina-Filho HP, Vanini, MMT, Oliveira AS, Gaspari-Pezzopane C, Fazuoli LC (2008) Genetic diversity of a *Coffea* germplasm collection assessed by RAPD markers. *Genetic Resources and Crop Evolution* 55:901-910.
- Souza FF and Santos MM (2009) Melhoramento genético do café canéfora em Rondônia In: Zambolim L (Ed) 10º Encontro sobre produção de Café com Qualidade: Produção de café conilon, viabilidade e perspectivas em áreas não tradicionais. Viçosa pp.
- Yu J, Pressoir G, Briggs WH, Vroh-Bi I, Yamasaki M, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203–208.

CAPÍTULO 3

Análise do desequilíbrio de ligação e estudo preliminar do mapeamento associativo em *Coffea canephora*

Linkage disequilibrium and preliminary association mapping studies in *Coffea canephora*

Abstract

Association mapping has been arisen as promise approach to overcome some limitations of traditional linkage mapping in vegetal species. This study aimed to investigate the population structure, examine the linkage disequilibrium (LD) between pair of simple sequence repeats (SSR) and perform a preliminary study of SSR-trait associations in a *C. canephora* collection. A set of 76 accessions were genotyped with 65 SSR and 15 phenotypic traits were evaluated. LD was measured using r^2 and P -value statistics and association analysis were carried out by means of general linear model (GLM) and mixed linear model, considering structure and kinship (Q+K MLM). Only 6.5% of SSR pairs presented significant LD ($r^2 \geq 0.05$ and $P \leq 0.05$). The levels of structure and kinship were low and six SSR presented significant association with five phenotypic traits. Here, we discussed about the potential of using association mapping for detection of QTLs in *C. canephora* and presented a series of recommendations to improve their chances of success in further studies.

Introduction

Coffea arabica L. and *Coffea canephora* Pierre ex. Froehner compound the main species which produce the beverage known as coffee. In commercial blends (mixed grains of both species), while *C. arabica* contributes with typical coffee aroma and flavor, *C. canephora* responses for solubility and body. Besides, the use of *C. canephora* in blends add profitability to industry, once its production costs are considerably lower than *C. arabica* ones. Moreover, *C. canephora* is the main matter of the instant coffees. Both species are originated from Africa and nowadays, they are cultivated in many places around the world, with remarkable importance to agribusiness in some countries. Worldwide *C. canephora* production was about 2.8 millions of tons, what corresponds about 35 to 40% coffee production (http://www.ico.org/trade_statistics.asp). Brazil is far the major producer of *C. arabica* and the second first of *C. canephora*

The notable performance of Brazilian coffee crops was crucially allied to the advances in genetic breeding programs conducted in that country. In the last 100 years, more than 100 cultivars were released to commercial cultivation (Carvalho, 2008). The most agronomic traits that have been targeted for the plant breeders have quantitative inheritance, being greatly influenced by a big amount of genes and very affected by environmental conditions. Comprehending their genetic basis has been the target of quantitative genetics, which relies on partitioning phenotypic variation within and among individuals with known levels of relatedness (Lynch and Walsh, 1998). Nevertheless, with the advance of molecular techniques a lots of markers have been developed and available, which became possible to associate genomic regions

containing these markers to the variations in complex traits (Hall et al. 2010) and consequently enabling the marker assisted selection. The most commonly tools used for dissecting those traits are linkage analysis and association mapping.

In plants, most of linkage mapping was carried out using highly structured populations with known pedigrees (e.g. F₂, backcrosses, double-haploids, recombinant isogenic lines). In the genus *Coffea* there are a few partial linkage maps available and others in construction. They have been elaborated based on different kinds of markers and populations, e.g. double-haploids (Paillard et al. 1996; Lashermes et al. 2001); backcross (Ky et al. 2000) pseudo-F₂ (Pearl et al. 2004) and test-cross (Cubry, 2009). However, to studies of quantitative trait loci (QTL), these populations present two major limitations: the low number of recombination events, resulting in a poor resolution for quantitative traits; and only two alleles at any locus can be studied simultaneously (Flint-Garcia et al, 2003).

Linkage analysis in plants typically localizes QTLs to 10 to 20 cM intervals (Zhu et al. 2008). Besides, this technique presents high running costs and can demand a long time to form populations, mainly in perennial species, as coffee trees. Therefore, despite the big amount of linkage analysis studies carried out in various plant species over past two decades, only a limited number of identified QTLs were cloned or tagged at the gene level (Price 2006). In *Coffea*, linkage analysis allowed identify a set of marker-discrete trait associations, as follows: a RFLP marker associated to the *S*-locus responsible to the auto-incompatibility system in *C. canephora* (Lashermes et al. 1996); Three RAPD markers associated to the *T* gene, that confers resistance to Coffee Berry Disease (CBD) (Agwanda et al. 1997); two AFLP markers responsible for pollen viability (Coulibaly et al. 2003a); nineteen AFLP markers linked to the coffee resistance to the root-knot nematode *Melodoygine*

exigua (Noir et al. 2003, Diniz et al. 2005); mapping of a caffeoyl-coenzyme A 3-O methyltransferase in a cross between *C. pseudozanguebariae* and *C. liberica* var. Dewevrei (Campa et al. 2003); ten sequence-characterized genetic markers closely linked to the SH₃ resistance leaf rust, originated from *C. liberica* (Prakash et al. 2004; Mahé et al. 2008); seven markers (5 AFLP and 2 SSR) associated with partial leaf rust resistance in lines of *C. arabica* introgressed with *C. canephora* (Herrera et al. 2009); three AFLP markers flanking leaf rust resistance gene (Brito et al. 2010); six sequence-characterized amplified region (SCAR) markers delimiting a chromosomal region of 9.45 cM and flanking the dominant gene of leaf rust resistance at 0.7 and 0.9 cM (Diola et al. 2011). Nevertheless, it is still lacking results involving quantitative traits.

Association mapping arose as a promise of overcome those limitations. This approach, initially developed for human genetics to detect QTL coding simple and complex diseases (Corder et al. 1994; Kerem et al. 1989), has been recently used in many studies with animal and vegetal species. Association mapping, also known as Linkage Disequilibrium (LD) mapping, is one of the many utilities of the linkage disequilibrium phenomenon.

In vegetal genetics, LD mapping allows the use of natural populations, accessions of germplasm collections or even breeding lines. This is an advantage, once these populations are ready for use and harbor a greater number of recombination events, accumulated generation after generation. Moreover, the technique relies on utilizing variation in diverse germplasm and therefore does not suffer from the lack of diversity that generally occurs on pedigree mapping populations (Abdurakhmonov and Abdugarimov, 2008).

Different from human and cattle populations, in vegetal species under breeding, random mating is an exception, while self is prevalent, moreover, effective population sizes are small and strong selection pressure is applied (Stich et al. 2005). On the other hand, due to easy production of large progenies from controlled crosses and conducting replicated trials with immortal individuals (clones, RILs, for instance), association mapping in plants may prove more promising than in human or animal genetics (Zhu et al. 2008). Another issue to be considered is that association mapping is only able of identifying phenotypic effects of alleles with reasonably high frequency in the population, once the presence of rare alleles usually increases variances, lacking statistical test power (Sorkhen et al. 2008).

Knowledge about the extent and genomic distribution of LD is essentially important to association mapping, because: 1) regions in LD need to be present in order to detect marker-phenotype associations and 2) the length of the regions limits the resolution of association mapping (Stich et al. 2005). The presence of long LD blocks permits the use of fewer markers to reasonable genome coverage (Flint-Garcia et al. 2003). Moreover, information about the causes generating LD is essential for drawing conclusions about the prospects of association mapping. LD conserved by linkage is useful for association mapping. The classification of selection and mutation forces generating useful and non-useful LD for association mapping depends on how the LD information is used. LD generated by relatedness, population stratification, and genetic drift causes spurious marker-phenotype associations (Stich et al. 2005). False positives generated by population structure or relatedness have been a barrier to association mapping and it has been difficult to replicate significant results in different studies (Zhu et al. 2008). The problem arises because any phenotypic trait that is also correlated with the underlying population

structure at neutral loci will show an inflated number of positive associations (Hall et al. 2010). So, is expected that the use of only one heterotic group eliminate population structure as linkage disequilibrium generating force (Stich et al. 2008a). Therefore, LD would have to be analyzed within single germplasm groups and not within the set of all populations. However, several methods have been developed to take in to account the population structure and relatedness and hence, overcoming prior problems.

Mixed linear model is the most recent and promising method to avoid false results in association mapping. This approach uses information of population structure and more cryptic relatedness among members of an association study to correct their own spurious effects (Yu et al. 2006). This information is incorporated in two matrices: a matrix of population effects (Q) and a matrix of relative kinship of individuals in a sample (K), both obtained in a sample of unlinked and putative neutral markers. Recent reports support that QK method is appropriate for association mapping in allogamous (Yu et al. 2006) and autogamous (Stich et al. 2008a). Q and K matrices may be obtained using many kinds of markers. However, SSRs markers are preferred due the highly polymorphism and the co-dominant segregation pattern. The vast majority of SNPs are bi-allelic and the information content per marker is therefore lower than in SSR markers (Hall et al. 2010).

As there is a representative sample of Brazilian *C. canephora* germplasm conserved at the gene bank of the Empresa Brasileira de Pesquisa Agropecuária – Embrapa, properly phenotyped and genotyped, this research aimed to: 1) investigate its population structure; and 2) examine the linkage disequilibrium between pair of SSRs; and 3) perform a preliminary study of SSR-trait associations.

Material and Method

Plant material

At first, a set of 185 accessions of *Coffea canephora*, representing the germplasm commercially cultivated, used in breeding programs, as so conserved at the main Brazilian gene banks, was used for genetic analysis with 20 SSRs, in a previous assay. The analysis of population structure demonstrated a strong stratification of those genotypes, obeying both phenotypic ('Robusta' vs 'Conilon') and geographical rules ('Rondônia' vs 'Espírito Santo' States) (Souza et al. unpublished). Generally, in these cases, to perform association studies, it is recommended to correct the analysis taking into account the population structure, or select one unstructured subpopulation (López-Gartner et al. 2009). Hence, 76 'Conilon' accessions (Table 1) representing Rondonian germplasm were chosen in order to carry out phenotypic evaluation, linkage disequilibrium analysis and a preliminary marker-phenotype association study.

Phenotypic evaluation

Phenotypic data was previously obtained in a field experiment carried out at the Embrapa Experimental Station, at Ouro Preto do Oeste, Rondônia, Brazil. This area is located in West Amazonia, under following geographic coordinates: 10°43'53"S and 62°15'24"W. The altitude is 244 m and the climate is Tropical wet and dry (Aw, conform Köppen classification). The accessions were established in a random block design with six replications and two plants (clones) by replication.

Table 1. List of *Coffea canephora* Rondonian accessions used on the studies of structure, linkage disequilibrium and association mapping.

Accession	County	Accession	County
Cpafro 001	Cacoal	Cpafro 089	Novo Horizonte
Cpafro 003	Cacoal	Cpafro 096	Novo Horizonte
Cpafro 004	Cacoal	Cpafro 098	Novo Horizonte
Cpafro 006	Cacoal	Cpafro 100	Novo Horizonte
Cpafro 010	Cacoal	Cpafro 101	Novo Horizonte
Cpafro 015	Cacoal	Cpafro 103	Novo Horizonte
Cpafro 016	Cacoal	Cpafro 119	Novo Horizonte
Cpafro 017	Cacoal	Cpafro 120	Novo Horizonte
Cpafro 018	Cacoal	Cpafro 125	Novo Horizonte
Cpafro 022	Cacoal	Cpafro 127	Novo Horizonte
Cpafro 024	Cacoal	Cpafro 130	Novo Horizonte
Cpafro 025	Cacoal	Cpafro 138	Nova União
Cpafro 026	Cacoal	Cpafro 139	Nova União
Cpafro 027	Cacoal	Cpafro 140	Nova União
Cpafro 030	Cacoal	Cpafro 141	Nova União
Cpafro 032	Cacoal	Cpafro 142	Nova União
Cpafro 035	Cacoal	Cpafro 146	Nova União
Cpafro 036	Cacoal	Cpafro 147	Nova União
Cpafro 038	Cacoal	Cpafro 149	Nova União
Cpafro 042	Cacoal	Cpafro 151	Nova União
Cpafro 043	Cacoal	Cpafro 155	Nova União
Cpafro 044	Cacoal	Cpafro 156	Nova União
Cpafro 045	Cacoal	Cpafro 160	Ji-paraná
Cpafro 047	Cacoal	Cpafro 161	Ji-paraná
Cpafro 049	Cacoal	Cpafro 164	Ji-paraná
Cpafro 052	Cacoal	Cpafro 167	Ji-paraná
Cpafro 056	Cacoal	Cpafro 171	Nova União
Cpafro 058	Cacoal	Cpafro 172	Nova União
Cpafro 059	Cacoal	Cpafro 183	Teixeirópolis
Cpafro 064	Cacoal	Cpafro 184	Teixeirópolis
Cpafro 072	Novo Horizonte	Cpafro 189	Teixeirópolis
Cpafro 073	Novo Horizonte	Cpafro 190	Vale do Paraíso
Cpafro 075	Novo Horizonte	Cpafro 193	Vale do Paraíso
Cpafro 076	Novo Horizonte	Cpafro 194	Vale do Paraíso
Cpafro 077	Novo Horizonte	Cpafro 196	Vale do Paraíso
Cpafro 085	Novo Horizonte	Cpafro 197	Vale do Paraíso
Cpafro 086	Novo Horizonte	Cpafro 199	Vale do Paraíso
Cpafro 088	Novo Horizonte	Cpafro 203	Jaru

The 76 *C. canephora* accessions collected in Rondônia were evaluated with regard to the following traits: plant height (PHT); number of plagiotropic branches per plant (NBP); length and width of leaves (LHT and LWT); number of fruits per inflorescence (NFI); length, width and weight of fruits (FLT, FWT and FWG); number of internodes per branch (NIB); length of internodes (ILT); length of branches (BLT); total weight of fruits per branch (PFB); production of green coffee per area, a long two seasons (YGC); The variables were submitted to analysis of variance and genotypic correlation coefficients were also estimated.

DNA extraction

For genomic total DNA extraction, young and completely extended leaves were collected from each accession, frozen at $-80\text{ }^{\circ}\text{C}$, lyophilized, ground until becoming a fine powder and kept at $-20\text{ }^{\circ}\text{C}$ until used. Approximately 50 mg of ground tissue was incubated at $65\text{ }^{\circ}\text{C}$ for 30 min in 1.5 ml of a mix of extraction [0.35M Sorbitol, 0.10M Tris-HCl, 5mM EDTA] and lyses [2% CTAB, 2M NaCl, 0.20M Tris-HCl, 50mM EDTA] buffer, plus Sarkosyl 5%, Sodium Bisulfite 1%, PVP 2% and active charcoal 0.1%. After centrifuged at 14,000 RPM for 10 min, the supernatant was transferred to another tube and 1 ml of chloroform:isoamyl (24:1) was added. The tube was inverted until form an emulsion and left to rest for about 10 min before centrifuged at 14,000 RPM for 10 min. The supernatant was transferred to a new tube, and DNA was precipitated with 1 ml of cold isopropanol. The tube was kept at $-80\text{ }^{\circ}\text{C}$ for 40 min, and then centrifuged for 20 min at 14,000 RPM at $4\text{ }^{\circ}\text{C}$. Alcohol was discarded, and the pellet, washed twice with 0.5 ml ethanol 70% and 95%, respectively, was left to dry for 15 min at room temperature. DNA was re-suspended with 200 μL of TE and 1.7 μL of RNAase at $37\text{ }^{\circ}\text{C}$ for 30

min. Then, 40 μL of NaCl and 200 μL of cold isopropanol were added and the tube was softly inverted until DNA becoming visible. Tube was kept at $-80\text{ }^{\circ}\text{C}$ for 40 min, centrifuged, washed and dried as reported before and DNA was finally re-suspended in 250 μL of TE and maintained at $-20\text{ }^{\circ}\text{C}$. DNA concentration and its quality were evaluated with spectrophotometer and agarose gel. All DNA samples were prepared to a final concentration of $25\text{ }\eta\text{g }\mu\text{L}^{-1}$.

SSR markers

A total of 201 SSR primers were tested in six divergent accessions. These SSR were synthesized from sequences available in literature (87 SRR) (Rovelli et al. 2000; Combes et al. 2000; Coulibaly et al. 2003b; Baruah et al, 2003; Moncada and McCouch, 2004; Poncet et al. 2004; Leroy et al. 2005; Bhat et al. 2005) and developed from non-redundant Express Sequence Tags (EST) of the Brazilian Coffee Genome Project (102 SRR) and also from enriched genomic library (12 SRR). From those 201 primers, 59 presented no amplified products in the gel and 38 were monomorphic. The 104 remain SSRs were polymorphic, but 27 offered serious difficulties in their evaluation. Thus, a set of 77 SSR were used to genotype all accessions. Finally, 12 SSR that presented rare alleles (frequency $< 5\%$) were excluded from the analysis. These minor alleles are biased to estimate LD loci pairwise, producing larger variances (Remington et al. 2001). So, a set of 65 SSR were used in the LD, structure, kinship and association analysis. Their characteristics are described in Tables 2 and 3.

Table 2. List of EST-SSR used on the study of structure and linkage disequilibrium in *Coffea canephora*

ID	Foward Primer ¹	Tm(°C)	Reverse Primer	Tm(°C)	Size (pb)
EST-SSR 001	XXXXXXXXXXXX	59,4	XXXXXXXXXXXX	59,4	151
EST-SSR 002	XXXXXXXXXXXX	57,3	XXXXXXXXXXXX	58,4	184
EST-SSR 003	XXXXXXXXXXXX	57,9	XXXXXXXXXXXX	57,9	244
EST-SSR 010	XXXXXXXXXXXX	49,6	XXXXXXXXXXXX	51,7	152
EST-SSR 013	XXXXXXXXXXXX	54,7	XXXXXXXXXXXX	49,6	141
EST-SSR 014	XXXXXXXXXXXX	53,7	XXXXXXXXXXXX	53,7	160
EST-SSR 018	XXXXXXXXXXXX	57,3	XXXXXXXXXXXX	59,4	195
EST-SSR 019	XXXXXXXXXXXX	58,4	XXXXXXXXXXXX	59,4	199
EST-SSR 022	XXXXXXXXXXXX	57,3	XXXXXXXXXXXX	57,3	155
EST-SSR 030	XXXXXXXXXXXX	53,7	XXXXXXXXXXXX	52,3	199
EST-SSR 036	XXXXXXXXXXXX	53,7	XXXXXXXXXXXX	54,7	271
EST-SSR 041	XXXXXXXXXXXX	56,7	XXXXXXXXXXXX	58,4	176
EST-SSR 046	XXXXXXXXXXXX	51,7	XXXXXXXXXXXX	52,3	119
EST-SSR 047	XXXXXXXXXXXX	52,9	XXXXXXXXXXXX	51,7	178
EST-SSR 048	XXXXXXXXXXXX	53,7	XXXXXXXXXXXX	53,7	126
EST-SSR 056	XXXXXXXXXXXX	51,7	XXXXXXXXXXXX	49,6	208
EST-SSR 058	XXXXXXXXXXXX	51,3	XXXXXXXXXXXX	51,7	201
EST-SSR 062	XXXXXXXXXXXX	53,7	XXXXXXXXXXXX	51,7	197
EST-SSR 067	XXXXXXXXXXXX	47,9	XXXXXXXXXXXX	53,7	100
EST-SSR 074	XXXXXXXXXXXX	52,9	XXXXXXXXXXXX	49,6	259
EST-SSR 096	XXXXXXXXXXXX	59,4	XXXXXXXXXXXX	57,3	152
EST-SSR 097	XXXXXXXXXXXX	49,6	XXXXXXXXXXXX	51,7	256
EST-SSR 108	XXXXXXXXXXXX	49,6	XXXXXXXXXXXX	52,9	121

¹Information momentarily confidential

Table 3. List of SSR used on the study of structure and linkage disequilibrium in *Coffea canephora*

Acesso N°	Forward primer (5'>3')	Reverse primer (5'>3')	Tm °C	size (pb)	Ref
SSR004	AGGCCTTCATCTCAAAAACC	AGCGTACTTGAGGCCAAAGA	60-55	217	1
SSR007	TGACATAGGGGGCTAAATTG	TTAATGGTGACGCTTTGATG	60-55	215	1
SSR008	CACTGGCATTAGAAAGCACC	GGCAAAGTCAATGATGACTC	60-55	198	1
SSR013	TGGCCGTGATAATAAACAGC	ATGTGGCAATCTAAAGCCAA	60-55	212	1
SSR016	AJ250252 ACCCGAAAGAAAAGAACCAAG	CCACACAACCTCCTCATT	60-55	140-146	2
SSR018	AJ250254 GGCTCGAGATATCTGTTAG	TTAATGGGCATAGGGTCC	60-55	132-166	2
SSR021	AJ250257 GACCATTACATTTACACAC	GCATTTTGTTCACACTGIA	60-55	103-122	2
SSR029	AJ308742 GGCTTCTGGGTGTCTGTGT	CCATTGGCTTTGTATTTCTGG	60-55	110	3
SSR030	AJ308745 ATGGGGCCAACCTGAATATG	CAGGGCATCTATCTACTTCTTT	60-55	220	4
SSR034	AJ308753 GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGT	58	294	4
SSR035	AJ308764 CTGGCATTAGAAAGCACCTTG	GCTTGGCTCACTGTAGACTG	60-55	158	4
SSR037	AJ308767 CAACACTATCTTGTATTTTCACT	CGTGCAAGTCACATACTTTACTAC	55-50	165	4
SSR040	AJ308782 AAAGGAAAATGTTGGCTCTGA	TCCACATACATTTCCAGCA	60-55	114	4
SSR043	AJ308790 TTTTCTGGGTTTTCTGTGTTCTC	TAACCTCCATTCCCGATT	60-55	134	4
SSR046	AJ308802 AATGAAGAAGAGGGTGGTG	CGAGGGTATTGTTTCCAG	60-55	228	4
SSR052	AJ308824 GACCAAATGTCAGCTCATTG	GCCGACTGCTCTTTAGTGT	60-55	251	4
SSR055	AJ308834 GCAGGTATTTGAAGGATGAACC	GTTAGGTGGTGCATGTGT	60-55	125	3
SSR056	AJ308836 AGCTATCTTTATCTCACACACACA	GTTAGTGTTCGATTGGTACTG	60-55	158	4
SSR057	AJ308837 CTCGCTTTCACGCTCTCTCT	CGGTATGTTCTCGTTCCTC	60-55	102	3
SSR059	AJ308840 CCAGCTCTCCTCACTTTTCA	GGTGGTGAGGGGTAATAGG	58	272	4
SSR061	AJ308847 GCACACATGAAAAGATGCTTG	TGATGGACAGGAGTTGATGG	60-55	194	3
SSR064	AJ308859 GTTAGTGTGCGACCGTGTGT	TTATGCCCTCCCATATCT	58	220	4
SSR065	AJ308864 CTCACACACACTGTCACG	CGAATGAGGCTCCATCAC	60-55	108	4
SSR068	AY220270 TGTGATGCCATTAGCCTAGC	TCCAACATGTGCTGGTGATT	60	177-229	5
SSR071	AY220273 GCTAAGTTCAATTGCCCTGT	GGGTTAATTTGATTGCGTGA	59	210-232	5
SSR074	AY220277 TGGGGAAAAGAAGGATATAGACAAGAG	GAGGGGGGCTAAGGGAATAACATA	55	95-129	5
SSR076	AY220279 GGTCCCCTCTCAAGCTGAA	GGCAATTGATTCTGGAACCT	59	139-157	5
SSR077	CFGA2 TCTCCTCTTTGCTGAGCC	AGATTACCCTTCAAGTTTCTC	55	186	6
SSR081	CFGA54 AGTAATGAACCTGCCGCTCTTT	TTGTCATTCTTGTTTTCATCC	55	230	6
SSR082	CFGA55 CAAAATGAAGGAGAAAAGTGGACA	TGGCTTCATCTCAACCTTCTTC	55	220	6
SSR084	CFGA75 AAGTAGATTGGTGAAAGGGGAAGC	TCCTTCATTTTCTCTTGGTT	55	130	6
SSR093	CFGA227 ATACAGCAATTTGAGAGGAGGAA	TCTGTGCCTTCCCGAGTCA	55	225	6
SSR100	CFGA465 ACCCTTACTACTTATTTACTCTC	ACATCCCCTTGCCATTTCTTC	55	157	6
SSR102	CFGA491 CTCAGAGCTGCGGTGGTGTCA	CCGGACGATCTTCTTTCTTCA	55	282	6
SSR106	AJ308755 CCCTCCCTCTTTCTCTCTC	TCTGGGTTTTCTGTGTTCTCG		184	2
SSR114	AJ871882 TAACAGAAGCACAAAACC	TCTAAACCCACCTCACAAAC		162	7
SSR119	AJ308795 TTGCCATCATCGTTCATTCT	GCATAGTGTGCGTTGTGTTGT	58	190	4
SSR121	AJ308854 CGACACTTCTTTGGCACTC	AGACACCCACCCATCCAC	55-50	177	4
SSR122	AJ308883 CGTCTCGTTTACGCTCTCT	GATCTGCATGTACTGGTGTCTC	60-55	237	4
SSR138	AJ308848 GCACTTCCCTCTCAACCAAC	ACTAGGACAGCAAATAGCATAACC	60-55	213	4
SSR151	AY705507 GGCCGAGGGGAAAAGAAGC	GGAAACCTCACGAGAAGATTACAAA	57	96-100	8
SSR156	CFGA260 CGGCTCAAAGGTGCTCAAT	TCCCAATCTTCTTTCTTTCTGT	55	135	6

References : 1-Rovelli et al (2000), 2 - Combes et al (2000), 3- Coulibaly et al (2003), 4 - Poncet et al (2004), 5 - Baruah et al (2003), 6 -Moncada e McCouch (2004), 7 - Leroy et al (2005), 8 - Bhat et al (2005).

PCR conditions, electrophoresis and molecular score

Each reaction was set to a final volume of 20 μ L, containing 50 ng of genomic DNA, 0.6 unit of *Taq* DNA polymerase, buffer 1x, 1mM of $MgCl_2$, 150 μ M of each dNTP and 0.1 μ M of each primer. PCR amplifications were carried out using touchdown proceeding, which comprises initial denaturation at 94 °C for 2 min, followed by 10 cycles of denaturation at 94 °C for 0.5 min, annealing at 67 °C for 0.5 min, decreasing 1°C after each cycle, and extension at 72 °C for 0.5 min. After that, another set of 30 cycles, comprising denaturation at 94 °C for 0.5 min, annealing at 55 °C for 0.5 min and extension at 72 °C for 0.5 min, was accomplished followed by a final 8 min extension time at 72 °C. Before running, PCR products were denatured in 8 μ L of denaturing dye (95% formamid) at 94 °C for 5 min and 7 μ L of sample were loaded on a standard 6% polyacrylamide gel at 50 °C and run at a constant power of 90 W for about 2 h.

Post-PCR multiplex, which involved the multi-loading of individual PCR assays (two to four SSRs per running), was performed spacing successive loads by 10 to 30 min during electrophoresis, depending on the prior information about fragment size. At last, the gel was treated with ethanol (10%) + acetic acid (1%), followed by nitrate acid (1,5%); stained with silver nitrate (4%), developed with sodium carbonate (3%) + formaldehyde (0.03%); and fixed with acetic acid (5%) and dried for posterior analysis at a transilluminator apparatus. SSR alleles were scored using a molecular weight standard ladder and compared to the expected size of the microsatellite sequence.

Analysis of population structure

Population structure of *C. canephora* was assessed using a Bayesian clustering analysis performed by the software STRUCTURE (Pritchard et al. 2000). In this approach, multilocus genotypes are used to assign fractions of genetic ancestry of each individual into genetic clusters based on a predefined number of clusters (k). The allelic frequencies within each cluster are in Hardy-Weinberg equilibrium and the loci are not linked. k was estimated using a Monte Carlo Markov Chain approach. For each run, the burn-in period was of 50,000 and a length of 100,000 iterations. The posterior log-likelihood was estimated for k from 2 to 12, with 20 independent runs at each k . It was used the admixture model of ancestry and specified that allele frequencies were correlated, without prior population information. The true number of genetic clusters (k) was estimated using the Δk method described by Evanno et al. (2005). The run with the higher likelihood was used to allocate individuals to a single cluster if their membership coefficient in that cluster was ≥ 0.70 .

Disequilibrium linkage analysis

Genome-wide LD between pairs of SSRs was estimated using r statistic. In SSR loci with two alleles, Fisher's exact tests were performed to calculate the probabilities of obtaining LD estimates under hypothesis of linkage equilibrium (P -values). In the case of more than two alleles per loci, P -values were calculated by repeatedly permuting the alleles at one of the loci, as described by Weir (1996). The software TASSEL 2.1 (Bradbury et al. 2007) was used to obtain LD estimates.

Association studies

Associations of individual SSR alleles with phenotypic trait values were estimated using the Q+K Mixed Linear Model proposed by Yu et al. (2006) and implemented in software TASSEL 2.1 (Bradbury et al. 2007). In this method,

random markers are used to estimate population structure (Q) and relative kinship (K) matrices, which are then fit into a mixed-model framework to test for marker-trait associations (Zhu et al. 2008). Q is an $n \times p$ co-ancestry matrix, obtained by using STRUCTURE software (Pritchard et al. 2000), where n is the number of individuals and p is the number of defined subpopulations. K was obtained by SPAGeDi software (Hardy and Vekemans, 2002) and consists in a $n \times n$ matrix, with off-diagonal elements being, a marker-based estimate of probability of identity by descent. The diagonal elements of K are “1” for inbreds and $0.5 \times (1 + F_x)$ for non-inbred individuals, where F_x is the inbreeding coefficient (Zhu et al. 2008). A general linear model method was also employed at the same data, in order to allow some comparison.

Results

Phenotypic traits

The analysis of variance revealed a remarkable phenotypic variability on the evaluated accessions. All traits presented significant differences among the accessions (Table 4) and the ranges between the minimum and maximum values were very wide. See for example, the yield of green beans in bags/ha, which ranged from 1.31 to 144.04 bags/ha.

The experimental precision was satisfactory. The coefficient of variation (Cv) ranged from 3.81% (ILT) to 31.45% (WFB); and the coefficient of genotypic determination (h^2) ranged from 44.73% (NPB) to 96.02% (YGC).

Table 4. Summary of the analysis of variance in 15 phenotypic traits evaluated in 76 accessions of *C. canephora*

Traits ¹	TMS ²		Min	Max	Mean	CV _E (%)	CV _G (%)	h(%)
LHT (cm)	15.8122 **		8.90	43.40	14.67	14.13	11.57	72.84
LWT (cm)	3.509 **		3.20	11.60	5.34	17.42	15.24	75.38
NFI (unit)	53.178 **		2.80	27.00	13.46	20.58	25.05	85.57
NRB (unit)	13.02 **		3.00	16.50	10.39	18.32	14.76	72.18
NIB (unit)	17.337 **		9.50	24.00	16.91	12.09	10.72	75.87
FLT (cm)	313.8865 **		34.30	93.50	60.31	12.74	13.24	81.20
FWT (cm)	0.0493 **		1.07	1.82	1.44	4.78	7.31	90.32
FWG (g)	0.0614 **		0.40	1.36	0.86	10.72	13.41	86.24
ILT (cm)	0.0202 **		0.83	1.45	1.19	3.81	5.68	89.87
BLT (cm)	0.5011 **		2.49	6.50	3.59	8.76	8.82	80.22
PHT (m)	0.0703 **		1.05	2.45	1.62	9.80	6.58	64.30
NPB (unit)	2536.9601 **		70.00	334.00	186.70	20.06	9.02	44.73
WFB (g)	3515.8848 **		4.60	203.30	73.42	31.45	37.19	84.84
YGC (bags/ha)	2741.0174 **		1.31	144.04	40.91	25.54	62.70	96.02

**Significant to 0.01 level of F test

¹LHT and LWT = length and width of leaves; NFI= number of fruits per inflorescence; NRB= number of inflorescence per branch; NIB= number of internodes per branch; FLT, FWT and FWG= length, width and weight of fruits; ILT= length of internodes; BLT= length of branches; PHT= plant height; NPB= number of plagiotropic branches per plant; WFB= total weight of fruits per branch; YGC= production of green coffee per area, a long two years.

²TMS= Treatment mean square; Min, Max and Mean= Minimum, Maximum and Mean value observed; CV_G = Coefficient of genetic variation; CV_E = Coefficient of environmental variation; h = coefficient of genotypic determination

High positive and significant genotypic correlations were observed among yield of green beans and its factors, *i.e.*: number of fruits per inflorescence (0.76), number of inflorescence per branch (0.61), and production of fruits per branch (0.77). Other characteristics, as length of the branches (0.47), plant height (0.43) and fruit length (0.43), also presented high correlations with the yield (Table 5).

Table 5. Genotypic correlations from 14 phenotypic traits evaluated in 76 accessions of *C. canephora*

	LHT ¹ (cm)	LWT (cm)	FWG (g)	NFI (unit)	NRB (unit)	NIB (unit)	FLT (cm)	FWT (cm)	ILT (cm)	BLT (cm)	PHT (m)	NPB (unit)	WFB (g)	YGC (bags/ha)
LHT	1,00	0,87	0,19	-0,07	0,30	0,51	0,39	0,39	0,48	0,35	0,53	0,59	-0,02	-0,07
LWT		1,00	0,03	-0,14	0,08	0,35	0,22	0,26	0,25	0,20	0,43	0,45	-0,19	-0,14
FWG			1,00	0,22	0,32	0,16	0,24	0,76	0,82	0,45	0,24	0,15	0,45	0,17
NFI				1,00	0,72	0,16	0,53	-0,04	0,27	0,58	0,52	-0,07	0,94	0,76
NRB					1,00	0,68	0,73	0,22	0,48	0,43	0,60	0,36	0,75	0,61
NIB						1,00	0,79	0,29	0,47	0,22	0,61	0,41	0,18	0,12
FLT							1,00	0,12	0,47	0,67	0,76	0,22	0,54	0,43
FWT								1,00	0,64	0,19	0,25	0,33	0,09	-0,09
ILT									1,00	0,58	0,57	0,43	0,42	0,23
BLT										1,00	0,71	0,23	0,60	0,47
PHT											1,00	0,37	0,45	0,43
NPB												1,00	-0,05	0,15
WFB													1,00	0,77
YGC														1,00

¹LHT and LWT = length and width of leaves; NFI= number of fruits per inflorescence; NRB= number of inflorescence per branch; NIB= number of internodes per branch; FLT, FWT and FWG= length, width and weight of fruits; ILT= length of internodes; BLT= length of branches; PHT= plant height; NPB= number of plagiotropic branches per plant; WFB= total weight of fruits per branch; YGC= production of green coffee per area, a long two years.

Pairwise linkage disequilibrium

In a total of 76 accessions of *C. canephora*, genotyped with 65 polymorphic SSRs, only 8.1%, 5.2% and 3.2% of marker pairs presented significant LD at $P \leq 0.01$, $P \leq 0.001$ and $P \leq 0.0001$, respectively. Based on r^2 estimates, the percent of SSR pairs in LD were 3.0% and 1.9%, considering $r^2 \geq 0.1$ and $r^2 \geq 0.2$, respectively. At significant reference threshold values ($r^2 \geq 0.05$ and $P \leq 0.05$), 6.5% of the SSR pairs were found in LD. This low level of LD is represented in Figure 1. As the most part of the SSRs are not plotted in any linkage map yet, it was not possible to estimate the magnitude of LD due to the strictly linkage. Nevertheless, the analysis of LD, in the same 76 accessions, using a smaller sample of 17 SSRs already mapped (Figure 1), showed that nine SSR pairs were in significant LD, and among them, one pair was composed by loci belonging to the same linkage group (GL). Or else, the most SSR pairs in LD were actually unlinked. So, this result suggests the action of other LD generating phenomena, besides linkage.

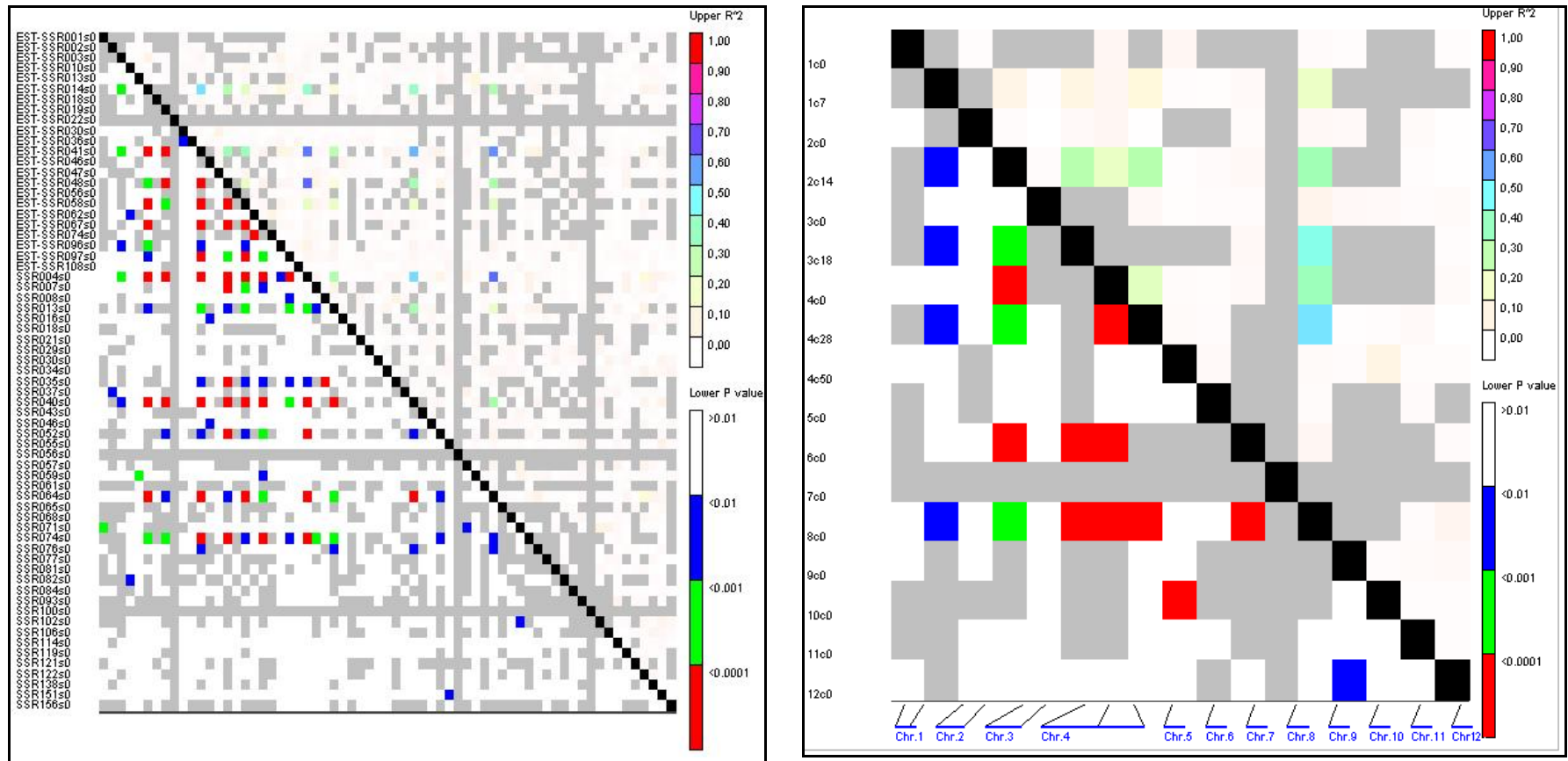


Figure 1. Matrix triangles representing loci pairwise disequilibrium linkage (LD) in 76 accessions of *C. canephora*. On the left, LD among 65 SSRs evaluated. On the right, LD among 17 SSR already plotted in a partial linkage map.

Population structure and kinship

Taking in to account the necessity of correcting the effects of structure and relatedness in the Rondonian population, we employed a Bayesian model approach to detect putative substructure. The number of group was defined using Δk Method (Evanno et al. 2005). Most accessions presented high level of shared ancestry, that can be seen on the bar plot (Figure 2). Three groups were found and they were comprised by 39, 29 and 4 accessions, respectively (Table 6).

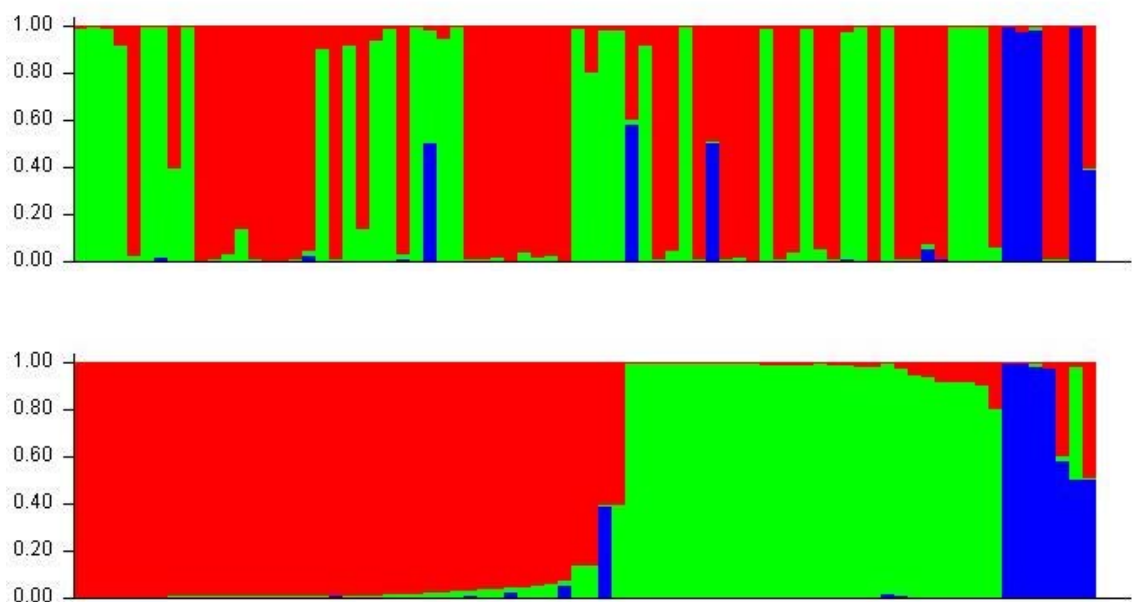


Figure 2. Population structure of 76 *C. canephora* accessions, according to Bayesian analysis. Each color represents a genetic group. Above: accessions sorted by original order; down: accessions sorted by membership value.

Table 6. Membership of 76 *Coffea canephora* accessions obtained by Bayesian analysis of 65 SSR.

Accession	I	II	III	Group	Accession	I	II	III	Group
Cpafro 001	0.013	0.985	0.002	2	Cpafro 089	0.199	0.798	0.003	2
Cpafro 003	0.005	0.993	0.001	2	Cpafro 096	0.016	0.982	0.001	2
Cpafro 004	0.008	0.990	0.002	2	Cpafro 098	0.018	0.981	0.002	2
Cpafro 006	0.081	0.916	0.003	2	Cpafro 100	0.398	0.021	0.581	Ad
Cpafro 010	0.975	0.023	0.002	1	Cpafro 101	0.085	0.913	0.002	2
Cpafro 015	0.006	0.992	0.002	2	Cpafro 103	0.986	0.010	0.004	1
Cpafro 016	0.004	0.978	0.018	2	Cpafro 119	0.955	0.042	0.004	1
Cpafro 017	0.605	0.392	0.004	Ad	Cpafro 120	0.004	0.994	0.002	2
Cpafro 018	0.007	0.991	0.002	2	Cpafro 125	0.992	0.005	0.003	1
Cpafro 022	0.996	0.003	0.001	1	Cpafro 127	0.489	0.008	0.503	Ad
Cpafro 024	0.991	0.005	0.004	1	Cpafro 130	0.989	0.007	0.004	1
Cpafro 025	0.970	0.028	0.003	1	Cpafro 138	0.984	0.014	0.002	1
Cpafro 026	0.862	0.137	0.001	1	Cpafro 139	0.994	0.004	0.002	1
Cpafro 027	0.989	0.008	0.002	1	Cpafro 140	0.013	0.984	0.003	2
Cpafro 030	0.995	0.003	0.002	1	Cpafro 141	0.988	0.010	0.002	1
Cpafro 032	0.996	0.003	0.001	1	Cpafro 142	0.963	0.034	0.003	1
Cpafro 035	0.991	0.007	0.002	1	Cpafro 146	0.010	0.988	0.002	2
Cpafro 036	0.956	0.020	0.024	1	Cpafro 147	0.944	0.053	0.003	1
Cpafro 038	0.093	0.905	0.002	2	Cpafro 149	0.991	0.004	0.005	1
Cpafro 042	0.988	0.009	0.002	1	Cpafro 151	0.028	0.958	0.014	2
Cpafro 043	0.083	0.914	0.003	2	Cpafro 155	0.005	0.993	0.002	2
Cpafro 044	0.860	0.138	0.002	1	Cpafro 156	0.994	0.004	0.002	1
Cpafro 045	0.062	0.936	0.002	2	Cpafro 160	0.005	0.993	0.002	2
Cpafro 047	0.008	0.990	0.002	2	Cpafro 161	0.990	0.007	0.002	1
Cpafro 049	0.969	0.020	0.011	1	Cpafro 164	0.990	0.007	0.002	1
Cpafro 052	0.004	0.993	0.002	2	Cpafro 167	0.925	0.019	0.056	1
Cpafro 056	0.018	0.479	0.503	2	Cpafro 171	0.987	0.005	0.008	1
Cpafro 058	0.053	0.945	0.002	2	Cpafro 172	0.007	0.987	0.006	2
Cpafro 059	0.006	0.992	0.002	2	Cpafro 183	0.007	0.992	0.001	2
Cpafro 064	0.987	0.011	0.002	1	Cpafro 184	0.003	0.996	0.001	2
Cpafro 072	0.986	0.013	0.001	1	Cpafro 189	0.942	0.056	0.002	1
Cpafro 073	0.981	0.017	0.002	1	Cpafro 190	0.002	0.002	0.995	3
Cpafro 075	0.993	0.003	0.004	1	Cpafro 193	0.022	0.006	0.973	3
Cpafro 076	0.962	0.035	0.003	1	Cpafro 194	0.003	0.014	0.983	3
Cpafro 077	0.979	0.019	0.002	1	Cpafro 196	0.992	0.006	0.001	1
Cpafro 085	0.973	0.024	0.002	1	Cpafro 197	0.987	0.009	0.003	1
Cpafro 086	0.995	0.004	0.001	1	Cpafro 199	0.003	0.003	0.994	3
Cpafro 088	0.008	0.990	0.002	2	Cpafro 203	0.606	0.006	0.388	Ad

Four remains accessions (Cpafro 017, Cpafro 100, Cpafro 127 and Cpafro 203) presented coefficients of membership distributed among the three groups and were classified as admixture. No clear relation was observed between model-based groups and original collect place. However, four (Cpafro 190, Cpafro 193, Cpafro 194 and Cpafro 199), from six accessions collected in 'Vale do Paraíso' county, presented differentiated ancestry and formed the group 3. The mean F_{ST} value, estimated by STRUCTURE, was higher on this group (0.34). Groups 1 and 2 presented F_{ST} 0.10 and 0.13, respectively. Divergences among groups, based on the allelic frequency were $G1 \times G2 = 0.02$; $G1 \times G3 = 0.13$ and $G2 \times G3 = 0.14$. These results suggest that accessions from G1 and G2 are very similar and differed from G3.

Overall kinship values, estimated in SPAGeDi, ranged from 0 to 0.96. More than 50% of the accession pairs presented $K=0$, and only 3% presented K higher than 0.25, indicating a low level of relatedness on the Rondonian subpopulation. In G1, kinship ranged from 0.0 to 0.43 and the percents of accession pairs that present $K=0$ and $K>0.25$ were 39.7% and 0.9%. In G2, kinship varied from 0.0 to 0.26 and 34.7% and 0.5% of the pairs of accessions, presented $K=0$ and $K>0.25$, respectively. On the other hand, in G3, kinship ranged from 0.65 to 0.96 and 100% of the pairs presented $K>0.25$. Thus, we may suppose that the elected sample of accessions presents a subtle structure, composed by three groups from which two are slightly separated and a third group is more distant and this isolation is due to the relatedness into its accessions, but the relatedness on the overall sample is low.

SSR-phenotypic traits associations

A preliminary and speculative analysis, using Q+K mixed linear model (MLM) approach, was performed, involving the 76 Rondonian accessions, the 65 high

polymorphic SSR and 15 phenotypic traits. General linear model (GLM) was also used with same data to allow a comparison with MLM results. The results indicated the occurrence of high significant associations ($P < 0.001$) between six SSRs (9% from all polymorphic SSR) and five phenotypic traits (Table 7). GLM and MLM presented very similar results. GLM presented only one significant association undetected by MLM.

Table 7. Putative SSR markers associated with some traits in *Coffea canephora*

Trait	Marker name	F-value	P-value	R ² marker
<i>Yield (bags/ha)</i>				
	SSR057	42.23	0.00000000	0.45
	SSR059	18.79	0.00000035	0.31
	EST-SSR059	8.82	0.00000362	0.37
	EST-SSR018	6.80	0.00001120	0.24
<i>Number of inflorescence per branch</i>				
	SSR055	7.60	0.0010	0.15
	EST-SSR019	3.06	0.0021	0.26
<i>Length internode</i>				
	EST-SSR019	3.14	0.0017	0.10
<i>Number of branches per plant</i>				
	EST-SSR019	3.03	0.0024	0.37
<i>Plant height</i>				
	EST-SSR002	2.96	0.0038	0.39

Yield of green beans measured in bags/ha, across two harvests, presented highly associated ($F > 3.0$ and $P < 0.01$) with four markers: SSR057, SSR059, EST-SSR059 and EST-SSR018. The locus SSR057 explained 45% of the phenotypic variation of this trait. It is important to note that the total of R² markers could be higher than 100%, once the MLM procedure is performed to one marker at a time. Linkage disequilibrium (LD) among markers could be one of the causes of R² sum inflation. Still based on the criterions of F and P statistics, other associations would be found between the number of inflorescence per branch and the markers SSR055

and EST-SSR019. The marker EST-SSR019 was also associated with internode length and number of branches per plant; and plant height was associated with EST-SSR002.

Discussion

Brazil is the second most important country in production of *Coffea canephora*. Part of this success is due to the performance of the germplasm introduced in the last century and the efforts in genetic breeding. In this study, we took a representative sample of *C. canephora* conserved in gene banks, cultivated in commercial plantations and also used in breeding programs of four Brazilian states, *i.e.*: São Paulo, Minas Gerais, Espírito Santo and Rondônia. Our previous analysis confirmed that this germplasm presents a wide diversity, but also revealed the existence of strong subgrouping (Souza et al. unpublished). The effects of structure over the association analysis are well known and have been extensively discussed (Pritchard et al. 2000; Flint-Garcia 2003; Gupta et al. 2005). Structured populations may increase the risk of achieving false positive associations. Aranzana et al. (2005), testing the feasibility of genome-wide association mapping in *Arabidopsis thaliana*, reported that removing accessions genetically distinct and phenotypically extreme from the association analysis reduced considerably the number false positive associations. Thus, in order to minimize this problem, among those four samples from different places, we selected the Rondonian sample to perform a first effort of association mapping in Brazilian *C. canephora* germplasm. The accessions from Rondônia represent a relevant part of the Conilon coffee cultivated in Brazil and encompass a satisfactory molecular diversity (Souza et al. unpublished).

Our results demonstrated that these accessions also harbor high phenotypic variability. This result is in accordance with *C. canephora* allogamous nature and the preferential system of propagation by seeds, very practiced on Rondônia state (Souza and Santos 2009). Most report of diversity analysis in this species around the world has indicated a wide phenotypic variability. Brazilian studies have followed this rule (Souza and Santos 2009; Ivoglo et al. 2008; Fonseca et al. 2006). In our results, Rondonian accessions presented yield of green beans, which is the main trait target in coffee breeding, ranged from 78 to 8,642 kg/ha. Considering the average yields of 918 kg/ha and 1316 kg/ha, exhibited by Rondonian and Brazilian plantations, respectively (CONAB 2011), it is possible to suppose the potential of breeding of this sample.

A low level of LD was observed on accessions from Rondônia. This result is also in accordance with the strict allogamous reproductive system of *C. canephora* and may be attributed to high recombination rates. Taking reference thresholds of LD parameters, used in similar studies ($r^2 \geq 0.05$ and $P \leq 0.05$), we found 6.5% of SSR pairs in significant LD. This value was lower than those observed in other outcross species, as maize, that ranged from 10% to 56% (Stich et al. 2005, 2006; Remington et al. 2001). Too much small LD rates could difficult the task of finding marker-phenotypic associations in a wide genome scan. Our estimates were low, but not prohibitive. Besides, recent LD analysis using African germplasm of *C. canephora* showed that LD magnitude and its extent along the genome may be very diverse, varying in function of the origin of population and the genomic region (Cubry 2009). Thus, it is expected that the use of distinct accessions and other set of markers may lead to achievement of higher levels of LD.

Although extracted from a still very precarious linkage map, we assessed some information about LD among unlinked loci. This estimate was high and indicates that other forces besides genetic linkage are acting as LD generator in Rondonian population. In vegetal species, most reports have demonstrated a remarkable participation of other phenomena in LD creation, instead linkage. Remington et al. (2001), studying structure, LD and phenotypic associations in a sample of inbred maize lines, reported that LD observed in that germplasm was generated by the selection to adaptive traits as flowering time. Association studies in sugar beet observed high LD due the presence of related individuals in the sample (Breseghelo and Sorrells 2006). In upland cotton, selection, co-selection, relatedness and structure were pointed as causative of the excessive LD (Abdurakhmonov et al. 2009). Also, African subpopulations of *C. canephora* presented high number of unlinked SSR in LD, which was attributed to high level of relatedness, promoted by the small effective size of wide *C. canephora* populations (Cubry 2009).

A low level of structure and relatedness were observed in Rondonian sample. Three groups were found, but two showed a slight genetic distance. The third group presented more differentiated and it was composed by more related accessions. These conditions allowed us to perform LD analysis and a preliminary association study. Furthermore, we used the Q+K MLM method (Yu et al. 2006), which may promote the necessary adjustments to reduce the bias due to the structure and relatedness. This method has been successfully used in similar studies with rapeseed, sugar beet, maize and *Arabidopsis* (Stich et al. 2008b), cotton (Abdurakhmonov et al. 2009) and other species. Recent study, where LD was tested in populations of *C. canephora* from Africa, pointed the Q+K MLM as the most appropriate method to reduce false positives, due population structure. In that study, 78 accessions from two

samples representing the Congolese subgroup 2 – SG2 (21 accessions) and Guinean group (57 accessions) were used (Cubry 2009). Here, our sample was restricted to a subpopulation belonging to the Congolese subgroup 1. Thus, the level of structure and relatedness were very low and its effects may be efficiently corrected by the Q+K MLM method.

A more restrict sampling resulted in a small sample size and could imply some loss in accuracy and reduces the power of statistical tests. For this reason we have faced this study as a preliminary and speculative trial. Nevertheless, some ability of detecting true associations persists, as can be seen in many studies involving plant species, in which samples about 100 individuals were used, and significant associations were detected with satisfactory levels of confidence (Yao et al. 2009, Jun et al. 2008, Kantarzi and Stewart, 2008, Agrama et al. 2007, Malosetti et al. 2007, Aranzana et al. 2005, Skøt et al. 2005, Thornsberry et al. 2001).

In our preliminary study, we found nine high significant associations between SSR markers and phenotypic traits, using a sample of 76 accessions and 65 polymorphic SSR. Two different and most powerful statistical methods were used (GLM and MLM) and only significant associations detected on both were considered. Yield presented significant associations with four markers (6%) and the number of inflorescence per branch was associated with two markers (3%). Cubry (2009), in a first effort to perform association analysis in African populations, used a structured sample of 78 accessions and a set of 108 SSR. That study detected five (4.6%) significant associations among markers and the mean size of beans, recorded during six years. From those significative associations, four were comprehended into the confidence interval established in a previous QTL linkage mapping. This number was similar those found in a lot of studies with many vegetal species. In cotton,

Abdurakhmonov et al. (2009), using a set of 95 SSR and 285 accessions, found between 6% and 13% of SSR associated with traits of fiber quality. Remington et al. 2001 estimated LD structure and phenotypic-marker associations in maize using a set of 47 SSR. They observed a higher percent of significant associations (36% to 64%) among a set of SSR and four phenotypic traits. Nevertheless, in this case, the authors stated that these excessive percents of associations would have been caused by the presence of LD over large chromosomal blocks due to selection, instead properly linkage. In that work, associations among SSR and phenotype were evaluated by simple regression, without considering structure and relatedness.

Taking into account the crucial importance of the yield in the development of new cultivars, the associations presented in our preliminary study, if confirmed in future, would play a key role in *C. canephora* breeding programs involving those accessions. Anyway, our study indicated the potential importance of using association mapping to detect QTLs in *C. canephora*. Thus, we encourage the accomplishment of further studies and present a series of recommendations to improve their chances of success.

To obtain bigger clearance about the disequilibrium linkage in Brazilian *Coffea canephora* germplasm, it will be fundamental to finish the current genetic map. The definition about linkage groups, marker positions and distances is fundamental to establish the magnitude of the disequilibrium linkage blocks. Besides, the resolution and the applicability (gene candidate or wide genome scan) of association studies depend on the knowledge of the extent and the structure of disequilibrium linkage in target populations.

Multi-allelic SSR in multiplex arrays still remains an interesting marker to perform association mapping. The number of required SSR is lower than the number

of SNPs to achieve the same task, once SSRs harbor more information content per marker (Hall et al. 2010). Nowadays, the amount of polymorphic SSR available to *Coffea* genus is still limited (about 1000) and a considerable part of them could be no useful due to problems of amplification and lack of polymorphism. However, depended on, the extent of LD on target population, this number could be quite enough to promote good genome coverage. Thus, we advice the use of a larger set of SSR.

Generally, a LD mapping experiment will have lower statistic power than a linkage mapping assay of equivalent size (Mackay and Powell 2007). Also, it has been shown that increasing the number of phenotyped individuals is far more efficient than to expand the number of markers for enhance the power of association studies (Yu et al. 2008). Hence, we recommend augmenting dramatically the number of accessions.

Another issue to consider is the choice of traits to be analyzed and their distribution in the sample of accessions selected to association studies. Some characters could be strongly correlated with presence of population substructure. In *C. canephora*, for instance, leaf rust resistance is naturally higher in Congolese genotypes (Cubry 2009), especially in SG2 (Robusta). Accessions from SG1 (e.g. Conilon), that are predominant in Brazil, are susceptible to rust, on the other hand, they are more tolerant to drought. So, additional care must be employed in the composition of sample to study traits like these.

So, it will be important to concentrate some efforts to obtain high quality phenotypic data, with phenotyping being carried out over many years, in multiple environments (Flint-Garcia et al. 2005). This experimental scenario is relatively easy to obtain in *Coffea canephora* due its perennial cycle of life and the feasible use of

vegetative propagation (clonal cuttings), that permits an accession to be genotyped once and phenotyped repeatedly, including different environments.

References

- Abdurakhmonov IY, Abdugarimov A (2008) Application of association mapping to understanding genetic diversity of plant germplasm resources. **International Journal of Plant Genomics**.
- Abdurakhmonov IY, Saha S, Jenkins JN, Buriev ZT, Shermatov SE, Scheffler BE, Pepper AE, Yu JZ, Kohel RJ, Abdugarimov A (2009) Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. **Genetica** 136:401–417
- Agwanda CO, Lashermes P, Trouslot P, Combes M, Charrier A. (1997) Identification of RAPD markers for resistance to coffee berry disease, *Colletotrichum kahawae*, in arabica coffee. **Euphytica** 97:241–248.
- Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. **Molecular Breeding** 19:341–356
- Aranzana MJ, Kim S, Zhao K, et al (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. **PLoS Genet.** 1:e60.
- Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* (L.) showing wide cross-species amplifications. **Molecular Ecology Notes**, 3: 647-650.

- Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (2005) Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* x *Coffea congensis*). **Molecular Ecology Notes** 5:80-83.
- Bradbury, PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Soft ware for association mapping of complex traits in diverse samples. **Bioinformatics** 23:2633–2635.
- Breseghele F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. **Genetics** 172:1165-1177.
- Brito GG, Caixeta ET, Gallina AP, Maciel-Zambolim E, Zambolim L, Diola V, Loureiro ME (2010) Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. **Euphytica** 173:255-264.
- Campa C, Noirot M, Bourgeois M, Pervent M, Ky CL, Chrestin H, Hamon S, Kochko A (2003) Genetic mapping of a caffeoyl-coenzyme A 3-Omethyltransferase gene in coffee trees. Impact on chlorogenic acid content. **Theoretical and Applied Genetics** 107:751-756.
- Carvalho CHS (2008) **Cultivares de café: origem, características e recomendações**. Brasília:Embrapa Café. 334p.il.
- Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular Ecology** 9:1171-1193.

CONAB 2011. Acompanhamento da Safra Brasileira Café Safra 2011, segunda estimativa, maio/2011. Companhia Nacional de Abastecimento. Brasília: Conab.

Coulibaly I, Louarn J, Lorieux M, Charrier A, Hamon S, Noirot M (2003a) Pollen viability restoration in a *Coffea canephora* P. and *C. heterocalyx* Stoffelen backcross. QTL identification for marker-assisted selection. **Theoretical and Applied Genetics** 106:311-316.

Coulibaly I, Revol B, Noirot M, Poncet V, Lorieux, Carasco-Lacombe, Minier J, Dufour M, Hamon P (2003b) AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* x *C. canephora*) x *C. canephora*]. **Theoretical and Applied Genetics** 107:1148-1155.

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmechel KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. **Nature Genetics** 7:180-184. 1994.

Cubry P (2009) **Structuration de la diversité génétique et analyse des patrons de déséquilibre de liaison de l'espèce *Coffea canephora* Pierre ex. Froehner.** Doctoral Thesis. Université Montpellier II. 242p.

Diola V, Brito GG, Caixeta ET, Zambolim EM, Sakiyama NS, Loureiro ME (2011) High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics & Genomes**, 7:1-10.

- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology** 14:2611–2620.
- Flint-Garcia SA, Thuillet A, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: A high-resolution platform for quantitative trait locus dissection. **Plant Journal** 44:1054–1064.
- Fonseca AFA, Sedyama T, Cruz CD, Sakiyama NS, Ferrão MAG, Ferrão RG and Bragança SM (2000) Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira** 41: 599–605.
- Gupta PK, Rustgi S, Kulwal PL, Linkage disequilibrium and association studies in higher plants : present status and future prospects. **Plant Molecular Biology** 57:461-485. 2005.
- Hall D, Tegström C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. **Briefings in Functional Genomics and Proteomics Advance Access**. P1-9.
- Hardy OJ, and Vekemans X (2002) SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. **Molecular Ecology Notes** 2:618–620.
- Herrera JC, Alvarado A, Cortina HA, Combes M-C, Romero G, Lashermes P (2009) Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk & Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. **Euphytica** 167:57-67

- Ivoglo MG, Fazuoli LC, Oliveira ACB, Gallo PB, Mistro JC, Silvarolla MB, Tomabraghini M (2008) Divergência genética entre progênies de café robusta. **Bragantia** 67:823–831.
- Jun T-H, Van K, Kim MY, Lee HS, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. **Euphytica** 162:179–91.
- Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox DK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of cystic fibrosis gene: genetic analysis. **Science** 245:1073-1080.
- Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetics** 101:669–676.
- Lashermes P, Couturon E, Moreau N, Paillard M, Louarn J (1996) Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. **Theor Appl Genet** 93:458–462
- Lashermes P, Combes MC, Prakash NS, Trouslot P, Lorieux M, Charrier A (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome** 44:589-596.
- Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LGE, Piffanelli P (2005) Construction and characterization of a *Coffea canephora* BAC library to

- study the organization of sucrose biosynthesis genes. **Theoretical and Applied Genetics** 111: 1032-1041.
- 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
- Lynch M, Walsh B (1998) **Genetics and Analysis of Quantitative Traits**. Sunderland, MA: Sinauer Associates.
- Kantartzi SK, Stewart JM (2008) Association analysis of fiber traits in *Gossypium arboreum* accessions, **Plant Breed** 127:173–179.
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science** 12:57-63.
- Mahé L, Combes MC, Várzea VMP, Guilhaumon C, Lashermes P (2008) Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.) **Molecular Breeding** 21:105-113.
- Malosetti, M., C.G. van der Linden, B. Vosman, and F.A. van Eeuwijk (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. **Genetics** 175:879–889.
- Moncada P, McCouch SR (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome** 47:501–509.
- Noir S, Anthony F, Bertrand B, Combes MC, Lashermes P (2003) Identification of a major gene (Mex-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea*. **Plant Pathology** 52:97-103.

- Paillard M, Lashermes P, Pétiard V (1996) Construction of a molecular linkage map in coffee. **Theoretical and Applied Genetics** 93:41-47.
- Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R (2004) Construction of a genetic map for arabica coffee. **Theoretical and Applied Genetics** 108:829-835.
- Poncet V, Hamon P, Minier J, Carasco-Lacombe C, Hamon S, Noirot M (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). **Genome** 47:1071–1081.
- 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 *Coffea arabica* L. **Theor Appl Genet** 109:1311–1317
- Price AH (2006) Believe it or not, QTLs are accurate! **Trends Plant Sci.** 11:213–216.
- Pritchard JK, Stephens M, and Donnelly P (2000) Inference of population structure using multilocus genotype data. **Genetics** 155:945–959.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Proc Natl Acad Sci** 98: 11479–84.
- Rovelli P, Mettullo 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**. Kluwer, Netherlands, pp 123–133.
- Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH, Sackville Hamilton NR (2005) An association mapping

approach to identify flowering time genes in natural populations of *Lolium perenne* (L). **Molecular Breeding** 15:233–45.

Sorkheh K, Malysheva-Otto LV, Wirthensohn MG, Tarkesh-Ssfahani S, Martínez-Gómez P (2008) Linkage disequilibrium, genetic association mapping and gene localization in crop plants. **Genetics and Molecular Biology**, 31, 4, 805-814.

Souza FF, Santos MM (2009) Melhoramento genético do café canéfora em Rondônia. In: Zambolim L. (edt). **Tecnologias para produção do café Conilon**. Viçosa, 2009, MG: Universidade Federal de Viçosa, p.175-200.

Souza FF, Caixeta ET, Ferrão LFV, Pena GF, Zambolim EM, Pereira AA, Oliveira ACB, Zambolim L, Cruz CD, Sakiyama NS. Molecular overview about *Coffea canephora* germplasm in Brazil. **(Unpublished)**.

Souza FF, Caixeta ET, Ferrão LFV, Pena GF, Zambolim EM, Pereira AA, Oliveira ACB, Zambolim L, Cruz CD, Sakiyama NS. Genetic structure of Brazilian *Coffea canephora* germplasm assessed by microsatellites. **(Unpublished)**

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. **Theoretical and Applied Genetics** 111:723-730.

Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, Peleman and Reif JC (2006) Comparison of Linkage Disequilibrium in Elite European Maize Inbred Lines using AFLP and SSR Markers. **Molecular Breeding** 3: 217-226.

Stich B, Möhring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008a) Comparison of mixed-model approaches for association mapping. **Genetics** 178:1745-1754.

- Stich B, Piepho HP, Schulz B, Melchinger AE (2008b) Multi-trait association mapping in sugar beet (*Beta vulgaris* L.) **Theoretical and Applied Genetics** 117:947-954.
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D and Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. **Nature Genetics** 28:286–289.
- Weir BS (1996) **Genetic Data Analysis II - Methods for Discrete Population Genetic Data**. Sinauer Associates, Sunderland.
- Yao J, Wang L, Liu L, Zhao C, Zheng Y (2009) Association mapping of agronomic traits on chromosome 2A of wheat. **Genetica** 137:67–75.
- Yu J, Pressoir G, Briggs WH, Vroh-Bi I, Yamasaki M, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nat. Genet.** 38:203–208.
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. **Genetics** 178:539–551
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Review & interpretation. **Plant Genome** 1:5-20

CONCLUSÃO GERAL

O germoplasma brasileiro de *C. canephora* apresenta grande diversidade, mas representa pequena porção da diversidade total da espécie. Novos esforços de coleta e introdução de acessos devem ser realizados visando a ampliar a base genética dos programas de melhoramento e, conseqüentemente, das lavouras nacionais.

A espécie apresenta forte estruturação, estando dividida em dois grupos genéticos, que correspondem aos subgrupos SG1 (Conilon) e SG2 (Robusta). Uma subestruturação em função do local de cultivo (Espírito Santo x Rondônia) também foi observada dentro dos acessos de Conilon.

Os genótipos do tipo Conilon, são predominantes nas lavouras brasileiras. O aumento de alelos de Robusta no melhoramento genético da cultura poderá trazer notáveis ganhos à cafeicultura nacional, pois agregaria características tais como: resistência à ferrugem, resistência aos nematóides, redução do teor de cafeína, aumento da peneira média, entre outros.

C. arabica e Híbrido de Timor apresentam maior similaridade com os genótipos do tipo Robusta do que com genótipos do tipo Conilon. Esse fato demonstra que a estrutura da espécie *C. canephora* deve ser levada em consideração nos estudos filogenéticos do gênero *Coffea*.

Ampla diversidade fenotípica, baixo nível de estruturação e baixo nível de parentesco foram observados entre os acessos de *C. canephora* coletados em Rondônia. Tais condições possibilitaram o emprego dessa subpopulação em estudos de mapeamento associativo.

Baixo nível de desequilíbrio de ligação (LD) foi verificado no germoplasma de *C. canephora* proveniente das lavouras de Rondônia. Parte desse LD pode ser atribuído a outras causas, além da ligação genética, uma vez que foi detectado considerável percentual de pares de locos microssatélites não ligados, que apresentaram significativo LD. Verificaram-se associações significativas entre alguns dos microssatélites avaliados e as características fenotípicas: altura de planta, número de ramos plagiotrópicos por planta, comprimento de internódio e número de inflorescências (rosetas) por ramo e produtividade.