



**ELISA SUSILENE LISBOA DOS SANTOS**

**DIVERSIDADE GENÉTICO-MOLECULAR DE CACAUEIROS DESCENDENTES DAS  
PRIMEIRAS INTRODUÇÕES OCORRIDAS NA BAHIA**

***MOLECULAR GENETIC DIVERSITY OF CACAO DESCENDANTS OF THE FIRST  
INTRODUCTIONS OCCURRED IN BAHIA***

**CAMPINAS**

**2014**





UNIVERSIDADE ESTADUAL DE CAMPINAS  
Instituto de Biologia

ELISA SUSILENE LISBOA DOS SANTOS

**DIVERSIDADE GENÉTICO-MOLECULAR DE CACAUEIROS DESCENDENTES DAS PRIMEIRAS  
INTRODUÇÕES OCORRIDAS NA BAHIA**

***MOLECULAR GENETIC DIVERSITY OF CACAO DESCENDANTS OF THE FIRST INTRODUCTIONS  
OCCURRED IN BAHIA***

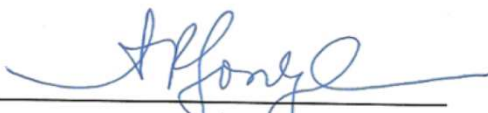
Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Doutora em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento

*Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Plant Genetics and Genetic Breeding*

Orientadora: ANETE PEREIRA DE SOUZA

Co-orientador: RONAN XAVIER CORRÊA

Este exemplar corresponde à versão final da tese defendida pela aluna Elisa Susilene Lisboa dos Santos e orientada pela profa. Dra. Anete Pereira de Souza



---

Anete Pereira de Souza

CAMPINAS  
2014

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Biologia  
Mara Janaina de Oliveira - CRB 8/6972

Santos, Elisa Susilene Lisboa, 1985-  
Sa59d      Diversidade genético-molecular de cacaueiros descendentes das primeiras introduções ocorridas na Bahia / Elisa Susilene Lisboa dos Santos. – Campinas, SP : [s.n.], 2014.

Orientador: Anete Pereira de Souza.  
Coorientador: Ronan Xavier Corrêa.  
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Variação genética. 2. Cacao - Melhoramento genético. 3. Marcadores moleculares. 4. Microsatélites (Genética). I. Souza, Anete Pereira de, 1962-. II. Corrêa, Ronan Xavier. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Molecular genetic diversity of cacao descendants of the first introductions occurred in Bahia

**Palavras-chave em inglês:**

Genetic variation

Cacao - Genetic breeding

Molecular markers

Microsatellites (Genetics)

**Área de concentração:** Genética Vegetal e Melhoramento

**Titulação:** Doutora em Genética e Biologia Molecular

**Banca examinadora:**

Anete Pereira de Souza [Orientador]

Antonio Augusto Franco Garcia

Fábio Gelape Faleiro

Maria Imaculada Zucchi

Luiz Roberto Martins Pinto

**Data de defesa:** 11-07-2014

**Programa de Pós-Graduação:** Genética e Biologia Molecular



Campinas - São Paulo, 11 de julho de 2014

**Banca Examinadora**

Profa. Dra. Anete Pereira de Souza  
(Orientadora)



Prof. Dr. Antonio Augusto Franco Garcia



Dr. Fábio Gelape Faleiro



Prof. Dr. Luiz Roberto Martins Pinto



Dra. Maria Imaculada Zucchi



Prof. Dr. André Olmos Simões



Prof. Dr. Rafael Silva Oliveira



Dr. Paulo de Souza Gonçalves





## Resumo

O cacaueiro é uma planta perene amplamente cultivada entre as latitudes 20°N e 20°S. O Brasil é o principal produtor das Américas, tendo 62% da sua produção concentrada em áreas de plantio do sudeste da Bahia. A introdução do cacau neste Estado ocorreu a partir de 1746 no município de Canavieiras, primeiro com a chegada da variedade 'Comum' e posteriormente com a chegada das variedades 'Pará' e 'Maranhão'. Descendentes destas introduções foram usadas para gerar plantios que predominaram na Bahia por mais de dois séculos e são denominados de 'cacau da Bahia' ou variedades locais baianas. Com a crise da vassoura-de-bruxa, a partir de 1989, as variedades locais baianas passaram a ser substituídas por clones resistentes selecionados em fazendas, coleções de germoplasma e programas de melhoramento genético. Embora tenham sido caracterizadas como susceptível à vassoura-de-bruxa, plantas de 'cacau da Bahia' ainda são encontradas em 50% da área produtiva, sendo cultivadas principalmente por pequenos produtores. Tendo em vista o exposto, o objetivo desta tese foi contribuir com informações genético-moleculares para o progresso dos programas de melhoramento do cacaueiro da Bahia. Para tanto, 17 marcadores microssatélite polimórficos foram obtidos a partir de bibliotecas genômicas enriquecidas. Estes marcadores foram somados a outros 13 presentes na literatura e empregados na avaliação da diversidade do 'cacau da Bahia', presentes em fazendas e em banco de germoplasma. A análise de 176 genótipos coletados em fazendas e institutos de pesquisa baianos (nos municípios de Canavieiras, Camacan, Uruçuca e Gandu) indicou a existência de dois grupos genéticos (denominados de cacau da Bahia (CB) I e II,  $G_{ST} = 0,22$ ). Os valores de diversidade apresentados por CB I e CB II foram baixos (riqueza alélica = 1,31 e 1,41 e  $H_E = 0,11$  e 0,15, respectivamente para CB I e CB II). Alto índice de fixação alélica foi também observado para os cacaueiros baianos ( $F = 0,28$ ). Análise de nove características de importância econômica a partir de frutos e grãos em 106 genótipos foi realizada e constitui informações preliminares para a escolha de plantas candidatas a novas avaliações. A avaliação de clones pertencentes

às seleções do Instituto do Cacau da Bahia (SIC) e do Instituto Agronômico do Leste (SIAL) (representantes do 'cacau da Bahia' em bancos de germoplasma), também indicaram baixa variabilidade genética (2,6 alelos por loco e  $H_E = 0,22$ ) e alto índice de fixação alélica (0,38). A variabilidade foi estruturada em dois grupos ( $G_{ST} = 0,27$ ), havendo uma tendência de separar os indivíduos de acordo com a série (SIC ou SIAL). A baixa diversidade observada para o 'cacau da Bahia' é reflexo, entre outros fatores, das restritas introduções de plantas que fundaram boa parte do cultivo no Estado. Análises moleculares conjuntas de todas as plantas neste trabalho indicaram que parte da diversidade dos cacauzeiros baianos presente em fazendas não está amostrada nas plantas do banco de germoplasma, sendo urgentemente necessário efetuar amostragens adicionais, para incremento da representatividade dos bancos no que tange o 'cacau da Bahia'. Os resultados obtidos ao longo do trabalho de doutorado trouxeram informações e perspectivas úteis tanto para incrementar a conservação do 'cacau da Bahia' quanto para o delineamento de ações visando implementar programas de seleção recorrente em programas de melhoramento direcionados à obtenção de 'cacau da Bahia' mais produtivo, com qualidade superior e também, tolerante às principais doenças que o acometem.

## Abstract

*Theobroma cacao* L. is a perennial plant cultivated in latitudes 20°N and 20°S. Brazil is the most important cacao-producing country in the Americas and 62% of the Brazilian cacao production is developed in Southern Bahia region (a Brazilian State). The introduction of cacao in Bahia started in 1746, in the municipality of Canavieiras; first with the arrival of the 'Comum' variety and next with the 'Pará' and 'Maranhão' varieties. Descendents of these introductions were naturalized as 'Bahian cacao' or Bahian local varieties and have been cultivated for over 200 years. With the witches' broom outbreak on the Bahian farms in 1989, 'Bahian cacao' have been replaced by more resistant clones, obtained by selection on farms, germplasm collection and from breeding programs. Even though local cultivars are susceptible to witches' broom disease, they are still being planted in 50% of farms, especially by smallholders. The goal of this thesis was to contribute to the breeding program of cacao using molecular and genetics characterization of 'Bahian cacao'. Seventeen polymorphic microsatellite markers for cacao were developed from genomic libraries and, in addition to other thirteen, were employed to characterize 'Bahian cacao' plants obtained from farms and in a germplasm collection. One hundred seventy six cacao genotypes of the Bahian cultivars were obtained from farms and institutes of four Bahian municipalities and the analysis indicated structure from genotypes in two groups ('Bahian cacao' (BC) I and II,  $G_{ST} = 0.22$ ). Lower genetic diversity were observed for BC I and BC II (allelic richness = 1.31 and 1.41; and  $H_E = 0.11$  and 0.15 for BC I and BC II, respectively). High fixation index was observed for 'Bahian cacao' (0.28). Phenotypic evaluations of nine characteristic of economic importance from fruits and seeds in 106 farm cacao plants were realized and constituted a preliminary approach for choice of candidate plants for additional analysis. Evaluation of clones representatives of 'Bahian cacao' in germplasm collections and selected by Bahian Cacao Institute and Agronomic Institute of East (these clones designated as SIC and SIAL, respectively), also indicated low genetic diversity (mean alleles per locus = 2.60 and  $H_E = 0.22$ ) and high fixation index ( $F =$

0.38). Structure levels were revealed in two groups ( $G_{ST} = 0.27$ ) formed according to clones selection (SIC and SIAL). The low genetic diversity observed for 'Bahian cacao' reflect the founder effect of introduced plants and that served as resource to start almost all 'Bahian cacao' plantations. Combined molecular analyzes of all the plants of 'Bahian cacao' used in this study indicated that part of the diversity present on farm is not sampled in plants of germplasms collections, being additional sampling needed. The results presented in this thesis are useful information both for conservation of 'Bahian cacao' plants and to the use of this in recurrent breeding programs in order to obtain 'Bahian cocoa' more productive, with superior quality and tolerant to major diseases.

## SUMÁRIO

|                                                                                                                                                                                                            |      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Resumo.....</b>                                                                                                                                                                                         | vii  |
| <b>Abstract .....</b>                                                                                                                                                                                      | ix   |
| <b>Dedicatória .....</b>                                                                                                                                                                                   | xiii |
| <b>Agradecimento .....</b>                                                                                                                                                                                 | xv   |
| <b>Organização da Tese.....</b>                                                                                                                                                                            | xvii |
| <b>INTRODUÇÃO GERAL .....</b>                                                                                                                                                                              | 1    |
| <b>OBJETIVO .....</b>                                                                                                                                                                                      | 31   |
| <b>CAPÍTULO 1. Desenvolvimento e caracterização de marcadores<br/>microsatélites para cacauero .....</b>                                                                                                   | 33   |
| New polymorphic microsatellite loci for <i>Theobroma cacao</i> :<br>isolation and characterization of microsatellites from enriched<br>genomic libraries .....                                             | 35   |
| <b>CAPÍTULO 2. Diversidade genética de cacaueros da Bahia I .....</b>                                                                                                                                      | 39   |
| Southern Bahia local cacao varieties from the Amelonado group:<br>genetic diversity and structure of populations established for 260<br>years and survivors of the epidemic of witches' broom disease .... | 41   |
| <b>CAPÍTULO 3. Diversidade genética de cacaueros da Bahia II .....</b>                                                                                                                                     | 71   |
| Molecular diversity and genetic structure of Amelonado cacao<br>from the Bahia, Brazil .....                                                                                                               | 73   |
| <b>RESULTADOS COMPLEMENTARES .....</b>                                                                                                                                                                     | 111  |
| <b>CONCLUSÕES .....</b>                                                                                                                                                                                    | 117  |
| <b>PERSPECTIVAS .....</b>                                                                                                                                                                                  | 119  |
| <b>LITERATURA CITADA .....</b>                                                                                                                                                                             | 121  |
| <b>ANEXOS .....</b>                                                                                                                                                                                        | 135  |
| Development and characterization of microsatellite markers for<br>the wild South American <i>Passiflora cincinnata</i> .....                                                                               | 137  |
| New microsatellite markers for wild and commercial species of<br><i>Passiflora</i> (Passifloraceae) and cross-amplification .....                                                                          | 141  |
| Declaração - Comissão Interna de Biossegurança.....                                                                                                                                                        | 147  |





## **Dedicatória**

À minha mãe Eunice pelo exemplo de força e trabalho que refletiram diretamente na minha busca pelo conhecimento, dedico.

Ao meu marido Carlos Bernard pelas palavras sábias e gentis atitudes, que transparecem seu caráter e dedicação ao trabalho e o faz uma pessoa notável, de ideias brilhantes e agradável companhia.



## **Agradecimentos**

À Deus pela oportunidade da formação acadêmica e pela a saúde e discernimento que me permitiram aproveitar as oportunidades dadas.

Ao meu marido Carlos Bernard, por compartilhar os sonhos comigo e não medir esforços para realizá-los.

Aos meus pais Eunice e Clóvis, e irmãos Heli e Junior, por entenderem a minha ausência e me apoiarem nas minhas escolhas.

À minha orientadora Profa. Dra. Anete Pereira de Souza, pela confiança depositada e pelos inúmeros ensinamentos.

Ao co-orientador Prof. Dr. Ronan Xavier Corrêa, por ter sido a ponte inicial entre mim e o doutorado na Universidade Estadual de Campinas (UNICAMP), pela disposição para as coletas e pelos ensinamentos.

Ao Prof. Dário Ahnert pelo auxílio prestado desde a proposição do projeto até as discussões científicas a respeito da história e do melhoramento genético do cacaueteiro.

Aos colegas e funcionários do Laboratório de Análise Genética Molecular (LAGM) do Centro de Biologia Molecular e Engenharia Genética (CBMEG) da UNICAMP, pelo apoio, companheirismo e pelos bons momentos vivenciados, em especial ao Dr. Juverlande e aos técnicos Aline, Patrícia e Danilo por garantirem o bom funcionamento das atividades laboratoriais.

Aos eternos amigos Melina, Fábio, Lívia, Gustavo, João Guilherme, Benício, Prianda, Rafaela, Américo e Guilherme pelo carinho no acolhimento e pela disponibilidade em auxiliar nas atividades científicas.

Aos amigos da Universidade Estadual do Sudoeste da Bahia (UESB), Murilo, Janaína, Obertal e Lígia pelo suporte dado nas atividades docentes e científicas que viabilizaram o meu afastamento temporário das atividades docentes para a finalização do doutorado.

Aos professores e funcionários do programa de Pós-Graduação em Genética e Biologia Molecular da UNICAMP pelo suporte administrativo.

Ao Dr. Paulo Gonçalves e à Dra. Maria Zucchi pelas contribuições realizadas durante o seminário de pré-banca de defesa de tese.

À UNICAMP, à UESB e à Universidade Estadual de Santa Cruz (UESC), pelo uso do espaço físico para realização das atividades de pesquisa.

Ao Centro de Pesquisas do Cacau da Comissão Executiva do Plano da Lavoura Cacaueira (CEPEC / CEPLAC), ao Instituto Federal de Educação Baiano de Uruçuca (IFBaiano) e aos fazendeiros que permitiram nosso acesso aos cacaueiros para a realização da pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) e à UESB pela concessão de bolsas de doutorado.

Ao CNPQ e à Fundação de Amparo à Pesquisa de São Paulo (FAPESP) pelo auxílio financeiro prestado aos projetos da Profa. Anete que possibilitaram a realização desta tese.

## Organização da Tese

Esta tese foi desenvolvida com intuito de contribuir para o programa de melhoramento genético do cacauzeiro na Bahia por meio de informações genético-moleculares sobre as variedades de cacauzeiro naturalizadas baianas. As etapas de cunho prático relativas às atividades da tese são apresentadas em forma de capítulos, estando estes organizados de forma a atender às normas específicas das revistas científicas para os quais foram submetidos. Todos os quesitos da tese refletem as normas estabelecidas pela Comissão Central de Pós-Graduação da UNICAMP. Neste sentido, itens como Introdução Geral e Objetivos precedem os capítulos, sendo que estes últimos podem ser didaticamente classificados de acordo com os objetivos específicos de cada um. Um dos capítulos está relacionado ao desenvolvimento de ferramentas moleculares e outros dois estão relacionados à aplicação destas ferramentas, bem como avaliações fenotípicas na caracterização da diversidade genética de cacauzeiros da Bahia.

No primeiro capítulo, o leitor encontrará a descrição acerca do desenvolvimento e caracterização de marcadores microssatélites para o cacauzeiro, sendo os resultados obtidos publicados na forma de nota científica no volume 56 do periódico *Biologia Plantarum*. Além de uma aplicação direta destes marcadores nos estudos de diversidade do cacau da Bahia, outras ações do grupo de pesquisa foram beneficiados pela disponibilização destes marcadores, sendo os mesmos aplicados na saturação de mapa genético para cacauzeiro, cujo trabalho foi desenvolvido por outra aluna de doutorado.

Nos capítulos seguintes a diversidade presente no cacau da Bahia pode ser contemplada por meio de avaliações moleculares, fenotípicas e das discussões dos resultados obtidos à luz das questões biológicas inerentes ao cacauzeiro, bem como dos eventos históricos relacionados tanto à sua introdução na Bahia quanto ao processo de dispersão ao longo das zonas de plantio. As peculiaridades dos tipos de amostras utilizada em cada capítulo permitem vislumbrar o panorama atual do cacau da Bahia

cultivado no presente, e incitam reflexões acerca dos efeitos da vassoura-de-bruxa sobre a diversidade genética (Capítulo 2). Ainda nesta linha, a representatividade da diversidade em bancos de germoplasma é discutida (Capítulo 3), sendo indicada a necessidade de novas coletas para amostrar de forma efetiva e realista a variabilidade genética disponível. Esta é uma medida importante para fins de conservação e para o que a diversidade existente no cacau da Bahia possa ser completamente aproveitada em programas de melhoramento do cacaueiro. Os artigos referentes aos Capítulos 2 e 3 encontram-se em processo de submissão para revistas científicas.

Com intuito de fornecer um panorama geral acerca da variabilidade presente no cacau da Bahia foi inserido, após o Capítulo 3, o quesito Resultados Complementares, sendo o espaço também utilizado para fomentar uma discussão geral acerca dos resultados obtidos. Logo após, o leitor terá disponível respectivamente as Conclusões da tese e as Perspectivas

Por fim, são apresentados, como Anexos, dois artigos publicados pelo grupo de pesquisa com objetivo de fornecer marcadores microssatélites úteis para estudos de diversidade do gênero *Passiflora*. Nestes artigos a autora da tese atuou na condição de colaboradora nas etapas de coleta, genotipagem e análise dos dados, sendo eles, portanto, parte de sua formação acadêmica.

## INTRODUÇÃO GERAL

### Características gerais do cacaueiro

O cacaueiro (*Theobroma cacao*) é uma planta diploide ( $2n = 10$ ) e perene, classificada na divisão angiosperma e classe dicotiledônea, sendo pertencente à família Malvaceae e ao gênero *Theobroma*, que engloba 22 espécies com ocorrência desde a Colômbia até a América Central. Devido a importância econômica dos seus frutos, cujos grãos constituem a matéria prima para fabricação do chocolate, o cacaueiro é amplamente cultivado, sendo a única espécie do gênero *Theobroma* cultivada intensamente ao redor do globo. Neste contexto, é possível encontrar plantações comerciais de cacau nas latitudes entre 20°N e 20°S, que oferecem as condições climáticas – quente e com precipitações regulares ao longo do ano – favoráveis para sua produção.

Em condições de cultivo as árvores de cacau podem alcançar até seis metros de altura, enquanto que em condições naturais, especialmente em florestas onde há disputa por luz com outras espécies, estas árvores podem alcançar até 20 metros (CARLETTO, 1973). Mesmo em condições naturais o cacaueiro é encontrado no sub-bosque, o que explica a sua adaptação aos ambientes com incidência indireta de luz solar. Essa característica é bastante relevante, haja vista que cacaueiros são encontrados em regiões consideradas *hot spots* de diversidade genética (GULTINAN, 2007) e a manutenção da espécie em consórcio com arbóreas de maior porte é uma alternativa viável ecológica e economicamente.

Do ponto de vista morfológico, as folhas da árvore do cacau apresentam variações, podendo ser mais estreitas ou mais largas, elípticas a ovaladas; bem como sua cor que, quando jovens, pode variar entre verde claro a matizes de vermelho. Além disso, variações quanto as bordas, ápices e bases foliares também são descritas (BARTLEY, 2005). O cacaueiro geralmente possui inflorescências do tipo cauliflora (Figura 1), com flores hermafroditas e pentâmeras (SCHAWWE et al., 2013), e cujo os cinco lóculos presentes no ovário produzem entre 30 e 70 óvulos (DIAS, 2001).



**Figura 1.** Flores, sementes e frutos de cacauero. Acima e à esquerda: botão floral e aspecto geral das flores. Acima e à direita: diversidade de cores e formas das sementes. Abaixo: Diferentes formas e cores de frutos de cacauero. Fonte: International Cocoa Germplasm Database <http://www.icgd.rdg.ac.uk/index.php>

Quando adultos, os cacaueiros produzem mais de 50 mil flores por ano (BHATTACHARJEE e KUMAR, 2007). Estas flores apresentam pétalas, sépalas, estames, carpelos e estaminódios. As pétalas e os estaminódios envolvem anteras e estigmas, respectivamente, formando uma arquitetura tal que dificulta a autopolinização e viabiliza a ocorrência de fecundação cruzada. Além disso, a presença de um sistema de autoincompatibilidade ajuda a justificar as altas taxas de fecundação cruzada



(variando entre 30 e 100%, SCHAWÉ et al. (2013); SILVA et al. (2011)) e, portanto, a discriminação de *T. cacao* enquanto espécie preferencialmente alógama. Cacaueiros autoincompatíveis são muito comuns em locais onde estas plantas ocorrem naturalmente (SILVA et al., 2011). Embora ainda não esteja totalmente elucidado para o cacaueiro, sabe-se que a autoincompatibilidade obedece a expressão dos alelos do gene S, que leva a falhas na fusão do óvulo com o pólen e culmina com a abscisão floral.

Do ponto de vista da polinização, pequenas moscas do gênero *Forcipomya* são reconhecidas como os principais agentes polinizadores das flores da árvore de cacau, podendo alcançar distâncias tão longas quanto 67 m (SILVA et al., 2011). Entretanto, há prevalência de polinização a curtas distâncias – entre 3 e 12 m, o que reflete o comportamento dos polinizadores (YAMADA e GURIES, 1998). Em condições de cultivo é observado que menos de 5% das flores produzidas pelas árvores de cacau são polinizadas, sendo que as não polinizadas caem dentro de 8 a 10 horas após emergência (ALVIM, 1984). Das flores fertilizadas, frequentemente, apenas 0,5 a 2,0 % vão gerar frutos, levando de 4 a 6 meses para que os mesmos estejam maduros (BHATTACHARJEE e KUMAR, 2007).

Os frutos do cacaueiro são indeiscentes, variando entre 12 e 30 cm, e embora esta seja uma característica bastante influenciada pelo ambiente, tamanhos específicos entre as variedades são notados (BARTLEY, 2005) havendo uma tendência dos frutos serem menores e em menor quantidade quando se trata de material silvestre (SCHAWÉ et al., 2013; LACHENAUD e ZHANG 2008). Além disso, como observa-se na Figura 1, os frutos apresentam aspectos variáveis quanto as formas e cores, além da superfície, que refletem características próprias a determinadas variedades. Neste sentido, a forma dos frutos de cacau pode ser desde esférica até alongada e a superfície pode ser completamente lisa ou de aspecto áspero, ou ainda, apresentar cristas cuja distância entre elas também gera um padrão de diversidade (BARTLEY, 2005). Ainda sobre a casca, estas variam desde espessas e rígidas, de difícil rompimento, até mais delgadas. Quanto às cores, os frutos podem ser desde verde claro até vermelho escuro, podendo ainda apresentar pigmentos vermelhos, amarelos ou alaranjados a depender da variedade e grau de maturidade (MONTEIRO et al., 2009).

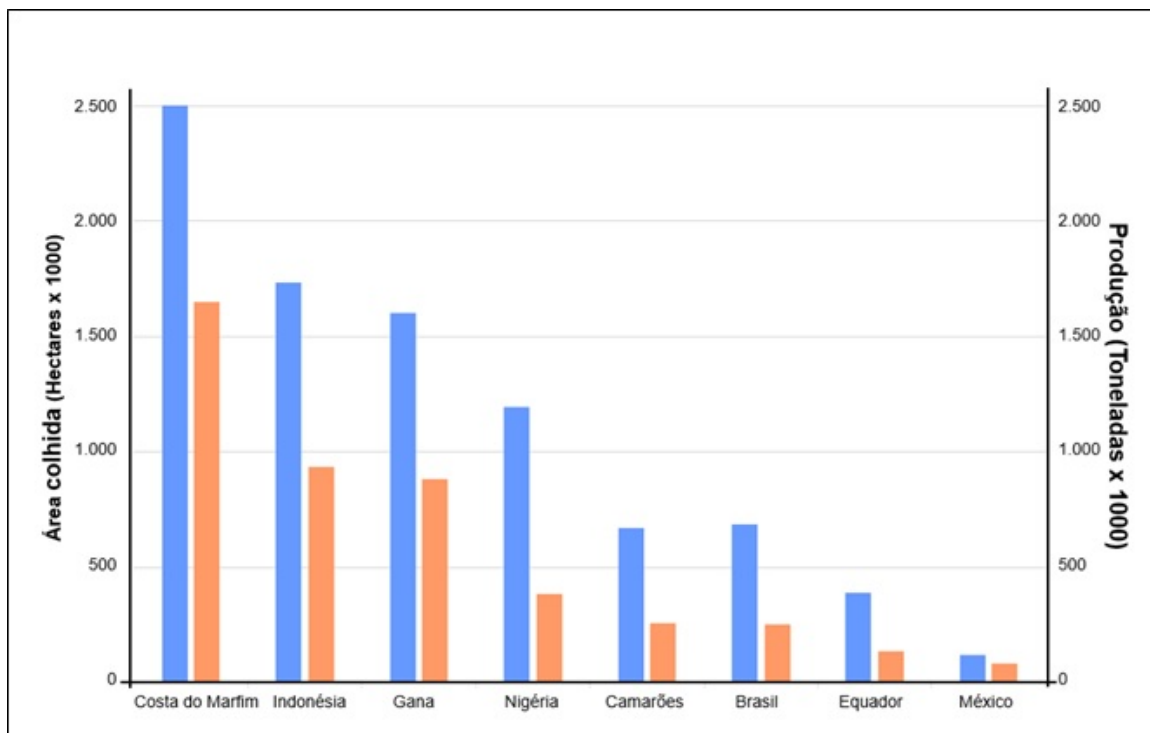
Assim como visto para os frutos, as sementes de cacau apresentam variações notáveis quanto à cor, forma e tamanho (Figura 1), sendo algumas destas características importantes para a indústria por influenciar diretamente processamento e sabor do chocolate. A cor da semente, por exemplo, varia de branco até púrpura, sendo a gradação definida pela presença e quantidade do pigmento antocianina. Às sementes sem este pigmento ou com pouca quantidade foram creditadas por muito tempo a produção de chocolate com sabor e aroma (flavor) mais agradável denominado de chocolate 'gourmet'. A mucilagem que envolve a semente tem pouca variação na cor, sendo na maioria das variedades de cor branca, mas varia amplamente no sabor, indo desde doce e palatável até ácido. Elevada acidez da mucilagem pode tornar este componente desagradável em algumas variedades de cacau e pode influenciar na dispersão das sementes. A dispersão das sementes de cacau é realizada especialmente por primatas, sendo relatados também a atuação de roedores e aves (ALLEN, 1988), além dos seres humanos que se destacam como os mais eficientes dispersores da espécie.

### **Importância econômica do cacaueiro**

O cacaueiro tem grande importância econômica especialmente por ser o principal ingrediente do chocolate, um dos alimentos mais apreciados do mundo. Além disso, os grãos de cacau e seus derivados são empregados na produção de bebidas, na indústria de cosméticos e mais recentemente, têm sido estudados pelo potencial benefício para a saúde humana, o que é em parte devido ao potencial antioxidante de alguns de seus componentes (LATIF, 2013).

É estimado que a produção de grãos de cacau nos anos de 2011/2012 tenha sido de cerca de 4,05 milhões de toneladas, sendo 58% deste montante obtido a partir de plantios da Costa do Marfim e Gana (ICCO, 2013). Estes países em conjunto com Camarões e Nigéria, são os líderes na produção de grãos de cacau (Figura 2), sendo responsáveis por 71% de toda a produção mundial. A América do Sul tem ocupado aproximadamente 15,5% deste mercado, sendo o Brasil e o Equador os países de

maior destaque na produção de cacau realizada no continente americano (<http://faostat3.fao.org>). A Indonésia vem experimentando a expansão do cultivo nos últimos anos, tendo alcançado o segundo lugar no ranking de produção mundial de grãos de cacau (<http://faostat3.fao.org>).



**Figura 2.** Área colhida (em azul) e quantidade de grãos de cacau produzida (em laranja) no ano de 2012. Fonte: <http://faostat3.fao.org>

Entre 5 e 6 milhões de fazendas de cacau estão distribuídas entre os quatro continentes produtores – África, América, Ásia e Oceania (<http://worldcocoafoundation.org>). Diferentes países destes continentes abrigam centenas de milhares de pessoas que dependem direta ou indiretamente da cadeia produtiva do cacau. Neste sentido, entre 80 e 90% da produção de cacau ocorre em pequenas propriedades, com até 3 hectares, geralmente sob um sistema de agricultura familiar (Lopes et al. 2011). Embora em campo experimental já tenha sido alcançada a produção de até 3.000 quilogramas de grãos por hectare (PANG THAU YIN, 2004),

fatores como componente genético das plantas de cacau utilizadas e a região de plantio influenciam diretamente no rendimento, levando a uma média nacional na ordem de 375 Kg/ha (IBGE, 2012) de grãos de cacau. Além disso, limitações na produção têm sido ocasionadas por variações climáticas e diferentes doenças, a exemplo da vassoura-de-bruxa e a Moníliase nas Américas (causadas pelos fungos *Moniliophthora perniciosa* e *M. roreri*, respectivamente), bem como da podridão-parda (causada por um complexo de espécies do gênero *Phytophthora*), que embora pandêmica, tem suas mais graves consequências em países africanos (GUILTINAN et al., 2008). Como os pequenos produtores são a maioria, a falta de orientação para o manejo adequado e baixo investimento nas lavouras são fatores que reduzem a otimização da produção.

No que tange ao processamento e consumo de cacau, países da Europa e América do Norte se destacam. Entretanto, a expansão do mercado consumidor em países emergentes como Brasil, China e Índia, tem ampliado a demanda na produção, que em conjunto com os consumidores tradicionais, supera 3 milhões de toneladas de grãos de cacau por ano (<http://worldcocoafoundation.org>). A despeito das variações nas preferências de consumo em cada país, um público exigente em termos de qualidade tem expandido a demanda por chocolate ‘gourmet’ e conseqüentemente, do cacau fino. Além da variedade produzida, a referência enquanto chocolate ‘gourmet’ está associado ao manejo sustentável do cacau, no que preze a ausência de agrotóxicos e a seleção cuidadosa dos frutos e grãos, bem como o método de processamento destes (SALTINI et al., 2013; SANTOS et al., 2012). Como na maioria das lavouras ampliar a produtividade tem se tornado uma tarefa difícil devido aos fatores bióticos e abióticos já citados, a produção do cacau fino tem se tornado uma alternativa para aumentar o lucro em até três vezes em comparação com o cacau comumente comercializado (SANTOS et al. 2012). Desta forma, algumas variedades de cacau já são reconhecidas internacionalmente pelo padrão ‘gourmet’, a exemplo da variedade Nacional que representa cerca de 30% da produção do Equador. Por outro lado, algumas variedades plantadas na Bahia vêm recentemente conquistando espaço no mercado externo de chocolate ‘gourmet’, bem como o reconhecimento de suas qualidades, com premiações em eventos especializados de escala nacional e internacional como o *Salon du Chocolat* ocorrido em 2011 na França (<http://www.cocoaofexcellence.org>).

## Origem, diversidade e domesticação do cacauero

Embora atualmente o cacau seja cultivado em diferentes países ao redor do globo, a sua origem, assim como é observado para outras espécies do gênero *Theobroma*, está associada à região Amazônica, onde também é encontrado ser o centro de diversidade (CHEESMAN, 1944). Discussões acerca da origem do cacauero são atuais e inconclusivas, ou seja, as informações obtidas até o momento não permitem apontar o local exato no qual a espécie tenha se originado (BARTLEY 2005; THOMAS et al., 2012). O centro de diversidade do cacauero é comumente referido como região Alto Amazônica. Embora este termo originalmente designe uma região restrita por onde correm dois rios peruanos (Ucayali e Marañon) até a fronteira com o Brasil, a região Alto Amazônica referida para o cacauero ultrapassa estes limites, não sendo portanto, claramente estabelecida (BARTLEY, 2005). Contudo, cacaueros designados como Alto Amazônicos apresentam diversidade bastante elevada, tanto do ponto de vista morfológico quanto molecular (CHEESMAN, 1944; MOTAMAYOR et al., 2008).

Os cacaueros Alto Amazônicos apresentam um padrão discreto de distribuição com populações variando geneticamente entre si (BARTLEY, 2005). Isto significa que não são observadas plantas de cacau dispostas continuamente e sim, grupos de plantas que formam ‘manchas’ ao longo da Amazônia. Este fato pode ser reflexo do último período glacial (cerca de 21.000 anos atrás) e das ações humanas nos períodos subsequentes de aquecimento (THOMAS et al., 2012). Neste sentido, durante os períodos mais críticos de congelamento e disponibilidade de água, plantas de cacauero podem ter sido restringidas a refúgios onde as condições climáticas eram melhores (especialmente na base dos Andes), ocasionando uma restrição do fluxo gênico e consequentemente, a diferenciação entre as populações de cacau isoladas. Posteriormente, nos períodos de aquecimento, quando houve aumento na precipitação em regiões secas, houve expansão dos habitats disponíveis para cacau em torno dos refúgios contribuindo para a geração do padrão de ‘manchas’ observado atualmente nas áreas naturais de ocorrência do cacauero (THOMAS et al., 2012).

Aliado ao fenômeno geológico supracitado, questões relacionadas à reprodução da espécie parecem ter servido de reforço para diferenciação entre as populações de cacaueiros. Neste sentido, o sistema de incompatibilidade e a curta distância percorrida por polinizadores, bem como os empecilhos na dispersão das sementes ocasionados por características de frutos (indeiscentes e fixos na árvore mesmo quando maduros, casca espessa de difícil de rompimento) e sementes (recalcitrantes e com germinação podendo ser iniciada dentro do próprio fruto) (LACHENAUD e ZHANG 2008), contribuíram sobremaneira na redução do fluxo gênico e aumento na estrutura genética entre as populações.

Se por um lado a distribuição primária das populações naturais foi influenciada por questões ambientais e biológicas, o padrão atual de distribuição do cacaueiro é creditado às ações antrópicas. Os mais antigos dados arqueológicos encontrados sobre o uso do cacau fornecem indícios da domesticação desta planta por povos Maias, acerca de 3.000 anos atrás no sul do México e América Central (HURST et al., 2002, GÓMEZ-POMPA et al., 1990). Embora se acredite que estes povos tenham sido atraídos originalmente pela polpa nutritiva do cacau sendo esta também utilizada para fabricação de bebidas (HENDERSON et al., 2007), o uso dos grãos na fabricação de uma bebida amarga em cerimônias religiosas foi sem dúvida, o principal emprego do cacau feito pelos povos Maias e Astecas. Dados consistentes sobre o uso do cacau pelos índios sul-americanos não são tão antigos quanto o observado para os Maias, mas ainda assim não é negligenciado o fato de que estes também devam ter atuado no processo de domesticação do cacau na floresta Amazônica (CLEMENT et al., 2010). Talvez a principal diferença na domesticação na região mais central e sul das Américas se dê pelo interesse que se tinha no cultivo, sendo os povos Maias interessados nos grãos enquanto os índios da Amazônia estavam atraídos pela mucilagem que as envolvem.

O processo de domesticação do cacau em regiões distintas levou a formação de grupos cultivados com clara distinção morfológica, o que influenciou diretamente na sua classificação. A priori cacaueiros cultivados pelos Maias e pelos povos da América do Sul foram caracterizados como subespécies distintas, sendo denominados popularmente como Crioulo e Forasteiro e cientificamente como *T.cacao* subesp. *cacao*

e *T. cacao* subesp. *sphaerocarpum*, respectivamente (CUETRACASAS, 1964). Sabe-se, entretanto, que a despeito da marcante variação encontrada, estas variantes não constituem subespécies e são apenas consideradas como grupos morfogeográficos distintos (CHEESMAN, 1944; BARTLEY, 2005). Neste sentido, o morfotipo Crioulo compõe um grupo uniforme, caracterizado por apresentar frutos vermelhos quando imaturos e amarelos quando maduros. Embora de baixa produção e pouca resistência a doenças, seus grãos estão relacionados com a produção de chocolate com sabor e flavor agradáveis, isto é, menos amargo.

Em contraste ao Crioulo, o morfotipo Forasteiro possui ampla variedade de populações, sendo estas definidas de acordo com a região geográfica onde são observadas, a exemplo dos tipos Baixo Amazonas e Parinari (DIAS, 2001). De acordo com o grau de maturidade em que seus frutos se encontram, cacauzeiros Forasteiro variam entre verde e amarelo ouro, quando imaturos e maduros, respectivamente. Os grãos geralmente são roxos para o cacau Forasteiro e suas árvores são notáveis pela produtividade e resistência a doenças.

Os cacauzeiros do tipo Forasteiro possuem diferentes habitats na região amazônica e por este motivo são classificados em dois subgrupos: Forasteiro Alto Amazônico e Forasteiro Baixo Amazônico. Os Alto Amazônicos concentram a maior parte da variabilidade do grupo Forasteiro (SERENO et al., 2006; ZHANG et al. 2009a). Já os Baixo Amazônicos englobam plantas silvestres das Guianas e do leste da Amazônia (SERENO et al., 2006) que teriam sido primariamente cultivadas no Pará, Brasil. O tipo Amelonado é um Baixo Amazônico caracterizado pelo formato de melão dos seus frutos, sendo que esta constitui a variedade mais cultivada no mundo (AIKPOKPODION et al. 2009; MOTAMAYOR et al., 2003). É interessante notar que assim como discutido acima para a região Alto Amazônica, a zona de abrangência do Baixo Amazonas referida para o cacau também não coincide exatamente com a região definida originalmente para este termo. Segundo Bartley (2005), a região verdadeiramente caracterizada como Baixo Amazonas é mais restrita que aquelas usualmente referidas para o cacauzeiro, compreendendo parte do Pará que vai desde a foz do rio Maicuru até a Ilha de Marajó.

Embora alguns estudos indiquem as Guianas e o norte do Brasil como os locais de origem silvestre e cultivada dos cacauzeiros do Baixo Amazonas, estas informações não são conclusivas, sendo a real região de origem ainda não estabelecida claramente. Sereno et al. (2006) a partir de estudos realizados com cacauzeiros da Amazônia brasileira, levantaram a hipótese de que sementes de frutos coletados na região Alto Amazônica teriam sido utilizadas como fundadoras de cultivos por povos mais primitivos no Pará, gerando assim o grupo denominado Baixo Amazônico. Este fato explicaria a baixa variabilidade genética observada no cacauzeiros Amelonado em uma gama de estudos. Estudos posteriores indicaram que ao menos parte dos Amelonados e de outro grupo classificado previamente como Alto Amazônico e atualmente como Marañon (MOTAMAYOR et al., 2008), parecem de fato terem sido originados no centro de diversidade genética do cacau (THOMAS et al., 2012).

Adicionalmente aos grupos Crioulo e Forasteiro supracitados, o grupo Trinitário surgiu em Trinidad a partir da hibridações naturais entre plantas de cacau Forasteiro (Alto e Baixo Amazonas) e Crioulo (YANG et al., 2013; MOTILAL et al., 2009). Os trinitários são conhecidos por sua alta variabilidade genética, potencializando as boas características herdadas dos seus genitores, o que fez deste grupo uma excelente opção para os plantios comerciais de cacau. Além dos três grupos principais citados, outras variedades cultivadas como Nacional e Refratário do Equador tem sido alvo de estudos, dada as características de flavor e resistência a doenças como a vassoura-de-bruxa e a monilíase do cacauzeiro causadas pelos fungos *Moniliophthora perniciosa* e *M. roreri*, respectivamente (SOLORZANO et al., 2012; LOOR et al., 2009; ZHANG et al., 2008; IWARO et al., 2003).

A classificação do cacauzeiro nos principais grupos supracitados tornou-se inconsistente com a diversidade observada, especialmente nos últimos anos, quando os estudos moleculares foram confirmando a ampla variação presente entre as diferentes populações. Desta forma, um estudo fazendo uso de marcadores microssatélites associados à análise Bayesiana de estrutura genética (PRITCHARD et al., 2000) e empreendido por MOTAMAYOR et al. (2008) indicou haver ao menos dez grupos genéticos estruturando a diversidade do cacau, sendo estes denominados como Marañon, Curaray, Criollo, Iquitos, Nanay, Contanama, Amelonado, Purús, Nacional e



Guiana. As denominações dadas para cada grupo refletem a região de origem dos clones ou a variedade mais representativa dentro de cada grupo. Sem dúvida esta classificação resolveu incongruências geográficas e populacionais quando da denominação de grupos que melhor representem as variações consideradas em conjunto como Alto Amazonas, enquanto reafirmou algumas antigas classificações, caso das variedades tradicionais Crioulo e Amelonado. Por outro lado, alguns grupos de clones no qual havia ambiguidade na classificação, como era o caso daqueles pertencentes à variedade Nacional do Equador (N'GORAN et al., 1994) foram confirmados como um grupo genético específico.

De qualquer forma, os grupos oriundos da subdivisão do Alto Amazonas, formados por maioria de indivíduos silvestres ou primitivos, apresentam maior variabilidade que aqueles grupos compostos por cultivares tradicionais (Crioulo, Amelonado e Nacional) (MOTAMAYOR et al., 2008). Este fato reflete o processo de domesticação a partir de uma estreita base genética (CLEMENT et al., 2010) dos quais descendem clones com baixa diversidade genética e altos níveis de homozigose (MOTILAL et al., 2013; MOTAMAYOR et al., 2008; LERCETEAU et al., 1997). Por melhor refletir a variação do cacaueiro, a classificação em dez grupos proporcionou um panorama mais abrangente para a conservação dos recursos genéticos da espécie, além de auxiliar em programas de melhoramento, por viabilizar que novos cruzamentos sejam implementados em busca de novas combinações heteróticas (MOTAMAYOR et al., 2008).

### **Conservação da diversidade genética do cacau**

A diversidade genética do cacaueiro pode ser mantida das seguintes formas: (i) *in situ* em áreas naturais preservadas, (ii) *ex situ* em bancos de germoplasma ou na manutenção de tecidos *in vitro* usando técnicas de cultura de tecidos e criopreservação e, (iii) *on farm* em plantios em fazendas e pomares comerciais (CacaoNet, 2012). A conservação de regiões onde as plantas de cacau ocorrem naturalmente é bastante importante, uma vez que a maior parte da diversidade do cacau ocorre nestes

ambientes naturais e parte considerável desta diversidade ainda é desconhecida. Neste sentido, a conservação *in situ* apresenta a vantagem de permitir que os processos evolutivos continuem ocorrendo de forma dinâmica sobre as plantas. Entretanto, a erosão genética causada pelas mudanças climáticas, catástrofes naturais, deflorestação e ações antrópicas diversas tem sido um problema na conservação da espécie, especialmente nas regiões onde é encontrada a maior diversidade genética do cacaueteiro. Em fazendas, a manutenção das variedades tradicionais e criação de novas por meio dos processos relacionados ao cultivo como a seleção, também tem sido considerada importante como meio para salvaguardar a diversidade em cacaueteiros (AIKPOKPODION et al., 2009).

Devido aos problemas supracitados de erosão genética em ambientes naturais a manutenção *ex situ* do cacaueteiro tem sido importante forma de amostrar e proteger a diversidade. Neste sentido, a principal forma de manutenção das plantas tem sido por meio do seu plantio em Bancos Ativos de Germoplasma (BAGs), isto porque devido à natureza recalcitrante das sementes é inviável a geração de bancos destas, sendo necessário o estabelecimento e a manutenção desse recurso genético em campo. Por estarem expostas às mesmas intempéries que as plantas de ambientes naturais, há a necessidade de um amplo cuidado para que materiais do banco não sejam perdidos. Este fato aliado à demanda de espaço para manutenção e duplicação do material, onera a manutenção. Entretanto, os BAGs têm se tornado indispensáveis para conservar plantas que amostram tanto a diversidade silvestre da espécie quanto a diversidade presente em variantes selecionadas em fazendas e produzidas por novas combinações (material elite) através de programas de melhoramento genético (MONTEIRO et al., 2009). Além de disponibilizar material biológico mais acessível para pesquisadores realizarem os mais diversos estudos (incluindo evolutivos e de diversidade), os BAGs estabelecem uma ligação entre a diversidade genética e seu uso sendo os que materiais ficam disponíveis para a realização de novos programas de melhoramento (MOTAMAYOR et al., 2008).

Como auxílio para a preservação *ex situ* da diversidade, a manutenção *in vitro* de tecidos tanto em meio de cultura quanto por criopreservação tem sido aprimorado para o cacaueteiro, sendo a cultura de tecidos útil para duplicação segura das plantas de campo e

rápida multiplicação de plantas livres de doenças. A criopreservação tem-se mostrado uma útil forma de manter embriões e uma barreira contra a disseminação de doenças virais do cacau como o Cocoa Swollen Shoot Virus (CACAONET, 2012).

Devido aos fatores já citados e, como uma forma de permitir a cooperação internacional entre as instituições de pesquisa, cerca de 40 coleções de germoplasma de cacau são mantidos ao redor do globo. As primeiras coleções de germoplasma para a espécie foram estabelecidas no início do século XX, a partir da necessidade de amostrar a variabilidade genética por meio de seleções em materiais locais (BARTLEY, 2005). O estabelecimento destas plantas nas coleções e suas informações obtidas durante este processo, passaram a ser reconhecidos como essenciais para o sucesso dos programas de melhoramento genético (BARTLEY, 2005). Expedições de coleta em ambientes naturais só ocorreram a partir do final da década de 1930, quando J. F. Pound coletou na região Alto Amazônica alguns exemplares de plantas potencialmente resistentes à vassoura-de-bruxa (ZHANG et al., 2009b). A partir de então, outras coletas foram realizadas, mas ainda assim estima-se que apenas 20% da variabilidade genética naturalmente exibida em ambientes naturais esteja representadas em bancos de germoplasma (BARTLEY, 2005), sendo necessárias novas expedições de coletas a fim de preencher as lacunas na diversidade amostrada.

A Tabela 1 apresenta informações acerca da distribuição dos principais bancos de germoplasma de cacau do mundo, com ênfase nos mais representativos quanto ao número de acessos mantidos. Parte da diversidade do cacau mantida nos BAGs está amostrada em dois bancos de germoplasma de cacau internacionais mantidos pelo Cocoa Research Unit (CRU) e pelo Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), localizados respectivamente em Trinidad e na Costa Rica. O banco de germoplasma do CRU é conhecido como ICG, T (International Cocoa Genebank, Trinidad) e o do CATIE como IC3 (International Cacao Collection at CATIE). Os diversos bancos de cacau existentes tanto dentro de regiões conhecidas como centros de diversidade da espécie quanto em outros países produtores mantêm mais de 24 mil acessos, sendo a maioria destes genótipos selecionados localmente e clones introduzidos a partir das duas coleções internacionais (CacaoNet, 2012). Algumas das coleções de grande representatividade da diversidade do cacau, estão localizadas nas

Américas (LOPES et al., 2011) (Tabela 1). No Brasil, duas coleções de germoplasma de cacau merecem destaque, sendo ambas mantidas pela Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC). Uma das coleções está localizada no estado do Pará, na Superintendência de Desenvolvimento da Região Cacaueira no Estado do Pará (SUEPA) e possui 90% dos 20 mil genótipos que o compõe (representando cerca de 2.200 acessos) formado por material silvestre da Amazônia brasileira. Por sua vez, a coleção do Centro de Pesquisas do Cacau (CEPEC) da Bahia (1.302 acessos) é formada especialmente por material internacional melhorado (LOPES et al., 2011; MONTEIRO et al., 2009).

A despeito do número de acessos mantidos nos BAGs, a obtenção de novos genótipos oriundos de regiões específicas e, ou, frutos de programas regionais de melhoramento genético deve contribuir para aumentar a representatividade da diversidade nestes locais. A caracterização deste material certamente é um caminho a ser trilhado tanto para conhecer a variabilidade existente quanto para fazer uso consciente desta em programas de melhoramento genético. Estas caracterizações incluem desde avaliações moleculares até avaliações morfológicas e agronômicas.

Um dos principais problemas que inviabiliza o sucesso no uso de plantas presentes em bancos de germoplasma, está nos erros de identificação das plantas em campo. Estes erros surgem por vezes durante a própria coleta do material ou ainda, durante as etapas do processo de transferência das plantas para o campo, o que inclui erros na marcação da posição das plantas no mapa ou na etiqueta de identificação das árvores (MOTILAL et al., 2012; IRISH et al., 2010). Marcadores moleculares vêm desempenhando papel fundamental nas atividades de identificação e classificação de genótipos e consequentemente, têm se tornado aliados na detecção de erros na identidade de plantas presentes em bancos de germoplasma (MOTILAL et al. 2012; ZHANG et al. 2009). De fato, é estimado que 40% do material depositado em coleções de germoplasmas de cacau em todo mundo apresentem ao menos um dos três tipos de erros seguintes: (i) homônimos - genótipos diferentes compartilhando uma mesma identificação; (ii) sinônimos - genótipos iguais com identificações diferentes no germoplasma, e; (iii) plantas que possuem um genótipo único na coleção de germoplasma, mas quando comparado com os padrões internacionais para aquela

identificação (conhecidos como genótipos referência), apresentam variações (IRISH et al., 2010).

Outra aplicação dos marcadores moleculares é na identificação de material redundante distribuído no germoplasma o que onera ainda mais a manutenção do banco. Neste sentido, alguns curadores estão identificando as plantas mais próximas entre si por meio de marcadores moleculares, sendo necessário posteriormente avaliações morfológicas destas plantas para averiguar se apresentam também semelhanças fenotípicas, e só então decidir em descartá-las ou não definitivamente da coleção (IRISH et al., 2010).

**Tabela 1.** Principais coleções (bancos de germoplasma) responsáveis pela manutenção *ex situ* da diversidade genética do cacau *Theobroma cacao*.

| Continente/Região | Nº de coleções | Nº total de acessos | Principais Coleções                                                                              |               |                   |
|-------------------|----------------|---------------------|--------------------------------------------------------------------------------------------------|---------------|-------------------|
|                   |                |                     | Instituição                                                                                      | Nº de acessos | País              |
| África            | 6              | 4303                | Centre national de recherche agronomique (CNRA)                                                  | 1.605         | Costa do Marfim   |
|                   |                |                     | Cocoa Research Institute of Ghana (CRIG)                                                         | 1.366         | Gana              |
|                   |                |                     | Cocoa Research Institute of Nigeria (CRIN)                                                       | 1.1           | Nigéria           |
| América Central   | 5              | 3870                | Centro Agronómico Tropical de Investigación y Enseñanza (CATIE)                                  | 1.146         | Costa Rica        |
|                   |                |                     | Cocoa Research Unit of the University of the West Indies (CRU/UWI)                               | 2.4           | Trinidad e Tobago |
| América do Norte  | 1              | 200                 | United States Department of Agriculture (USDA)                                                   | 200           | Estados Unidos    |
| América do Sul    | 13             | 10362               | Centro de Pesquisas do Cacau / Comissão Executiva do Plano da Lavoura Cacaueira (CEPEC / CEPLAC) | 1.302         | Brasil            |
|                   |                |                     | Superintendência de Desenvolvimento Região Cacaueira no Estado do Pará (SUEPA / CEPLAC)          | 2.504         | Brasil            |
|                   |                |                     | Instituto Nacional Autónomo de Investigaciones Agropecuarias (INIAP)                             | 2.332         | Equador           |
| Ásia              | 5              | 3607                | Malaysian Cocoa Board (MCB)                                                                      | 2.263         | Malásia           |
| Europa            | 2              | 533                 | International Cocoa Quarantine Centre, University of Reading (ICQC,R)                            | 395           | Reino Unido       |
| Oceania           | 4              | 1495                | Cocoa and Coconut Institute (CCI)                                                                | 1.2           | Papua Nova Guiné  |
| <b>Total</b>      | <b>36</b>      | <b>24370</b>        | <b>-</b>                                                                                         | <b>17.813</b> | <b>-</b>          |

## **Caracterizações Genéticas e o Melhoramento Genético do Cacaueiro**

Como premissa básica para o sucesso de um programa de melhoramento está a necessidade de uma representativa base genética disponível para o melhorista, sendo de suma importância, não apenas a existência dos bancos de germoplasma, mas também que as plantas que o compõem tenham seguras informações acerca das características biológicas, agrônomicas e genéticas (LOPES et al., 2011). Estudos de variabilidade genética têm sido realizados para cacaueiros sendo estes desenvolvidos no âmbito dos bancos de germoplasma a partir de acessos de fazendas e silvestres. Enquanto estudos de variabilidade genética obtidos diretamente a partir de genes de interesse ainda são incipientes para o cacaueiro, o uso de marcadores moleculares como estimativa indireta da diversidade dos genes vem sendo uma ferramenta bastante utilizada (MONTEIRO et al., 2009). Desta forma, marcadores moleculares têm sido empregados intensamente desde meados da década de 1990, com intuito de conhecer de forma geral a identidade das plantas e suas relações genéticas.

A aplicação dos marcadores moleculares em estudos com cacaueiros além de contribuírem na eliminação de erros de identificação em germoplasma (MOTILAL et al. 2012; ZHANG et al. 2009; FALEIRO et al., 2002), vêm permitindo identificar plantas, promover estudos de taxa de cruzamento (SCHawe et al., 2013; SILVA et al., 2011), caracterizar a diversidade genética (JI et al., 2013; YAMADA et al., 2009; BOZA et al., 2009; SERENO et al., 2006; ZHANG et al., 2006; FALEIRO et al., 2004; MARITA et al., 2001; LERCETEAU, 1997), classificar genótipos (LIMA et al., 2013; TROGNITZ et al., 2013); identificar estruturas genéticas (TROGNITZ et al., 2011; MOTAMAYOR et al., 2008), rastrear a origem das variedades cultivadas (MOTAMAYOR et al., 2002; 2003), construir mapas de ligação e mapear genes de interesse (SANTOS et al., 2013; KUHN et al., 2012; LIMA et al., 2010; ARAÚJO et al., 2009; STEVEN BROWN et al., 2007; FALEIRO et al., 2006; LANAUD et al., 2004; 2009). Mais recentemente tem sido mostrado o potencial uso dos marcadores SNP (Single Nucleotide Polymorphism) para rastrear em grãos de cacaueiros de variedades comercializadas como cacau fino, a presença de outras variedades, consideradas neste caso, como contaminantes (FANG et al., 2014).

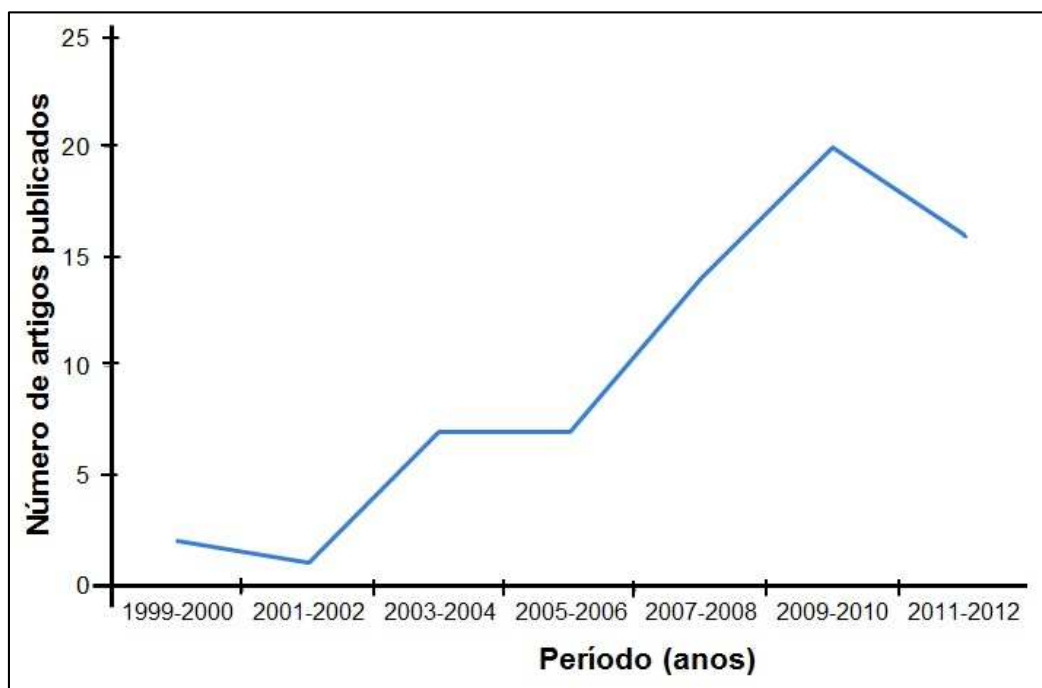
Dentre os marcadores moleculares amplamente utilizados nas mais diversas análises genéticas em cacaueiros estão os microssatélites (Figura 3). Marcadores microssatélites ou SSR (*Simple Sequence Repeat*) são caracterizados como repetições de 2 a 6 nucleotídeos lado a lado no genoma. Estes marcadores são altamente distribuídos no genoma; eles possuem caráter codominante, isto é, permitem identificar os diferentes alelos em um heterozigoto. Além disso, o avanço nas técnicas de desenvolvimento destes marcadores e de genotipagem permitiram que os dados fossem obtidos de forma rápida e relativamente barata, o que justifica seu crescente uso nos últimos anos.

Os marcadores microssatélites são utilizados em cacaueiros desde o final do século XX, quando o primeiro conjunto de primers foi disponibilizado para a espécie (LANAUD et al., 1999). Posteriormente, novos marcadores microssatélites foram identificados para cacau (SANTOS et al., 2012; LIMA et al., 2008; ARAÚJO et al., 2007; PUGH et al., 2004) tendo grande contribuição na elucidação de importantes questões genéticas, a exemplo da classificação das variantes de cacau em 10 grupos genéticos (MOTAMAYOR et al., 2008), bem como a confirmação da região Alto Amazônica como berço da diversidade da espécie (ZHANG et al., 2009b; SERENO et al., 2006). Mais recentemente, com o sequenciamento do genoma do cacaueiro (ARGOUT et al., 2011) foi possível a identificação de grande quantidade de marcadores microssatélites (409) e SNPs (1.536) *in silico*, sendo tais marcadores já utilizados para construir um mapa genético de alta densidade para a espécie (ALLEGRE et al., 2012). Estes estudos possuem impacto direto no melhoramento do cacaueiro, por permitir através do conhecimento sobre a estrutura da diversidade e da identificação de regiões de interesse no genoma, escolher indivíduos prioritários para cruzamentos visando ganhos genéticos (ALLEGRE et al., 2012; MOTAMAYOR et al., 2008).

Aliado aos estudos moleculares, as caracterizações fenotípicas também devem ser encaradas como fundamentais para conhecer a variação existente e torna-la útil no melhoramento genético. Para orientar nas avaliações, descritores botânicos e agrônômicos para cacaueiros são indicados e empregados em estudos desde a década de 1970 (AIKPOKPODION, 2010; BEKELE et al., 2006; IWARO et al., 2003; ENGELS et al., 1980). Programas de melhoramento vêm sendo desenvolvidos para cacau e têm



o foco especialmente em características relacionadas à produção e resistência a doenças e pragas. O rendimento em cacau é mensurada pelo peso de grãos secos por planta ou área, sendo esta medida denominada por índice de fruto; e seus componentes incluem o número de frutos por árvore ou área, o número de grãos por fruto e o peso de um grão (LOPES et al., 2011). Variedades de cacauzeiros que produzam a partir de 1,5 kg de grãos secos por árvore por ano são consideradas com boa produtividade. Mais recentemente, a qualidade dos grãos tem sido foco no melhoramento, de modo que as avaliações físicas e químicas destes têm ganhado destaque. A característica física do grão mais importante no melhoramento é o peso, sendo requerido pela indústria o peso mínimo de 1 grama por grão seco. A qualidade inclui ainda o conteúdo e dureza dos ácidos graxos, sendo candidatos à seleção as plantas com alta quantidade de ácidos graxos e maior dureza da manteiga do cacau (PIRES, 1998; 2003) por serem características que interferem positivamente no sabor e odor do chocolate.



**Figura 3.** Número de artigos publicados com uso de microsatélites em cacauzeiro desde 1999, ano em que foi disponibilizado o primeiro conjunto de primers microsatélites para a espécie.

As doenças causam prejuízos expressivos na cacauicultura. Estima-se que cerca de 30% da produção de cacau seja perdida anualmente por ocorrência de fungos, vírus e insetos. Neste sentido, perdas históricas na produção de cacaueiros se configuram em casos de grande importância para a agricultura, a exemplo do efeito da vassoura-de-bruxa na Bahia e da podridão-parda em países da África, sendo relatadas perdas de até 90% na produção (BOWERS et al., 2001).

Como forma de controle das doenças causadas por pragas e patógenos, o melhoramento genético é vislumbrado como a alternativa mais eficaz, barata e sustentável, sendo alvo generalizado dos programas desde meados do século XX (MONTEIRO et al., 2009). Características como compatibilidade sexual e arquitetura das plantas têm sido alvos recentes dos programas de melhoramento em prol de plantas autocompatíveis (já que plantas autoincompatíveis tendem a ser mais vigorosas e menos produtivas) e de mais fácil manejo (LOPES et al., 2011; MONTEIRO et al., 2009). Para o futuro, características relacionadas ao conteúdo de polifenóis e outras substâncias benéficas para a saúde devem estar entre os objetivos dos programas de melhoramento.

A avaliação de bancos de germoplasma tem reforçado a importância da manutenção destes locais enquanto refúgio útil da variabilidade genética, sendo neles encontrado materiais de origens distintas, os quais são importantes fontes de genes. Desta forma, plantas das expedições realizadas por Pound à região Alto Amazônica compõem algumas das séries mais utilizadas no melhoramento, como o IMC (Iquitos mixed Cabacillos), Scavinas (SCA), Parinari, Morona e Nanay (JOHNSON et al., 2009), contribuindo com genes para resistência e características relacionadas a produção e qualidade do grão. Plantas obtidas em fazendas a exemplo daquelas das séries Imperial College Selection (ICS) e United Fruit Selection (UF) coletadas em produções comerciais de Trinidad e Costa Rica, respectivamente, também foram e continuam sendo utilizados em programas de melhoramento devido principalmente às boas características de produção. Outros exemplos são os clones selecionados em plantações da Bahia pelo Instituto do Cacau (série SIC) e pelo Instituto Agrônomo do Leste (SIAL) sendo caracterizados como bastante produtivos e por isso empregados na produção de híbridos e como genitores em programas de seleção recorrente (VELLO,

GARCIA e MAGALHÃES, 1972). A seleção de material elite previamente obtido a partir de programas de melhoramento também têm sido uma alternativa de grande valor para emprego em novos ciclos de cruzamento e seleção em busca de plantas com bom desempenho. As séries TSH e TSA (*Trinidad Selection Hybrids* e *Trinidad Selection Amazon*, respectivamente) foram selecionadas em Trinidad a partir dos cruzamentos de plantas alto amazônicas entre si, como alguns SCA e IMC, ou entre estas e trinitários, a exemplo de clones da série ICS, e são amplamente utilizados em programas de melhoramento (MONTEIRO et al. 2009).

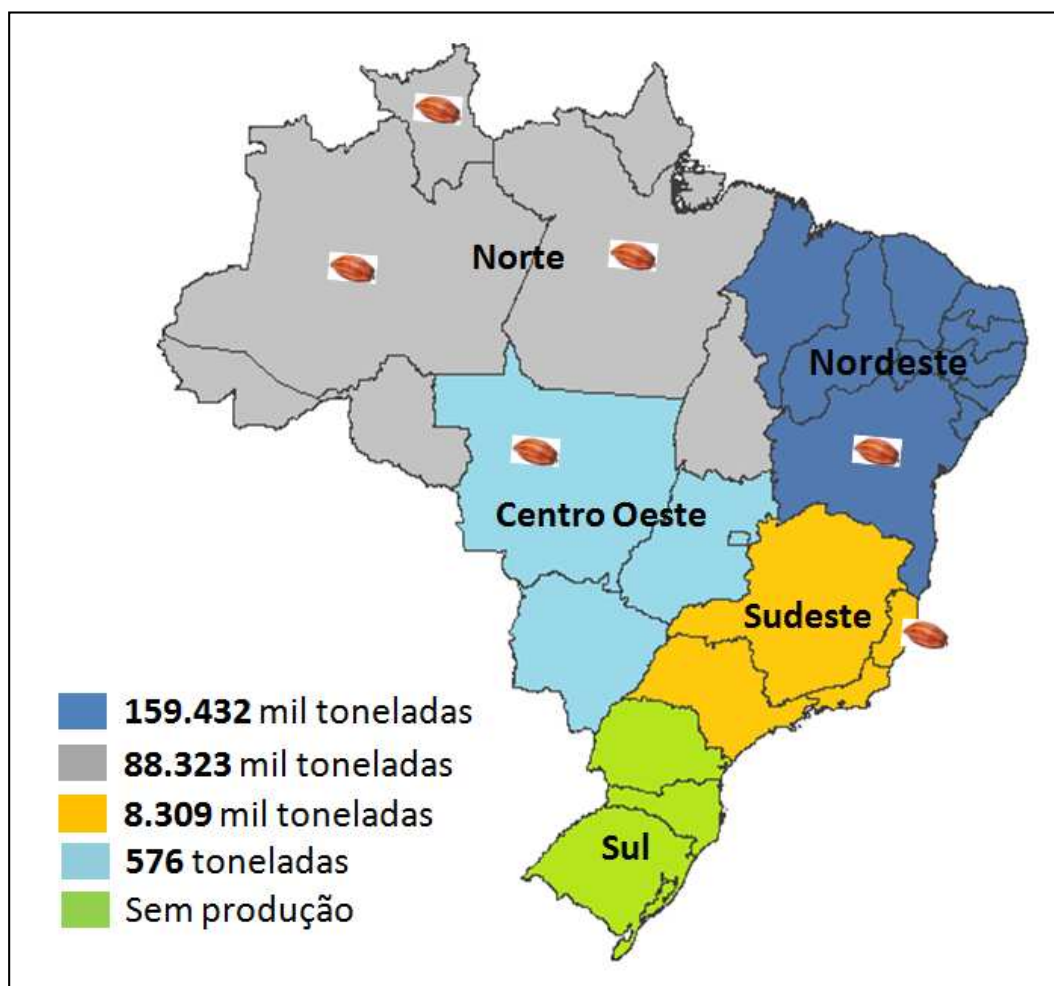
É interessante notar que por muito tempo os programas de melhoramento do cacauero foram baseados exclusivamente em apenas um ciclo de seleção. Neste contexto, os acessos clonais presentes em bancos de germoplasma eram cruzados entre si e os híbridos ou famílias eram caracterizados e selecionados e já recomendados para plantio. Além disso, os clones selecionados em fazendas ou dentro progênes, com características de interesse são também indicados para plantio comercial (Monteiro et al. 2009). Programas participativos de seleção de plantas com características de produtividade e resistência têm sido estratégicos na identificação de plantas promissoras em fazendas (CACAONET, 2012). Neste tipo de programa, os fazendeiros e extensionistas recebem treinamento dos pesquisadores e atuam diretamente na seleção das plantas que posteriormente são submetidas a testes para confirmar e caracterizar seu potencial agrônomo. Programas de caráter participativo já foram empregados na Bahia (LOPES et al. 2011) e estão sendo utilizados em países do oeste da África, tendo bons resultados principalmente no que diz respeito à seleção de genótipos resistentes a doenças, já que é justamente nas fazendas que tende a existir grande severidade das enfermidades, o que aumenta a chance do sucesso na seleção para tais características.

A seleção de clones e híbridos foi ou ainda é bastante utilizada em programas de melhoramento do Brasil, Equador, Trinidad, Papua Nova Guiné, Malásia e Costa do Marfim (CacaoNet, 2012). Entretanto, mais recentemente, a estratégia de seleção recorrente está ganhando espaço em programas de melhoramento para o cacauero, a fim de ampliar os ganhos genéticos com o avanço nos ciclos de cruzamento e prover informações sobre a capacidade de combinação dos parentais utilizados. Esta

estratégia vem sendo empregada no Brasil, Malásia e Costa do Marfim, a partir das características de produtividade, qualidade do grão e resistência a doenças (LOPES et al., 2011).

### **A cultura do cacau na Bahia**

O Brasil é o maior produtor de cacau das Américas, sendo as regiões Nordeste e Norte responsáveis por aproximadamente 62 e 34% da produção brasileira, respectivamente (Figura 4) (IBGE, 2012). Toda a produção registrada para a região Nordeste vem do Estado da Bahia e, portanto, este é o maior produtor de cacau do Brasil. São contabilizados mais de 540 mil hectares plantados com cacaueiros na Bahia (IBGE, 2012) em 29 mil fazendas de 107 municípios, sendo a maior parte destas na região sul e sudeste do Estado (<http://sim.sei.ba.gov.br>). A produção baiana em grãos de cacau em 2012 alcançou 159 mil toneladas, com rendimento médio de 300 Kg/ha (IBGE, 2012). Parte das plantações de cacau na Bahia (70%) ocorre num sistema agroflorestal denominado Cabruca (ARAÚJO, 1998), sendo caracterizado pelo plantio do cacaueiro no sub-bosque de plantas nativas da Mata Atlântica. Embora tenha sido retirada parte das plantas do bioma Mata Atlântica para o plantio de cacaueiros, as Cabrucas se mostraram eficientes devido ao menor impacto causado ao meio ambiente do que seria esperado caso outras culturas com plantio a pleno sol fossem empregadas. Neste sentido, as Cabrucas viabilizaram a manutenção de um refúgio para plantas e animais silvestres (FARIA et al., 2007; SAMBUICHI, 2006).



**Figura 4.** Macrorregiões brasileiras e valores correspondentes em toneladas de grãos de cacau produzidas em cada uma. As imagens dos frutos de cacau representam os estados responsáveis pela produção. Mapa obtido com SpeciesMapper tool (<http://splink.cria.org.br/>) e dados de produção disponíveis em: [http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam\\_2012\\_v39\\_br.pdf](http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam_2012_v39_br.pdf)

A história do cultivo do cacau na Bahia teve início em meados do século XVIII quando as primeiras sementes foram introduzidas na região. Desde então a cacaucultura baiana experimentou o ápice da sua produção, sendo a principal responsável pelo Brasil figurar entre os maiores exportadores de cacau do mundo (<http://faostat3.fao.org>), e o declínio no final da década de 1980, por ocorrência da vassoura-de-bruxa, que entrou para a história dos grandes desastres de culturas de importância econômica mundial.

As primeiras sementes de cacau foram levadas para a Bahia em 1746 pelo francês Luiz Frederico Warneaux, com o intuito de quebrar a hegemonia da Espanha na produção do cacau (YOUNG, 2007). Nesta ocasião, o fazendeiro Antônio Dias Ribeiro as plantou na fazenda Cubículo, que ainda resiste ao tempo no município de Canavieiras, litoral sul da Bahia. As 'sementes fundadoras' foram trazidas do estado do Pará, região norte do Brasil e foram oriundas da variedade 'Comum', uma designação para diferenciar as plantas originadas da primeira introdução, daquelas que ocorreriam posteriormente.

Outras introduções são relatadas terem ocorrido pelas mãos de um colonizador alemão chamado Steiger em 1874 e 1876 (BARTLEY 2005), sendo as sementes também trazidas do Pará. Os descendentes destas introduções são designados como variedades Pará e Maranhão, que também se tornaram de típico cultivo na Bahia diferindo entre si e do cacau comum por um conjunto de propriedades fenotípicas, especialmente no que tange os caracteres físicos dos frutos. É importante notar que embora os relatos tenham apontado para a origem do cacau da Bahia no estado do Pará, a região precisa de onde surgiram os acessos fundadores permanece não esclarecida (MOTAMAYOR et al., 2008; BARTLEY, 2005). Sabe-se no entanto, por análise moleculares e fenotípicas que estas plantas possuem background genético próprio das originadas na região classificada para o cacau como Baixo Amazônica, fazendo parte do grupo Amelonado (MOTAMAYOR et al., 2008; 2003). As características que sustentam tal classificação se encontram no formato de melão dos frutos da variedade 'Comum' em adição a alta homogeneidade das plantas com altas taxas de homozigose decorrentes da capacidade de autofecundação (MOTILAL et al., 2013; N'GORAN et al., 2000) e, muito provavelmente, como reflexo do efeito fundador sob o qual o cacaueiro Amelonado parece ter sido formada (SERENO et al., 2006).

Devido a pequenas variações nos frutos, principalmente relacionadas ao tamanho e à textura da casca (VELLO e GARCIA, 1971), as variedades Pará e Maranhão podem ser subdividas em Pará e Parazinho para o primeiro e Maranhão Liso e Maranhão Rugoso para o segundo. Além destas, outras variantes foram sendo descobertas ao longo do tempo nas plantações de cacau da Bahia, algumas contendo grande interesse comercial, tendo sido originadas a partir de processos espontâneos de

mutação e recombinação (FIGUEIREDO, 2008). Entre estas variantes estão: (i) aquelas que apresentam alguma alteração nos componentes florais, a exemplo do cacau Maracujá e do cacau Macho; (ii) aquelas cujas variações notáveis estão no formato foliar, como é o caso do cacau Rui e do Jaca; (iii) além do cacau Sem Vidro que apresenta, entre outras características, frutos partenocárpicos (FIGUEIREDO, 2008). Além dos tipos citados outras duas variedades de cacaueiros da Bahia são destaques, tendo sido descobertos na década de 1930, e apesar de ser do tipo Forasteiro não possuem antocianina nas suas sementes, flores e folhas. Estes cacaueiros são denominados como Catongo e Almeida. Chocolate produzido a partir do clone Catongo SR16 mostrou ser menos adstringente e de sabor mais agradável (LEITE et al., 2013), sendo portanto esta variedade interessante para a fabricação de chocolate “gourmet”.

Durante dois séculos as sementes derivadas das plantações originais de cacaueiros na Bahia (especialmente da variedade comum) foram utilizadas para estabelecer praticamente todos os plantios comerciais desta região (MONTEIRO et al., 2009). Como resultado, uma população extremamente homogênea e com baixa variabilidade é esperada. Neste sentido, se em primeira instância uma alta endogamia é esperada nos cacaueiros da Bahia como consequência da origem genética a partir de populações do Baixo Amazonas (LACHENAUD e ZHANG, 2008), é também esperado que essa endogamia seja acentuada pelo efeito fundador das poucas sementes introduzidas, refletindo em populações com tendência à baixa variabilidade genética e elevado índice de fixação. O fato das plantas baianas terem excelente desempenho em produtividade e razoável tamanho de grão (VELLO e GARCIA, 1971), acabou sendo um ponto favorável na seleção e dispersão destas por quase todas as fazendas de cacau da Bahia, em tempos onde os altos valores pagos pelos grãos fizeram da cacaucultura uma excelente fonte de riquezas.

Acompanhando o cenário internacional no que diz respeito ao desenvolvimento de programas de melhoramento genético, as primeiras seleções em variedades baianas foram iniciadas em 1930 pelo Instituto do Cacau da Bahia (VELLO et al. 1972) em plantas presentes na Estação Experimental de Água Preta, no município de Urucuá, atual unidade do Instituto Federal de Educação e Tecnologia da Bahia, e foram baseadas especialmente em características de produção. Estas primeiras seleções

deram origem aos clones da série SIC (Seleção do Instituto do Cacau). Posteriormente, na década de 1940, novas seleções foram realizadas gerando as séries denominadas de Seleção do Instituto Agrônomo do Leste (SIAL) e Estação Experimental de Goitacazes (EEG) sendo os clones SIAL obtidos a partir de plantas de diferentes fazendas do sul da Bahia, e multiplicadas na Estação Experimental Dr. Gileno Amado, no município de Juçari também na Bahia (MONTEIRO et al., 2009; PEREIRA et al., 1987). Os clones EEG foram, por sua vez, selecionados na Estação Experimental de Linhares, que embora estivesse sediada no estado do Espírito Santo, tiveram na base da seleção plantas das variedades baianas e, portanto, possui uma origem genética comum com SIC e SIAL.

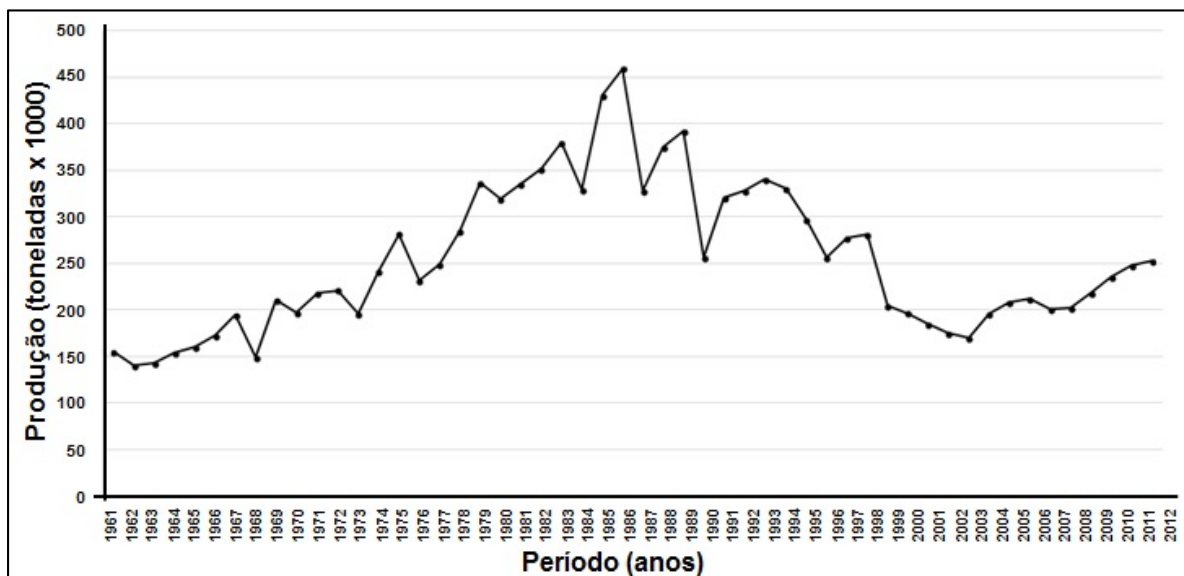
Com a implantação da CEPLAC na segunda metade da década de 1950 no município de Ilhéus, Bahia, clones das séries SIC e SIAL foram submetidos a avaliações fenotípicas tanto de caracteres botânicos quanto relacionados à produção (CASTRO e BARTLEY, 1983; 1985; PEREIRA et al., 1987). Embora possuam boas características relacionadas à produtividade SIC, SIAL e EEG apresentam baixa variabilidade fenotípica para algumas características, com pouca resistência às principais doenças do cacauzeiro como a podridão-parda e a vassoura-de-bruxa (MONTEIRO et al., 2009; PIRES, 2003). Paralelamente, e a partir de 1960 clones foram introduzidos na região sul baiana, especialmente vindos da Costa Rica e Trinidad, a fim de prover aos agricultores variedades de cacau superiores tanto por meio do uso direto como cultivar quanto através do emprego em programas de melhoramento genético (MONTEIRO et al., 2009). Neste sentido, e com a descoberta da heterose em cacauzeiros desde meados do século XX, a expectativa de ganhos genéticos com a produção de híbridos ficaram maiores e cruzamentos em busca de plantas superiores foram postos em prática. Diferentes combinações foram testadas nos cruzamentos utilizando como parentais os clones locais (SIC e SIAL), clones Alto e Baixo Amazônicos, bem como Crioulos e Trinitários. Cerca de 600 progênies híbridas foram produzidas sendo as seleções de variedades focadas na produtividade, precocidade, resistência a podridão-parda e qualidade industrial (VELLO et al., 1972). Segundo Monteiro et al. (2009) híbridos entre os clones locais X Scavinas apresentaram alta produtividade, mas grãos de peso inferior a 1 grama; e aqueles obtidos pelo



cruzamento clones locais X UF 613 apresentaram boa produtividade e alta incidência de podridão-parda causada por *Phytophthora palmivora*. De uma forma geral, os clones locais combinados entre si, apresentaram boa produtividade, superior às combinações locais X Trinitários (MONTEIRO et al., 2009).

A exemplo do relatado para outros países, as primeiras variedades híbridas liberadas para os agricultores eram bastante heterogêneas o que acabou levando a competição entre as plantas e a redução na produção ao longo do tempo, culminando na insatisfação de alguns produtores que optaram novamente pelo plantio das variedades tradicionais (PIRES, 2003) propagadas seminalmente. Com melhores informações sobre as plantas híbridas, as seleções recomendadas passaram a ser baseadas em plantas mais parecidas no intuito de evitar competição entre estas.

Em meados da década de 1980, o Brasil foi beneficiado por uma safra recorde e com os altos preços do cacau no mercado (<http://faostat3.fao.org>). Neste momento o país despontou como o segundo maior produtor de cacau do mundo, o que de fato parecia indicar uma era de grande prosperidade. Entretanto, no ano de 1989 a primeira ocorrência da vassoura-de-bruxa foi observada em uma fazenda no município de Uruçuca (PEREIRA et al., 1989) e em pouco tempo boa parte das plantações de cacau da Bahia estava comprometida, gerando uma crise sem precedentes na cacauicultura baiana com impactos sociais diretos. A drástica redução na produção do cacau na Bahia teve reflexo direto na produção brasileira como pode ser acompanhado no gráfico apresentado na Figura 5. A dispersão da vassoura-de-bruxa foi beneficiada por condições climáticas favoráveis à proliferação do fungo, além da alta susceptibilidade das plantas no campo. Como não houve melhoramento preventivo para a doença, até então exótica na Bahia, a grande devastação nas lavouras cacaueiras fez com que o foco basicamente na produtividade e resistência à podridão-parda, fosse desviado para a busca de clones mais tolerantes à vassoura-de-bruxa. Por este motivo, o programa de seleção de famílias híbridas, anteriormente relatado, foi descontinuado (MONTEIRO et al., 2009).



**Figura 5.** Produção de grãos de cacau no Brasil entre os anos 1961 e 2012. Fonte: <http://faostat3.fao.org>

Após a chegada da vassoura-de-bruxa na Bahia a busca por clones foi baseada em seleção de genótipos em áreas comerciais, coleções de germoplasma e progênies híbridas (PIRES, 2003). O foco do programa de melhoramento foi mantido na ampliação da base genética para a resistência e associação com outras características de interesse agrônomo, sendo utilizado como critério para seleção dos parentais, informações genéticas de diversidade molecular (a fim de aumentar a diversidade) em conjunto com informações agrônomicas. Como resultado do esforço inicial em busca de variedades com tolerância à vassoura-de-bruxa, em 1995 o primeiro híbrido resistente foi lançado na Bahia, sendo este gerado a partir dos clones Scavina 6 e ICS 1, caracterizados pelas positivas qualidades de resistência e produtividade, respectivamente. Esta variedade foi denominada de Theobahia e foi seguida pelo lançamento dos igualmente híbridos Theobahia 1 (Sca 6 X ICS 6) e Theobahia 2 (Sca 6 X ICS 8) (MONTEIRO et al., 2009). As duas últimas variedades não obtiveram muito sucesso, notadamente pela elevada susceptibilidade à murcha-de-Ceratocystis (causada pelo fungo *Ceratocystis cacaofunesta*) e a estreita base de resistência à

vassoura-de-bruxa (MONTEIRO et al., 2009). Além disso, assim como foi observado no Equador, a primeira variedade Theobahia teve sua resistência quebrada por variedades do fungo *M. pernicioso*, tornando esta ineficaz enquanto variedade resistente para os plantios baianos.

Como estratégia para obtenção de clones resistentes, a CEPLAC iniciou um programa participativo de seleção em plantações comerciais envolvendo diretamente fazendeiros e extensionistas treinados na identificação de plantas resistentes a vassoura-de-bruxa e produtivas (CacaoNet, 2012). Uma pré-seleção de 5.000 plantas foi realizada e após testes, uma parcela foi liberada tanto para uso imediato nas fazendas quanto para aplicação nos programas de melhoramento. PS 13.19, CEPEC 2002 e PH 16 são exemplos de clones originados a partir de fazendas do sul da Bahia e que compõem a lista de plantas atualmente recomendadas para plantio pelo CEPEC/CEPLAC (Comunicado CEPLAC/CEPEC 02/2014, acessado em: [www.ceplac.gov.br](http://www.ceplac.gov.br)).

A partir da década de 1990, foi iniciado também um programa de melhoramento genético por meio de seleção recorrente recíproca, tendo como base uma população formada por material elite, plantas de germoplasma e seleções de fazendas, no qual o foco principal foi o melhoramento simultâneo de características como resistência a doenças, produtividade e qualidade (PIRES 2003). Neste programa, com o avanço dos ciclos de cruzamentos, buscando plantas cada vez mais melhoradas, foi possível mensurar as capacidades de combinação geral e específica dos parentais. O esforço empreendido no programa de melhoramento do cacauzeiro na Bahia alcançou números tão grandes quanto a geração de mais de 30.000 árvores, teste de 500 clones e liberação de 39 destes para plantio comercial. Estas plantas foram e continuam sendo úteis na substituição daquelas mais susceptíveis, estimando-se já terem sido plantadas em mais de 150 mil hectares (LOPES et al., 2011). Atualmente, os clones destas seleções são produzidos e comercializados pela Biofábrica da CEPLAC que é incumbida de propagar as plantas vegetativamente por meio de estacas enraizadas ou como enxertos.

O plantio dos clones recomendados pela CEPLAC é uma das ações que compõem o manejo integrado da lavoura, a fim de mantê-la mais produtiva e resistente

à vassoura-de-bruxa com o mínimo custo. Neste sentido, além da resistência genética, são recomendados a poda fitossanitária, o uso racional de químicos e o controle biológico (MEDEIROS et al., 2010). Para este último, o primeiro produto desenvolvido para cacaueiros - o Tricovab - feito a partir do fungo *Trichoderma stromaticum*, naturalmente antagonista do *M. pernicioso*, já está sendo utilizado em plantações comerciais de cacau da Bahia.

O melhoramento genético realizado para cacaueiros na Bahia atualmente, mantém o foco em ampliar a diversidade genética do material empregado nos cruzamentos, buscando diferentes fontes de variação para as características envolvidas, incluindo a resistência à vassoura-de-bruxa. Além disso, a busca de materiais autocompatíveis constitui uma meta. Como estratégia de melhoramento preventivo, plantas caracterizadas como resistentes à Monilíase (causada por *M. royeri*) também fazem parte do programa, com intuito de preparar a lavoura baiana para uma eventual introdução do fungo (LOPES et al., 2011).

---

**OBJETIVO GERAL**

Contribuir com informações genético-moleculares para a conservação de germoplasma e o melhoramento genético do cacaueteiro.

**OBJETIVOS ESPECÍFICOS**

- Desenvolver marcadores moleculares SSR para *Theobroma cacao*;
- Caracterizar a diversidade genética de genótipos de cacaueteiro da Bahia presentes em fazendas, a partir de avaliações moleculares e fenotípicas;
- Caracterizar a diversidade genética dos primeiros clones de cacaueteiro selecionados na Bahia, representativos das variedades baianas em bancos de germoplasma.



## **CAPÍTULO I**

**Desenvolvimento e caracterização de  
marcadores microssatélites para cacaueiro**

**Artigo Publicado em 2012 na revista *Biologia Plantarum***

Doi 10.1007/s10535-012-0134-y

**New polymorphic microsatellite loci for *Theobroma cacao*:  
isolation and characterization of microsatellites from  
enriched genomic libraries**

**Santos E.S.L.**

Cerqueira-Silva C.B.M.

Mori G.M.

Ahnert D.

Corrêa R.X.

**Souza A.P**





## BRIEF COMMUNICATION

**New polymorphic microsatellite loci for *Theobroma cacao*: isolation and characterization of microsatellites from enriched genomic libraries**

E.S.L. SANTOS<sup>1,2</sup>, C.B.M. CERQUEIRA-SILVA<sup>1,2</sup>, G.M. MORI<sup>2</sup>, D. AHNERT<sup>3</sup>, R.X. CORRÊA<sup>3</sup>  
and A.P. SOUZA<sup>2,4\*</sup>

*Departamento de Estudos Básicos e Instrumentais, Universidade Estadual do Sudoeste da Bahia, Rodovia BR415, Km03, s/n, Itapetinga, 45700-000, Bahia, Brazil<sup>1</sup>.*

*Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, CP 6010, Campinas, 13083-875, São Paulo, Brazil<sup>2</sup>.*

*Centro de Biotecnologia e Genética, Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna Km 16, Ilhéus, 45662-900, Bahia, Brazil<sup>3</sup>.*

*Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas, 13083-970, São Paulo, Brazil<sup>4</sup>.*

**Abstract**

Seventeen polymorphic microsatellite markers were isolated from enriched genomic libraries for *Theobroma cacao*, providing additional tools for studying the genetic diversity and map saturation of this species. These markers were characterized in 32 accessions of the *T. cacao* germplasm collection from the Centro de Pesquisas do Cacau. The number of alleles at each locus varied from 2 to 8, with an average of 4.41 alleles per locus. The polymorphism information content varied from 0.060 to 0.695, with an average of 0.333. The markers characterized in this study will be employed in map saturation studies and diversity assessments of cacao genotypes.

*Additional key words:* cacao, genetic diversity, molecular markers, PCR.

*Theobroma cacao* L. (2n=20) is a neotropical plant that is native to the northern regions of South America and has been domesticated in Central America (Motamayor *et al.* 2002). Among the 22 species of the *Theobroma* genus cacao is the only species cultivated on a large scale for chocolate production (Luz and Silva 2001). International genetic breeding programs for cacao have been developed as a means of obtaining cultivars with improved productivity, increased resistance and good industrial quality (Guiltinan *et al.* 2008). As a useful tool for genetic studies, molecular markers, especially microsatellite markers, have been developed for different species, such as *Jatropha curcas* (Phumichai *et al.* 2011) and *Aegiphila sellowiana* (Ruas *et al.* 2011). Specifically

in cacao, some simple sequence repeat (SSR) markers have already been developed (Lanaud *et al.* 1999, Pugh *et al.* 2004, Araújo *et al.* 2007, Lima *et al.* 2008) and have been used for diversity studies and breeding programs (Aikpokpodion *et al.* 2009, Lanaud *et al.* 2009, Yamada *et al.* 2009). To study the genetic diversity and genetic map saturation of *T. cacao*, we isolated and characterized a new set of polymorphic microsatellites from *T. cacao*. These markers may also be used for future genetic studies of this species.

To identify and characterize microsatellites, two microsatellite-enriched libraries for *T. cacao* were developed following the protocol described by Billotte *et al.* (1999) from two phenotypically and genetically

Received 14 July 2011, accepted 20 December 2011.

**Abbreviations:** CTAB - cetyltrimethylammonium bromide; HWE - Hardy-Weinberg equilibrium; LD - linkage disequilibrium; PCR - polymerase chain reaction; PIC - polymorphism information content; SSR - simple sequence repeat; TBE - Tris/borate/EDTA.  
**Acknowledgments:** The authors are grateful to Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, PROCAD-NF2008) for financial support, to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing graduate fellowships to G.M. Mori and E.S.L. Santos, respectively, and a research fellowship (from CNPq) to A.P. Souza.

\* Corresponding author: fax: (+55) 19 35211089, e-mail: anete@unicamp.br

E.S.L. SANTOS *et al.*

distinct clones: TSH 1188 (Upper Amazon rainforest of Trinidad) and GU 261 (Lower Amazon rainforest of French Guiana). These clones were sampled from the germplasm collection at the Centro de Pesquisas do Cacau/Comissão Executiva do Plano da Lavoura Cacaueira (CEPEC/CEPLAC) in Itabuna, Bahia, Brazil. Genomic DNA was isolated from fresh leaves from each clone using a cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1990) with modifications specific for cacao (Faleiro *et al.* 2002). The DNA samples (5 µg) were digested with *AfaI* and ligated to the double-stranded adapters 5'-CTCTTGCTTACGCGTGGACTA-3' and 5'-TAGTCCACGCGTAAGCAAGAGCACA-3'. The enrichment was performed using a hybridization-based capture with (GT)<sub>8</sub> and (CT)<sub>8</sub> biotin-linked probes and magnetic-coated streptavidin beads (streptavidin *MagneSphere* paramagnetic particles, *Promega*, Madison, WI, USA). Selected fragments were then cloned into a *pGEM-T Easy* (*Promega*) plasmid vector and transformed into *Escherichia coli* XL1-Blue competent cells (*Stratagene*, La Jolla, CA, USA). Recombinant

colonies were selected by blue/white screening. Clones were randomly selected and were double-sequenced using the T7 and SP6 primers and the *v3.1 Big Dye* terminator kit (*Applied Biosystems*, Foster City, CA, USA) on an *ABI PRISM 377* automated DNA sequencer (*Applied Biosystems*). Every obtained sequence was aligned, edited and eliminated if redundant using *SeqMan* software (*DNASTAR*, Madison, WI, USA). *MICROSAT* software (A.M. Risterucci, CIRAD, personal communication) was used to eliminate adapters and restriction sites from the sequences. The simple sequence repeat identification tool (*SSRIT*) (Temnykh *et al.* 2001) was used to identify the following microsatellites: di- and trinucleotides with four or more repeats; and tetra- and pentanucleotides with three or more repeats.

Primer pairs were designed using the *PrimerSelect* (*DNASTAR*) and *Primer3Plus* (Untergasser *et al.* 2007) programs. The polymerase chain reaction (PCR) was conducted in a final volume of 0.015 cm<sup>3</sup> containing 15 ng of template DNA, 1× PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 1.5 mM MgCl<sub>2</sub>, 0.2 µM of

Table 1. Characteristics of the forward (F) and reverse (R) primers, repeat motif, and amplification conditions (AC) of 17 microsatellite loci from *Theobroma cacao* L. \* - TD65-55 indicates touchdown PCR with temperatures ranging from 65 to 55 °C.

| Locus         | GenBank acc. | Primer sequence (5'-3')                               | Repeat motif                                              | AC*     |
|---------------|--------------|-------------------------------------------------------|-----------------------------------------------------------|---------|
| mTc-UNICAMP01 | JF812968     | F: TCATGCAAAGCAAAGTGAAG<br>R: ACGGGAAACTCATCATTACA    | (AG) <sub>11</sub>                                        | TD65-55 |
| mTc-UNICAMP02 | JF812969     | F: GGTCCCTCCAAGCTGAGTAACA<br>R: CTCCCTATTTGCATCGCATT  | (TC) <sub>15</sub>                                        | TD65-55 |
| mTc-UNICAMP03 | JF812970     | F: GCGATGCCAAAAGTTGTGTGA<br>R: AGAGAAGGGATGGGTGTGTG   | (AG) <sub>16</sub>                                        | TD65-55 |
| mTc-UNICAMP04 | JF812971     | F: GCAGCGAGAGACAAAGATA<br>R: TTTGACTGAAATGGTGGTAA     | (TC) <sub>21</sub>                                        | 60 °C   |
| mTc-UNICAMP05 | JF812972     | F: AGCTGTTTATGATTCACATCC<br>R: GAAGCAGCAATTGTAACCAC   | (TA) <sub>7</sub> (CA) <sub>6</sub>                       | 60 °C   |
| mTc-UNICAMP06 | JF812973     | F: AACTCTGGCATTTCATTGG<br>R: GTGATGGATGCGAGGTGGAAT    | (TC) <sub>16</sub> TGT(CA) <sub>10</sub>                  | 60 °C   |
| mTc-UNICAMP07 | JF812974     | F: CATGCCTGAGCTATGAACA<br>R: TCAGATATGGGTAAAGGAAGAG   | (GA) <sub>11</sub> ...(GA) <sub>5</sub>                   | TD65-55 |
| mTc-UNICAMP08 | JF812975     | F: GGGAAACTGGGCATCACTTA<br>R: TTAATGGGTGAGGAGAGCAT    | (AC) <sub>5</sub> AG(AT) <sub>9</sub>                     | 60 °C   |
| mTc-UNICAMP09 | JF812976     | F: TTCGGCAGTTCGATCTATGA<br>R: ATCCACCGTAAGCCTTTCCT    | (TAAAG) <sub>7</sub>                                      | TD65-55 |
| mTc-UNICAMP10 | JF812977     | F: AACTGTCGCATACGCTCATC<br>R: TCAAAAAGTGAAGCCCAACCAT  | (AG) <sub>9</sub>                                         | 60 °C   |
| mTc-UNICAMP11 | JF812978     | F: GCTTGTTCCTCGACACACAT<br>R: TGGAAGCTAGTTAGGAATCACA  | (CA) <sub>6</sub>                                         | 60 °C   |
| mTc-UNICAMP12 | JF812979     | F: GCCATCTCGCTATTTCATGTG<br>R: GTCATTGCTAACGTAACCAGAA | (AG) <sub>16</sub>                                        | 60 °C   |
| mTc-UNICAMP13 | JF812980     | F: ACTTGTCGCTTGCTTCCTGT<br>R: TTCCTCAACCTCTCCGAATG    | (AT) <sub>6</sub>                                         | TD65-55 |
| mTc-UNICAMP14 | JF812981     | F: ACATGATGCCTGAGCTGTTG<br>R: CTCCAATTTTGACCCAGAA     | (AG) <sub>11</sub>                                        | TD65-55 |
| mTc-UNICAMP15 | JF812982     | F: TGCTTTGAGGCACCTGTCC<br>R: TGGCATTTCATTGAGAGGTGA    | (TG) <sub>12</sub>                                        | TD65-55 |
| mTc-UNICAMP16 | JF812983     | F: CAGGAAGGATACTTCTTAAAGG<br>R: AGTAGAGTCGAGTGGCTTGA  | (CT) <sub>12</sub> (CA) <sub>9</sub> ...(CT) <sub>4</sub> | TD65-55 |
| mTc-UNICAMP17 | JF812984     | F: CTGCACAGCTTCATGGACTC<br>R: TGATGATCAGGTGGTTTCTCA   | (CA) <sub>9</sub>                                         | TD65-55 |



## NEW POLYMORPHIC MICROSATELLITE LOCI

Table 2. Characteristics of the product size range in base pairs, number of effectively evaluated genotypes (N), number of alleles ( $N_A$ ), observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity, polymorphism information content (PIC) and  $P$  value of Hardy-Weinberg equilibrium (HWE) for 17 microsatellite loci from *Theobroma cacao* (\* - departs significantly from HWE at  $P < 0.003$  after the Bonferroni correction, † - significant values for 'null allele' occurrence after the Bonferroni correction).

| Locus         | Size range | N  | $N_A$ | $H_o$ | $H_e$ | PIC   | $P$ value |
|---------------|------------|----|-------|-------|-------|-------|-----------|
| mTc-UNICAMP01 | 190-208    | 29 | 7     | 0.269 | 0.286 | 0.273 | 0.194     |
| mTc-UNICAMP02 | 260-290    | 31 | 8     | 0.381 | 0.746 | 0.695 | 0.000*†   |
| mTc-UNICAMP03 | 282-291    | 31 | 5     | 0.433 | 0.719 | 0.661 | 0.000*    |
| mTc-UNICAMP04 | 162-186    | 30 | 7     | 0.786 | 0.742 | 0.695 | 0.024     |
| mTc-UNICAMP05 | 233-242    | 29 | 4     | 0.269 | 0.544 | 0.454 | 0.003†    |
| mTc-UNICAMP06 | 140-170    | 32 | 4     | 0.107 | 0.123 | 0.120 | 0.032     |
| mTc-UNICAMP07 | 265-312    | 28 | 4     | 0.240 | 0.225 | 0.211 | 1         |
| mTc-UNICAMP08 | 278-286    | 32 | 5     | 0.581 | 0.673 | 0.605 | 0.044     |
| mTc-UNICAMP09 | 148-190    | 31 | 6     | 0.032 | 0.667 | 0.596 | 0.000*†   |
| mTc-UNICAMP10 | 230-236    | 32 | 3     | 0.094 | 0.091 | 0.088 | 1         |
| mTc-UNICAMP11 | 290-296    | 32 | 2     | 0.226 | 0.204 | 0.180 | 1         |
| mTc-UNICAMP12 | 240-290    | 32 | 6     | 0.226 | 0.369 | 0.347 | 0.008†    |
| mTc-UNICAMP13 | 140-144    | 30 | 2     | 0     | 0.066 | 0.064 | 0.017†    |
| mTc-UNICAMP14 | 254-258    | 28 | 3     | 0.174 | 0.166 | 0.156 | 1         |
| mTc-UNICAMP15 | 292-302    | 32 | 2     | 0.067 | 0.064 | 0.060 | 1         |
| mTc-UNICAMP16 | 244-254    | 31 | 4     | 0.207 | 0.255 | 0.240 | 0.153     |
| mTc-UNICAMP17 | 296-300    | 29 | 3     | 0.174 | 0.236 | 0.213 | 0.311     |

each dNTP, 0.5  $\mu$ M of each primer and 1 U of *Taq* DNA polymerase (*Invitrogen*, Carlsbad, CA, USA). All primers were first evaluated in a PCR amplification protocol with a 60 °C annealing temperature as follows: 94 °C for 5 min; 34 cycles of 94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min; and a final extension at 72 °C for 10 min. Later, if necessary, touchdown PCR was performed. The touchdown protocol consisted of 5 steps as follows: a) initial denaturation (94 °C for 5 min); b) 10 cycles of denaturation (94 °C for 1 min), annealing (65 - 1 °C per cycle for 1 min) and extension (72 °C for 1 min); c) step b repeated; d) 14 cycles of denaturation (94 °C for 1 min), annealing (55 °C for 1 min) and extension (72 °C for 1 min); e) a final extension (72 °C for 10 min). The amplification products were visualized in 0.5× TBE/3 % agarose gels stained with ethidium bromide prior to vertical electrophoresis in 1× TBE/6 % denaturing polyacrylamide silver nitrate-stained gels. Product sizes were determined by comparison with a 10 bp DNA ladder (*Invitrogen*). Statistical analyses were performed with *GENEPOP* on the web (Raymond and Rousset 1995), *PIC Calculator* (Kemp 2002) and *MICRO-CHECKER* (Oosterhout *et al.* 2004).

In all, 192 positive colonies were sequenced and 163 microsatellites were obtained. Dinucleotide motifs were the most abundant, followed by tri-, tetra- and penta-nucleotide motifs (with proportions of approximately 94, 3, 2 and 1 %, respectively). Thirty-five primer pairs were designed such that the product size ranged between 100 and 300 bp for accurate genotyping using vertical electrophoresis in 1× TBE/6 % denaturing polyacrylamide gels. Twenty-five of these primer pairs generated consistent patterns of amplification, with fragments matching size expectations based on the sequence information, and these pairs were used for

further characterization. For polymorphism analysis, 32 DNA samples of *T. cacao* were extracted and analyzed from the germplasm collection of *CEPEC/CEPLAC*. To sample greater genetic variation, we selected genotypes collected from the Amazon in addition to Bahia cacao farms. Of these 25 primer pairs evaluated, 17 tagged polymorphic loci whose numbers of alleles varied (Table 1). To measure the frequency of identity between the polymorphic loci presented here and those described in the literature, we used nucleotide *BLAST* (Altschul *et al.* 1990) (using the megablast algorithm). We observed identity (varying from 95 to 97 %) to four sequences (loci: mTc-UNICAMP03, 04, 07, 09). However, these primers had annealing sites different from those described previously. Of these four loci, three (mTc-UNICAMP03, 04 and 07) were similar to microsatellite loci published by Pugh *et al.* (2004) and are included here to evaluate statistical parameters such as polymorphism information content (PIC) and Hardy-Weinberg equilibrium (HWE). No previous paper has described locus mTc-UNICAMP09. The mean number of alleles was 4.41 (range: 2 to 8 alleles), and the PIC varied from 0.060 to 0.695. Genotypic frequencies of the 17 polymorphic microsatellite markers were tested for HWE and linkage disequilibrium (LD) using the Bonferroni correction. The analysis revealed significant deviations from HWE at three microsatellite loci (Table 1), possibly due to the occurrence of null alleles (as observed for mTc-UNICAMP02 and 09) or LD (as observed for mTc-UNICAMP02, 03 and 09). Significant LD was observed among 19 of 136 possible pairs of markers (mTc-UNICAMP01 and 04, 05, 11, 12; mTc-UNICAMP03 and 07, 11; mTc-UNICAMP05 and 14; mTc-UNICAMP12 and 03, 04, 05, 11; mTc-UNICAMP16 and 01, 02, 03, 06, 07, 09, 11, 12).



E.S.L. SANTOS *et al.*

The probability of 'stuttering', 'large allele dropout' and 'null alleles' was also evaluated. No microsatellite loci showed a significant probability of 'stuttering' or 'large allele dropout'. Five loci (mTc-UNICAMP02, 05, 09, 12, 13) showed significant values for 'null alleles' according to the Bonferroni correction ( $P < 0.003$ ). Eight

other microsatellite markers were monomorphic for the evaluated population.

Microsatellite markers characterized in this study may be used in future genetic map saturation and genetic diversity studies of cacao genotypes.

## References

- Aikpokpodion, P.O., Motamayor, J.C., Adetimirin, V.O., Adu-Ampomah, Y., Ingelbrecht, I., Eskes, A.B., Schnell, R.J., Kolesnikova-Allen, M.: Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. - *Tree Genet. Genom.* **5**: 699-711, 2009.
- Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D.: Basic local alignment search tool. - *J. mol. Biol.* **215**: 403-410, 1990.
- Araújo, I.S., Intorne, A.C., Pereira, M.G., Lopes, U.V., De Souza Filho, G.A.: Development and characterization of novel tetra-, tri- and di-nucleotide microsatellite markers in cacao (*Theobroma cacao* L.). - *Mol. Breed.* **20**: 73-81, 2007.
- Billotte, N., Lagoda, P.J.R., Risterucci, A.M., Baurens, F.C.: Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. - *Fruits* **54**: 277-288, 1999.
- Doyle, J.J., Doyle, J.L.: Isolation of plant DNA from fresh tissue. - *Focus* **12**: 13-15, 1990.
- Faleiro, F.G., Santos, I.S., Bahia, R.C.S., Santos, R.F., Yamada, M.M., Anheer, D.: Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando obtenção de marcadores RAPD. [Optimization of DNA extraction and amplification of *Theobroma cacao* L. aiming obtainment of RAPD markers.] - *Agrotrópica* **14**: 31-34, 2002. [In Port.]
- Guiltinan, M.J., Verica, J., Zhang, D., Figueira, A.: Genomics of *Theobroma cacao*, "the food of the Gods". - In: Moore, P.H., Ming, R. (ed.): *Genomics of Tropical Crop Plants*. Pp. 145-170. Springer, New York 2008.
- Kemp, S.: PIC Calculator. Extra. <http://www.genomics.liv.ac.uk/animal/Pic1.html>. 2002 <Accessed 08 march 2011>
- Lanaud, C., Fouet, O., Clément, D., Boccara, M., Risterucci, A.M., Surujdeo-Maharaj, S., Legavre T., Argout, X.: A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. - *Mol. Breed.* **24**: 361-374, 2009.
- Lanaud, C., Risterucci, A.M., Pieretti, I., Falque, M., Bouet, A., Lagoda, P.J.L.: Isolation and characterization of microsatellites in *Theobroma cacao* L. - *Mol. Ecol.* **8**: 2141-2143, 1999.
- Lima, L.S., Gramacho, K.P., Gesteira, A.S., Lopes, U.V., Gaiotto, F.A., Zaidan, H.A., Pires, J.L., Cascardo, J.C.M., Micheli, F.: Characterization of microsatellites from cacao-*Moniliophthora perniciosa* interaction expressed sequence tags. - *Mol. Breeding* **22**: 315-318, 2008.
- Luz, E.D.M.N., Silva, S.D.V.M.: Podridão-parda dos frutos, cancro e outras doenças causadas por *Phytophthora* no cacauero. [Black pod, cancer and other diseases caused by *Phytophthora* in cocoa.] - In: Luz, E.D.M.N., Santos, A.F., Matsuoka, K., Bezerra, J.L. (ed.): *Doenças Causadas por Phytophthora no Brasil*. [Diseases Caused by *Phytophthora* in Brazil.] Pp. 175-265. Rural, Campinas 2001. [In Port.]
- Motamayor, J.C., Risterucci, A.M., Lopez, P.A., Ortiz, C.F., Moreno, A., Lanaud, C.: Cacao domestication I: the origin of the cacao cultivated by the Mayas. - *Heredity* **89**: 380-386, 2002.
- Oosterhout, C.V., Hutchinson, W.F., Wills, D.P.M., Shipley, P.: MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. - *Mol. Ecol. Notes* **4**: 535-538, 2004.
- Phumichai, C., Phumichai, T., Kongsiri, N., Wongkaew, A., Sripichit, P., Kaveeta, R.: Isolation of 55 microsatellite markers for *Jatropha curcas* and its closely related species. - *Biol. Plant.* **55**: 387-390, 2011.
- Pugh, T., Fouet, O., Risterucci, A.M., Brottier, P., Abouladze, M., Deletrez, C., Courtois, B., Clément, D., Larmande, P., N'Goran, J.A., Lanaud, C.: A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. - *Theor. appl. Genet.* **108**: 1151-1161, 2004.
- Raymond, M., Rousset, F.: *Genepop* (version 1.2): population genetics software for exact tests and ecumenicism. - *J. Hered.* **86**: 248-249, 1995.
- Ruas, E.A., Damasceno, J.O., Conson, A.R.O., Costa, B.F., Rodrigues, L.A., Reck, M., Santos Vieira, A.O., Ruas, C.F., Medri, C., Ruas, P.M.: Isolation and characterization of eleven polymorphic microsatellite loci in *Aegiphila sellowiana* and their transferability. - *Biol. Plant.* **55**: 396-399, 2011.
- Temnykh, S., De Clerck, G., Lukashova, A., Lipovich, L., Cartinhou, S., McCouch, S.: Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. - *Genome Res.* **11**: 1441-1452, 2001.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M.: Primer3Plus, an enhanced web interface to Primer3. - *Nucl. Acids Res.* **35**: 71-74, 2007.
- Yamada, M.M., Faleiro, F.G., Flores, A.B., Lopes, U.V., Pires, J.L., Corrêa, R.X., Santos, R.F.: Microsatellite diversity and heterozygosity of parents of a cocoa breeding population. - *Crop Breed. appl. Biotechnol.* **9**: 17-22, 2009.

## **CAPÍTULO II**

### **Diversidade genética de cacaueiros da Bahia I**

**Artigo em processo de submissão**

**Southern Bahia local cacao varieties from the Amelonado group:  
genetic diversity and structure of populations established for 260  
years and survivors of the epidemic of witches' broom disease**

**Santos E.S.L.**

Cerqueira-Silva C.B.M.

Melo D.L.N.

Corrêa R.X.

Ahnert D.

**Souza A.P**



## **Southern Bahia local cacao varieties from the Amelonado group: genetic diversity and structure of populations established for 260 years and survivors of the epidemic of witches' broom disease**

Elisa S. L. Santos<sup>1,2</sup>, Carlos Bernard M Cerqueira-Silva<sup>1,2</sup>, Durval Libânio Netto Mello<sup>3</sup>, Ronan X. Corrêa<sup>4</sup>, Dário Ahnert<sup>4</sup>, Anete P. Souza<sup>2,5</sup>

Departamento de Ciências Exatas e Naturais, Universidade Estadual do Sudoeste da Bahia, Rodovia BR415, Km03, s/n, Itapetinga, 45700-000, Bahia, Brazil<sup>1</sup>.

Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, CP 6010, Campinas, 13083-875, São Paulo, Brazil<sup>2</sup>.

Instituto Federal de Educação, Ciência e Tecnologia Baiano, Rua Dr. João Nascimento Centro, Uruçuca, 45680-000, Bahia, Brasil<sup>3</sup>

Centro de Biotecnologia e Genética, Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna Km 16, Ilhéus, 45662-900, Bahia, Brazil<sup>4</sup>.

Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas, 13083-970, São Paulo, Brazil<sup>5</sup>.

### **Abstract**

Southern Bahia local cacao cultivars: Comum, Pará, Parazinho, Maranhão have been cultivated in the region since eighteenth/nineteenth century and are still predominant in the plantations nowadays. Comum variety was introduced to Bahia in 1746 and Para and Maranhão around one century latter. They belong to the Amelonado cacao group, are self-compatible and were introduced from the Lower Amazon region, most probably from the Para State, Brazil. During 260 of cultivation in Southern Bahia, many mutants arose, for example, Catongo and Almeida, which have white seeds, Laranja and Jaca, which have fruits in shape of orange and leaves in shape of jackfruit, respectively. Even though these cultivars are no resistant to witches' broom disease, they are very much appreciated by farmers because of their elevated productivity and quality features. With the identification of witches' broom disease in 1989 and the advent of the witches' broom resistant clonal cultivars this germplasm is being replaced in many farms. This study aimed to characterize the genetic variability of the local cacao cultivars. A sample of 176 plants was taken in four representative municipalities: Canavieiras, Uruçuca, Camacan and Gandu. We used 29 microsatellite markers and nine agronomic traits of fruits (weight, length, diameter and shell thickness) and seeds (total number and weight per fruit, weight of 10 seeds and length and width). The molecular evaluations allowed for the identification of 119 alleles, and Bayesian statistics indicated the presence of three genetic groups. One group was formed by six genotypes and was not considered representative of Bahian cacao local cultivars. The remaining two groups were composed of 71 and 75 individuals and were termed Bahian cacao I

(allele richness = 1.31,  $H_0$  = 0.054 and  $H_E$  = 0.106) and Bahian cacao II (allele richness = 1.41,  $H_0$  = 0.067 and  $H_E$  = 0.149), respectively. Rogers' modified genetic distance calculations revealed a trend of the grouping of plants of the Maranhão type, which is used for the production of fine chocolate. Phenotypic evaluations indicated significant differences between genotypes ( $p < 0.01$ ) for each of the nine traits. Mahalanobis distance measurements indicated a relatively small genetic distance between genotypes and a tendency of plants that were collected from regions that were closer to one another to group together. These results indicate that although the genetic variability of Bahian local cacao is restricted, these plants have potential for use in breeding programs because they typically have good production rates and are flavorful. The various events that may have reduced the genetic diversity of cacao in Bahia and the importance of maintaining data derived from farm-grown plants in gene banks are discussed.

**Keywords:** molecular markers, microsatellite, agronomic evaluations, genetic breeding, witches' broom.

## Introduction

*Theobroma cacao* L. is a neotropical perennial plant that originated in the Amazon region, which is also its diversity center, and was first domesticated by the Mayan people in Mesoamerica approximately 3,000 years ago. Considering the origins of domestication of the different cacao varieties, two morphogeographical groups are recognized according to Cheesman (1944), including Criollo, which was domesticated from southern Mexico to northernmost South America and western Ecuador, and Forastero, which was originally cultivated in the Amazon basin. Morphological differences between these two cacao groups include the colors of the fruits and inner seeds, and varying degrees of disease resistance are also observed. Forastero cacao is characterized by producing purple seeds and green immature fruits. Additionally, this group is recognized to be more heterogeneous with vigorous and disease resistant plants. Criollo cacao is characterized by producing white seeds and red immature fruits. Additionally, this group is recognized to be more homogeneous with less vigorous and disease susceptible trees. A third group, Trinitario, originated from natural hybridization between Criollo and Forastero in the Trinidad islands, and it is characterized by characteristics intermediate between those of the former two groups (Cheesman, 1944).



Molecular and phenotypic evaluations have indicated high levels of genetic diversity in *T. cacao*, which were underestimated due to the restrictive classifications of the three aforementioned geographical groups (Motilal et al. 2013; Faleiro et al. 2004; N'Goran et al. 2000). Only recently, classifications that were based on microsatellite markers indicated the occurrence of at least ten genetic groups in the cacao germplasm, which were named according to their geographical locations or most prominent traditional cultivars in the particular clusters, including Mara  n, Curaray, Criollo, Iquitos, Nanay, Contamana, Amelonado, Pur  s, Nacional and Guiana (Motamayor et al. 2008). Amelonado refers to those plants that originated in the Lower Amazon region with characteristic melon fruit shape. This group encompasses auto-compatible plants with genotypes that have been extensively cultivated on cacao farms and are especially productive (Bhattacharjee and Kuma 2007). The incipient domestication (Clement et al. 2010) and successive generation over time of plants of this group have resulted in uniform individuals with homozygous genotypes (Bhattacharjee and Kuma 2007, Bartley 2005). It is possible that the Amelonado group arose from a few Upper Amazon Forastero plants (Clement et al. 2010; Sereno et al. 2006) that may have been domesticated in tributaries of the Brazilian Para River (Motamayor et al. 2008).

Brazil is the most important cacao-producing country in the Americas, yielding 253,211 t of cacao beans in 2012 (<http://faostat3.fao.org>). Sixty-two percent of the Brazilian cacao production is developed in Southern Bahia region, which has approximately 540.000 ha and yield on average 300 kg/ha (IBGE 2012). In Southern Bahia, approximately 70% of cacao cultivation is performed in cabruca agroforestry systems (Ara  jo 1998), which are characterized by understory tree species in Atlantic forests. This system has allowed the conservation of native flora and fauna (Faria et al. 2007; Sambuichi 2006). In this region, average temperature ranges from 20 to 27  C, and annual rainfall may reach 2,350 mm (<http://www.inpe.br/>). Due to the abundant rainfall, cacao is produced for approximately nine months out of the year, which is divided into two harvest periods, including the main period from November to February and the secondary period (locally known as tempor  o) from May to September.

The first introduction of cacao to Southern Bahia occurred in 1746, when seeds of the Amelonado group were imported by a Frenchman and planted in the municipality of

Canavieiras, on the River Pardo margins (Vello and Garcia 1971). This first introduction originated the variety named as Cacao Comum and over the course of two centuries descendants of these plants spread and founded nearly all Bahian cacao plantations in addition to commercial plantations in Espírito Santo, other Brazilian states and African countries (Young 2007). Additional introductions were accomplished by a German individual between 1874 and 1876, and these varieties were named Maranhão and Pará. They differed from each other and the Comum variety mainly by their fruit characteristics (Bartley 2005, Vello and Garcia 1971). Spontaneous mutants of these cacao founders arose forming many mutants, for example, the Almeida and Catongo varieties, which possess white seeds (Vello et al. 1972). Due to the unique characteristics of the aforementioned varieties, they were naturalized as Bahian cacao or local cacao cultivars, and this state is currently considered to be a secondary area of cacao genetic diversity (Bartley 2005).

Following years of cultivation and selections by farmers in Bahia, productive and homogenous genotypes were generated (Ahnert and Ferreira 2013) that were considered to possess low levels of resistance to certain diseases, such as pod rot (caused by the *Phytophthora* species) and witches' broom (caused by *Moniliophthora perniciosa*). In fact, they still occupy approximately 50% of the cultivated areas of cacao in Bahia (Pires 2003). Before the witches' broom outbreak on the Bahian farms in 1989 (Pereira et al. 1989), Cacao Comum, Pará and Maranhão were the predominant varieties. The decline in position from second to sixtieth in world production depicts the consequences of diseases on Bahia's economy and justifies the replacement of some of the originally introduced plant varieties with those that are more disease resistant. Additionally, the low prices that were associated with the cacao market and specific environmental factors contribute to the cacao crisis in Bahia.

To overcome this crisis in cacao production, which is also occurring in other cacao-producing countries (Trogitz et al. 2011), some Bahian farmers are replacing the local cacao varieties by resistant clonal cultivars. Additionally, they are investing in the 'gourmet' chocolate market, in which flavorful cacao is worth three times the price of standard cacao. For producing fine chocolate, it is necessary to take into account intrinsic genetic characteristics in addition to on-farm processing. The Maranhão and

Catongo varieties are being employed in the fine cacao market, and they have been reported to produce less astringent, more flavorful chocolate (Leite et al. 2013).

In general, the local cacao varieties are encountered mainly in smallholder farms and constitute important genetic resources that may be applied in breeding programs as locally adapted material. It is plausible that these plants, although descendants of the original population that was planted in Bahia, are genetically different than those that were originally introduced. Their molecular and morphological characteristics are thus important to elucidate the genetic and structural diversities that exist and to identify plants that are optimal for use in breeding programs. For this purpose, this study aimed to assess the genetic diversities and structures of a sample of Bahian cacao local cultivars that were obtained from farms in different municipalities using microsatellite markers and the assessment of the physical characteristics of fruits and seeds in association with production rates. The employment of Bahian cacao in genetic breeding programs and the representative germplasms of these plants are discussed.

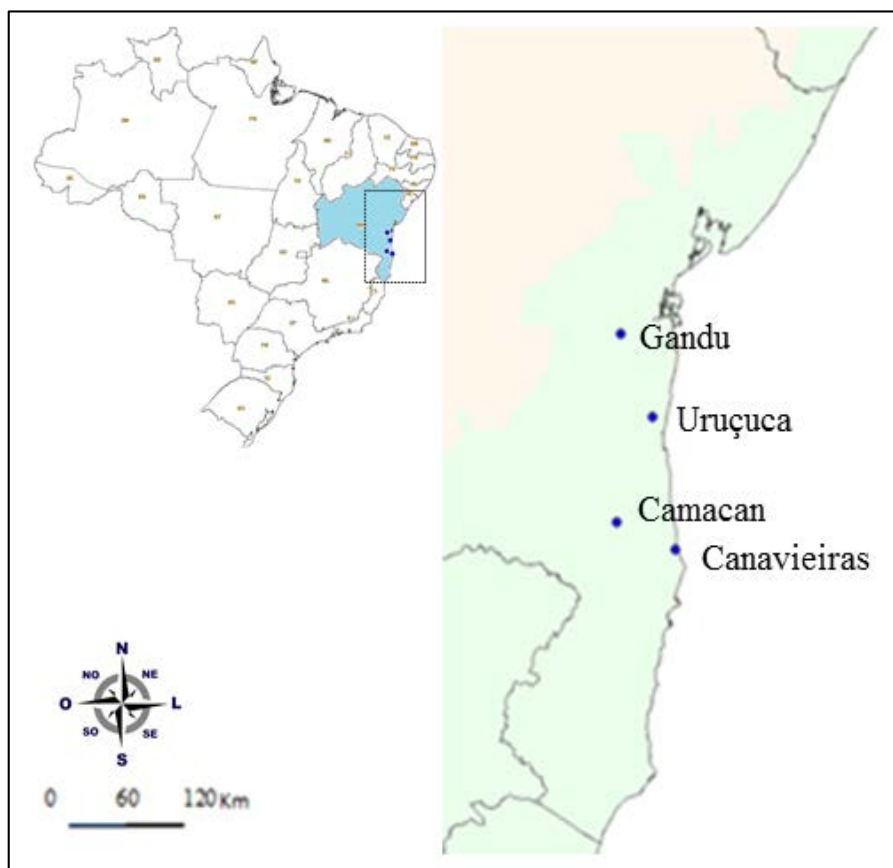
## **Materials and Methods**

### **Plant material and DNA isolation**

Young leaves of a sample of 176 plants belonging to the local cacao cultivars were obtained from nine locations in four municipalities of Southern Bahia, including three in Canavieiras, two in Uruçuca, three in Gandu and one in Camacan (Fig. 1). With the exception of the Bahian Federal Institute of Education Science and Technology of Uruçuca and Executive Committee of Cacao Farming Plan in Canavieiras, the remaining seven collection points were farms. Healthy, mature fruits of 106 genotypes were also collected between August and September of 2013 for morphological analysis.

The plants were chosen randomly according to identifications as Bahian cacao varieties, which were based on fruit characteristics and information provided by the farmers. Table 1 shows the fruit characteristics of the three main Bahian cacao varieties: Comum; Pará; and Maranhão. Currently, accurate distinction among varieties according to botanical characteristics is not always possible because hybridizations occurred among the plants at some point, and, in general, the fruits shared characteristics of two

or more types. Notes were recorded describing the fruits and other characteristics that were observed for these plants for future reference. Mutants of Bahian cacao plants were also included in this study, including one cacao jaca genotype and fifteen cacao laranja genotypes (the Portuguese terms for jackfruit and orange, respectively). The cacao jaca exhibited small, rounded leaves that are similar to those of jackfruit trees (Bartley 2005) in addition to three plagiotropic branches instead of five and abnormalities in flower formation (Figueiredo 2008). Cacao laranja is a subvariety of Para and is characterized by rounded, very small fruits similar to orange fruits. DNA from each plant was obtained using the CTAB protocol according to Faleiro et al. (2002).



**Fig. 1.** Map of Brazil indicating the state of Bahia in blue. The southern region of Bahia is highlighted, which includes the four municipalities where the cacao leaves and fruits were collected.

**Table 1.** Characteristics of Bahian local cacao varieties based on fruit features.

| Varieties | Size         | Shape                                                            | Shell              |                          |                     |
|-----------|--------------|------------------------------------------------------------------|--------------------|--------------------------|---------------------|
|           |              |                                                                  | Thickness          | Ridges                   | Texture             |
| Comum     | Medium       | Amelonado without or with slight basal constriction              | Thick              | Deep and paired          | Not rough and rough |
| Para      | Medium       | Rounded                                                          | Medium             | Superficial and unpaired | Flat and smooth     |
| Parazinho | Small        | Rounded                                                          | Thin               | Superficial and unpaired | Flat and smooth     |
| Maranhão  | Medium/Large | Amelonado, but more elongated with pronounced basal constriction | Thick - very thick | Superficial and paired   | Not rough and rough |

### **SSR markers and polymerase chain reaction (PCR) amplifications**

Twenty-nine microsatellite markers were used to genotype the plant materials (Table 2). Twenty-six of them were genomic markers developed by Santos et al. (2012) and Lanaud et al. (1999), and the remaining three were functional markers developed by Lima et al. (2008). A polymerase chain reaction (PCR) was performed in a final volume of 0.020 cm<sup>3</sup> containing 6 ng of template DNA, 1× PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 1.5 mM MgCl<sub>2</sub>, 0.2 μM of each dNTP, 0.5 μM of each primer and 1 U of Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA). The amplification reactions for each locus were conducted according to the protocols described by the authors of the respective articles. The amplification products were resolved using vertical electrophoresis on 1× TBE/6% denaturing polyacrylamide gels at 75 W for approximately 2 h and then stained with silver nitrate according Creste et al. (2001). For each run, eight DNA samples were repeated as controls. The product sizes were determined by comparison with a 10-bp DNA ladder (Invitrogen).

### **Analyses of population structure**

To assess the genetic diversities among the farm genotypes, three different approaches were used, including the Bayesian model-based approach, G<sub>ST</sub> statistics and Rogers' modified genetic distance. In the first, genetic structures were inferred using the STRUCTURE software, v.2.3.4 (Pritchard et al. 2000). The admixture ancestry model was implemented, and the burn-in periods and replications that were used were 100,000 and 400,000, respectively. The number of clusters (K) that were tested ranged from 2 to 10, and 20 independent runs were applied for each K. The Structure Harvester v.0.6.93 software (Earl and von Holdt 2012) was used to analyze the STRUCTURE output, and the most optimal K values were calculated using Evanno's methods (Evanno et al. 2005). The best alignment over the 20 runs for the optimal K values was obtained using the Cluster Matching and Permutation Program v1.1.2 (CLUMPP; Jakobsson and Rosenberg 2007), and the results were visualized using the DISTRUCT software (Rosenberg 2004). The levels of structuring among the genetic groups as indicated by STRUCTURE were also estimated using an analog of F<sub>ST</sub> (Nei's G<sub>ST</sub>) with 999

permutations and 1000 bootstraps, which was implemented with the GenAEx v.6.1 software (Peakall and Smouse 2012).

The genetic relationships among the genotypes were obtained using Rogers' genetic distance (1972) as modified by Wright (Wright 1978) from the Tools for Population Genetic Analyses (TFPGA; Miller 1997). The relationships among the genotypes were visualized using a neighbor-joining (Saitou and Nei 1987) tree function that was built into the DARwin software (Perrier and Jacquemoud-Collet 2006).

### **Polymorphism analyses**

Descriptive statistics for the genotype data were generated by estimations of the total number of alleles, the allelic frequencies at each locus and the observed and expected heterozygosities using GenAEx v.6.1 (Peakall and Smouse, 2012). These parameters were determined for the global data set, which included 176 genotypes, in addition to independent measurements for delimited groups according to the STRUCTURE software. Allelic richness was also calculated using the FSTAT software (Goudet 1995). Additionally, the levels of polymorphism in each locus that was used in this study - the polymorphic information contents (PIC) were calculated using the MTools application (Stephen Park, <http://animalgenomics.ucd.ie/sdepark/ms-toolkit/>). Calculations of deviations from Hardy–Weinberg equilibrium and fixation indices for each locus were also assessed using GenAEx v.6.1 (Peakall and Smouse 2012).

Unique allele numbers were obtained for the groups according to the STRUCTURE software. Additionally, the CoreFinder software (Cipriani et al. 2010), which adopted 99 iterations, was used to assess the minimum number of genotypes representing all of the allelic diversity that was present in the genotypes that were evaluated. For this measurement, two runs were established; the first included all 176 studied cacao genotypes, and the other was performed according the results that were reported by the STRUCTURE software.

**Table 2.** Summary statistics of allelic diversity across 30 microsatellite loci in cacao genotypes.

| Marker                                 | Na      | Ar      | H <sub>O</sub> | H <sub>E</sub> | F         | PIC    |
|----------------------------------------|---------|---------|----------------|----------------|-----------|--------|
| mTc-UNICAMP01                          | 3       | 2.99    | 0.02           | 0.03           | 0.39      | 0.03*  |
| mTc-UNICAMP02                          | 3       | 2.88    | 0.02           | 0.02           | -0.01     | 0.02   |
| mTc-UNICAMP03                          | 6       | 6       | 0.27           | 0.52           | 0.48      | 0.47*  |
| mTc-UNICAMP04                          | 8       | 7.99    | 0.17           | 0.26           | 0.34      | 0.25*  |
| mTc-UNICAMP05                          | 7       | 6.87    | 0.29           | 0.5            | 0.41      | 0.41*  |
| mTc-UNICAMP06                          | 3       | 2.9     | 0.03           | 0.03           | -0.01     | 0.03   |
| mTc-UNICAMP07                          | 4       | 3.87    | 0.03           | 0.03           | -0.01     | 0.03   |
| mTc-UNICAMP08                          | 4       | 2.9     | 0.25           | 0.38           | 0.36      | 0.32*  |
| mTc-UNICAMP09                          | 4       | 4       | 0              | 0.51           | 1         | 0.40*  |
| mTc-UNICAMP10                          | 3       | 2.98    | 0.02           | 0.02           | -0.01     | 0.02   |
| mTc-UNICAMP12                          | 5       | 4.89    | 0.03           | 0.08           | 0.66      | 0.08*  |
| mTc-UNICAMP13                          | 2       | 2       | 0.09           | 0.26           | 0.68      | 0.23*  |
| mTc-UNICAMP14                          | 3       | 2.88    | 0.02           | 0.02           | -0.01     | 0.02   |
| mTc-UNICAMP15                          | 2       | 2       | 0.04           | 0.04           | -0.02     | 0.04   |
| mTc-UNICAMP16                          | 3       | 3       | 0.03           | 0.05           | 0.36      | 0.05*  |
| mTc-UNICAMP17                          | 4       | 4       | 0.1            | 0.26           | 0.59      | 0.23*  |
| msEstTsh-1                             | 3       | 2.89    | 0.03           | 0.03           | -0.01     | 0.03   |
| msEstTsh-6                             | 4       | 3.91    | 0.04           | 0.03           | -0.01     | 0.03   |
| msEstTsh-7                             | 4       | 3.99    | 0.31           | 0.51           | 0.39      | 0.40*  |
| mTcCir 3                               | 6       | 5.74    | 0.15           | 0.42           | 0.63      | 0.35*  |
| mTcCir6                                | 4       | 3.88    | 0.03           | 0.03           | 0.16      | 0.03*  |
| mTcCir7                                | 3       | 2.95    | 0.04           | 0.04           | -0.02     | 0.04   |
| mTcCir8                                | 3       | 2.89    | 0.02           | 0.02           | -0.01     | 0.02   |
| mTcCir11                               | 4       | 3.93    | 0.13           | 0.29           | 0.57      | 0.25*  |
| mTcCir13                               | 6       | 6       | 0.11           | 0.21           | 0.49      | 0.21*  |
| mTcCir15                               | 8       | 7.68    | 0.19           | 0.3            | 0.38      | 0.28*  |
| mTcCir18                               | 2       | 2       | 0.05           | 0.05           | -0.02     | 0.05   |
| mTcCir25                               | 5       | 4.99    | 0.07           | 0.08           | 0.13      | 0.08** |
| mTcCir26                               | 3       | 2.92    | 0.02           | 0.02           | 0.24      | 0.02*  |
| <b>Mean</b>                            | 4.1     | 3.99    | 0.09           | 0.17           | 0.28      | 0.15   |
| <b>Confidence interval<sup>#</sup></b> | 3.5-4.5 | 3.4-4.5 | 0.06-0.12      | 0.11-0.23      | 0.18-0.37 |        |
| <b>Groups</b>                          |         |         |                |                |           |        |
| BC I (71)                              | 1.72    | 1.31    | 0.05           | 0.11           | 0.37      | -      |
| BC II (75)                             | 1.9     | 1.41    | 0.07           | 0.15           | 0.39      | -      |
| NB (6)                                 | 3.55    | 3.37    | 0.55           | 0.54           | -0.06     | -      |

Na, number of alleles; Ar, Allelic richness; H<sub>O</sub>, observed heterozygosity; H<sub>E</sub>, expected heterozygosity; F, fixation index; PIC, polymorphism information content.

The numbers in parentheses indicate the sample sizes for each group.

Departs significantly from Hardy-Weinberg equilibrium after Bonferroni correction. \*p < 0.05 and \*\*p < 0.001; <sup>#</sup>Confidence interval estimated by bootstrap (100,000 simulations).



### **Phenotypic evaluations**

The production features were assessed in four to ten fruits per plant from 106 genotypes of Bahian cacao. Measurements of weight, length, diameter and minimum and maximum shell thickness were performed for the mature and healthy fruits. The wet seeds were evaluated for total number and weight per fruit, weight of 10 seeds and length and width of four seeds.

Descriptive statistics was obtained for all evaluated characteristics, and relationships among the accessions were explored in terms of Mahalanobis distance using the Genes software (Cruz 2006) because this program is used to build dendrograms by the unweighted pair group method with an arithmetic mean (UPGMA). Correlations between Mahalanobis and Rogers' modified distance matrix were assessed by Mantel's correlation test with 10,000 simulations using the Genes software (Cruz 2006).

## **Results**

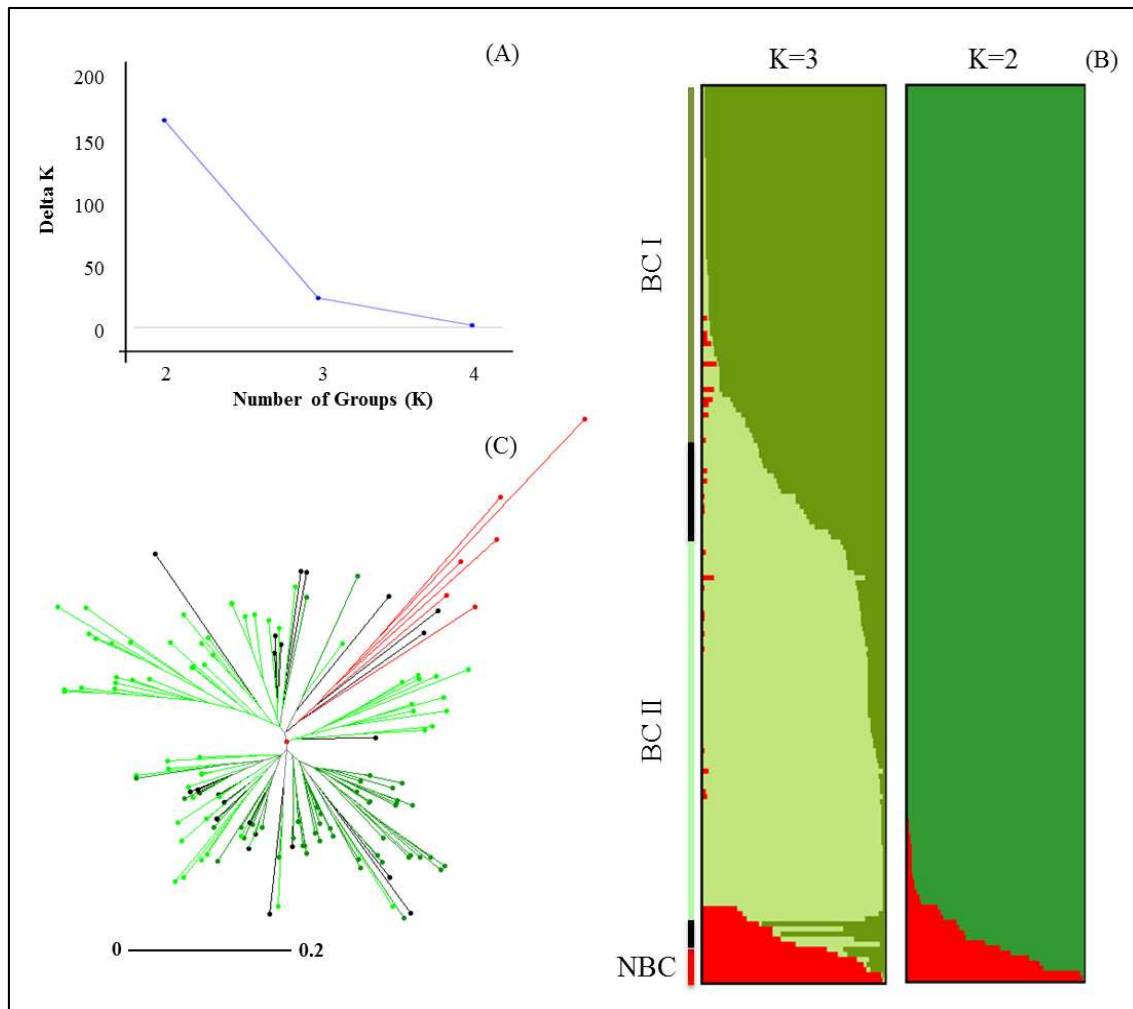
### **Relationship evaluations**

Plant material from a sample of 176 cacao plants from the Southern Bahia local cultivars were collected on farms, genotyped, and the data subjected to genetic structure analysis with a Bayesian approach using STRUCTURE software. The output was then assessed with the Structure Harvester software, which showed that the highest  $\Delta K$  values were  $K = 2$  and  $K = 3$  (Fig. 2). In this study, the plants were considered as belonging to a particular cluster if they presented with at least 70% identity (Q value) for it. Therefore, genotypes with identity values of less than 70% for a cluster were treated as having mixed ancestries. For  $K = 2$ , consensus of the 20 STRUCTURE runs that were obtained with CLUMPP and submitted to DISTRUCT revealed a homogenous and predominant cluster with 166 genotypes in contrasting to the small cluster with four genotypes (Fig. 2). The remaining plants (six) presented with mixed ancestries. Because of the high homogeneities that were obtained for the major cluster (compatible with an Amelonado origin), the genotypes were inferred as having Bahian cacao ancestry, and for  $K = 3$ , analyses revealed that the clusters were substructured into two groups

(Bahian cacao I [BC I] and Bahian cacao II [BC II]). However, the minor cluster was not considered a descendant of the Bahian varieties, and no evidence of substructuring was detected. Specific associations between farms and clusters were not clear for all of the plants, but some did show associations. Seventy-one plants formed the BC I cluster, of which, 35% were obtained from two farms in Canavieiras, which is where cacao first was introduced in Bahia. However, this cluster showed lower proportions of Gandu genotypes, in contrast to BC II, in which 53% of the plants from this location represented half of the cluster (52%) that was formed by 75 plants in total. One Gandu farm was not represented in the BC I cluster, and 92% of the plants from this farm were present in BC II. This group also contained 87% of the laranja genotypes in addition to the representative jaca varieties.

The levels of structure between the groups as detected by the aforementioned Bayesian approach were quantified using Nei's  $G_{ST}$ , and relatively high values were observed between the Bahian cacao clusters themselves (BC I x BC II = 0.221) and between each Bahian cluster and that of the non-Bahian cacao (NB; Table 3). For the latter, a higher value was observed in comparison with Bahian cacao I (BC I x NB = 0.245), which was also high in comparison with Bahian cacao II (BC II x NB = 0.215).

The diversities between the groups that were observed by the neighbor-joining tree shared some similarities with the STRUCTURE results. However, additional tendencies to form substructures that reflected the Maranhão variety plants could clearly be observed in the neighbor-joining tree.



**Fig. 2.** Groupings obtained by analyses of 176 cacao genotypes and 29 microsatellite markers. **(A)** The graph depicts the most probable cluster numbers (K), **(B)** histogram for K = 2 and K = 3, **(C)** neighbor-joining tree built by Rogers' modified genetic distance measurements. In the histograms and dendrogram, the color green represents plants of Bahian cacao ancestry, red represents plants of non-Bahian cacao ancestry, and black indicates those considered to be of mixed ancestry, as indicated by the STRUCTURE software.

**Table 3.** Genetic structures as assessed by Nei's  $G_{ST}$  (lower diagonal) between groups using STRUCTURE software (Pritchard et al. 2000) with  $p < 0.001$ .

|       | BC I  | BC II |
|-------|-------|-------|
| BC II | 0,221 |       |
| NBC   | 0,245 | 0,215 |

*BC* Bahian cacao *NBC* Non-Bahian cacao

### Polymorphism analyses

Polymorphisms were assessed for the global data and for groups that formed according to the STRUCTURE software, which are shown in Table 2. The 29 markers were polymorphic, revealing a total of 119 alleles. The number of alleles per locus and allelic richness presented with a minimum of 2 and maximum of 8 and 7.99, respectively. The means for these parameters were also similar, including 4.1 alleles per locus and 3.99 for allelic richness. Small values for the observed ( $H_0$ ) and expected ( $H_E$ ) heterozygosities were detected in almost all of the loci with averages of 0.089 and 0.174 for the  $H_0$  and  $H_E$ , respectively. The aforementioned small values for the polymorphisms in the global data were largely influenced by the Bahian cacao genotypes. Thus, the diversity parameters that were evaluated for BC I (allele richness of 1.31,  $H_0$  of 0.054 and  $H_E$  of 0.106) and BC II (allele richness of 1.41,  $H_0$  of 0.067 and  $H_E$  of 0.149) were very low, mainly in comparison with the non Bahian cacao group (allele richness of 3.37,  $H_0$  of 0.546 and  $H_E$  of 0.539).

For the 176 genotypes, the PIC values ranged from 0.02 for mTcCir8 and mTc-UNICAMP02, 10 and 14 to 0.47 for mTc-UNICAMP03 with a mean of 0.15 (Table 2). Testing for Hardy–Weinberg equilibrium indicated that most of the loci were in disequilibrium at  $p < 0.05$  (18 loci) and  $p < 0.001$  (17 loci) after Bonferroni corrections.

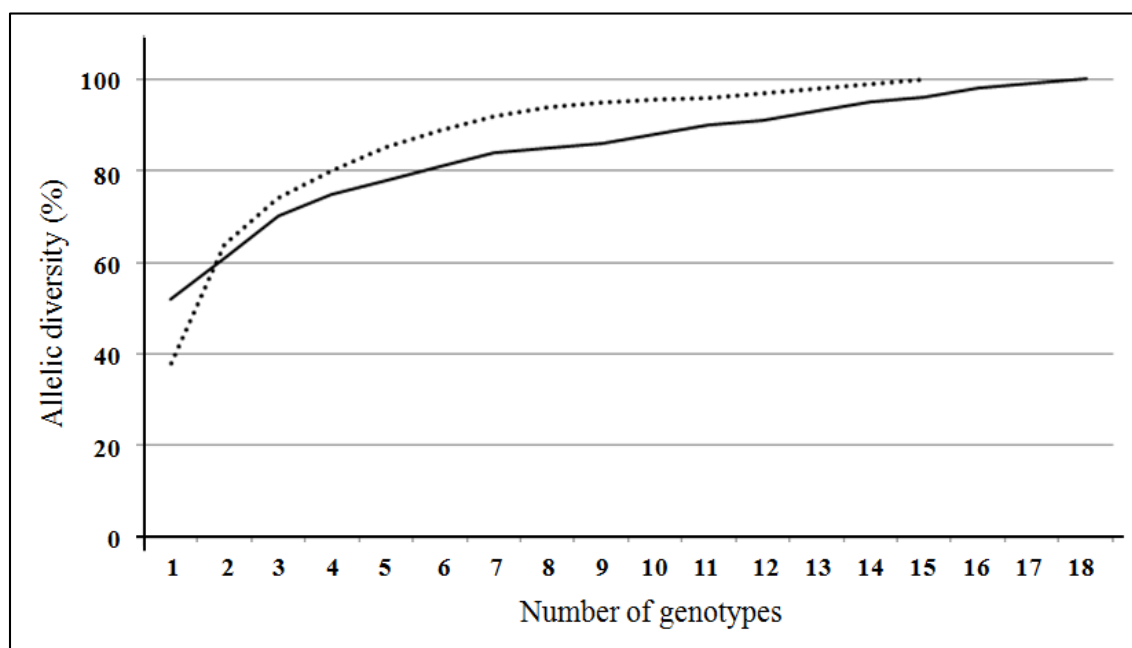
The positive fixation index mean value (0.280) indicated the deficiency of heterozygotes in the total analyzed data. This mean was surpassed for the Bahian cacao groups that presented with values of 0.365 and 0.393 for BC I and BC II, respectively. In contrast, for the NBC, a negative mean of -0.056 was observed. Lower

fixation index values for the non Bahian cacao are opposite of those of the Bahian cacao, indicating the influences of the latter on the global data with regard to the paucity of the heterozygotes that were observed.

Private alleles were assessed using two different analyses, including one for all of the cacao groups that were appointed by the Structure software and the other for only the Bahian cacao groups (data not shown). In the first analysis, 55 alleles were obtained, of which, 48 were from the non-Bahian cacao groups. Out of the seven alleles from the Bahian cacao groups, three were present in BC I and four in BC II, and only one (present in BC II) was present in over 5% of the plants. If only Bahian cacao plants from different groups were assessed, the number of unique alleles that were observed rose to 10 and 15 for BC I and BC II, respectively, and out of the seven alleles that were present in at least 5% of the genotypes, six originated in BC II.

To assess the maximum allelic diversity of the samples from a minimum number of genotypes, a graph was constructed using CoreFinder software, and the results are shown in Fig. 3. From the 176 genotypes, at least 15 individuals were necessary to sample 119 alleles. Thus, it is important to note that most of the individuals (nine genotypes) were not part of the Bahian cacao groups, and therefore, comprising plants with 70% of non-Bahian cacao ancestry (five genotypes) or of mixed ancestry, constituting more than 30% of this group (four genotypes). Moreover, the five genotypes of non-Bahian cacao ancestry were mainly responsible for the core diversity, possessing 86% of the allelic variations, which corresponded to 103 alleles.

A core collection was also built considering only plants of Bahian cacao ancestry. Therefore, genotypes were considered that corresponded with each group of Bahian cacao and those with at least 70% of mixed ancestry between them. From the 164 plants that were used in this analysis, 70 alleles were observed and sampled from 18 genotypes. Plants from this core were collected from different farms, and seven were obtained from two farms in Canavieiras. The low number of sampled alleles and higher genotype numbers most likely reflect the low genetic diversities of these plants in addition to their high homozygosities.



**Fig. 3.** Minimum number of clones to sample the allelic diversity considering the 176 cacao plants evaluated using microsatellite markers (dotted line) and only considering the 164 plants classified with 70% (q values) of the Bahian cacao I and/or Bahian cacao II groups using the STRUCTURE software (Pritchard et al. 2000) (continuous line)

### Phenotypic evaluation

Descriptive results of the phenotypic evaluations are shown in Table 4. The lower coefficient of variation of 5.27% and the higher value of 19.37% were observed for fruit width and seed weight per fruit, respectively. Correlations were positive and highly significant between most variables (Table 5) with the exception of the number of seeds per fruit, which was not significantly correlated with the fruit length and width, seed length and width or weight of 10 seeds per fruit. Furthermore, fruit length was not correlated with fruit or seed width, and shell thickness was not correlated with total seed weight per fruit or seed width. With the exception of shell thickness and seed number per fruit, the significant correlations that were observed were positive and varied from 0.23 to 0.79 for shell thickness and fruit length and for weight of 10 seeds per fruit and seed width, respectively.

**Table 4.** Descriptive statistics obtained from physical analyses of fruits and seeds for 106 cacao genotypes collected from Bahian farms.

|                          | Fruit               |                     |                    |                              | Wet seed                   |               |                                 |                     |                    |
|--------------------------|---------------------|---------------------|--------------------|------------------------------|----------------------------|---------------|---------------------------------|---------------------|--------------------|
|                          | Weight <sup>1</sup> | Length <sup>2</sup> | Width <sup>2</sup> | Shell thickness <sup>3</sup> | Weight/ fruit <sup>1</sup> | Number/ fruit | Weight of 10 seeds <sup>1</sup> | Length <sup>3</sup> | Width <sup>3</sup> |
| Minimum                  | 304.44              | 9.05                | 7.59               | 9.24                         | 46.41                      | 22.63         | 11.81                           | 16.25               | 10.56              |
| Maximum                  | 871.33              | 20.8                | 10.08              | 17.74                        | 163.78                     | 49.5          | 24.23                           | 27.47               | 15.43              |
| Mean                     | 451.65              | 14.26               | 8.51               | 13.7                         | 96.8                       | 38.25         | 15.76                           | 24.04               | 12.44              |
| Standard Error           | 7.85                | 0.16                | 0.04               | 0.15                         | 1.82                       | 0.52          | 0.22                            | 0.16                | 0.07               |
| Coefficient of variation | 17.90%              | 11.42%              | 5.27%              | 11.36%                       | 19.37%                     | 14.11%        | 14.08%                          | 6.63%               | 5.96%              |

Measurements in <sup>1</sup>grams, <sup>2</sup>centimeters and <sup>3</sup>millimeters

**Table 5.** Correlation matrix between fruit and seed variables evaluated in 106 cacao genotypes collected from Bahian farms.

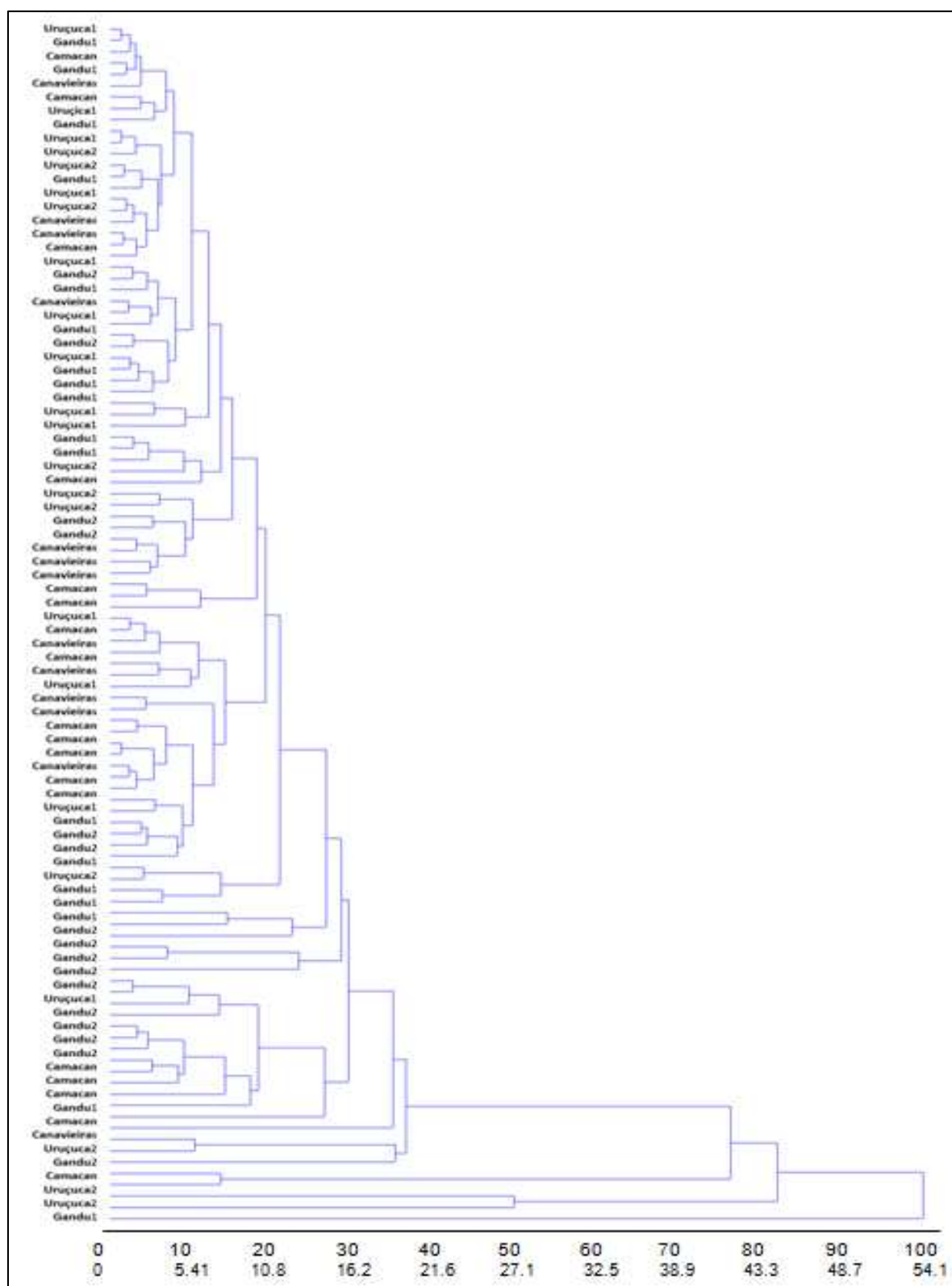
|                 |                    | Fruit   |         |         |                 | Wet seed      |               |                    |         |
|-----------------|--------------------|---------|---------|---------|-----------------|---------------|---------------|--------------------|---------|
|                 |                    | Weight  | Length  | Width   | Shell thickness | Weight/ fruit | Number/ fruit | Weight of 10 seeds | Length  |
| <b>Fruit</b>    | Length             | 0.57*** |         |         |                 |               |               |                    |         |
|                 | Width              | 0.75*** | 0,08    |         |                 |               |               |                    |         |
|                 | Shell thickness    | 0.62*** | 0.23*   | 0.56*** |                 |               |               |                    |         |
|                 | Total weight/fruit | 0.68*** | 0.35*** | 0.51*** | 0,07            |               |               |                    |         |
| <b>Wet seed</b> | Number/fruit       | 0.24**  | 0,16    | 0,16    | -0.22*          | 0.64***       |               |                    |         |
|                 | Weight of 10 seeds | 0.55*** | 0.33*** | 0.48*** | 0.29**          | 0.55***       | -0,02         |                    |         |
|                 | Length             | 0.48*** | 0.26*** | 0.40*** | 0.29**          | 0.52***       | 0,10          | 0.62***            |         |
|                 | Width              | 0.30*** | 0,17    | 0.31*** | 0,07            | 0.47***       | 0,12          | 0.79***            | 0.47*** |

Significant at \*p < 0.05, \*\*p < 0.01 and \*\*\*p < 0.001

Analyses including all 106 plants using Mahalanobis distance as visualized in the UPGMA dendrogram revealed a reduced distance between individuals (Fig. 4). The UPGMA figure showed two grouping trends, including one that was related to the region of collection and another that was related to the type of variety that was identified. In the first, a group was formed only by plants from Camacan and Canavieiras, which are the southernmost municipalities of Sothern Bahia. In the second, plants that were collected from a Gandu farm and classified as 'pure' Maranhão were grouped together.

A low correlation value between the genotyping and phenotyping distance matrix (built using Rogers' modified and Mahalanobis distances, respectively) of 0.18 was observed. For the Mantel correlation with 10,000 simulations, this value was lower, being equal to 0.12.





**Fig. 4.** Mahalanobis dendrogram built from physical fruit and seed characteristics obtained from 106 genotypes from Bahian farms.

## Discussion

In this study, molecular markers and agronomic characteristics were employed to evaluate a sample of Sothern Bahia local cacao varieties from the Amelonado group. Amelonado Bahian varieties have been cultivated around 260 years, and during this time, Bahia came to be the second most important cacao producer worldwide. Although the outbreak of witches' broom disease in 1989 forced the replacement of the local varieties with more resistant clones, they continue to be cultivated, particularly by smallholders. These local cultivars are very appreciated by the farmers and currently occupy half of the planting areas.

### Molecular diversity and structure

The SSR markers used in this study effectively assessed the diversities and structures of the cacao plants sampled in Sothern Bahia. Despite the low levels of genetic diversity observed, they could be structured by a Bayesian approach into two and three genetic groups, which were in both cases one group formed by plants considered to be of non-Bahian varieties. The low levels of diversity that were revealed here contrast with previous reports for most of the cacao groups (Motamayor et al. 2008; Bartley 2005; Bennet 2003) but are in agreement with reports involving traditional cacao cultivars, such as Amelonado from the Lower Amazon region (Aikpokpodion et al. 2009; Motamayor et al. 2002). Although it is not clearly delimited, the Lower Amazon is designated by the Pará region, which is a Brazilian state that extends from Marajó Island to the Maicuru River, which is a tributary of the Amazon River (Bartley 2005). The origin of the wild plants in the Lower Amazon is uncertain, but it is possible that some populations originated from the limited introductions of plants from the Upper Amazon (Sereno et al. 2006). The Upper Amazon area has not been very well established but is largely used as a diversity center for cacao plants. This fact could explain the genetic proximities between the Upper and Lower Amazon plants as detected by Sereno et al. (2006) in populations from the Brazilian Amazon, in addition to the narrow genetic base of the varieties that were derived from Amelonado.

Cacao was introduced to Southern Bahia in the eighteenth and nineteenth centuries, including the Comum, Pará and Maranhão varieties, and although these plants were thought to have been brought from Pará, their precise origin is uncertain, and the limited introductions may have potentiated the intrinsic low levels of diversity that are observed in the Amelonado group. In Bahia they were planted in large scale over all the cultivated area (Monteiro et al. 2009) and new types have naturally emerged by mutation or meiotic events originating many varieties, such as laranja cacao, Catongo and Almeida. Laranja cacao is derived from the Pará variety and characterized by small fruits and seeds; Catongo and Almeida varieties, which were discovered in 1930 and possess white seeds and are considered of good quality for chocolate production (Bondar 1938); and many others (Bartley 2005)

In addition to their restricted origins, the self-compatibility of Amelonado cacao plants has led to high levels of homozygosity and homogeneity. This fact was demonstrated in this study by the presence of lower  $H_E$  values and higher positive fixation index values for all of the plants that were analyzed, which was observed to be intensified in the Bahian groups. For this reason, the groups that escaped the expected diversity patterns were considered to be of different origins than the plants that were naturalized as Bahian cacao, indicating these plants may have suffered introgression of alleles from other sources of germplasm during the cultivation process. Although differences among the original Bahian cacao varieties that were employed in this study were based mainly on fruit characteristics, which were clearly defined (Table 1), the artificial and natural outcrosses that occurred over time among themselves and with other plants that were introduced makes it potentially difficult to precisely distinguish each Bahian cacao variety and to discriminate them from the non-Bahian cacao plants. However, some of the plants were easily distinguished, indicating lower levels of hybridization and introgression of alleles between the different varieties. For some of the plants that were classified as 'pure' Maranhão, their grouping on a unique branch of the neighbor-joining tree suggests the possibility of separate Bahian cacao varieties using SSR markers. For other plants, such as laranja cacao, although they were identified relatively easily due their orange fruits, no associations could be made between their morphological characteristics and grouping from the molecular analyses.

The narrow genetic base of the Bahian cacao cultivars was also reflected in the core collection, which was mainly composed of plants that were not targets of this study. In contrast, the presence of private alleles in the Bahian groups, although small in quantity, indicated the specificity of this group with regard to the other plants that were evaluated. A substructure that was observed in the Bahian cacao groups also contributed to the preservation of important variations in these plants.

### **Phenotypic characterizations**

The phenotypic characterizations employed in this study are important both for the elucidation of variations that are present in cacao genotypes from farms, which are recognized as diversity reservoirs (Aikpokpodion 2010), and for the in-depth evaluation of a select number of plants that may be potentially used in breeding or prebreeding programs. Although the features that were evaluated here are of agronomic importance, they are recommended over other botanical characteristics for the classification of cacao plants.

Interesting data were revealed in this study with regard to the economic potentials of the plants. For example, the mean number of seeds per fruit (38.3) was similar to that observed for cultivated clones and reported in the International Cacao Genebank (ICG) of Trinidad ([mean = 38.8]; Bekele et al. 2006) and was lower than the mean observed for cultivated plants from Nigerian farm fields ([39.5]; Aikpokpodion 2010). The overall weight of the seeds that were obtained (1.5 g) was lower than the 1.7 g that was reported by the Nigerian farms. These two features are very important for the cacao market because they directly influence yields and chocolate production. Homogeneous seed sizes are required to ensure uniform roasting and avoid the burnt flavoring of the chocolate (Dias 2001), with a minimum allowed size of at least 1 g.

Although correlations only exceeded 70% for two of the pairwise variables (0.75 for fruit weight and width and 0.79 for seed width and weight of 10 seeds per fruit), the significant and positive correlations between the characteristics facilitates breeding because the choice of one feature does not negatively affect the other. Solely negative variables that were significantly correlated (-0.22) were also of interest because the selection for thin shell, for example, tends to not affect the number of seeds per fruit.

The low Mantel correlation that was observed between the Mahalanobis and Rogers' genetic distance was not surprising because the molecular markers that were used were not selected from regions related to genes. However, we did observe some plants from the core collection that presented with optimal phenotypic features, and it can thus be useful to also choose plants with allelic variations to potentially be used in breeding programs.

### **Bahian local cacao cultivars: conservation and breeding**

The cultivation of cacao in Southern Bahia has been of great economic and historical importance because this crop has supplied the base of the economy of the region since 1800s and helped to shape the social and cultural aspects of the local people. The identification of witches' broom in 1989 associated with the reduction of the cocoa prices in the international market forced the region to reorganize the system of cacao cultivation. In some farms, cacao plants have been completely replaced by other crops or transformed in consortia with valuable plants. Over the years, there is an increased tendency to replace Bahian cacao cultivars with clonal cultivars or modern cultivars. The modern cultivars were mainly derived from selections made on farms and based upon resistance to witches' broom and production rate. Analyses of microsatellite markers in these clonal cultivars (data not shown) have indicated that there are differences between them and the Bahian varieties. If, on one hand, the cultivars that are currently recommended for planting are more diverse among themselves, this diversity not encompasses those present in Bahian cacao, indicating that some alleles own by the latter group may not be represented in the modern material. In Nigeria, a similar situation was reported by Aikpokpodion (2010), in which the local plants, which are known as African Amelonado (descendants of the local varieties from Bahia), have been replaced by plants that are more resistant to swollen shoot virus disease, which cause great damage to plantations in this country.

Evaluations of accessions from the series SIC (Selection of the Cacao Institute) and SIAL (Selection of the Cacao Institute of East) maintained at CEPLAC's germplasm bank indicated that although SIC and SIAL are susceptible to witches' broom, they presented variability in terms of fruit and seed features and were remarkable in terms of

productivity (Pires 2003). These series were selected in the 1930s and 1940s and are considered to represent the spectra of diversity of the Bahian local cacao cultivars. It is possible that the genetic variability sampled for this present study includes alleles from these series once some accessions were used to produce seeds to deliver to farmers. Indeed, the cacao introduced in the eighteenth and nineteenth century in Bahia has been submitted to a continues process of selection by farmers and local governmental organizations, therefore the present variability represent the result of the action of deliberated selection for important fruit features, production, adaptation to different environments of the region.

Even though local cultivars are susceptible to witches' broom disease they are still being planted by some farmers because they appreciate some features these cultivars have. In general, all cultivars present high productivity considering the level of management provided by farmers. The Maranhão and Catongo cultivars have gained prominence in the international fine chocolate market because of their fine flavor and quality. Considering that the striking effects of witches' broom will persist, the current scenario of Bahian cacao production indicates that most of the local varieties will be replaced by the modern cultivars. Therefore, specific materials like the Maranhão cacao from Gandu as well as from other regions may disappear, representing genetic erosion of this important source of germplasm.

It is likely that the representativeness of Bahian cacao in germplasm banks are underestimated because the SIC and SIAL selections were obtained from a restricted sampling area (Vello et al. 1972), and their collections were focused on optimal features for genetic improvement.

Thus, considering that the farms are important reservoirs of genetic diversity, the combined analyses of farm-derived and germplasm plants of the Bahian varieties may allow for a clearer picture of the variability that truly characterizes these plants. In addition to its utility for the genetic improvement of the species, this collection can assist in further studies that seek insight into the Amazon populations from which Bahian cacao originated.

## References

- Aikpokpodion PO, Motamayor JC, Adetimirin VO, Adu-Ampomah Y, Ingelbrecht I, Eskes AB, Schnell RJ, Kolesnikova-Allen M (2009) Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genetics and Genomes* 5: 699–711.
- Aikpokpodion PO (2010) Variation in agro-morphological characteristics of cacao, *Theobroma cacao* L., in farmers' fields in Nigeria. *New Zealand Journal of Crop and Horticultural Science*, 38: 157-170.
- Ahnert D, Ferreira ACR. Melhoramento Genético e Variedades do Sul da Bahia. In: Mello DLN, Gross E (org) Guia de Manejo do Agroecossistema Cacau Cabruca. Vol. 1: Instituto Cabruca, Ilhéus: 24-27, 2013.
- Araújo M, Alger K, Rocha R, Mesquita CAB (1998) A Mata Atlântica do sul da Bahia: situação atual, ações e perspectivas. Reserva da Biosfera da Mata Atlântica – MAB – UNESCO. Caderno 8:1–36
- Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI Publishing, Wallingford.
- Bekele FL, Bekele I, Butler DR, Bidaisee GG (2006) Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genetic Resources and Crop Evolution* 53: 933-948.
- Bennett AB (2003) Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Science* 8: 561-563.
- Bhattacharjee R, Kumar PL. Cacao. In: Kole C (ed) Genome Mapping and Molecular Breeding. Vol. 7: Technical Crops, Springer, Heidelberg, Berlin, New York, Tokyo. Pp. 127-142, 2007.
- Bondar GA. cultura do cacau na Bahia. Boletim técnico do Instituto de Cacau da Bahia. n. 1, p. 7-39. Salvador: Instituto de Cacau da Bahia, 1938.
- Cheesman E (1944) Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture* 21: 144– 159.

Cipriani G, Spadotto A, Jurman I, Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A, Testolin R (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages and reveals a large admixture amongst varieties of different geographic origin. *Theor Appl Genet* 121:1569-1585.

Clement CR, Cristo-Araújo M, d'Eeckenbrugge GC, Pereira AA, Picanço-Rodrigues D (2010) Origin and Domestication of Native Amazonian Crops. *Diversity* 2: 72-106

Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Molecular Biology Report* 19: 299–306.

Cruz CD. Programa GENES: biometria. Viçosa: UFV, 2006. 382p.

DIAS, L.A.S. Melhoramento genético do cacauzeiro. Viçosa: Funape, 2001. 578p.

Earl DA, vonHoldt BM (2012) Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.

Faleiro FG, Pires JL, Monteiro WR, Lopes UV, Yamada MM, Piedra AG, Moura AD, Arévalo-Gardini E, Marques JRB, Gramacho KP, Faleiro ASG, Santos MCM (2004) Variability in cacao accessions from the Brazilian, Ecuadorian, and Peruvian Amazons based on molecular markers. *Crop Breeding and Applied Biotechnology* 4: 227-233.

Faleiro FG, Santos IS, Bahia RCS, Santos RF, Yamada MM, Anhert D. (2002) Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando obtenção de marcadores RAPD. *Agrotrópica* 14: 31-34.

Faria D, Paciencia MLB, Dixo M, Laps RR, Baumgarten J (2007) Ferns, frogs, lizards, birds and bats in forest fragments and shade cacao plantations in two contrasting landscapes in the Atlantic forest, Brazil. *Biodiversity Conservation* 16: 2335–2357.

Figueiredo GSF, 2008Análise Cariotípica em cacauzeiro (*Theobroma cacao* L.) e em mutantes espontâneos.Tese de Doutorado. Ilhéus, UESC, XX p.



Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. J Hered 86: 485-486.

IBGE - Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal, 2012. [http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam\\_2012\\_v39\\_br.pdf](http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam_2012_v39_br.pdf)

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJJ (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Molecular Ecology 8: 2141–2152.

Leite PB, Maciel LF, Opretzka LCF, Soares SE, Bispo ES (2013) Phenolic compounds, methylxanthines and antioxidant activity in cocoa mass and chocolates produced from “witch broom disease” resistant and non resistant cocoa cultivars. Ciência e agrotecnologia 37: 244-250.

Lima LS, Gramacho KP, Gesteira AS, Lopes UV, Gaiotto FA, Zaidan HA, Pires JL, Cascardo JCM, Micheli F. (2008) Characterization of microsatellites from cacao–*Moniliophthora perniciosa* interaction expressed sequence tags. Molecular Breeding 22: 315–318.

Miller MP (1997) Tools for Population Genetic Analyses (TFPGA) 1.3: A Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

Monteiro WR, Lopes UV, Clement D (2009) Genetic Improvement in Cocoa. In: S.M. Jain, P.M. Priyadarshan (eds.). Breeding Plantation Tree Crops: Tropical Species. Springer: 589-626.

Motamayor JC, Lachenaud Pe, Mota JW, Llor R, Kuhn DN, Brown JS, Schnell RJ. (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). Plos One 3: e3311.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity 89: 380-386.

Motilal LA, Zhang D, Mischke S, Meinhardt LW, Umaharan P (2013) Microsatellite-aided detection of genetic redundancy improves management of the International Cacao Genebank, Trinidad. *Tree Genetics and Genomes* 9:1395–1411.

N’Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000) The genetic structure of cacao populations (*Theobroma cacao* L.) revealed by RFLP analysis. *Euphytica* 115: 83–90.

Peakall R, Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28:2537-2539. doi:10.1093/bioinformatics/bts460.

Pereira JL, Ram A, Figueiredo JM, Almeida LCC (1989) Primeira ocorrência de vassoura-de-bruxa na principal região produtora de cacau do Brasil. *Agrotrópica* 1: 79-81.

Perrier X, Jacquemoud-Collet JP (2006). DARwin software <http://darwin.cirad.fr/>

Pires JL (2003) Avaliação Quantitativa e molecular de germoplasma para o melhoramento do cacaueiro com ênfase na produtividade, qualidade de frutos e resistência a doenças. Tese de Doutorado. Viçosa, UFV, 342 p.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945-959.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4: 137–138.

Sambuichi RHR (2006) Estrutura e dinâmica do componente arbóreo em área de cabruca na região cacaueira do sul da Bahia, Brasil. *Acta botanica brasiliensis* 20: 943-954.

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425

Santos ESL, Cerqueira-Silva CBM, Mori GM, Ahnert D, Corrêa RX, Souza AP. (2012) New polymorphic microsatellite loci for *Theobroma cacao* L.: Isolation and characterization of microsatellites from enriched genomic libraries. *Biologia Plantarum* 56: 789-792.

Sereno ML, Albuquerque PSB, Vencovsky RAF, 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.

Trognitz B, Scheldeman X, Hansel-Hohl K, Kuant A, Grebe H, Hermann M (2011) Genetic Population Structure of Cacao Plantings within a Young Production Area in Nicaragua. *PLoS ONE* 6: e16056.

Vello F, Garcia JR (1971) Características das principais variedades de cacau cultivadas na Bahia. *Theobroma* 1:3-10, 1971.

Vello F, Garcia JR, Magalhães WS (1972) Produção e Seleção de Híbridos na Bahia. *Theobroma* 2:15-35.

Wright S (1978) *Evolution and the genetics of populations*, vol.4. Variability within and among natural populations. University of Chicago Press, Chicago. 580pp.

Young AM. *The chocolate tree: a natural history of cacao*. UPF, Gainesville, 2007



## **CAPÍTULO III**

### **Diversidade genética de cacaueiros da Bahia II**

**Artigo em processo de submissão**

#### **Molecular diversity and genetic structure of Amelonado cacao from the Bahia, Brazil**

**Santos, E.S.L.**

Cerqueira-Silva, C.B.M.

Mori, G.M.

Corrêa R.X.

Pires, J.L.

Ahnert, D.

**Souza, A.P.**



## Molecular diversity and genetic structure of Amelonado cacao from the Bahia, Brazil

Elisa Susilene Lisboa dos Santos<sup>1,2</sup>, Carlos Bernard Moreno Cerqueira-Silva<sup>1,2</sup>,  
Gustavo Maruyama Mori<sup>2</sup>, Ronan Xavier Corrêa<sup>3</sup>, José Luis Pires<sup>4</sup>, Dário Ahnert<sup>3</sup>,  
Anete Pereira de Souza<sup>2,5</sup>.

Departamento de Ciências Exatas e Naturais, Universidade Estadual do Sudoeste da Bahia, Rodovia BR415, Km03, s/n, Itapetinga, 45700-000, Bahia, Brazil<sup>1</sup>.

Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, CP 6010, Campinas, 13083-875, São Paulo, Brazil<sup>2</sup>.

Centro de Biotecnologia e Genética, Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna Km 16, Ilhéus, 45662-900, Bahia, Brazil<sup>3</sup>.

Centro de Pesquisas do Cacau, Comissão Executiva do Plano da Lavoura Cacaueira, Rodovia Ilhéus-Itabuna Km 22, Itabuna, 45600-970, Bahia, Brazil<sup>4</sup>.

Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas, 13083-970, São Paulo, Brazil<sup>5</sup>.

### Abstract

Bahia is the most important cacao-producing state in Brazil, where cultivation was initiated in the eighteenth century when the first seeds were introduced from the Lower Amazon region and planted in this southern state. The plants that were obtained from this introduction and from another introduction in the twentieth century formed Bahian Amelonado cacao, which is cultivated in Bahia. In the 1930s and 1940s, selections were performed in farmers' fields in Bahia, generating the SIC and SIAL clones, respectively. In general, these selections have low genetic diversity, but no studies of codominant markers that are specific for these clones have been reported. With the aim of characterizing the genetic diversity and identifying the population substructure, we genotyped the SIC and SIAL clones along with reference clones (totaling 103 clones) using 30 simple sequence repeat (SSR) markers. The results confirmed the low genetic diversity of the Bahian clones (mean alleles per locus = 2.60, observed and expected heterozygosities of 0.116 and 0.218, respectively) compared to the reference clones (mean alleles per locus = 6.67,  $H_0 = 0.502$  and  $H_E = 0.603$ ). In addition, high positive fixation index values for the SIC and SIAL clones were observed, indicating inbreeding, which is typical for these self-compatible clones. Substructure levels were revealed using clustering strategies based on traditional and Bayesian methods to separate the clones. Despite the low genetic diversity, Bahian clones remain important sources of alleles both for conservation and for use in genetic breeding programs and have important economic features.

**Keywords:** *Theobroma cacao*. Cacao improvement. Germplasm. Conservation. Genetic groups.

## Introduction

*Theobroma cacao* ( $2n = 20$ ) is a neotropical, perene and allogamous tree that is cultivated in approximately 60 countries around the world (<http://faostat3.fao.org>). Despite its origin in the Amazon (Cheesman 1944), cacao was domesticated by the Mayans in Central America (Motamayor et al. 2002) approximately 3,000 years ago and has become one of the world's most important commodities, especially for chocolate production.

Approximately 90% of cacao beans are produced by smallholders, mostly less than 3 hectares per producer (Lopes et al. 2011). Currently, West African countries are the most notable producers of cacao beans worldwide (approximately 70%). Brazil is the sixth-largest cacao-producing country, although it was the second-largest producer in the world (<http://faostat3.fao.org>) until the occurrence of witches' broom (caused by the basidiomycete *Moniliophthora perniciosa*) in the late 1980s in the state of Bahia.

Cacao breeding is an important strategy for reducing production costs and improving economic features, such as bean quality, disease resistance and productivity. To this end, there has been an international effort to collect and preserve cacao germplasm to offer genetic variability in quantity and quality to breeders. Approximately 5,000 accessions of cacao are conserved in worldwide germplasm banks as living trees because of the recalcitrant nature of its seeds (Guiltinan et al. 2008). These accessions include clones that are derived from 10 cacao groups (Marañon, Curaray, Criollo, Iquitos, Nanay, Contamana, Amelonado, Purús, Nacional and Guiana) that have been defined by Motamayor et al. (2008).

The cacao-growing areas in Bahia, the main cacao-producing region of Brazil, were established initially through a limited introduction of genetic materials that were derived from the Lower Amazon region starting in 1746 (Bondar 1938). The fruits of the Lower Amazon plants are characterized by a melon shape; therefore, they are denominated as Amelonado. A second introduction, also from the Amazon, occurred approximately 100 years later (Bartley 2005). The two introductions formed the genetic base of the Amelonado cacao plantations in Bahia and Espírito Santo, other Brazilian state. In the 1930s and 1940s, the first selections were performed in farmers' fields in



Bahia to obtain more productive clones, resulting in the SIC and SIAL series, respectively, in Bahia and EEG in Espírito Santo (Monteiro et al. 2009). These clones are considered representatives of the genetic materials that were introduced and disseminated throughout the Bahian cacao plantations (Pereira et al. 1987; Yamada et al. 2001). The SIC and SIAL clones were initially subject to phenotypic evaluation for floral, fruit, leaf and seed component characteristics (Castro and Bartley 1983; 1985), which detected good productivity and variability for some traits. This initial evaluation was performed to identify the best clones for introduction into the cacao genetic breeding program of the Cocoa Research Center (CEPEC) of the Executive Committee of the Plan of Cacao Farming (CEPLAC). Thus, the SIC and SIAL clones, along with clones that were introduced from other countries, were used in breeding to generate hybrid cultivars with high productivity, resistance to *Phytophthora palmivora* and good bean quality for industry (Vello et al. 1972). However, with the outbreak of witches' broom in Bahia in 1989, the development of hybrid cultivars stopped and gave way to the selection and breeding of resistant clonal cultivars through recurrent selection or the selection of resistant plants in the farmers' fields (Monteiro and Ahnert 2007).

Although SIC and SIAL are important genetic resources for breeding of Bahian cacao, investigations into their molecular genetic diversity are restricted to the use of dominant markers (Cascardo et al. 1993; Yamada et al. 2001), which allow for the simultaneous evaluation of multiple loci but mask the allelic diversity at each locus, or to the use of a few representative clones (Johnson et al. 2009; Irish et al. 2010). Obtaining information on molecular genetic variability is important for practical ex situ conservation and is considered a crucial step in pre-breeding programs with any species.

The objectives of this study were 1) to assess the genetic diversity of the first cacao clones that were obtained in Bahia using microsatellite markers and 2) to characterize the genetic relationships of these clones with other cultivated and wild clones. The substructure of the Bahian cacao clones and their use in breeding programs are also discussed.

## **Materials and Methods**

### **Plant material and DNA isolation**

A total of 103 cacao clones were used (Table 1). Young leaves were collected from an institution and a farm in southern Bahia, Brazil: the CEPEC/CEPLAC and the Farms of Juliana's Valley in the respective municipalities of Ilhéus and Igrapiúna. Fifty-one samples comprised clones representing the first selections of cacao in Bahia, which were obtained from the original plants that founded the cacao production areas in this region. Of these clones, 31 were selected from production fields by the former Cacao Institute of Bahia (currently the Bahian Federal Institute of Education, Science and Technology) in the 1930s and are designated as the SIC clones, and 20 were selected by the now defunct Agronomic Institute of East in the 1940s and comprise the SIAL clones. Additionally, 52 accessions were included and are denoted hereafter as the reference clones. These clones are from different geographical regions as shown in Table 1, and eleven are clonal cultivars (also known as 'modern cultivars') planted in some Bahian farms – CCN 51, CEPEC 2002, FA 13, PH 09, PH 15, PH 16, PS 1319, SJ 02, CEPEC 2004, BN 34 and PS 1030. The genomic DNA was extracted from each of the 103 clones using the cetyltrimethyl ammonium bromide (CTAB) protocol (Doyle and Doyle 1990) with modifications for cacao as described by Faleiro et al. (2002).

### **Simple sequence repeat (SSR) markers and polymerase chain reaction (PCR) amplifications**

Thirty microsatellite markers were used for genotyping the genetic materials (Table 2). Twenty-seven of these markers are genomic, seventeen of which were characterized by Santos et al. (2012), and ten were characterized by Lanaud et al. (1999). The remaining three markers are functional markers that were characterized by Lima et al. (2008). Polymerase chain reaction (PCR) was performed in a final volume of 0.020 cm<sup>3</sup> containing 6 ng of template DNA, 1× PCR buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 1.5 mM MgCl<sub>2</sub>, 0.2 μM of each dNTP, 0.5 μM each primer and 1 U of Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA). The amplification reactions for each locus were conducted according to the articles in which they were described. The amplification

products were resolved using vertical electrophoresis on 1× TBE/6% denaturing polyacrylamide gels at 75 W for approximately 2 h and then stained with silver nitrate according Creste et al. (2001). For each run, eight DNA samples were repeated as controls. The product sizes were determined by comparison with a 10-bp DNA ladder (Invitrogen).

**Table 1** List of the 103 cacao clones used for SSR genotyping and their origin locations.

| Clone                   | Origin       | Clone                 | Origin       | Clone                | Origin          |
|-------------------------|--------------|-----------------------|--------------|----------------------|-----------------|
| APA 4                   | Colombia     | PH15 <sup>a</sup>     | Bahia-Brazil | SIC 20               | Bahia-Brazil    |
| BE 10                   | Brazil       | PH16 <sup>a</sup>     | Bahia-Brazil | SIC 21               | Bahia-Brazil    |
| BE 2                    | Brazil       | PS1030 <sup>a</sup>   | Bahia-Brazil | SIC 22               | Bahia-Brazil    |
| BN34 <sup>a</sup>       | Bahia-Brazil | PS1319 <sup>a</sup>   | Bahia-Brazil | SIC 23 <sup>b</sup>  | Bahia-Brazil    |
| C SUL 9                 | Brazil       | RB 39                 | Brazil       | SIC 24 <sup>b</sup>  | Bahia-Brazil    |
| CA 6                    | Brazil       | RB 47                 | Brazil       | SIC 250              | Bahia-Brazil    |
| CAB 10                  | Brazil       | RIM 15                | Mexico       | SIC 329 <sup>b</sup> | Bahia-Brazil    |
| CAB 14                  | Brazil       | RIM 52                | Mexico       | SIC 4 <sup>b</sup>   | Bahia-Brazil    |
| CAB 3                   | Brazil       | RIM 76                | Mexico       | SIC 433 <sup>b</sup> | Bahia-Brazil    |
| CCN 51 <sup>a</sup>     | Ecuador      | SCA 6                 | Peru         | SIC 5                | Bahia-Brazil    |
| CEPEC 2002 <sup>a</sup> | Brazil       | SIAL 105 <sup>b</sup> | Bahia-Brazil | SIC 628              | Bahia-Brazil    |
| CEPEC 2004 <sup>a</sup> | Brazil       | SIAL 12               | Bahia-Brazil | SIC 662 <sup>b</sup> | Bahia-Brazil    |
| CHORONI 28              | Venezuela    | SIAL 13 <sup>b</sup>  | Bahia-Brazil | SIC 680              | Bahia-Brazil    |
| CHUAO 120               | Venezuela    | SIAL 130              | Bahia-Brazil | SIC 747              | Bahia-Brazil    |
| CJ 8                    | Brazil       | SIAL 15               | Bahia-Brazil | SIC 765              | Bahia-Brazil    |
| EEG 29                  | Brazil       | SIAL 163              | Bahia-Brazil | SIC 801              | Bahia-Brazil    |
| EET 392                 | Ecuador      | SIAL 164              | Bahia-Brazil | SIC 802 <sup>b</sup> | Bahia-Brazil    |
| EET 399                 | Ecuador      | SIAL 169              | Bahia-Brazil | SIC 806 <sup>b</sup> | Bahia-Brazil    |
| FA 13 <sup>a</sup>      | Bahia-Brazil | SIAL 244 <sup>b</sup> | Bahia-Brazil | SIC 812              | Bahia-Brazil    |
| GS 29                   | Grenada      | SIAL 253 <sup>b</sup> | Bahia-Brazil | SIC 813 <sup>b</sup> | Bahia-Brazil    |
| GS 36                   | Grenada      | SIAL 325              | Bahia-Brazil | SIC 831              | Bahia-Brazil    |
| ICS 1                   | Trinidad     | SIAL 407              | Bahia-Brazil | SIC 842              | Bahia-Brazil    |
| ICS 6                   | Trinidad     | SIAL 495              | Bahia-Brazil | SIC 848              | Bahia-Brazil    |
| ICS 8                   | Trinidad     | SIAL 512              | Bahia-Brazil | SIC 864              | Bahia-Brazil    |
| ICS 9                   | Trinidad     | SIAL 578              | Bahia-Brazil | SIC 872              | Bahia-Brazil    |
| ICS 95                  | Trinidad     | SIAL 644              | Bahia-Brazil | SIC 891 <sup>b</sup> | Bahia-Brazil    |
| IMC 67                  | Peru         | SIAL 698              | Bahia-Brazil | SIC 961              | Bahia-Brazil    |
| MA 15                   | Brazil       | SIAL 84               | Bahia-Brazil | SJ02 <sup>a</sup>    | Bahia-Brazil    |
| MOC 1                   | Brazil       | SIAL 910              | Bahia-Brazil | SPA 5                | Colombia        |
| NA 312                  | Peru         | SIAL 93               | Bahia-Brazil | TSH1188              | Descendente SCA |
| OC 67                   | Venezuela    | SIC 17                | Bahia-Brazil | UF 12                | Costa Rica      |
| P 7                     | Peru         | SIC 18                | Bahia-Brazil | UF 221               | Costa Rica      |
| 22 P                    | Mexico       | SIC 19                | Bahia-Brazil | UF 613               | Costa Rica      |
| PA 30                   | Peru         | SIC 2                 | Bahia-Brazil | UF 667               | Costa Rica      |
| PH 09 <sup>a</sup>      | Bahia-Brazil |                       |              |                      |                 |

<sup>a</sup>Modern cultivar used in commercial plantations on Bahia farms

<sup>b</sup>Clones of the SIAL and SIC clones that were selected to form the core collection sampling of all 78 of the alleles that are present in these groups

### **Allelic distribution and polymorphisms**

Summary statistics were calculated from the genotyping data to estimate the total number of alleles, the allele frequencies for each locus and the observed and expected heterozygosity values using GenAEx v.6.5 (Peakall and Smouse 2012). Each of the aforementioned genetic parameters was determined for the complete dataset (103 samples) and independently for the SIC and SIAL clones and the reference clones. Additionally, to assess the level of polymorphism that was associated with each locus used in this study, the polymorphic information content (PIC) was calculated using the PIC Calculator program (Kemp 2002). Tests for deviation from Hardy-Weinberg equilibrium and for the fixation index for each locus were also performed using GenAEx v.6.5 (Peakall and Smouse 2012).

### **Genetic relationships among accessions**

The genetic structure was inferred using Bayesian clustering analysis with STRUCTURE v.2.3.4 software (Pritchard et al. 2000) using an admixture ancestry model with 400,000 replications after a burn-in period of 100,000. A first STRUCTURE run was conducted with the global dataset (103 samples) looking for any existing substructure within the groups of clones, while the other two runs were performed as follows: one using the 51 SIC and SIAL clones and another using the 52 reference clones. For the first analysis (global data), the number of tested clusters (K) ranged from 2 to 10, and 20 independent runs were applied for each tested K. For the other runs, the K ranged from one to five (SIC and SIAL clones) and from one to eight (reference clones). STRUCTURE HARVESTER v.0.6.93 (Earl and von Holdt 2012) was used to analyze the results of the Structure output, and the most optimal K value was calculated using Evanno's methods (Evanno et al. 2005). For each most probable K, the optimum alignment over the 20 iterations was matched by permutation using CLUMPP (Cluster Matching and Permutation Program) v1.1.2 software (Jakobsson and Rosenberg 2007), and the outputs were used in DISTRUCT v.1.1 (Rosenberg 2004) to create the histogram.

The population differentiation for the evaluation of polymorphisms was analyzed using an analog of  $F_{st}$  (Nei's  $G_{ST}$ ) with 999 permutations and 1000 bootstraps.

Furthermore, an analysis of molecular variance (AMOVA) was conducted to obtain the percentages of variability among the groups. Both Nei's  $G_{ST}$  and the AMOVA were calculated using GenAlEx v.6.5 software (Peakall and Smouse, 2012), and the groups were formed by common origin, as suggested by the STRUCTURE results.

The genetic relationships among all of the clones that were used in this study were obtained using cluster analysis. For this analysis, TFPGA software (Miller 1997) was used to calculate Rogers' modified genetic distance (Wright 1978), and the dendrogram was built using the Neighbor Joining method (Saitou and Nei 1987) as implemented on DARwin 5.0 software (Perrier and Jacquemoud-Collet 2006). For the purpose of comparative analysis, Roger's modified distance and Nei's  $G_{ST}$  were also calculated among the SIC, SIAL and modern cultivars that were planted in Bahia. The dispersion of the clones of these groups in a bidimensional plan was obtained by Principal Coordinate Analysis (PCoA), whose distance and projection were calculated using GenAlEx v.6.5 software (Peakall and Smouse 2012).

### **Private alleles and core collection**

GenAlEx v.6.5 software (Peakall and Smouse 2012) was used to detect the private alleles in each group. In addition, the core collection that represents all of the genetic diversity that was present in the 103 analyzed clones was determined using COREFINDER software (Cipriani et al. 2010), adopting 99 iterations. A new run was performed combining the SIC clones, the SIAL clones and the modern Bahian clones. These strategies were used mainly to identify individuals of these clones that represent the maximum level of diversity.

## **Results**

### **Allelic distribution and polymorphism**

The list of SSRs that were used and the results of the statistical analysis for each locus are shown in Table 2. All 30 of the SSR loci that were used were polymorphic in the 103 clones, producing scorable bands. The analysis revealed a total of 205 alleles

and a mean of 6.83 alleles per locus ranging from two for the mTc-UNICAMP16 locus to fifteen for the mTc-UNICAMP13 locus. For almost all of the analyzed loci, the expected heterozygosity ( $H_E$ ) was higher than the observed heterozygosity ( $H_O$ ), except for mTc-UNICAMP09 and mTc-UNICAMP16, which showed slightly higher observed than expected values. Thus, the  $H_O$  and  $H_E$  ranged from 0.021 (mTc-UNICAMP05) and 0.132 (mTc-UNICAMP06) to 0.663 (mTc-UNICAMP12) and 0.787 (mTc-UNICAMP03), with an average of 0.308 and 0.456 for the observed and expected heterozygosities, respectively.

The PIC values ranged from 0.125 and 0.126 for mTc-UNICAMP06 and mTc-UNICAMP16, respectively, to 0.754 for mTc-UNICAMP03, with a mean of 0.432 (Table 2). The fixation index is also presented in Table 2, and the obtained mean was  $F = 0.291$ , indicating a deficiency of heterozygotes. The fixation index values ranged from -0.079 in mTc-UNICAMP16 to 0.971 in mTc-UNICAMP05. The mean values of the fixation index were higher (0.317) and lower (0.156) for the SIC and SIAL clones and reference clones, respectively, than the observed mean of the global data. Testing for Hardy-Weinberg equilibrium in the total population revealed that most of the loci were in disequilibrium, with  $p < 0.002$  (21 loci) and  $p < 0.0003$  (18 loci), after Bonferroni correction.

Independent analyses of the diversity parameters were performed for the SIC and SIAL clones and for the reference clones (Table 2). The results of these analysis indicated lower mean values for the SIC and SIAL clones compared to the reference clones (for the number of alleles per locus,  $H_O$  and  $H_E$ ). Furthermore, the mean  $H_O$  and  $H_E$  values for the reference clones were higher than the mean values that were obtained for all 103 samples (Table 2). Thus, considering all of the loci examined, for the SIC and SIAL clones, the average number of alleles per locus was 2.60 (total of 78 alleles), and the  $H_O$  and  $H_E$  were 0.116 and 0.218, respectively. The reference clones had a mean of 6.7 alleles per locus (total of 200 alleles) and mean values of 0.502 and 0.603 for  $H_O$  and  $H_E$ , respectively. Furthermore, an independent analysis of the SIC and SIAL clones indicated that of the thirty loci that were used for genotyping, seven were monomorphic only for these clones. After removing these loci, the new mean values for the number of alleles per locus and the  $H_O$  and  $H_E$  were 3.09, 0.15 and 0.28, respectively

(Supplementary Material S1) which were still lower than the values that were observed for the reference clones used in this study.

**Table 2** Summary statistics of allelic diversity across 30 microsatellite loci in the clones of cacao.

| Locus                 | Tm       | Size range (bp) | Na      | H <sub>o</sub> | H <sub>E</sub> | F         | PIC                |
|-----------------------|----------|-----------------|---------|----------------|----------------|-----------|--------------------|
| mTc-UNICAMP01         | TD65-55  | 174-214         | 8       | 0.333          | 0.368          | 0.095     | 0.357              |
| mTc-UNICAMP02         | TD65-55  | 260-300         | 11      | 0.456          | 0.683          | 0.332     | 0.667 <sup>b</sup> |
| mTc-UNICAMP03         | TD65-55  | 270-291         | 8       | 0.515          | 0.787          | 0.345     | 0.754 <sup>b</sup> |
| mTc-UNICAMP04         | 60 °C    | 262-312         | 5       | 0.267          | 0.295          | 0.094     | 0.277 <sup>a</sup> |
| mTc-UNICAMP05         | 60 °C    | 148-161         | 7       | 0.021          | 0.725          | 0.971     | 0.681 <sup>b</sup> |
| mTc-UNICAMP06         | 60 °C    | 292-302         | 3       | 0.071          | 0.132          | 0.465     | 0.125 <sup>b</sup> |
| mTc-UNICAMP07         | TD65-55  | 244-254         | 4       | 0.284          | 0.381          | 0.253     | 0.357 <sup>b</sup> |
| mTc-UNICAMP08         | 60       | 296-308         | 6       | 0.320          | 0.354          | 0.095     | 0.335              |
| mTc-UNICAMP09         | TD65-55  | 254-260         | 5       | 0.306          | 0.304          | -0.006    | 0.287              |
| mTc-UNICAMP10         | 60       | 140-144         | 3       | 0.069          | 0.203          | 0.658     | 0.184 <sup>b</sup> |
| mTc-UNICAMP11         | 60       | 140-170         | 9       | 0.167          | 0.316          | 0.473     | 0.308 <sup>b</sup> |
| mTc-UNICAMP12         | 60       | 262-286         | 10      | 0.663          | 0.775          | 0.144     | 0.744 <sup>a</sup> |
| mTc-UNICAMP13         | TD65-55  | 148-196         | 15      | 0.579          | 0.760          | 0.238     | 0.737 <sup>b</sup> |
| mTc-UNICAMP14         | TD65-55  | 230-252         | 10      | 0.450          | 0.719          | 0.374     | 0.667              |
| mTc-UNICAMP15         | TD65-55  | 230-236         | 3       | 0.149          | 0.158          | 0.058     | 0.151              |
| mTc-UNICAMP16         | TD65-55  | 290-296         | 2       | 0.146          | 0.135          | -0.079    | 0.126              |
| mTc-UNICAMP17         | TD65-55  | 240-292         | 10      | 0.365          | 0.608          | 0.401     | 0.584 <sup>b</sup> |
| msEstTsh-1            | 56.8     | 192-200         | 4       | 0.214          | 0.244          | 0.123     | 0.229              |
| msEstTsh-6            | TD 60–48 | 208-224         | 5       | 0.310          | 0.370          | 0.163     | 0.353              |
| msEstTsh-7            | TD 60–48 | 153-161         | 5       | 0.564          | 0.654          | 0.136     | 0.595 <sup>a</sup> |
| mTcCIR3               | 46       | 210-248         | 8       | 0.453          | 0.586          | 0.226     | 0.555 <sup>b</sup> |
| mTcCIR6               | 46       | 228-250         | 7       | 0.245          | 0.356          | 0.313     | 0.336 <sup>b</sup> |
| mTcCIR7               | 51       | 154-166         | 6       | 0.263          | 0.400          | 0.344     | 0.379 <sup>b</sup> |
| mTcCIR8               | 46       | 290-308         | 4       | 0.250          | 0.361          | 0.307     | 0.332 <sup>b</sup> |
| mTcCIR11              | 46       | 290-318         | 8       | 0.363          | 0.609          | 0.405     | 0.585 <sup>b</sup> |
| mTcCIR13              | 46       | 246-270         | 9       | 0.196          | 0.639          | 0.693     | 0.611 <sup>b</sup> |
| mTcCIR15              | 46       | 236-258         | 10      | 0.424          | 0.748          | 0.433     | 0.713 <sup>b</sup> |
| mTcCIR18              | 51       | 330-350         | 4       | 0.265          | 0.340          | 0.222     | 0.295              |
| mTcCIR25              | 46       | 134-168         | 9       | 0.258          | 0.374          | 0.311     | 0.363 <sup>b</sup> |
| mTcCIR26              | 46       | 286-318         | 7       | 0.265          | 0.307          | 0.138     | 0.287 <sup>b</sup> |
| <b>Mean</b>           |          |                 | 6.83    | 0.308          | 0.456          | 0.291     | 0.432              |
|                       |          |                 |         | 0.25-          |                |           |                    |
| °Confidence interval  |          |                 | 5.8-7.7 | 0.36           | 0.38-0.52      | 0.22-0.36 |                    |
| <b>Group</b>          |          |                 |         |                |                |           |                    |
| SIC/SIAL (51)         |          |                 | 2.6     | 0.116          | 0.218          | 0.317     | -                  |
| Reference clones (52) |          |                 | 6.7     | 0.502          | 0.603          | 0.156     | -                  |

Na number of alleles, H<sub>E</sub> expected heterozygosity, H<sub>o</sub> observed heterozygosity, PIC polymorphism information content, and F fixation index

The numbers in parentheses are the sample sizes for each group

Departs significantly from Hardy-Weinberg equilibrium after Bonferroni correction: <sup>a</sup>p<0.002 and p<0.0003

°Confidence interval estimated by bootstrap (100,000 simulations).



### Genetic diversity and relationships among the clones

A Bayesian analysis performed using STRUCTURE and based on the higher  $\Delta K$  values (Evanno's method) separated the 103 evaluated genotypes into three and four groups (Fig. 1a), respectively. For further analysis, the clones were considered of a single recognized cacao cluster if their inferred ancestry (Q value) was at least 70% for this cluster, and each group was denominated according the clone types that were quantitatively more representative of this. Therefore, the three groups that were formed were named SIC/SIAL, Upper Amazon/Hybrids and Trinitario/Criollo. The SIC/SIAL cluster, which is represented by blue in the STRUCTURE histogram (Fig. 1b), was formed by the 51 SIC and SIAL clones together with the three reference clones CAB 14, BE 10 and EEG 29. The remaining individuals were separated into two groups: one containing the majority of the Upper Amazon/Hybrids and the other formed mainly by the Trinitario/Criollo clones (represented by red and orange colors, respectively, in Fig. 1b). The majority of the modern Bahian cultivar clones were grouped into the Upper Amazon/Hybrid group; therefore, those clones showed the same inferred ancestry. Three individuals (UF 613, PH 09 and EET 399) presented Q values lower than 70% for any cluster and were classified as of mixed ancestry.

For  $K = 4$ , where the lowest peak was observed (Fig. 1a), the Upper Amazon/Hybrid cluster was subdivided into two groups, one of which presented nine of the eleven Bahian modern cultivar clones (data not shown). Both the SIC/SIAL and Trinitario/Criollo clusters remained independent groups that were composed of the same clones that were observed when  $K = 3$ .

Analyses were conducted to obtain substructures, and a run for SIC and SIAL clones and another analysis for the reference clones were conducted using STRUCTURE software. For the latter, the highest  $\Delta K$  values were observed for  $K = 2$  and  $K = 3$ . In both of these K values, the genetic structure with regard to the distribution of the reference clones was similar to that observed for the global data for  $K = 3$  and  $K = 4$ , respectively.

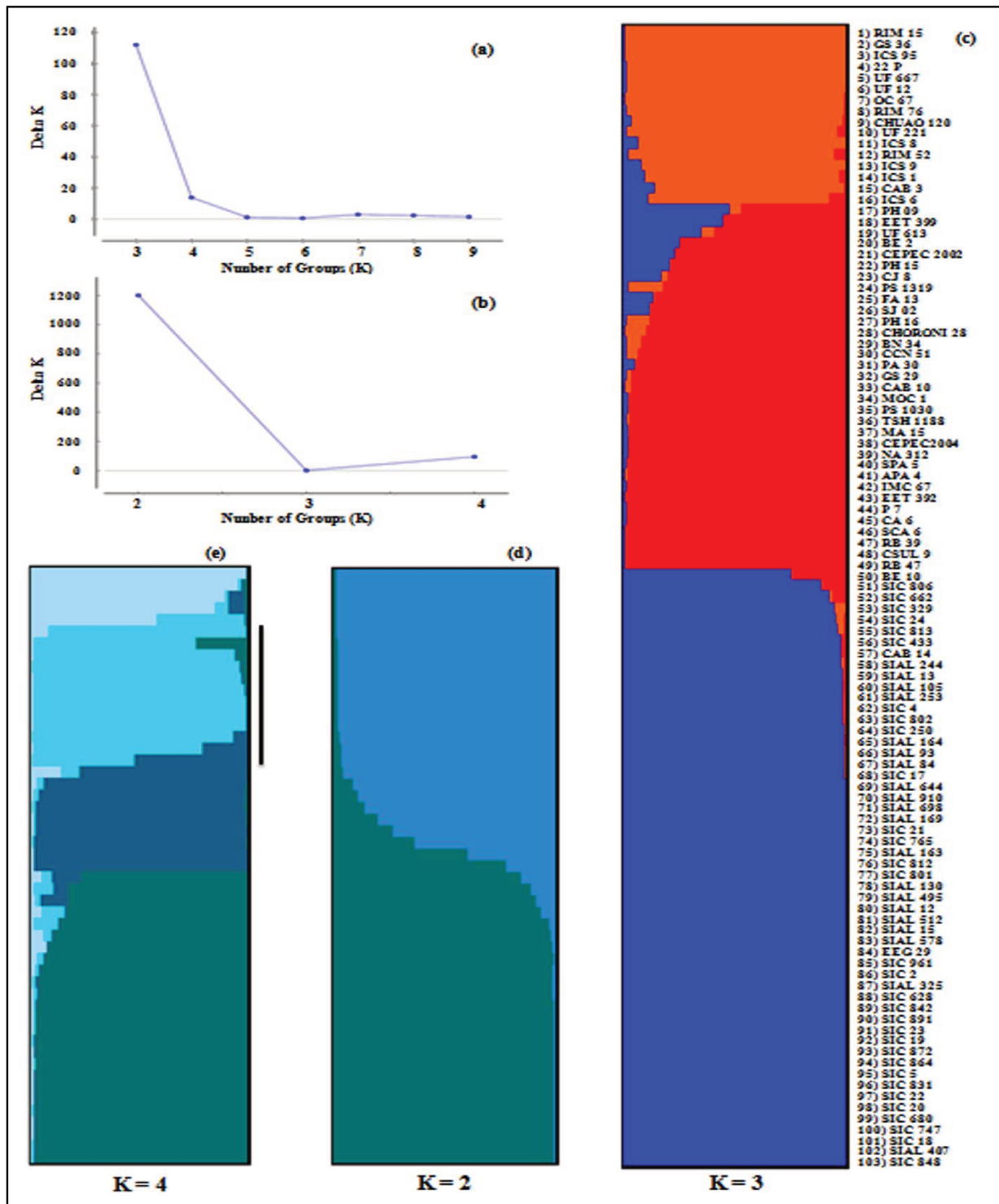
In contrast to the reference clones, the STRUCTURE run with the SIC and SIAL clones showed substructures that were not observed in the global data analysis. High  $\Delta K$  values were met for  $K = 2$  followed by  $K = 4$  (Fig. 1c), and there was a tendency to

separate the SIC of the SIAL clones (Fig. 1d). For  $K = 2$ , of the 20 SIAL clones that were used in this study, 18 were grouped into the same cluster along with eight SIC clones and are denominated here as the SIAL group. In addition, the group that was denominated as SIC, was formed by 24 clones, including two of the SIAL series. Of the analyzed clones, only SIC 812 and SIC 813 were considered to be of mixed ancestry. In the second level of substructure  $K = 4$  (Fig. 1e), while no striking difference was observed for the SIAL group compared with  $K = 2$ , the SIC group was subdivided into three clusters, one of which was formed only by the SIC clones that were tagged with numbers greater than 800 at the time of selection. This tag reflects the cacao variety, known as Catongo, with white seeds that were selected from a mutant of the common cacao, noticed for its potential to produce fine chocolate. Except for SIC 806, all of the cacao Catongo cultivars were grouped into the same cluster, including genotypes that were considered as mixed ancestry (SIC 812 and SIC 813) or were allocated into the SIAL group (SIC 961) at  $K = 2$ .

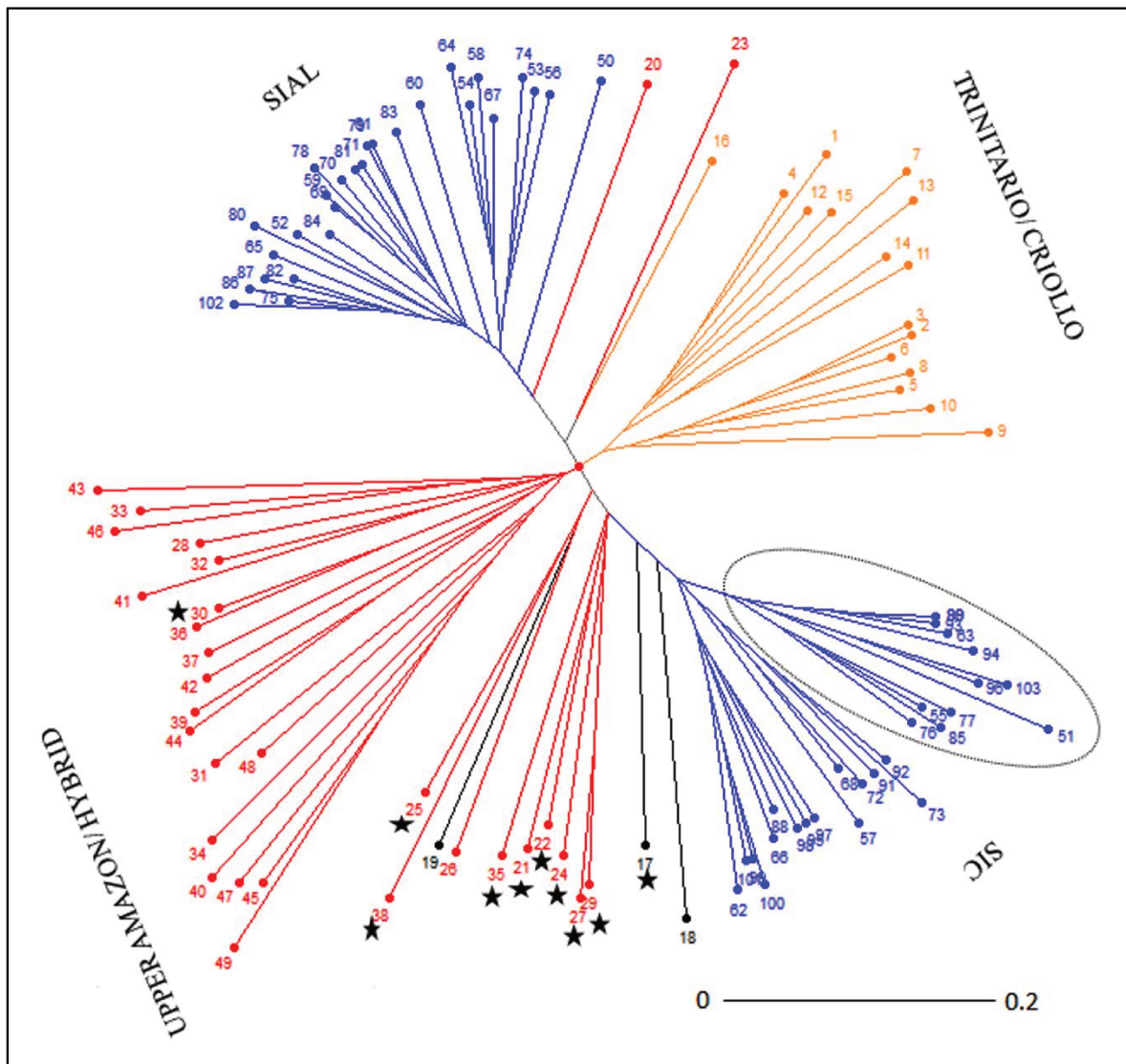
The traditional clustering Neighbor Joining method (Fig. 2) using Roger's modified genetic distance and estimates of Nei's  $G_{ST}$  (Table 3) corroborated the previous STRUCTURE results. The branches of the Neighbor Joining tree were tagged with color according to the inferred STRUCTURE cluster. The Neighbor Joining analyses indicate a clear tendency of individuals with the same ancestry to group together, representing the main structure. Furthermore, the Trinitario/Criollo, Upper Amazon/Hybrid and SIC and SIAL groups separate into a substructure that can discern subgroups such as the Catongo among SIC individuals and the modern Bahian cultivars in a specific Upper Amazon/Hybrid subgroup. Additional substructures that were not seen in the STRUCTURE results were visualized using the Neighbor Joining method, as seen with the Trinitario/Criollo group (Fig. 2).

Using the same component groups as indicated by the STRUCTURE software, Nei's  $G_{ST}$  values for the SIC/SIAL, Upper Amazon/Hybrid and Trinitario/Criollo groups ranged from moderate to relatively high (Table 3). A higher structuring was obtained between the SIC/SIAL and Trinitario/Criollo clusters (0.295) and a lower structuring between the Upper Amazon/Hybrid and Trinitario/Criollo clusters (0.122). The contrast

between the SIC and SIAL clusters (also formed based on the STRUCTURE results) shows Nei's  $G_{ST}$  values indicating high structure (0.273).



**Fig. 1** Structure analysis results. **(A)** Graphic with the most probable cluster numbers (K) for the global data (103 clones) and **(B)** histogram for K = 3 (Blue: SIAL/SIC; Red: Upper Amazon/Hybrid and Orange: Trinitario/Criollo). **(C)** The most probable K values and histogram only for the SIAL and SIC clones considering **(D)** K = 2 (green: SIAL and blue: SIC) and **(E)** K = 4 (green: SIAL and blue tones: subgroups of SIC clones). The black bar indicates the substructure in SIC corresponding to the Catongo clones



**Fig. 2** Neighbor Joining tree obtained using Roger's modified distance from 103 cacao clones. The colors of each group and the numbers of branches correspond to the groups and clones, respectively, that were present in the Structure histogram for the global data in  $K = 3$  (Fig. 1). The star symbols correspond to the modern cultivar clones, and the circle highlights the Catongo clones.

**Table 3** Structuring level and the percentages of differences between the groups and a subgroup from Nei's  $G_{ST}$  and an analysis of molecular variance. The group components were defined using Structure software (Pritchard et al. 2000).

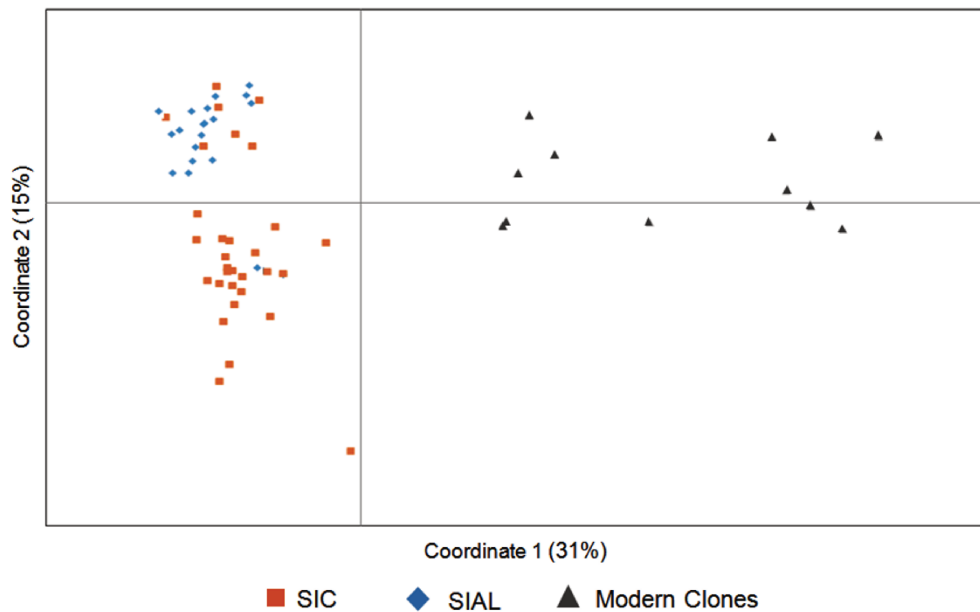
| Clusters                                 | $G_{ST}^a$ | AMOVA <sup>a</sup> |
|------------------------------------------|------------|--------------------|
| SIC/SIAL × Upper Amazon/Hybrid           | 0.204      | 21%                |
| SIC/SIAL × Trinitario/Criollo            | 0.295      | 31%                |
| Upper Amazon/Hybrid × Trinitario/Criollo | 0.122      | 10%                |
| SIC × SIAL                               | 0.273      | 25%                |

<sup>a</sup>Considering 999 permutations and  $p < 0.001$

An analysis contrasting the SIC, SIAL and Bahian modern cultivars detected differences among these cultivars (Table 4). Higher Nei's  $G_{ST}$  and Roger's modified distance values (0.273 and 0.382, respectively) were observed between the modern cultivars and the SIAL clones. The discrepancy of the  $G_{ST}$  values between SIC and SIAL in Tables 3 and 4, respectively, reflects groups that are considered peculiar for each analysis, specifically, clusters indicated by the STRUCTURE program and groups formed according with origin of selection, i.e., selection of Cacao Institute (SIC) or selection of Agronomic Institute of the East (SIAL). Lower values were detected between the SIC and SIAL series for both Nei's  $G_{ST}$  (0.201) and Roger's modified distance (0.149) compared to modern cultivars. In fact, when dispersion was evaluated in a PCoA of the modern cultivars and the SIC and SIAL clones, a high distribution of the modern cultivars and a relationship between the SIC and SIAL clones were observed. Despite the greater diversity that was present in the modern cultivars, clear differences were detected in the PCoA between all three of the analyzed groups (Fig. 3).

**Table 4** Genetic structure and the distances between the SIAL and SIC groups and the cultivar clones that are currently employed in Bahia for commercial planting, as accessed by Roger's modified distance (Lower diagonal) and Nei's  $G_{ST}$  (Upper diagonal).

|           | SIAL  | SIC   | Cultivars |
|-----------|-------|-------|-----------|
| SIAL      | ---   | 0.149 | 0.273     |
| SIC       | 0.201 | ---   | 0.224     |
| Cultivars | 0.382 | 0.354 | ---       |

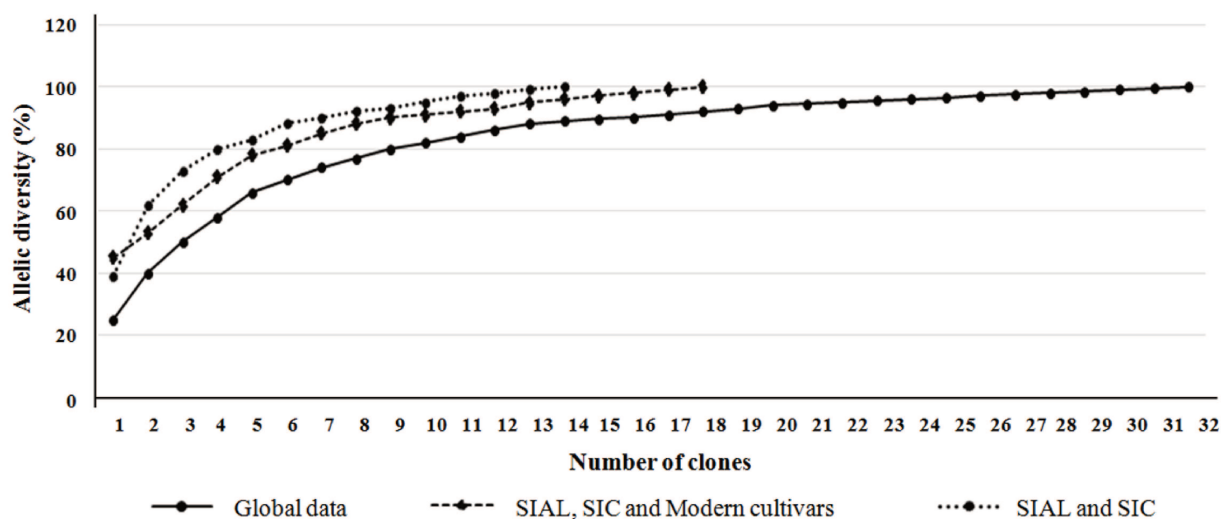


**Fig. 3** Principal coordinates analysis between the SIC, SIAL and modern clones (at last currently cultivated on Bahia farms).

### Private alleles and core collection

One hundred thirty-two private alleles were obtained from the global data (103 clones), only five of which were of the SIC and SIAL clones: two specific for SIAL and three for the SIC clones (Supplementary Material S2). The two private SIC alleles are present in approximately 30% of the individuals and can be on selection in this group. Seven alleles occur in the SIAL clones but not in the SIC clones, and SIC presents 15 private alleles that do not occur in the SIAL clones. When the SIC, SIAL and modern clones are evaluated together, the number of private alleles is 5, 8 and 66, respectively.

Submitting all 103 clones to the core collection analyses, a total of 32 clones (Fig. 4) were identified as representing the total allelic diversity (205 alleles) that was present in the data. This core was formed mostly by the reference clones, five from Bahian modern cultivars and six from the SIAL and SIC clones. However, when the SIC and SIAL clones were evaluated individually, 78 alleles were represented in fourteen individuals, and only four of these alleles were SIAL clones (Table 1; Fig. 4). Five SIC and SIAL clones were represented in the core collection in both the analysis of all 103 clones and that of only SIC and SIAL. When 51 SIC / SIAL clones were evaluated along with 11 modern clones, the core collection (Fig. 4) was formed by 18 clones with 120 alleles, representing the majority (10) of modern cultivars.





**Fig. 4** Number of clones representing all of the allelic diversity present in the different evaluated groups – global data (from 103 clones); SIAL, SIC plus modern cultivars (from 62 clones) and only SIAL and SIC (from 51 clones).

## Discussion

### Allelic distribution and polymorphisms

Molecular markers have been used in several studies in *T. cacao* and support previous statements based on morphological characteristics that cacao exhibits high levels of genetic diversity (Motamayor et al. 2008; Bartley 2005; Bennet 2003). In this study, we focused our analyses on structure and diversity status of the SIC and SIAL clones, the first ones to be selected in Bahia's cacao plantations in the 1930s and 1940s, respectively. The set of 30 SSR markers revealed polymorphisms among the clones and relationships among the genetic groups, reflecting the different genetic origins of the plants. The genetic parameters for the total data that were evaluated here presented relatively low values for  $H_E$  (0.406),  $H_0$  (0.308), and the number of alleles per locus (6.83), compared with the values observed in other studies (Aikpokpodion et al. 2009; Irish et al. 2010; Boza et al. 2013). The influence of the SIC and SIAL clones on these low values became evident when these clones were evaluated independently of the remaining clones. In this case, the values for SIC and SIAL were 2.60 alleles per locus,  $H_E = 0.218$  and  $H_0 = 0.116$ . For the reference clones, these values could become equal to or greater than 6.67 alleles per locus,  $H_0 = 0.502$  and  $H_E = 0.603$ .

Low genetic diversity, based on molecular markers, has been reported for the Amelonado group (Lanaud 1987; N'Goran et al. 2000; Motamayor et al. 2003) and was also found for SIC and SIAL clones using dominant markers (Cascardo et al. 1993). This fact could reflect the domestication history of Amelonado cacao. Low diversity has also been observed in other varieties, such as Criollo (Motamayor et al. 2002) and Nacional (Solorzano et al. 2012; Loo de la Cruz et al. 2009). These varieties shared a similar domestication process, which was marked by genetic bottlenecks likely caused by humans seeking nutritional benefits of cacao (Clement et al. 2010). In this sense, although the origin of the Amelonado group is not clear (Bartley 2005), the hypothesis that the Lower Amazon

populations originated from the Upper Amazon by few introductions (Serenó et al. 2006), could explain the lower diversity for the Lower Amazon plants and consequently for the Amelonado group.

In addition to the domestication history, the mating system of Amelonado cacao may also have influenced the genetic parameters that were observed in this study. Similar to other traditional cultivated varieties, Amelonado presents self-compatibility that enhances inbreeding, as detected by positive and high fixation index values in the global data. This fact has been reported in other studies for the Lower Amazon cacao (Motilal et al. 2013; Aikpokpodion et al. 2009; Sereno et al. 2006). In the specific case of the SIC and SIAL clones, the relatively narrow genetic base caused by the restricted introduction of the ancestors of these plants also increases the endogamy. For the reference clones, the positive mean fixation index may be caused by the Wahlund effect, reflecting the origin of these clones from different populations with different allele frequencies, thus reducing the overall heterozygosity and consequently increasing the fixation index.

### **Group relationships**

The genetic diversity that was present in the 103 clones is structured into at least three genetic groups, here classified as SIC/SIAL, Upper Amazon/Hybrids and Trinitario/Criollo. The STRUCTURE results indicate that SIC/SIAL and Trinitario/Criollo are more homogeneous than the Upper Amazon/Hybrid cluster, which is the only cluster that showed substructure in the first level of analysis (103 clones together) and in a second level of analysis using the reference clones. This result corroborates the hypothesis that the cacao diversification center is in the Upper Amazon region (Cheesman 1944; Sereno et al. 2006; Motamayor et al. 2008; Thomas et al. 2012), which maintains a germplasm with a wide genetic diversity.

It is interesting to note that the observed substructure in the Upper Amazon/Hybrid clones was strongly influenced by the presence of the Bahian modern clonal cultivars: CCN 51; CEPEC 2002; FA 13; PH 09; PH 15; PH 16; PS 1319; SJ 02; CEPEC 2004; BN 34 and PS 1030. The majority of these cultivars were obtained by selection from local farmers, except for CCN 51, which was selected in Ecuador. The grouping by Bayesian method of the majority of these clonal cultivars along with SCA 6

( $K = 3$ , Fig. 1) indicate a common ancestry, which is plausible because this clone was the main source of resistance used to obtain many hybrids that were distributed by CEPEC/CEPLAC for commercial planting (Yamada et al. 2001). However, in this and other studies (Leal et al. 2008; Yamada et al. 2005), the clustering method based on genetic distance used does not show a common group for SCA 6 and modern cultivars, indicating a considerable distance among these cultivars, possibly because the SCA clone was used as a progenitor in the initial crosses, contributing only small portions of its genomes to these clonal cultivars.

Although some differences, such as the aforementioned between the STRUCTURE and Neighbor Joining analyses, were observed, the main results were similar to each other. These findings are consistent with those of Odong et al. (2011), in which no significant differences were detected between traditional clustering analyses and the Bayesian model that was implemented in STRUCTURE. In this study, both of these methods were complementary, and while the Structure analysis was useful for clearly delimiting groups based on the inferred ancestry, the genetic diversity among the clones could be easily visualized using Roger's distance as shown in the Neighbor Joining tree.

In addition to the aforementioned methods, Nei's  $G_{ST}$  values permitted the quantification of the structuring level among the three main genetic groups, with a higher structure between SIC/SIAL and Trinitario/Criollo. Motamayor et al. (2003) suggested that the Brazilian Amelonado and Criollo varieties are involved in the origin of Trinitario; therefore, a relatively closer genetic approximation between these parental lines to the hybrid would have been expected but was not detected here. Furthermore, the closer relationship between SIC and SIAL and the Upper Amazon/Hybrid group (Nei's  $G_{ST} = 0.204$ ) is consistent with the hypothesis of Sereno et al. (2006), who posited that a restricted number of plants of the Upper Amazon founded the Lower Amazon populations.

### **Core collection**

Private alleles and core collection analyses are useful when determining the distribution of the alleles and the specific differences among the groups. The core collection reflects the distribution of variability in the data to select the minimal number of clones that represents all of the allelic diversity, including private alleles, and is a good tool for selecting plants for a work collection. Therefore, when constructing a work collection for cacao breeding representing traditional Bahian cacao, priority should be given to the selection of plants with economical features (pod and seed size, plant height, productivity, quality features, etc), combined with the maximum genetic diversity, as in the fourteen plants of SIC and SIAL that were indicated for the core collection. Indeed, these work's collection should include accessions of the different cultivars present in the SIC and SIAL collection based both, on genetic diversity obtained via molecular markers in this study and on the different types of cultivars defined on the base of fruit format, been features and productivity. Therefore, this collection should include representatives of the SIC and SIAL cultivars from the types Cacao Common, Para, Parazinho, Maranhão, Catongo and Almeida. As these cultivars have been previously selected for yield, they may be delivered for farmers as representatives of the specific type with a specific feature, as example, the white seeds for fine chocolate. Also, they may participate in crossing cycles with plant material such as Trinitario and Upper Amazon to explore hybrid vigor and disease resistance, once SIC and SIAL are very adapted to this region and bear important agronomic traits.

### **The cacao clones of Bahia**

The SIC and SIAL clones are representative of Bahian cacao (Pereira et al. 1987; Yamada et al. 2001) and are maintained in several worldwide cacao ex situ collections. These clones have historical importance because they represent the first cacao plants that were cultivated on a large scale in Brazil and have great potential for use in breeding programs. These clones are a part of the ancestry of the hybrid cultivars in past and in modern cultivars, such as CEPEC 2002, LP 06, FA 13 and many others (Monteiro and Ahnert 2007).

In this study, we identified a high degree of homogeneity among Bahian clones in relation to STRUCTURE clustering, supporting the fact that the Lower Amazon Amelonado is a distinct genetic group with its own characteristics (Motamayor et al. 2008; Johnson et al. 2009). Other genotypes (BE 10, CAB 14 and EEG 29) that are grouped in this cluster are also characterized as Lower Amazon. Although EEG clones were selected in Espírito Santo, these clones are descendants of the original plants that were introduced into Bahia (Vello et al. 1972); therefore, they have a common genetic basis with Bahian cacao.

Despite the homogeneity of SIC and SIAL plants, the substructure analysis in this study revealed differences within the groups, agreeing with early morphological accounts and suggesting differences in some features, such as pod weight, pod diameter and the wet and dry seed weights (Pereira et al. 1987; Vello et al. 1972). However, in contrast to the observations in this study, early reports indicated that SIAL clones were more variable than SIC clones. These differences were attributed to the way each group was selected: the SIC clones were selected from a restricted area of cacao planting, and the SIAL clones were obtained from different farmers in southern Bahia, thus covering a large selection area (Pereira et al. 1987). Nevertheless, in this study, we can conclude that even though the SIAL clones were obtained from a more widely distributed area than were the SIC clones, this difference did not lead to a broader genetic base for the SIAL clones compared with the SIC clones. This conclusion is supported by the structure and genetic distance analyses (mean genetic diversities of 0.46 and 0.72, respectively for SIAL and SIC clones), and this difference is reflected in the core collection. In fact, although the selection of the SIC clones focused especially in productivity, quality were also considered and seems to have influenced in the diversity of these plants. In this sense, reports of SIC selection indicate the following criteria for identifying the clones according their type: (i) clones identified from 1 to 100 were selected directly from the experimental area, (ii) clones identified from 100 to 800 were derived from self-pollinations of the first one group (i), and; (iii) clones with numbers greater than 800 correspond to the Catongo cultivar.

As previously mentioned, the SIC and SIAL clones were independently selected in Bahia. These clones, along with some introduced Upper Amazon clones, show

superiority as hybrid parents for features concerning productivity, resistance to diseases, bean quality and productivity (Monteiro and Ahnert 2007; Vello et al. 1972). Agronomically superior hybrid seeds were distributed to farmers until the 1980s, but with the outbreak of witches' broom in Bahia, the program was stopped, and clonal cultivars that were potentially more resistant were used (Monteiro et al. 2009). Over 25 years of occurrence of witches' broom in Bahia, clones have been selected mainly on Bahian farms with high productivity and/or resistance and employed for commercial planting. After hard work done by the farmers and researchers, just now the clonal cultivars have been well tested in various environments and became an important technology to recover the plantation in Bahia. Besides that, the recurrent selection strategy has been used in Bahian cacao breeding programs, looking for accumulate genes for resistance, productivity and quality. Diversity of the parents used in crossing has been also considered, and pre-breeding characterizations is important to choose plants.

Currently, other alternatives that have been explored to keep farmers in business are organic and fine cocoa production, which adds value in the international market. In this sense, Bahian cacao as a Catongo variety has been appointed with superior flavor and as functional food, due higher concentration of antioxidants and theobromine associated with blood pressure reduction (Leite et al. 2013). Also the Maranhão cultivar is known to produce fine chocolate and is now being explored by some farmers.

This work is the first reported effort to characterize, via codominant microsatellite markers, cacao clones selected in the Bahia state of Brazil, known as the SIC and SIAL clones. As these plants were originally selected for use as progenitors to produce seeds for delivering to farmers, it is possible that these plants do not actually represent the diversity that is present in naturalized Bahian cocoa. Additional evaluations in plants of this origin that are cultivated on farms are being performed and comparing them with SIAL and SIC plants, may help in selection. Bahian cacao series have important features for conservation and breeding. This study also demonstrates that a deeper level of substructure can be revealed by careful analysis of informative loci and that this information can be used in breeding programs to select local material with the advantage of being well adapted to environmental conditions.

### **Acknowledgments**

The authors are grateful to the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, 2008/52197-4), Coordenação do Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Programa Nacional de Cooperação Acadêmica Novas Fronteiras (Procad NF 2008) for financial support, to the Centro de Pesquisa do Cacau (CEPEC/CEPLAC) and the Farms of Juliana's Valley for providing plant material; to the Universidade Estadual do Sudoeste da Bahia (UESB) for providing a graduate fellowship to ESL Santos and CBM Cerqueira-Silva and, to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for research fellowship to AP Souza and RX Corrêa. This work is part of the doctoral thesis of the first author in the Programa de Pós-graduação em Genética e Biologia Molecular of Universidade Estadual de Campinas (UNICAMP).

## References

- Aikpokpodion PO, Motamayor JC, Adetimirin VO, Adu-Ampomah Y, Ingelbrecht I, Eskes AB, Schnell RJ, Kolesnikova-Allen M (2009) Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genet Genomes* 5:699–711
- Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI Publishing, Wallingford
- Bennett, AB (2003) Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Sci* 8:561-563
- Boza EJ, Irish BM, Meerow AW, Tondo CL, Rodríguez AO, Ventura-López M, Gómez JA, Moore JM, Zhang D, Motamayor JC, Schnell RJ (2013) Genetic diversity, conservation, and utilization of *Theobroma cacao* L.: genetic resources in the Dominican Republic. *Genet Resour Crop Evol* 60:605–619
- Bondar G (1938) A cultura do cacau na Bahia. Instituto de Cacau da Bahia 1:7-39
- Cascardo JC, Pires JL, Figueira A (1993) Estimating genetic diversity of cacao in Southern Bahia using RAPDs. In: Annals of the Brazilian Plant Biotechnology Meeting. REDBIO, Brasília
- Castro GCT, Bartley BGD (1983) Caracterização dos recursos genéticos do cacaueiro. 1. Folha, fruto e semente de seleções da Bahia das series SIC e SIAL. *Theobroma* 13:263-273
- Castro GCT, Bartley BGD (1985) Caracterização dos recursos genéticos do cacaueiro. II. Flor de seleções da Bahia das series SIC e SIAL. *Theobroma* 15:49-55



Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. Trop Agric 21:144–159

Cipriani G, Spadotto A, Jurman I, Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A, Testolin R (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages and reveals a large admixture amongst varieties of different geographic origin. Theor Appl Genet 121:1569-1585

Clement CR, Cristo-Araújo M, d'Eeckenbrugge GC, Pereira AA, Picanço-Rodrigues D (2010) Origin and Domestication of Native Amazonian Crops. Diversity 2:72-106

Creste S, Tulmann Neto A, Figueira A (2001) Detection of Single Sequence Repeat Polymorphisms in Denaturing Polyacrylamide Sequencing Gels by Silver Staining. Plant Mol Biol Rep 19:299–306

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15

Earl DA, vonHoldt BM. (2012) Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour 4:359-361

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

Faleiro FG, Santos IS, Bahia RCS, Santos RF, Yamada MM, Anhert D (2002) Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando obtenção de marcadores RAPD. Agrotrópica 14:31-34

Guiltinan MJ, Verica J, Zhang D, Figueira A. (2008) Genomics of *Theobroma cacao*, "the Food of the Gods". In: Moore PH, Ming R (eds) Genomics of Trop Crop Plants. Springer, New York, pp 145-170

Irish BM, Goenaga R, Zhang D, Schnell R, Steve Brown J, Motamayor JC (2010) Microsatellite Fingerprinting of the USDA-ARS Tropical Agriculture Research Station Cacao (*Theobroma cacao* L.) Germplasm Collection. Crop Sci. 50:656–667

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806

Johnson ES, Bekele FL, Brown SJ, Song Q, Zhang D, Meinhardt LW, Schnell RJ (2009) Population Structure and Genetic Diversity of the Trinitario Cacao (*Theobroma cacao* L.) from Trinidad and Tobago. Crop Sci. 49:564-572

Kemp, S (2002) PIC Calculator. <http://www.genomics.liv.ac.uk/animal/Pic1.html>. Accessed 08 march 2014

Lanaud C, Sounigo O, Amefia YK, Paulin D, Lachenau P, Clément D (1987). Nouvelles données sur le fonctionnement du système d'incompatibilité du cacaoyer et ses conséquences pour la élection. Café, Cacao, Thé 31:267–277

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJJ (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol Ecol 8:2141–2152

Lima LS, Gramacho KP, Gesteira AS, Lopes UV, Gaiotto FA, Zaidan HA, Pires JL, Cascardo JCM, Micheli F (2008) Characterization of microsatellites from cacao-*Moniliophthora perniciosa* interaction expressed sequence tags. Mol Breeding 22:315–318

Loor Solorzano RG, Risterucci AM, Courtois B, Fouet O, Jeanneau M, Rosenquist E, Amores F, Vasco A, Medina M, Lanaud C (2009) Tracing the native ancestors of modern *Theobroma cacao* L. population in Ecuador. *Tree Genet Genomes* 5 (3): 421-433

Lopes UV, Monteiro WR, Pires JL, Clement D, Yamada MM, Gramacho KP (2011) Cacao breeding in Bahia, Brazil - strategies and results. *Crop Breed Appl Biotechnol* S1: 73-81

Miller MP (1997) Tools for Population Genetic Analyses (TFPGA) 1.3: A Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author

Monteiro WR, Lopes UV, Clement D (2009) Genetic Improvement in Cocoa. In: Jain SM, Priyadarshan PM (eds) *Breeding Plantation Tree Crops: Tropical Species*. Springer, New York, pp 589-626

Monteiro WR, Ahnert D (2007) Melhoramento Genético do Cacaueiro. In: Valle RR (ed) *Ciência Tecnologia e Manejo do Cacaueiro*. Vital Ltda, Itabuna, Bahia

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89: 380-386

Motamayor JC, Risterucci AM, Heath M, Lanaud C (2003) Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. *Heredity* 91: 322-330

Motamayor JC, Lachenaud Pe, Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS One* 3: e3311

Motilal LA, Zhang D, Mischke S, Meinhardt LW Umaharan P (2013) Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa. Tree Genet Genomes 9:1395–1411

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, 83–90

Peakall R, Smouse PE (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics28:2537-2539

Pereira MG, Carletto GA, Castro GCT (1987) Variabilidade genética das características de fruto e semente em *Theobroma cacao* L., clones SIC e SIAL. Theobroma 17:209-218

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/>. Accessed 05 March 2014

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

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137-138

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Santos ESL, Cerqueira-Silva CBM, Mori GM, Ahnert D, Corrêa RX, Souza AP (2012) New polymorphic microsatellite loci for *Theobroma cacao* L.: Isolation and characterization of microsatellites from enriched genomic libraries. Biol Plantarum 56: 789-792

Sereno ML, Albuquerque PSB, Vencovsky RAF, Figueira A (2006) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conserv Genet* 7:13-24

Leal JB, dos Santos LM, dos Santos CAP, Pires JL, Ahnert D, Corrêa RX (2008) Diversidade genética entre acessos de cacau de fazendas e de banco de germoplasma na Bahia. *Pesq Agropec Bras* 43:851-858

Leite PB, Bispo ES, Santana LRR (2013) Sensory profiles of chocolates produced from cocoa cultivars resistant to *Moniliophthora perniciosa*. *Rev Bras Frutic* 35:594-602

Loor Solorzano RG, Fouet O, Lemainque A, Pavék S, Boccara M, Argout X, Amores F, Courtois B, Risterucci AM, Lanaud C (2012) Insight into the wild origin, migration and domestication history of the fine flavour nacional *Theobroma cacao* L. variety from Ecuador. *PLoS ONE* 7(11): e48438.

Thomas E, van Zonneveld M, Loo J, Hodgkin T, Galluzzi G, van Etten J (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS ONE* 7(10): e47676.

Vello F, Garcia JR, Magalhães WS (1972) Produção e Seleção de Híbridos na Bahia. *Theobroma* 2:15-35

Wright S (1978) *Evolution and the genetics of populations: Variability within and among natural populations*. University of Chicago Press, Chicago

Yamada MM, Faleiro FG, Lopes UV, Bahia RC, Pires JL, Gomes LMC, Melo GRP (2001) Genetic variability in cultivated cacao populations in Bahia, Brazil, detected by isozymes and RAPD markers. *Crop Breed Appl Biotechnol* 1:377-384

Yamada MM, Flores A B, Faleiro FG, Melo GRP, Macedo MM, Lopes UV, Corrêa RX, Santos RF (2005) Identificação e variabilidade genética de acessos de cacaueiros autocompatíveis selecionados para resistência à vassoura-de-bruxa em fazendas produtoras. Agrotropica 17:17-22



## Supplementary Material

**Supplementary Material S1** Summary statistics of allelic diversity across 30 microsatellite loci in the SIC and SIAL clones.

| Loci              | SIC/SIAL Clones |                |                |       |
|-------------------|-----------------|----------------|----------------|-------|
|                   | Na              | H <sub>o</sub> | H <sub>E</sub> | F     |
| mTc-UNICAMP01     | 2               | 0,06           | 0,06           | -0,03 |
| mTc-UNICAMP02     | 4               | 0,20           | 0,49           | 0,60  |
| mTc-UNICAMP03     | 4               | 0,30           | 0,71           | 0,58  |
| mTc-UNICAMP04     | 7               | 0,45           | 0,54           | 0,17  |
| mTc-UNICAMP05     | 4               | 0,22           | 0,66           | 0,67  |
| mTc-UNICAMP06     | 2               | 0,00           | 0,04           | 1,00  |
| mTc-UNICAMP07     | -               | -              | -              | -     |
| mTc-UNICAMP08     | 7               | 0,63           | 0,73           | 0,14  |
| mTc-UNICAMP09     | 2               | 0,00           | 0,50           | 1,00  |
| mTc-UNICAMP10     | -               | -              | -              | -     |
| mTc-UNICAMP11     | -               | -              | -              | -     |
| mTc-UNICAMP12     | 2               | 0,02           | 0,02           | -0,01 |
| mTc-UNICAMP13     | 2               | 0,06           | 0,19           | 0,69  |
| mTc-UNICAMP14     | -               | -              | -              | -     |
| mTc-UNICAMP15     | 2               | 0,02           | 0,02           | -0,01 |
| mTc-UNICAMP16     | 2               | 0,02           | 0,02           | -0,01 |
| mTc-UNICAMP17     | 3               | 0,16           | 0,15           | -0,08 |
| msEstTsh-1        | 3               | 0,18           | 0,24           | 0,27  |
| msEstTsh-6        | 2               | 0,02           | 0,02           | -0,01 |
| msEstTsh-7        | 2               | 0,35           | 0,50           | 0,29  |
| mTcCir 3          | 3               | 0,26           | 0,42           | 0,40  |
| mTcCir6           | -               | -              | -              | -     |
| mTcCir7           | 3               | 0,04           | 0,04           | -0,02 |
| mTcCir8           | -               | -              | -              | -     |
| mTcCir11          | 3               | 0,16           | 0,29           | 0,46  |
| mTcCir13          | 5               | 0,16           | 0,31           | 0,50  |
| mTcCir15          | 3               | 0,16           | 0,53           | 0,71  |
| mTcCir18          | -               | -              | -              | -     |
| mTcCir25          | 2               | 0,02           | 0,02           | -0,01 |
| mTcCir26          | 2               | 0,02           | 0,02           | -0,01 |
| Total             | 71              | -              | -              | -     |
| Mean <sup>a</sup> | 3,09            | 0,15           | 0,28           | 0,32  |

<sup>a</sup>Mean obtained considering only polymorphic loci  
*Na* number of alleles, *H<sub>o</sub>* observed heterozygosity, *H<sub>E</sub>* expected heterozygosity and *F* fixation index



**Supplementary Material S2** Frequency of private alleles per locus obtained in each group considering SIC, SIAL and Modern clones.

| Group         | Locus         | Allele | Frequency |
|---------------|---------------|--------|-----------|
| SIAL          | mTcCir7       | 154    | 0,025     |
| SIAL          | mTcCir26      | 296    | 0,025     |
| SIC           | mTc-UNICAMP02 | 273    | 0,129     |
| SIC           | mTc-UNICAMP06 | 151    | 0,032     |
| SIC           | mTc-UNICAMP05 | 232    | 0,017     |
| Modern Clones | mTc-UNICAMP01 | 174    | 0,010     |
| Modern Clones | mTc-UNICAMP01 | 190    | 0,039     |
| Modern Clones | mTc-UNICAMP01 | 198    | 0,069     |
| Modern Clones | mTc-UNICAMP01 | 205    | 0,029     |
| Modern Clones | mTc-UNICAMP01 | 208    | 0,078     |
| Modern Clones | mTc-UNICAMP01 | 214    | 0,059     |
| Modern Clones | mTc-UNICAMP02 | 260    | 0,077     |
| Modern Clones | mTc-UNICAMP02 | 277    | 0,106     |
| Modern Clones | mTc-UNICAMP02 | 285    | 0,058     |
| Modern Clones | mTc-UNICAMP02 | 290    | 0,106     |
| Modern Clones | mTc-UNICAMP02 | 295    | 0,067     |
| Modern Clones | mTc-UNICAMP02 | 300    | 0,048     |
| Modern Clones | mTc-UNICAMP02 | 305    | 0,029     |
| Modern Clones | mTc-UNICAMP03 | 270    | 0,031     |
| Modern Clones | mTc-UNICAMP03 | 275    | 0,184     |
| Modern Clones | mTc-UNICAMP03 | 277    | 0,010     |
| Modern Clones | mTc-UNICAMP03 | 283    | 0,020     |
| Modern Clones | mTc-UNICAMP07 | 262    | 0,059     |
| Modern Clones | mTc-UNICAMP07 | 288    | 0,216     |
| Modern Clones | mTc-UNICAMP07 | 298    | 0,020     |
| Modern Clones | mTc-UNICAMP07 | 312    | 0,039     |
| Modern Clones | mTc-UNICAMP09 | 154    | 0,222     |
| Modern Clones | mTc-UNICAMP09 | 158    | 0,222     |
| Modern Clones | mTc-UNICAMP09 | 178    | 0,022     |
| Modern Clones | mTc-UNICAMP09 | 188    | 0,144     |
| Modern Clones | mTc-UNICAMP09 | 198    | 0,022     |
| Modern Clones | mTc-UNICAMP15 | 282    | 0,010     |
| Modern Clones | mTc-UNICAMP16 | 244    | 0,069     |
| Modern Clones | mTc-UNICAMP16 | 248    | 0,176     |
| Modern Clones | mTc-UNICAMP17 | 296    | 0,088     |
| Modern Clones | mTc-UNICAMP17 | 304    | 0,069     |
| Modern Clones | mTc-UNICAMP17 | 318    | 0,020     |

|               |               |     |       |
|---------------|---------------|-----|-------|
| Modern Clones | mTc-UNICAMP14 | 254 | 0,200 |
| Modern Clones | mTc-UNICAMP14 | 256 | 0,070 |
| Modern Clones | mTc-UNICAMP14 | 260 | 0,060 |
| Modern Clones | mTc-UNICAMP14 | 262 | 0,010 |
| Modern Clones | mTc-UNICAMP13 | 141 | 0,010 |
| Modern Clones | mTc-UNICAMP06 | 140 | 0,049 |
| Modern Clones | mTc-UNICAMP06 | 153 | 0,098 |
| Modern Clones | mTc-UNICAMP06 | 158 | 0,010 |
| Modern Clones | mTc-UNICAMP06 | 162 | 0,078 |
| Modern Clones | mTc-UNICAMP06 | 165 | 0,039 |
| Modern Clones | mTc-UNICAMP06 | 168 | 0,010 |
| Modern Clones | mTc-UNICAMP06 | 170 | 0,049 |
| Modern Clones | mTc-UNICAMP08 | 262 | 0,032 |
| Modern Clones | mTc-UNICAMP08 | 266 | 0,032 |
| Modern Clones | mTc-UNICAMP08 | 276 | 0,021 |
| Modern Clones | mTc-UNICAMP04 | 152 | 0,011 |
| Modern Clones | mTc-UNICAMP04 | 172 | 0,023 |
| Modern Clones | mTc-UNICAMP04 | 174 | 0,023 |
| Modern Clones | mTc-UNICAMP04 | 176 | 0,045 |
| Modern Clones | mTc-UNICAMP04 | 186 | 0,011 |
| Modern Clones | mTc-UNICAMP04 | 190 | 0,023 |
| Modern Clones | mTc-UNICAMP04 | 194 | 0,011 |
| Modern Clones | mTc-UNICAMP04 | 196 | 0,011 |
| Modern Clones | mTc-UNICAMP05 | 230 | 0,010 |
| Modern Clones | mTc-UNICAMP05 | 238 | 0,030 |
| Modern Clones | mTc-UNICAMP05 | 244 | 0,080 |
| Modern Clones | mTc-UNICAMP05 | 246 | 0,030 |
| Modern Clones | mTc-UNICAMP05 | 248 | 0,010 |
| Modern Clones | mTc-UNICAMP05 | 252 | 0,010 |
| Modern Clones | mTc-UNICAMP10 | 232 | 0,100 |
| Modern Clones | mTc-UNICAMP10 | 236 | 0,070 |
| Modern Clones | mTc-UNICAMP11 | 296 | 0,144 |
| Modern Clones | mTc-UNICAMP12 | 240 | 0,190 |
| Modern Clones | mTc-UNICAMP12 | 250 | 0,040 |
| Modern Clones | mTc-UNICAMP12 | 258 | 0,130 |
| Modern Clones | mTc-UNICAMP12 | 260 | 0,110 |
| Modern Clones | mTc-UNICAMP12 | 262 | 0,030 |
| Modern Clones | mTc-UNICAMP12 | 280 | 0,020 |
| Modern Clones | mTc-UNICAMP12 | 288 | 0,010 |
| Modern Clones | mTc-UNICAMP12 | 292 | 0,010 |
| Modern Clones | msEstTsh-7    | 149 | 0,190 |
| Modern Clones | msEstTsh-7    | 153 | 0,150 |
| Modern Clones | msEstTsh-7    | 155 | 0,060 |

|               |            |     |       |
|---------------|------------|-----|-------|
| Modern Clones | msEstTsh-1 | 214 | 0,019 |
| Modern Clones | msEstTsh-6 | 208 | 0,184 |
| Modern Clones | msEstTsh-6 | 210 | 0,102 |
| Modern Clones | msEstTsh-6 | 224 | 0,061 |
| Modern Clones | mTcCir 3   | 210 | 0,167 |
| Modern Clones | mTcCir 3   | 214 | 0,038 |
| Modern Clones | mTcCir 3   | 228 | 0,128 |
| Modern Clones | mTcCir 3   | 230 | 0,051 |
| Modern Clones | mTcCir 3   | 238 | 0,051 |
| Modern Clones | mTcCir6    | 228 | 0,178 |
| Modern Clones | mTcCir6    | 234 | 0,022 |
| Modern Clones | mTcCir6    | 238 | 0,011 |
| Modern Clones | mTcCir6    | 240 | 0,022 |
| Modern Clones | mTcCir6    | 248 | 0,011 |
| Modern Clones | mTcCir6    | 250 | 0,189 |
| Modern Clones | mTcCir7    | 160 | 0,194 |
| Modern Clones | mTcCir7    | 162 | 0,143 |
| Modern Clones | mTcCir7    | 166 | 0,010 |
| Modern Clones | mTcCir8    | 292 | 0,143 |
| Modern Clones | mTcCir8    | 300 | 0,020 |
| Modern Clones | mTcCir8    | 308 | 0,276 |
| Modern Clones | mTcCir11   | 290 | 0,087 |
| Modern Clones | mTcCir11   | 300 | 0,196 |
| Modern Clones | mTcCir11   | 302 | 0,033 |
| Modern Clones | mTcCir11   | 306 | 0,022 |
| Modern Clones | mTcCir11   | 318 | 0,120 |
| Modern Clones | mTcCir15   | 236 | 0,135 |
| Modern Clones | mTcCir15   | 238 | 0,292 |
| Modern Clones | mTcCir15   | 240 | 0,021 |
| Modern Clones | mTcCir15   | 244 | 0,010 |
| Modern Clones | mTcCir15   | 246 | 0,010 |
| Modern Clones | mTcCir15   | 248 | 0,031 |
| Modern Clones | mTcCir15   | 258 | 0,042 |
| Modern Clones | mTcCir25   | 134 | 0,054 |
| Modern Clones | mTcCir25   | 138 | 0,076 |
| Modern Clones | mTcCir25   | 142 | 0,011 |
| Modern Clones | mTcCir25   | 144 | 0,022 |
| Modern Clones | mTcCir25   | 156 | 0,109 |
| Modern Clones | mTcCir25   | 160 | 0,022 |
| Modern Clones | mTcCir25   | 168 | 0,011 |
| Modern Clones | mTcCir26   | 286 | 0,039 |
| Modern Clones | mTcCir26   | 288 | 0,029 |
| Modern Clones | mTcCir26   | 300 | 0,235 |

|               |          |     |       |
|---------------|----------|-----|-------|
| Modern Clones | mTcCir26 | 302 | 0,029 |
| Modern Clones | mTcCir26 | 318 | 0,010 |
| Modern Clones | mTcCir13 | 248 | 0,010 |
| Modern Clones | mTcCir13 | 250 | 0,078 |
| Modern Clones | mTcCir13 | 262 | 0,324 |
| Modern Clones | mTcCir13 | 264 | 0,020 |
| Modern Clones | mTcCir18 | 330 | 0,382 |
| Modern Clones | mTcCir18 | 335 | 0,010 |
| Modern Clones | mTcCir18 | 350 | 0,029 |



## RESULTADOS COMPLEMENTARES

O cacau da Bahia tem grande valor para a conservação dado o fato da Bahia ser considerada uma zona de reserva secundária da diversidade do cacau (BARTLEY, 2005). Estas plantas possuem ainda qualidades úteis para aplicação no melhoramento genético do cacaueiro como produtividade e flavor apreciado na produção de chocolate 'gourmet' (PIRES, 2003; MONTEIRO e AHNERT, 2007). Tendo em vista o fato de que o cacau possui sementes recalcitrantes e sua conservação *ex situ* se dá pela manutenção da planta *in vivo* em bancos de germoplasma, a avaliação da diversidade genético-molecular a fim de otimizar a utilização do espaço com plantas que amostrem ao máximo a variabilidade existente não pode ser negligenciada. Neste sentido, a disponibilidade de plantas divergentes geneticamente amplia as opções para o melhorista no que tange a seleção das plantas que serão utilizadas nos ciclos de cruzamento. Como consequência, plantas melhoradas e com base genética diversificada reduz as chances de perdas expressivas por fatores inesperados como doenças.

Com intuito de permitir uma visão geral da diversidade presente nas 215 plantas de cacaueiro classificadas nesta tese enquanto cacau da Bahia e promover uma reflexão acerca da sua conservação a luz da sua representatividade em bancos de germoplasma, foram apresentados resultados de algumas análises realizadas a partir da genotipagem de 279 plantas utilizando 30 marcadores microssatélites referidos na tese (Tabela 2, capítulo 2). Das plantas mantidas nas avaliações empreendidas ao longo deste trabalho, 176 foram genótipos coletados em fazendas de quatro municípios (Canavieiras, Uruçuca, Camacan e Gandu, [ver Figura 1, Capítulo 2] e classificados a priori como cacau da Bahia. Após análise por meio do programa STRUCTURE (PRITCHARD et al., 2000) apenas 164 destas foram de "fato consideradas" como cacau da Bahia, outros seis como genótipos não baianos e seis como de origem híbrida entre os dois primeiros. Outras 103 plantas constituíram-se de clones silvestres e cultivados obtidos no banco de germoplasma do Centro de Pesquisas do Cacau na Comissão Executiva do Plano da Lavoura Cacaueira (CEPEC/CEPLAC) com sede em Ilhéus e nas Fazendas Reunidas Vale do Juliana, no município de Igrapiúna. Desta forma, além

das 164 plantas de fazendas, foram considerados como cacau da Bahia 51 clones (obtidos no CEPEC/CEPLAC) tidos como representantes das variedades baianas em bancos de germoplasma, sendo 31 pertencentes à série SIC (Seleção do Instituto do Cacau) e 20 à série SIAL (Seleção do Instituto Agrônômico do Leste) (ver Tabela 1, Capítulo 3). Para fins de comparação, 58 plantas de cacau ‘não baiano’ serão aqui empregados, sendo 6 genótipos obtidos em fazendas e 52 clones (aqui denominados como clones referência) obtidos no CEPEC/CEPLAC e nas Fazendas Reunidas do Vale do Juliana. Além disso, os seis genótipos considerados como de origem misturada (Capítulo 2) serão também utilizados em algumas análises.

A análise frequentista obtida por meio do programa GenAlEx v.6.1 (PEAKALL e SMOUSE, 20012) e mostrada na Tabela 2 indicou haver baixa variabilidade para o grupo de plantas de cacau da Bahia, bem como para os subgrupos destes indicados como cacau de fazenda, SIC e SIAL. Entre os valores que mais chamam atenção estão a estimativa de heterozigosidade esperada ( $H_E$ ) e o índice de fixação ( $F$ ). Para o estimador de diversidade  $H_E$  os valores abaixo de 0,5 indicam uma reduzida expectativa de heterozigotos, o que reflete diretamente a baixa quantidade de alelos observados no conjunto das plantas avaliadas. Portanto, mesmo considerando o cacau da Bahia como um único grupo, os valores de  $H_E$  estão bastante baixos. No entanto, é interessante notar que o valor de  $H_E$  do grupo SIC é ainda maior do que o valor médio observado para todos os clones de cacau da Bahia.

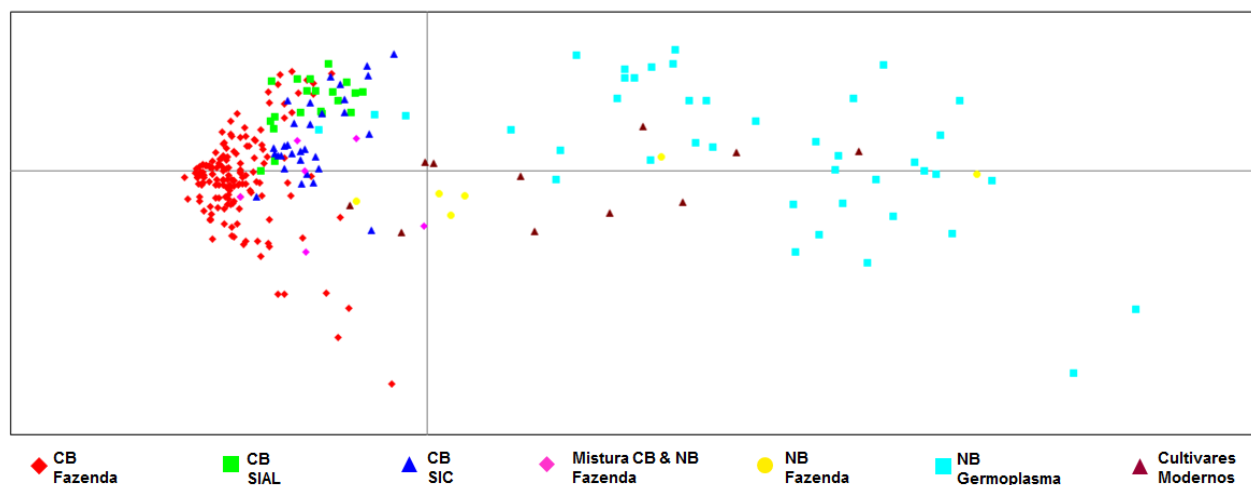
**Tabela 2.** Estatística descritiva obtida a partir de dados de genotipagem de cacaueiros com 30 marcadores microssatélites.

| <b>Grupos / Subgrupos</b>          | <b>N</b> | <b>Na</b> | <b>H<sub>0</sub></b> | <b>H<sub>E</sub></b> | <b>F</b> |
|------------------------------------|----------|-----------|----------------------|----------------------|----------|
| Cacau da Bahia                     | 215      | 3,30      | 0,07                 | 0,17                 | 0,36     |
| <b>Subgrupos do cacau da Bahia</b> |          |           |                      |                      |          |
| Fazenda                            | 164      | 2,23      | 0,06                 | 0,14                 | 0,38     |
| SIAL                               | 20       | 2,10      | 0,11                 | 0,16                 | 0,21     |
| SIC                                | 31       | 2,37      | 0,12                 | 0,22                 | 0,37     |
| <b>Cacau não baiano</b>            |          |           |                      |                      |          |
| Fazenda – Outros                   | 6        | 3,47      | 0,53                 | 0,52                 | -0,06    |
| Clones referência                  | 52       | 6,70      | 0,50                 | 0,60                 | 0,16     |

*N* número de plantas avaliadas; *Na* número médio de alelos por loco; *H<sub>0</sub>* heterozigosidades observada; *H<sub>E</sub>* heterozigosidade esperada; *F* índice de fixação.

Por outro lado, o valor estimado para  $F$  é positivo e relativamente alto para o cacau da Bahia (0,36) em contraste com o material não baiano de fazenda (-0,06) e para o grupo de clones referência (0,16). Como já relatado na literatura, os valores positivos de  $F$  são indicativos de déficit de heterozigotos observados, sendo este um reflexo da endogamia. Neste caso, o valor mais baixo de  $F$  obtido para SIAL, entre os grupos de cacau da Bahia analisados, reflete a maior proximidade entre os valores de  $H_E$  e  $H_0$ . O conjunto de resultados observados até aqui pode ser explicado a luz da história de domesticação que originou o grupo genético de cacau denominado como Amelonado (do qual as variedades baianas são originadas) (SERENO et al., 2006), bem como a restrita base genética sob a qual foi fundado o cacau da Bahia. Adicionalmente, o fato das plantas cultivadas de cacau geralmente serem autocompatíveis, explica também os valores de  $F$  relativamente altos e positivos, encontrados entre as plantas do grupo Amelonado.

As 279 plantas empregadas nos estudos de diversidade genética ao longo da tese foram utilizadas em uma Análise de Componentes Principais a partir do programa GenAEx v.6.1 (PEAKALL e SMOUSE, 2012). Como pode ser visualizado na Figura 6, houve uma tendência de agrupamento dos materiais considerados como cacau da Bahia, tanto de fazenda quanto de germoplasma (SIC e SIAL), o que comprova a proximidade destes materiais.

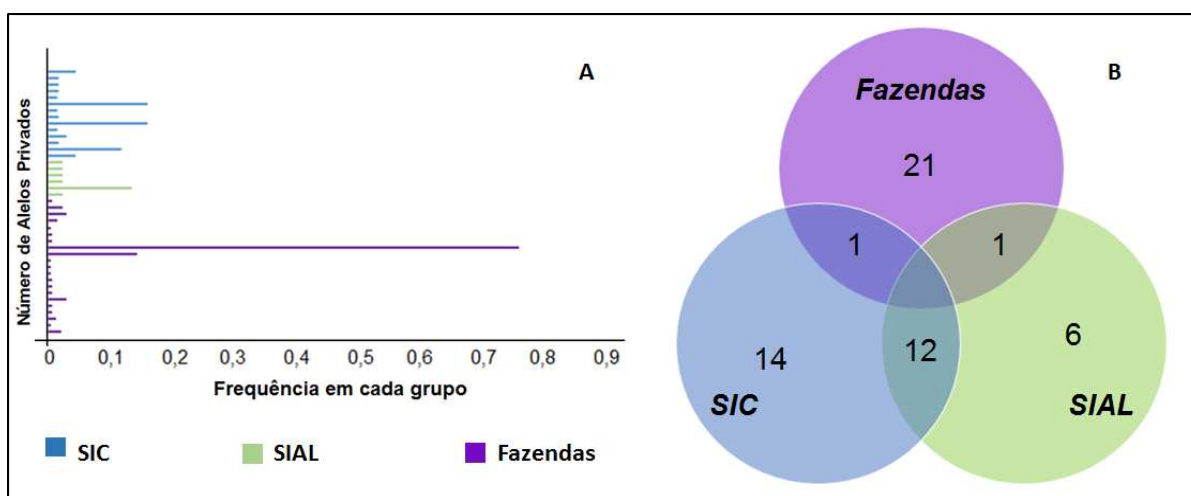


**Figura 6.** Análise de Coordenadas principais a partir de dados de genotipagem de 279 cacaueiros com 30 marcadores microsatélites.

**CB** cacau da Bahia; **NB** Cacau não baiano



Por outro lado, a dispersão deste grupo em relação às plantas consideradas como cacau não baiano deixa claro a menor variabilidade do cacau da Bahia corroborando os resultados obtidos pela análise bayesiana, sendo que seis plantas coletadas em fazendas não são cacau da Bahia. É imperativo notar que, a despeito da maior proximidade entre os clones SIC e SIAL e das plantas de cacau da Bahia obtidas em fazendas, estes grupos não estão completamente relacionados, isto é, a região de dispersão dos clones SIC e SIAL é mais restrita do que a alcançada pelas plantas de fazenda. A análise de alelos privados entre os principais grupos de cacau da Bahia – Fazendas, SIC e SIAL (Figura 7), também corrobora com ideia de que a variabilidade do cacau da Bahia não se encontram regularmente distribuída entre os grupos considerados.



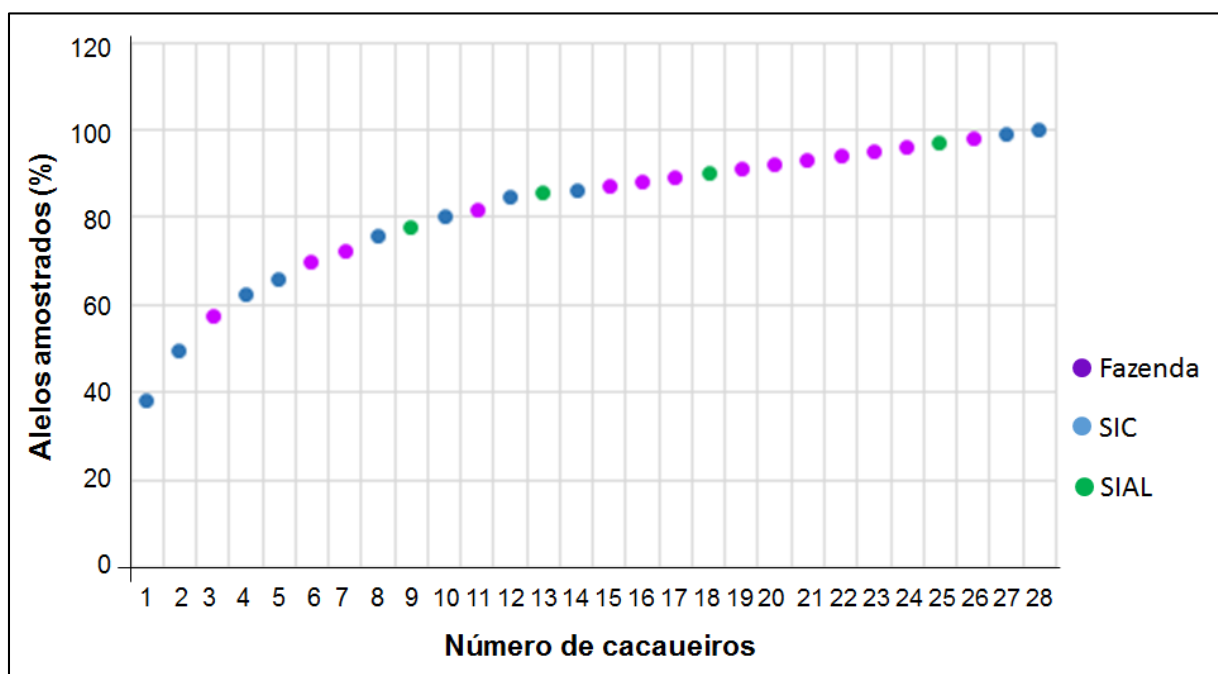
**Figura 7.** Alelos privados observados nos grupos formados por cacau da Bahia. **(A)** Gráfico com a frequência de para cada alelo privado. **(B)** Distribuição dos alelos entre os grupos de cacau da Bahia.

Neste sentido, é possível inferir que a variabilidade genética amostrada atualmente em bancos de germoplasma de cacau da Bahia contempla apenas uma fração daquela realmente existente. Coletas direcionadas de plantas de cacau presentes em fazendas permitirão uma amostragem mais real da variabilidade do cacau da Bahia. Por outro lado, a pequena fração de variabilidade ocorrendo nas plantas de

germoplasma e não presente em fazendas pode ser consequência da seleção de plantas por ocasião da vassoura-de-bruxa, a partir de 1989, tendo sido mantidas apenas aquelas mais tolerantes a esta enfermidade. Assim, alguns alelos mesmo que não associados à resistência, podem ter sido indiretamente perdidos da população de plantas presentes em fazendas.

Em atividades que envolvam a seleção a priori de plantas para compor uma coleção de trabalho, a amostragem da diversidade genética pode ser um ponto relevante a ser considerado. Deste modo, pensando em plantas de cacau da Bahia para compor uma coleção nuclear, o programa COREFINDER (CIPRIANI et al., 2010) foi empregado tendo sido avaliadas as 215 plantas em questão (Figura 8). Os resultados apontaram para a necessidade de no mínimo 28 plantas de cacaueiros para amostrar 100% da variabilidade existente no cacau da Bahia. A partir da análise destas plantas foi observado que precisamente metade são plantas amostradas em fazenda e metade de germoplasma, sendo 4 destas plantas da série SIAL e 10 da série SIC.

Os resultados obtidos durante o desenvolvimento deste trabalho são relevantes para o conhecimento da base genética do cacau da Bahia, constituindo uma fonte de informação importante para a conservação e melhoramento genético do cacaueiro.



**Figura 8.** Coleção nuclear amostrando o total de alelos presentes em 215 cacaueiros da Bahia obtidos em fazendas do sul da Bahia e coleção de germoplasma (SIC e SIAL).

## CONCLUSÕES

Os marcadores microssatélites utilizados neste estudo foram eficientes para reconstruir a história do cacau da Bahia, tanto por possibilitar acessar sua diversidade quanto por permitir encontrar estruturas compatíveis com o histórico de coleta e com o tipo de material avaliado.

O cacau da Bahia apresenta baixa diversidade genético-molecular, sendo as plantas que compõem este grupo, bastante relacionadas. Este fato reflete as restritas introduções ocorridas a partir de sementes trazidas da Amazônia para a Bahia, o que acabou servindo de base para a formação de quase todas as fazendas de cacau do estado.

A despeito da baixa diversidade genético-molecular observada, os resultados obtidos fornecem indícios de que as plantas de cacau da Bahia presentes em fazendas possuem características agronômicas de interesse. Neste sentido, os resultados embora preliminares, possibilitam indicar plantas prioritárias para avaliações adicionais e seu provável futuro emprego em programas de melhoramento genético do cacaueiro.

A diversidade do cacau da Bahia em bancos de germoplasma está subrepresentada, uma vez que parte da análise indicou haver composições alélicas em fazendas não amostrada em bancos de germoplasma, sendo portanto, justificável e altamente recomendado a realização de novas coletas de plantas divergentes presentes em fazendas.

A variação encontrada no cacau da Bahia presente em germoplasma e não representada em plantas de fazenda, permite inferir acerca da influência da vassoura-de-bruxa na manutenção de plantas de cacau com maior nível de tolerância à doença nas fazendas, em detrimento a plantas mais susceptíveis presentes no germoplasma. Neste sentido, é possível que alguns alelos, mesmo não relacionados à doença tenham sido perdidos por deriva genética.

A partir dos resultados obtidos, associado às informações obtidas com os produtores durante as coletas, bem como ao relatado na literatura, é possível sugerir o uso das plantas do cacau da Bahia para emprego em programas de melhoramento genético devido às suas boas características físico-químicas, sua produtividade e sua adaptação ao ambiente de cultivo baiano.

## PERSPECTIVAS

Enquanto parte dos objetivos do grupo de pesquisa em genética e melhoramento do cacau, interessado em informações genéticas acerca do cacau, cujo intuito é conhecer sua base genética e estabelecer o melhoramento da cultura, o trabalho desenvolvido nesta tese traz informações que servirão como alicerce para que sejam alcançadas as metas do grupo.

Neste sentido, pretende-se ampliar as avaliações moleculares para todos os clones das séries SIC e SIAL presentes no banco de germoplasma do CEPEC/CEPLAC. Além disso, novas avaliações fenotípicas deverão ser realizadas em plantas de cacau da Bahia presentes em fazendas, cujos resultados obtidos nesta tese apontaram como candidatas a serem incluídas nos programas de melhoramento genético. Estas plantas e outras que em conjunto amostram a variabilidade genética do cacau da Bahia em fazenda serão ainda utilizadas para formar uma coleção núcleo (*core collection*) a ser estabelecida na Universidade Estadual de Santa Cruz em Ilhéus, Bahia, como material base para conservação e trabalhos de melhoramento genético do cacau da Bahia, de importância fundamental para o futuro da cacauicultura na região Sudeste da Bahia.



**LITERATURA CITADA**

AIKPOKPODION PO, MOTAMAYOR JC, ADETIMIRIN VO, ADU-AMPOMAH Y, INGELBRECHT I, ESKES AB, SCHNELL RJ, KOLESNIKOVA-ALLEN M (2009) Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genetics & Genomes* 5:699–711.

AIKPOKPODION PO (2010) Variation in agro-morphological characteristics of cacao, *Theobroma cacao* L., in farmers' fields in Nigeria. *New Zealand Journal of Crop and Horticultural Science*, 38: 157-170.

ALLEGRE M, ARGOUT X, BOCCARA M, FOUET O, ROGUET Y, BÉRARD A, THÉVENIN JM, CHAUVEAU A, RIVALLAN R, CLEMENT D, COURTOIS B, GRAMACHO K, BOLAND-AUGÉ A, TAHI M, UMAHARAN P, BRUNEL D, LANAUD C (2012) Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. *DNA Research* 19: 23–35.

ALLEN JB Geographical variation and population biology in wild *Theobroma cacao*, Ph.D. thesis, University of Edinburgh, 1988.

ARAÚJO, I.S.; SOUZA FILHO, G.A.; PEREIRA, M.G.; FALEIRO, F.G.; QUEIROZ, V.T.; GUIMARÃES, C.T.; MOREIRA, M.A.; BARROS, E.G.; MACHADO, R.C.R.; PIRES, J.L.; SCHENELL, R.; LOPES, U.V (2009). Mapping of Quantitative Trait Loci for Butter Content and Hardness in Cocoa Beans (*Theobroma cacao* L.). *Plant Mol Biol Rep* 27:177–183.

ARAÚJO, I.S.; INTORNE, A.C.; PEREIRA, M.G.; LOPES, U.V.; SOUZA FILHO, G.A (2007) Development and characterization of novel tetra-tri- and di-nucleotide microsatellite markers in cacao (*Theobroma cacao* L.). *Mol Breeding*, 20: 73–81



ARGOUT, X; SALSE, J; AURY, J; GUILTINAN, MJ; DROC G. et al (2011) The genome of *Theobroma cacao*. Nature Genetics, doi:10.1038/ng.736.

ARAÚJO M, ALGER K, ROCHA R, MESQUITA CAB (1998) A Mata Atlântica do sul da Bahia: situação atual, ações e perspectivas. Reserva da Biosfera da Mata Atlântica – MAB – UNESCO. Caderno 8:1–36

ALVIM PT (1984) Flowering of cocoa. Cocoa Growers Bull 35:22–31

BARTLEY GD (2005) The genetic diversity of cacao and its utilization. Wallingford, UK: CABI Publishing.

BEKELE FL, BEKELE I, BUTLER DR, BIDAISEE GG (2006) Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. Genetic Resources and Crop Evolution 53: 933-948.

BHATTACHARJEE R, KUMAR PL CACAO. In: C. KOLE (Ed.) Genome Mapping and Molecular Breeding in Plants, Vol. 6. Springer-Verlag: Berlin Heidelberg, 2007.

BOWERS, J. H., BAILEY, B. A., HEBBAR, P. K., SANOGO, S., LUMSDEN, R. D. 2001. The impact of plant diseases on world chocolate production. Online. Plant Health Progress doi:10.1094/PHP-2001-0709-01-RV.

CacaoNet. 2012. A Global Strategy for the Conservation and Use of Cacao Genetic Resources, as the Foundation for a Sustainable Cocoa Economy (B. Laliberté, compiler). Bioversity International, Montpellier, France.

CARLETTO GA. Expedição Internacional à Amazônia Equatoriana para coleta de material botânico de cacau. Theobroma 3:41-47, 1973.

BOZA EJ, Irish BM, MEEROW AW, TONDO CL, RODRÍGUEZ AO, VENTURA-LÓPEZ M, GÓMEZ JA, MOORE JM, ZHANG D, MOTAMAYOR JC, SCHNELL RJ (2013) Genetic diversity, conservation, and utilization of *Theobroma cacao* L.: genetic resources in the Dominican Republic. *Genet Resour Crop Evol* 60:605–619.

CHEESMAN EE (1944). Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop Agricult* 21: 144–159.

CLEMENT CR, CRISTO-ARAÚJO M, D'EECKENBRUGGE GC, PEREIRA AA, PICANÇO-RODRIGUES D (2010) Origin and Domestication of Native Amazonian Crops. *Diversity* 2010, 2, 72-106.

CUATRECASAS J (1964). Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contrib US Herbarium* 35: 379–614.

DIAS LAS (2001) Melhoramento genético do cacaueiro. Viçosa: Funape - UFP.

ENGELS JMM, BARTLEY BGD and ENRIQUEZ GA (1980) Cacao descriptors, their states and modus operandi. *Turrialba (Costa Rica)* 30: 209–218.

FALEIRO FG, SANTOS IS, BAHIA RCS, SANTOS RF, YAMADA MM, ANHERT D (2002) Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando obtenção de marcadores RAPD. *Agrotrópica* 14: 31-34.

FALEIRO FG, PIRES JL, MONTEIRO WR, LOPES UV, YAMADA MM, PIEDRA AG, MOURA AD, ARÉVALO-GARDINI E, MARQUES JRB, GRAMACHO KP, FALEIRO ASG, SANTOS MCM (2004) Variability in cacao accessions from the Brazilian, Ecuadorian, and Peruvian Amazons based on molecular markers. *Crop Breeding and Applied Biotechnology* 4: 227-233.

FALEIRO FG, QUEIROZ VT, LOPES UV, GUIMARÃES CT, PIRES JL, YAMADA MM, ARAÚJO IS, PEREIRA MG, SOUZA Filho GA, BROWN JS, SCHNELL R, FERREIRA CF, BARROS EG and MOREIRA MA (2006) Mapping QTLs for witches' broom (*Crinipellis pernicioso*) resistance in cacao (*Theobroma cacao* L.). *Euphytica* 149: 227-235.

FANG W, MEINHARDT LW, MISCHKE S, BELLATO CM, MOTILAL L, ZHANG D (2014) Accurate determination of genetic identity for a single cacao bean, using molecular markers with a Nanofluidic System, ensures Cocoa Authentication. *Journal of Agriculture Food Chemical* 62: 481–487.

FARIA D, PACIENCIA MLB, DIXO M, LAPS RR, BAUMGARTEN J (2007) Ferns, frogs, lizards, birds and bats in forest fragments and shade cacao plantations in two contrasting landscapes in the Atlantic forest, Brazil. *Biodiversity Conservation* 16: 2335–2357.

FIGUEIREDO GSF, 2008 Análise Cariotípica em cacaueiro (*Theobroma cacao* L.) e em mutantes espontâneos. Tese de Doutorado. Ilhéus, UESC, 57 p.

GÓMEZ-POMPA A, FLORES JS, FERNANDEZ MA (1990). The sacred cacao groves of the Maya. *Latin Am Antiquity* 1: 247–257.

GUILTINAN M (2007) Cacao. In: Pua EC, Davey MR (eds) *Biotechnology in Agriculture and Forestry - Transgenic Crops VI*. Springer-Verlag, Berlin Heidelberg (in press).

GUILTINAN MJ, VERICA J, ZHANG D, FIGUEIRA A. (2008) Chapter 6. Genomics of *Theobroma cacao*, "the Food of the Gods". In: MOORE PH, MING R (eds.). *Genomics of Tropical Crop Plants*.

HENDERSON JS, JOYCE RA, HALL GR, HURST WJ, MCGOVERN PE (2007) Chemical and archaeological evidence for the earliest cacao beverages. Proc. Natl. Acad. Sci. 104: 18937-18940.

HURST WJ, TARKA SM, POWIS TG, VALDEZ F, HESTER RT (2002) Archaeology: cacao usage by the earliest Maya civilization. Nature 418: 289-290.

IBGE - Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal, 2012. [http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam\\_2012\\_v39\\_br.pdf](http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam_2012_v39_br.pdf)

ICCO (International Cocoa Organization). Annual Report, 2013, London United Kingdom [www.icco.org](http://www.icco.org)

IWARO AD, BEKELE FL, BUTLER DR (2003) Evaluation and utilisation of cacao (*Theobroma cacao* L.) germplasm at the international cacao genebank, Trinidad. Euphytica 130:207–221

IRISH BM, GOENAGA R, ZHANG D, SCHNELL R, STEVE BROWN J, MOTAMAYOR JC (2010) Microsatellite Fingerprinting of the USDA-ARS Tropical Agriculture Research Station Cacao (*Theobroma cacao* L.) Germplasm Collection. Crop Sci. 50:656–667.

JI, K; ZHANG, D.; MOTILAL, LA; BOCCARA, M; LACHENAUD, P; MEINHARDT, LW (2013) Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) Markers. Genet Resour Crop Evol, 60: 441-453.

JOHNSON ES, BEKELE FL, BROWN SJ, SONG Q, ZHANG D, MEINHARDT LW, SCHNELL RJ (2009) Population Structure and Genetic Diversity of the Trinitario Cacao (*Theobroma cacao* L.) from Trinidad and Tobago. Crop Sci. 49:564-572.

KUHN DN.; LIVINGSTONE III D.; MAIN D.; ZHENG, P.; SASKI, C.; FELTUS, FA.; MOCKAITIS, K.; FARMER, AD.; MAY, GD.; SCHNELL, RJ.; MOTAMAYOR, JC (2012).

Identification and mapping of conserved ortholog set (COS) II sequences of cacao and their conversion to SNP markers for marker-assisted selection in *Theobroma cacao*.and comparative genomics studies. Tree Genetics & Genomes 8:97–111.

LANAUD C, RISTERUCCI AM, PIERETTI I, FALQUE M, BOUET A, LAGODA PJL (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. Molecular Ecology 8: 2141–2152.

LANAUD C.; FOUET O.; CLEMENT D.; BOCCARA M.; RISTERUCCI AM.; SURUJDEO-MAHARAJ S.; Legavre T.; Argout X. (2009) A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. Mol Breeding. DOI 10.1007/s11032-009-9297-4

LANAUD C, RISTERUCCI AM, PIERETTI I, N'GORAN JAK, FARGEAS D (2004) Characterisation and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao* L.). Mol Breed 13:211–227. doi:10.1023/B:MOLB. 0000022515.23880.1b

LACHENAUD P, ZHANG D (2008) Genetic diversity and population structure in wild stands of cacao trees (*Theobroma cacao* L.) in French Guiana. Ann. For. Sci. 65 (2008) 310p1-p7.

LACHENAUD P, ZHANG D (2008) Genetic diversity and population structure in wild stands of cacao trees (*Theobroma cacao* L.) in French Guiana. Ann. For. Sci. 65-310 (7p). DOI: 10.1051/forest:2008011.

LATIF, R (2013) Chocolate/cocoa and human health: a review. The Journal of Medicine. 71: 63-68.

LEITE PB, MACIEL LF, OPRETZKA LCF, SOARES SE, BISPO ES (2013) Phenolic compounds, methylxanthines and antioxidant activity in cocoa mass and

chocolates produced from “witch broom disease” resistant and non resistant cocoa cultivars. *Ciência e agrotecnologia* 37: 244-250.

LERCETEAU E, ROBERT T, PETIARD V, CROUZILLAT D (1997) Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. *Theoretical and Applied Genetics* 95: 10-19.

LIMA EM; PEREIRA NE; PIRES JL; BARBOSA AMM; CORRÊA RX (2013) Genetic molecular diversity, production and resistance to witches’ broom in cacao clones. *Crop Breeding and Applied Biotechnology* 13: 127-135.

LIMA LS, GRAMACHO KP, PIRES JL, CLEMENT D, LOPES UV, CARELS N, GESTEIRA AS, GAIOTTO FA, CASCARDO JC, MICHELI F (2010) Development, characterization, validation, and mapping of SSRs derived from *Theobroma cacao* L. *Moniliophthora perniciosa* interaction ESTs. *Tree Genetics & Genomes* 6: 663-676.

LIMA LS, GRAMACHO KP, GESTEIRA AS, LOPES UV, GAIOTTO FA, ZAIDAN HA, PIRES JL, Cascardo JCM, Micheli F (2008) Characterization of microsatellites from cacao–*Moniliophthora perniciosa* interaction expressed sequence tags. *Molecular Breeding* 22: 315–318.

LOPES UV, MONTEIRO WR, PIRES JL, CLEMENT D, YAMADA MM, GRAMACHO KP (2011) *Crop Breeding and Applied Biotechnology* S1: 73-81.

LOOR SOLORZANO RG, RISTERUCCI AM, COURTOIS B, Fouet O, JEANNEAU M, Rosenquist E, Amores F, Vasco A, Medina M, Lanaud C (2009) Tracing the native ancestors of modern *Theobroma cacao* L. population in Ecuador. *Tree genetics and genomes*, 5 (3): 421–433.

MARITA, JM.; NIENHUIS, J.; PIRES, JL.; AITKEN, JM (2001). Analysis of Genetic Diversity in *Theobroma cacao* with Emphasis on Witches’ Broom Disease Resistance. *Crop Sci.* 41:1305–1316.

MEDEIROS FHV, POMELLA AWW, DE SOUZA JT, NIELLA GR, VALLE R, BATEMAN RP, FRAVEL D, VINYARD B, HEBBAR PK (2010) A novel, integrated method for management of witches' broom disease in Cacao in Bahia, Brazil. *Crop Protection* 29: 704-711.

MONTEIRO WR, AHNERT D. Melhoramento Genético do Cacaueiro. In: RR. Valle (Ed). **Ciência Tecnologia e Manejo do Cacaueiro**, Editora Vital Ltda. Itabuna, Bahia, 2007.

MONTEIRO WR, LOPES UV, CLEMENT D (2009) Genetic Improvement in Cocoa. In: S.M. JAIN, P.M. PRIYADARSHAN (eds.). *Breeding Plantation Tree Crops: Tropical Species*. Springer: 589-626.

MOTILAL, LA.; ZHANG D.; UMAHARAN P.; BOCCARA M.; MISCHKE S.; SANKAR A.; MEINHARDT LW (2012) Elucidation of genetic identity and population structure of cacao germplasm within an international cacao genebank. *Plant Genetic Resources: Characterization and Utilization* 10: 232–241

MOTILAL LA, ZHANG D, MISCHKE S, MEINHARDT LW UMAHARAN P (2013) Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa. *Tree Genetics & Genomes* 9:1395–1411.

MOTAMAYOR JC, RISTERUCCI AM, LOPEZ PA, ORTIZ CF, MORENO A, LANAUD C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89:380–386

MOTAMAYOR JC, RISTERUCCI AM, HEATH M, LANAUD C (2003) Cacao domestication II: progenitor germplasm of Trinitario cacao cultivar. *Heredity* 91:322–330

MOTAMAYOR JC, LACHENAUD P, SILVA e MOTA JW, LOOR R, KUHN DN, BROWN JS, SCHNELL RJ (2008) Geographic and Genetic Population Differentiation of the Amazonian Chocolate Tree (*Theobroma cacao* L). *PLoS ONE* 3(10): e3311.

MOTAMAYOR, JC; LANAUD C (2002) Molecular analysis of the origin and domestication of *Theobroma cacao* L. In: ENGELS, JMM, RAMANATHA Rao, V, BROWN, AHD, JACKSON MT, editors. Managing Plant Genetic Diversity. Rome. 77–87.

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, 83–90.

N'GORAN JAK, LAURENT V, RISTERUCCI AM, LANAUD C (1994). Comparative genetic analysis of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity* 73: 589–597.

PANG THAU YIN J (2004) Rootstock effects on cocoa in Sabah, Malaysia. *Experimental Agriculture*:40.

PEREIRA J L, FIGUEIREDO JM, ALMEIDA LCC. The first occurrence of witches' broom disease in the principal cocoa growing region of Brazil. *Tropical Agriculture, Inglaterra*, v. 67, p. 188-189, 1989.

PEAKALL R, SMOUSE PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28:2537-2539. doi:10.1093/bioinformatics/bts460

PINNEY S (2009) Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. *Tropical Plant Biology* 2: 23–27. doi:10.1007/s12042-008-9016-z



PIRES JL (2003) Avaliação quantitativa e molecular de germoplasma para o melhoramento do cacau com ênfase na produtividade, qualidade de frutos e resistência a doenças. Tese de Doutorado. Universidade Federal de Viçosa, 342 p.

PIRES JL, CASCARDO JCM, LAMBERT SV, FIGUEIRA A (1998) Increasing cocoa butter yield through genetic improvement of *Theobroma cacao* L: seed fat content variability, inheritance, and association with seed yield. *Euphytica* 103:115–121.

PRITCHARD JK, STEPHENS M, DONNELLY P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

PEREIRA MG, CARLETTO GA, CASTRO GCT (1987) Variabilidade genética das características de fruto e semente em *Theobroma cacao* L., clones SIC e SIAL. *Theobroma* 17:209-218.

PEREIRA JL, RAM A, FIGUEIREDO JM, ALMEIDA LCC (1989) Primeira ocorrência de vassoura-de-bruxa na principal região produtora de cacau do Brasil. *Agrotrópica* 1: 79-81.

PIRES JL (2003) Avaliação Quantitativa e molecular de germoplasma para o melhoramento do cacau com ênfase na produtividade, qualidade de frutos e resistência a doenças. Tese de Doutorado. Viçosa, UFV, 342 p.

PUGH T, FOUET O, RISTERUCCI AM, BROTTIER P, ABOULADZE M, DELETREZ C, COURTOIS B, CLEMENT D, LARMANDE P, N'GORAN JAK and LANAUD C (2004) A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theoretical and Applied Genetics* 108: 1151–1161.

SALTINI R, AKKERMAN R, FROSCH S (2013) Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control* 29: 167-187.

SAMBUICHI RHR (2006) Estrutura e dinâmica do componente arbóreo em área de cabruca na região cacaueira do sul da Bahia, Brasil. *Acta botanica brasiliensis* 20: 943-954.

SANTOS, RMF, CLEMENT D, LIMA LS, LEGAVRE T, LANAUD C, SCHNELL RJ, PIRES JL, LOPES UV, MICHELI F, MICHELI F, GRAMACHO KP (2013) Identification, characterization and mapping of EST-derived SSRs from the cacao-*Ceratocystis cacaofunesta* interaction. *Tree Genetics & Genomes* 9: 117-127.

SANTOS ESL, CERQUEIRA-SILVA CBM, MORI GM, AHNERT D, CORRÊA RX, SOUZA AP (2012) New polymorphic microsatellite loci for *Theobroma cacao* L.: Isolation and characterization of microsatellites from enriched genomic libraries. *Biologia Plantarum* 56: 789-792. doi 10.1007/s10535-012-0134-y

SCHawe CC; DURKA W; TSCHARNTKE T; HENSEN I; KESSLER M (2013) Gene flow and genetic diversity in cultivated and wild Cacao (*Theobroma cacao* ) in bolivia. *American Journal of Botany* 100: 2271–2279.

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 for microsatellite markers. *Cons Genet* 7: 13–24

SILVA CRS, ALBUQUERQUE PSB, ERVEDOSA FR, Mota JWS, FIGUEIRA A, SEBBENN AM (2011) Understanding the genetic diversity, spatial genetic structure and mating system at the hierarchical levels of fruits and individuals of a continuous *Theobroma cacao* population from the Brazilian Amazon. *Heredity*: 106, 973–985.

SOLORZANO, RGL; FOUET, O.; LEMAINQUE, A.; Pavek S.; BOCCARA M.; ARGOUT X.; AMORES F.; COURTOIS B.; RISTERUCCI AM.; LANAUD C (2012). Insight into the Wild Origin, Migration and Domestication History of the Fine Flavour Nacional

*Theobroma cacao* L. Variety from Ecuador. PLoS ONE 7(11): e48438. doi:10.1371/journal.pone.0048438.

STEVEN BROWN, J.; PHILLIPS-MORA, W.; POWER, EJ.; KROL, C.; CERVANTES-MARTINEZ, C.; MOTAMAYOR, JC.; SCHNELL, RJ (2007). Mapping QTLs for Resistance to Frosty Pod and Black Pod Diseases and Horticultural Traits in *Theobroma cacao* L. Crop Sci. 47:1851 – 1858.

TROGNITZ B, CROS E, ASSEMAT S, DAVRIEUX F, FORESTIER-CHIRON N, et al. (2013) Diversity of Cacao Trees in Waslala, Nicaragua: Associations between Genotype Spectra, Product Quality and Yield Potential. PLoS ONE 8: e54079. doi:10.1371/journal.pone.0054079.

TROGNITZ B, SCHELDEMAN X, HANSEL-HOHL K, KUANT A, GREBE H, HERMANN M (2011) Genetic Population Structure of Cacao Plantings within a Young Production Area in Nicaragua. PLoS ONE 6: e16056.

THOMAS E, VAN ZONNEVELD M, Loo J, HODGKIN T, GALLUZZI G, et al. (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. PLoS ONE 7(10): e47676.

VELLO F, GARCIA JR (1971) Características das Principais Variedades de Cacau Cultivadas na Bahia. Revista Theobroma 1(2): 3-10.

VELLO F, GARCIA JR, MAGALHÃES WS (1972) Produção e seleção de cacaueiros híbridos na Bahia. Revista Theobroma 2(3): 15-35.

YAMADA, MM, GURIES (1998) Mating system analysis in cacao (*Theobroma cacao* L.). Agrotrópica, 10: 165-176.

YAMADA, MM, BARTLEY BGD (2009). Compatibilidade de clones de cacau oriundos da ilha do careiro amaonas. *Agrotropica*, 21: 109-112.

YOUNG AM. The chocolate tree: a natural history of cacao. UPF, Gainesville, 2007

YANG, JY.; SCASCITELLI, M.; MOTILAL, LA.; SVEINSSON, S.; ENGELS, JMM.; KANE, NC.; DEMPEWOLF, H.; ZHANG, D.; MAHARAJ, K.; CRONK, QCB (2013). Complex origin of Trinitario-type *Theobroma cacao* (Malvaceae) from Trinidad and Tobago revealed using plastid genomics. *Tree Genetics & Genomes*, 9: 829–840

ZHANG D, AREVALO-GARDINI E, MISCHKE S, ZÚÑIGA-CERNADES L, BARRETO-CHAVEZ A, ADRIAZOLA DEL AGUILA J (2006) Genetic diversity and structure of managed and semi-natural populations of cocoa (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. *Ann Bot* 98:647–655.

ZHANG D, BOCCARA M, MOTILAL L, BUTLER DR, UMAHARAN P, MISCHKE S and MEINHARDT L (2008) Microsatellite variation and population structure in the “Refractario” cacao of Ecuador. *Conservation Genetics* 9: 327–337.

ZHANG D, BOCCARA M, MOTILAL L, MISCHKE S, JOHNSON ES, BUTLER DR, BAILEY B, MEINHARDT L (2009a) Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from Upper Amazon using microsatellite DNA markers. *Tree Genet Genom* 5:595–607.

ZHANG D, MISCHKE BS, JOHNSON ES, MORA A, PHILLIPS-MORA W, MEINHARDT LW (2009b) Molecular characterization of an International cacao collection using microsatellite markers. *Tree Genet Genom* 5:1–10.



## ANEXOS

## ARTIGO 1.

C.B.M. Cerqueira-Silva; E.S.L. Santos; A.M. Souza; G.M. Mori; E.J. Oliveira; R.X. Corrêa; A.P. Souza (2012). **Development and characterization of microsatellite markers for the wild South American *Passiflora cincinnata* (Passifloraceae).** *American Journal of Botany*. [doi:10.3732/ajb.1100477](https://doi.org/10.3732/ajb.1100477)

## ARTIGO 2.

C.B.M. Cerqueira-Silva; E.S.L. Santos; J.G.P. Vieira; G.M. Mori; O.J. Nunes; R.X. Corrêa; A.P. Souza (2014). **New microsatellite markers for wild and commercial species of *Passiflora* (Passifloraceae) and cross-amplification.** *Applications in Plant Sciences*. [doi.org/10.3732/apps.1300061](https://doi.org/10.3732/apps.1300061)

## DECLARAÇÃO

Comissão Interna de Biossegurança



## DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE MARKERS FOR THE WILD SOUTH AMERICAN *PASSIFLORA* *CINCINNATA* (PASSIFLORACEAE)<sup>1</sup>

CARLOS B. M. CERQUEIRA-SILVA<sup>2,3</sup>, ELISA S. L. SANTOS<sup>2,3</sup>, ALISSON M. SOUZA<sup>3</sup>, GUSTAVO M. MORI<sup>2</sup>, EDER J. OLIVEIRA<sup>4</sup>, RONAN X. CORRÊA<sup>5</sup>, AND ANETE P. SOUZA<sup>2,6,7</sup>

<sup>2</sup>Centro de Biologia Molecular e Engenharia Genética and Instituto de Biologia, Universidade Estadual de Campinas, CP 6010, 13083-875 Campinas, São Paulo, Brazil; <sup>3</sup>Departamento de Estudos Básicos e Instrumentais, Universidade Estadual do Sudoeste da Bahia, 45700-000 Itapetinga, Bahia, Brazil; <sup>4</sup>Empresa Brasileira de Pesquisa Agropecuária, Embrapa Mandioca e Fruticultura Tropical, Cruz das Almas, 44380-000 Bahia, Brazil; <sup>5</sup>Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna Km 16, 45662-900 Ilhéus, Bahia, Brazil; and <sup>6</sup>Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, 13083-875 Campinas, São Paulo, Brazil

- **Premise of the study:** We developed the first set of microsatellite markers for the passion fruit, *Passiflora cincinnata*, to provide tools for further study of its genetic diversity and to support current conservation and genetic studies.
- **Methods and Results:** We used a microsatellite-enriched library approach to isolate and characterize 25 new molecular markers. Seven primer pairs successfully amplified polymorphic microsatellite regions, with an average of five alleles per locus. The mean values of expected and observed levels of heterozygosity were 0.516 and 0.525, respectively.
- **Conclusions:** The microsatellite markers identified in this study may be valuable tools for population genetic studies, and this set of markers also may be useful in the design of a genetic conservation strategy, mating system, and hybridization studies.

**Key words:** genetic diversity; genomic microsatellite-enriched library; molecular markers; *Passiflora cincinnata*; Passifloraceae; wild passion fruit.

The family Passifloraceae originated in tropical America and comprises approximately 580 species and 18 genera (Bernacci, 2003). The genus *Passiflora* L. consists of approximately 400 species, and Brazil is considered a center of diversity, presenting nearly one-third of the *Passiflora* species (approximately 129 species) (Cervi et al., 2010). *Passiflora cincinnata* Mast. is a widely distributed South American passion fruit that has been found from northeastern Brazil to eastern Paraguay and western Bolivia. This species is locally cultivated, and its fruits are sold commercially in small cities in Brazil. The aerial parts of this plant are traditionally used for ethnopharmacological and ornamental purposes. Although this species is not seriously threatened by human activities within its habitat, there are current conservation and breeding endeavors focused on this species. Molecular genetic diversity studies of *P. cincinnata* are limited to the work of Cerqueira-Silva et al. (2010) using dominant markers. The development of microsatellite markers for the genetic studies of *Passiflora* was first reported for *P. edulis* Sims

(Oliveira et al., 2005) and *P. alata* Curtis (Pádua et al., 2005). However, these loci were not successfully cross-amplified or were not found to be polymorphic in *P. cincinnata*. Thus, to address this problem and to evaluate the molecular diversity that is present in *P. cincinnata* populations, we isolated and characterized a set of microsatellite markers from this species.

### METHODS AND RESULTS

A microsatellite-enriched library was developed using DNA from the *P. cincinnata* (Pc UESB-01) accession collected in the Active Collection of *Passiflora* Work Germplasm (CAGT-*Passiflora*) of the Universidade Estadual do Sudoeste da Bahia (UESB) in Vitória da Conquista, Bahia, Brazil. Genomic DNA was isolated from fresh leaves using the cetyltrimethyl ammonium bromide (CTAB) method, and the library construction followed the methodology of Billote et al. (1999). The DNA samples (5 µg) were digested with *AfaI* and ligated into the double-stranded adapters, 5'-CTCTTGCTTACGCGTGGACTA-3' and 5'-TAGTCCACGCGTAAGCAAGAGCACA-3'. The enrichment procedure was performed using a hybridization-based capture with (GT)<sub>8</sub> and (CT)<sub>8</sub> biotin-linked probes and streptavidin-coated magnetic beads (Streptavidin MagneSphere Paramagnetic Particles, Promega, Madison, Wisconsin, USA). The selected fragments were cloned into a pGEM-T Easy (Promega) vector and used to transform *Escherichia coli* XL1-Blue competent cells (Stratagene, La Jolla, California, USA). The recombinant colonies were selected using white/blue screening. We randomly selected a total of 192 positive clones, which were then double-sequenced using an ABI PRISM 377 automated DNA sequencer (Applied Biosystems, Foster City, California, USA). Every sequence was aligned and edited using SeqMan software (DNASTAR, Madison, Wisconsin, USA) analysis. MICROSAT software (A. M. Risterucci, Centre de Coopération Internationale en Recherche Agronomique pour le Développement [CIRAD], personal communication) was used to eliminate the adapters and restriction sites from the sequences. The Simple Sequence Repeat Identification Tool

<sup>1</sup>Manuscript received 22 September 2011; revision accepted 6 November 2011.

The authors would like to thank the Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, PROCAD-NF2008) for financial support, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a research fellowship to A.P.S., and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for a graduate fellowship to G.M.M.

<sup>7</sup>Author for correspondence: anete@unicamp.br

doi:10.3732/ajb.1100477



TABLE 1. Characteristics of the 25 microsatellite markers that were developed for *Passiflora cincinnata*.

| Marker        | Primer sequences (5′–3′)                              | Allele size (bp) <sup>a</sup> | Repeat motif                                          | AC <sup>b</sup> | GenBank accession no. |
|---------------|-------------------------------------------------------|-------------------------------|-------------------------------------------------------|-----------------|-----------------------|
| mPc-UNICAMP01 | F: AGACAGACATTCAGTCCGGTC<br>R: TCAAGAAGCTCCTCCAACCTGA | 300–310                       | (AG) <sub>4</sub>                                     | TD65–55         | JF815377              |
| mPc-UNICAMP02 | F: CATTTCAAAGCGTGGTTCT<br>R: AAGCGCCATGGTTTACAG       | 180–184                       | (TCC) <sub>4</sub>                                    | TD65–55         | JF815378              |
| mPc-UNICAMP03 | F: CTAGCCATATCCCGCATCAT<br>R: GGGGATTACATTCACCATCG    | 281–284                       | (CT) <sub>5</sub>                                     | TD65–55         | JF815379              |
| mPc-UNICAMP04 | F: ATGACAACGTGGTTAGGGAA<br>R: CACGCGTGTGTGGAAATAC     | 270–340                       | (AC) <sub>16</sub> (AT) <sub>4</sub>                  | TD65–55         | JF815380              |
| mPc-UNICAMP05 | F: TAAGCTTCCCTCCAAACCAA<br>R: AGGAAGGAGGAAGCAGGAAG    | 178–182                       | (AC) <sub>4</sub>                                     | TD65–55         | JF815381              |
| mPc-UNICAMP06 | F: AAGCACCCTAGCAGACTCAAC<br>R: TTCTCAGGTGACAGGGGAAC   | 230–252                       | (TG) <sub>8</sub> (TG) <sub>6</sub> (TG) <sub>7</sub> | TD65–55         | JF815382              |
| mPc-UNICAMP07 | F: GCGACTAATGTGCCACTGTCT<br>R: ATTGTGGTGTGCGTAGAGC    | 260–330                       | (CAT) <sub>5</sub>                                    | TD65–55         | JF815383              |
| mPc-UNICAMP08 | F: CACAGCAAACGAAAACGAGA<br>R: ACGTGTGACTGTTGGCACTT    | 103                           | (AT) <sub>4</sub>                                     | TD65–55         | JF815384              |
| mPc-UNICAMP09 | F: GAGAAAGAGTTTACCAGGCTA<br>R: TGCACCCATACTGCTCTTTG   | 203                           | (GAGC) <sub>3</sub>                                   | TD65–55         | JF815385              |
| mPc-UNICAMP10 | F: GCGATCAACACGACTAC<br>R: CCTGAGAAACCGGATGAAG        | 230                           | (GAG) <sub>6</sub>                                    | TD65–55         | JF815386              |
| mPc-UNICAMP11 | F: TGCCACACAACGTACCTCAT<br>R: CGAGAGCGCCAGAGAAGTAG    | 194                           | (TACA) <sub>3</sub>                                   | TD65–55         | JF815387              |
| mPc-UNICAMP12 | F: AGTCCCGACCTCCAACCTTT<br>R: TTCCTGAGACTGGAGAGCAAC   | 113                           | (GAT) <sub>6</sub>                                    | TD65–55         | JF815388              |
| mPc-UNICAMP13 | F: GCACAGTCTTCAGCCAGTGTT<br>R: ATGCTCCTCCTCTGGGGTAT   | 231                           | (CCAG) <sub>3</sub>                                   | TD65–55         | JF815389              |
| mPc-UNICAMP14 | F: GCGATCAACGAGCACTATCA<br>R: GTCGCTCTGCGAAGTAATCC    | 129                           | (GAG) <sub>5</sub>                                    | TD65–55         | JF815390              |
| mPc-UNICAMP15 | F: GCACGTGAGTGGGTTTACA<br>R: AGTCTGTGCTGGGCAGTT       | 161                           | (AC) <sub>4</sub>                                     | TD65–55         | JF815391              |
| mPc-UNICAMP16 | F: TAAGCTTCCCTCCAAACCAA<br>R: AGGAAGGAGGAAGCAGGAAG    | 253                           | (AC) <sub>4</sub>                                     | TD65–55         | JF815392              |
| mPc-UNICAMP17 | F: GTCCACCATATCCTGGCATC<br>R: CTGAAGGAGGCTAGACAGCTT   | 176                           | (CT) <sub>4</sub>                                     | TD65–55         | JF815393              |
| mPc-UNICAMP18 | F: CAAAACCATACAAGGCGTCA<br>R: TATGGCACTTCTGTATCAGCAT  | 209                           | (AT) <sub>4</sub>                                     | TD65–55         | JF815394              |
| mPc-UNICAMP19 | F: ACGAGACAGAATTCACCGG<br>R: AAGCTCAGATCTCCATGGACA    | 109                           | (GAGC) <sub>3</sub>                                   | TD65–55         | JF815395              |
| mPc-UNICAMP20 | F: CGAGAAGATTCTGTAAGCAGGC<br>R: GCATTATCTCGGTGTGATGGT | 141                           | (TG) <sub>4</sub>                                     | TD65–55         | JF815396              |
| mPc-UNICAMP21 | F: CCGTATTCCCTGCACCTTTAAG<br>R: GTGGCTTCGTCTGATGTGAA  | 151                           | (TA) <sub>5</sub> (TG) <sub>12</sub>                  | TD65–55         | JF815397              |
| mPc-UNICAMP22 | F: TCTCCTCCTCCTCCTCCTTC<br>R: TTTTGAGCTAGCCCGTGAGT    | 189                           | (TTC) <sub>4</sub> (TC3)                              | TD65–55         | JF815398              |
| mPc-UNICAMP23 | F: CGCCAAACAACATGAAAGC<br>R: CCCCAGAGAGTGCCAAAGA      | 105                           | (TGG) <sub>4</sub>                                    | TD58–48         | JF815399              |
| mPc-UNICAMP24 | F: GTTTCAGCCCAACGACTT<br>R: ACATATTGAACAGTCCGAACAT    | 225                           | (TTAA) <sub>3</sub> (TAAT) <sub>3</sub>               | TD65–55         | JF815400              |
| mPc-UNICAMP25 | F: CCTTTATATCTCGCAACCGC<br>R: GGAGTCGGTAGGCTCAAGAC    | 139                           | (CGCT) <sub>3</sub>                                   | TD58–48         | JF815401              |

<sup>a</sup>Fragment size ranges for polymorphic loci and expected product sizes based on the clone fragment for monomorphic markers.

<sup>b</sup>Optimal PCR amplification conditions (AC). TD65–55 and TD58–48 indicate touchdown PCR programs with temperatures ranging from 65°C to 55°C and 58°C to 48°C, respectively.

(SSRIT) (Temnykh et al., 2001) was used to identify the microsatellites. Of the 62 microsatellites found, dinucleotide motifs were the most abundant (60%), followed by trinucleotide and tetranucleotide motifs (22% and 18%, respectively). The number of repetitions of dinucleotide microsatellites varied from four to 33 (average: 6.4). Primer pairs were designed using PrimerSelect (DNASTAR) and Primer3Plus (Untergasser et al., 2007). The main criteria adopted for the design of the primers were as follows: annealing temperature 45–65°C (with a maximum difference of 3°C between the forward and reverse primers), GC content 40–70%, and amplified product sizes that vary from 100 and 300 bp for accurate genotyping using vertical electrophoresis in 1× TBE buffer and 6% polyacrylamide gels.

PCR was conducted using a final volume of 15 µL containing 15 ng of template DNA, 1× PCR buffer (20 mM Tris HCl [pH 8.4] and 50 mM KCl),

1.5 mM MgCl<sub>2</sub>, 0.2 µM of each dNTP, 0.5 µM of each primer, and 1 U of *Taq* DNA polymerase (Invitrogen, Carlsbad, California, USA). All of the primers were evaluated using three PCR amplification protocols with an annealing temperature of 60°C, as follows: 94°C for 5 min, 34 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 10 min, and two different touchdown PCR protocols, TD 65–55°C and TD 58–48°C. The touchdown protocols consisted of the following five steps: (1) initial denaturation (94°C for 5 min); (2) 10 cycles of denaturation (94°C for 1 min), annealing (65°C – 1°C per cycle for 1 min for TD65–55 or 58°C – 1°C per cycle for 1 min for TD58–48), and extension (72°C for 1 min); (3) step 2 was repeated; (4) 14 cycles of denaturation (94°C for 1 min), annealing (55°C or 48°C for 1 min), and extension (72°C for 1 min); and (5) a final extension (72°C for 10 min). Superior results were obtained using the touchdown programs. The amplification products

TABLE 2. Results of the initial screening of polymorphic microsatellite markers in one population of *Passiflora cincinnata* ( $N = 31$ ).

| Marker        | A | $H_o$ | $H_e$ | PIC   | HWE                 |
|---------------|---|-------|-------|-------|---------------------|
| mPc-UNICAMP01 | 2 | 0.428 | 0.342 | 0.28  | 0.30                |
| mPc-UNICAMP02 | 3 | 0.620 | 0.472 | 0.411 | 0.15                |
| mPc-UNICAMP03 | 2 | 0.000 | 0.063 | 0.061 | 0.02 <sup>†</sup>   |
| mPc-UNICAMP04 | 9 | 0.576 | 0.838 | 0.803 | 0.001* <sup>†</sup> |
| mPc-UNICAMP05 | 3 | 0.880 | 0.522 | 0.397 | 0.001*              |
| mPc-UNICAMP06 | 8 | 0.535 | 0.567 | 0.537 | 0.35                |
| mPc-UNICAMP07 | 8 | 0.636 | 0.813 | 0.769 | 0.01                |

Note: A = number of alleles;  $H_e$  = expected heterozygosity;  $H_o$  = observed heterozygosity; HWE = Hardy–Weinberg equilibrium; PIC = polymorphism information content.

\*Departs significantly from HWE after Bonferroni correction ( $P < 0.007$ ).

<sup>†</sup>Significant values for the occurrence of null alleles.

were visualized on ethidium bromide–stained agarose gels (3%) in 0.5× TBE before vertical electrophoresis on 6% denaturing polyacrylamide gels run in 1× TBE and staining with silver nitrate. The product sizes were determined using a 10-bp DNA ladder (Invitrogen). Statistical analyses were performed using GENEPOP (Raymond and Rousset, 1995), PIC Calculator (Kemp, 2002), and MICRO-CHECKER (van Oosterhout et al., 2004) software. We designed 33 primer pairs that matched the predefined criteria, and 25 generated consistent patterns of amplification that matched the expected band sizes based on the size of the sequenced fragment; these were used for further characterization (Table 1). The remaining eight loci were discarded due to amplification failure in the preliminary tests.

For the polymorphism analysis, 31 *P. cincinnata* accessions from the germplasm collection of the Brazilian Agricultural Research Corporation (EMBRAPA—Mandioca e Fruticultura Tropical) were analyzed. Among the 25 microsatellite markers that were developed, seven markers were polymorphic, with varying degrees of polymorphism (Table 2). The mean number of alleles was five (with a range from two to nine alleles). The mean values of expected and observed levels of heterozygosity were 0.516 (0.063–0.838) and 0.525 (0–0.88), respectively, and the polymorphism information content varied from 0.061 to 0.803. A small number of polymorphic microsatellites have been reported in the literature for *Passiflora* species, and the low number of polymorphic loci seems to be characteristic of the genus. Of the microsatellites developed for *P. cincinnata*, 21% were polymorphic, which is similar to the percentage of polymorphic microsatellites reported for *P. edulis* (20%; Oliveira et al., 2008) and for *P. alata* (26%; Pereira, 2010).

The analysis revealed significant deviations from Hardy–Weinberg equilibrium at only two microsatellite loci (mPc-UNICAMP03 and mPc-UNICAMP04) ( $P < 0.007$  after Bonferroni correction). These results might be due to the probable occurrence of null alleles or the linkage disequilibrium (LD) that is observed for the mPc-UNICAMP04 locus (Table 1). With the exception of one pair of markers, mPc-UNICAMP04 and mPc-UNICAMP07, none of the other 21 marker pairs showed a significant LD. Furthermore, none of the microsatellite loci exhibited a significant probability of stuttering or large allele dropout (Table 2). Two of the loci (mPc-UNICAMP03 and mPc-UNICAMP04) showed significant values for the occurrence of null alleles using Bonferroni correction. Eighteen microsatellite markers were monomorphic for the evaluated sample (Table 2). However, these markers might be polymorphic among other *P. cincinnata* populations or other *Passiflora* species.

## CONCLUSION

The microsatellite markers characterized in this study are the first to be developed for *P. cincinnata* and are potentially useful for genetic studies on other wild populations of this species.

Considering the low frequency of polymorphic loci previously reported and the great difficulty associated with obtaining informative microsatellites for *Passiflora* species, the markers developed in this study might be important tools to assess the genetic diversity of other species from this genus.

## LITERATURE CITED

- BERNACCI, L. C. 2003. Passifloraceae. In M. G. L. Wanderley, G. J. Shepherd, A. M. Giulietti, and T. S. Melhem [eds.], *Flora Fanerogâmica do Estado de São Paulo*, 247–248. RIMA/FAPESP, São Paulo, Brazil.
- BILLOTE, N., P. J. L. LAGODA, A. M. RISTERUCI, AND F. C. BAURENS. 1999. Microsatellite-enriched libraries: Applied methodology for the development of SSR markers in tropical crops. *Fruits* 54: 277–288.
- CERQUEIRA-SILVA, C. B. M., L. D. H. C. S. CONCEIÇÃO, E. S. L. SANTOS, C. B. CARDOSO-SILVA, A. S. PEREIRA, A. C. OLIVEIRA, AND R. X. CORRÊA. 2010. Genetic variability in wild genotypes of *Passiflora cincinnata* based on RAPD markers. *Genetics and Molecular Research* 9: 2421–2428.
- CERVI, A. C., M. A. MILWARD DE AZEVEDO, AND L. C. BERNACCI. 2010. Passifloraceae. In *Lista de Espécies da Flora do Brasil*. Jardim Botânico do Rio de Janeiro, Rio de Janeiro [online]. Website <http://floradobrasil.jbrj.gov.br/2010/FB000182> [accessed 17 October 2011].
- KEMP, S. 2002. PIC Calculator Extra. Website <http://www.genomics.liv.ac.uk/animal/pic.html> [accessed 10 April 2011].
- OLIVEIRA, E. J., J. G. PADUA, M. I. ZUCCHI, L. E. A. CAMARGO, M. H. P. FUNGARO, AND M. L. C. VIEIRA. 2005. Development and characterization of microsatellite markers from the yellow passion fruit (*Passiflora edulis* f. *flavicarpa*). *Molecular Ecology Notes* 5: 331–333.
- OLIVEIRA, E. J., M. L. C. VIEIRA, A. A. F. GARCIA, C. F. MUNHOZ, G. R. A. MARGARIDO, L. CONSOLI, F. P. MATTÁ, AND M. C. MORAES. 2008. An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. *Journal of the American Society for Horticultural Science* 133: 35–41.
- PÁDUA, J. G., E. J. OLIVEIRA, M. I. ZUCCHI, G. C. X. OLIVEIRA, L. E. A. CAMARGO, AND M. L. C. VIEIRA. 2005. Isolation and characterization of microsatellite markers from the sweet passion fruit (*Passiflora alata* Curtis: Passifloraceae). *Molecular Ecology Notes* 5: 863–865.
- PEREIRA, G. S. 2010. Desenvolvimento de Marcadores SSR, M-AFLP e SNP visando à integração de mapas genético-moleculares de *Passiflora alata* Curtis. M.Sc. dissertation, Universidade Estadual de São Paulo, Escola Superior de Agricultura ‘Luiz de Queiroz’, Piracicaba, São Paulo, Brazil.
- RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Heredity* 86: 248–249.
- TEMNYKH, S., G. CLERCK, A. LUKASHOVA, L. LIPOVICH, S. CARTINHOUR, AND S. MCCOUCH. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* 11: 1441–1452.
- UNTERGASSER, A., H. NIJVEEN, X. RAO, T. BISSELING, R. GEURTS, AND J. A. M. LEUNISSEN. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35: 71–74.
- VAN OOSTERHOUT, C., W. F. HUTCHINSON, D. P. M. WILLS, AND P. SHIPLEY. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.





## NEW MICROSATELLITE MARKERS FOR WILD AND COMMERCIAL SPECIES OF *PASSIFLORA* (PASSIFLORACEAE) AND CROSS-AMPLIFICATION<sup>1</sup>

CARLOS B. M. CERQUEIRA-SILVA<sup>2,3</sup>, ELISA S. L. SANTOS<sup>2,3</sup>, JOÃO G. P. VIEIRA<sup>3</sup>, GUSTAVO M. MORI<sup>2</sup>, ONILDO N. JESUS<sup>4</sup>, RONAN X. CORRÊA<sup>5</sup>, AND ANETE P. SOUZA<sup>2,6,7</sup>

<sup>2</sup>Centro de Biologia Molecular e Engenharia Genética, Instituto de Biologia, Universidade Estadual de Campinas, CP 6010, 13083-875 Campinas, São Paulo, Brazil; <sup>3</sup>Departamento de Ciências Exatas e Naturais, Universidade Estadual do Sudoeste da Bahia, 45700-000 Itapetinga, Bahia, Brazil; <sup>4</sup>Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa em Mandioca e Fruticultura Tropical, 44380-000 Cruz das Almas, Bahia, Brazil; <sup>5</sup>Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna Km 16, 45662-900 Ilhéus, Bahia, Brazil; and <sup>6</sup>Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, 13083-875 Campinas, São Paulo, Brazil

- **Premise of the study:** We developed the first microsatellites for *Passiflora setacea* and characterized new sets of markers for *P. edulis* and *P. cincinnata*, enabling further genetic diversity studies to support the conservation and breeding of passion fruit species.
- **Methods and Results:** We developed 69 microsatellite markers and, in conjunction with assessments of cross-amplification using primers available from the literature, present 43 new polymorphic microsatellite loci for three species of *Passiflora*. The mean number of alleles per locus was 3.1, and the mean values of the expected and observed levels of heterozygosity were 0.406 and 0.322, respectively.
- **Conclusions:** These microsatellite markers will be valuable tools for investigating the genetic diversity and population structure of wild and commercial species of passion fruit (*Passiflora* spp.) and may be useful for developing conservation and improvement strategies by contributing to the understanding of the mating system and hybridization within the genus.

**Key words:** genetic diversity; genomic microsatellite-enriched library; molecular markers; *Passiflora*; simple sequence repeats; wild passion fruit.

The genus *Passiflora* L. (Passifloraceae) comprises approximately 400 species, of which at least 30% are distributed within Brazilian forests (Cervi et al., 2010). Species such as *P. edulis* Sims are important because of the economic value of their fruit (Faleiro et al., 2005). Certain wild species, including *P. setacea* DC. and *P. cincinnata* Mast., are of interest because of their potential use in genetic breeding. However, the limited number of molecular genetic diversity studies of this genus (Faleiro et al., 2005; Cerqueira-Silva et al., 2012) attests to the need for and relevance of novel molecular tools for studies of its populations and mating system.

Although diversity studies of passion fruit began in the late 1990s, efforts to use microsatellites only began in 2005 (Oliveira et al., 2005; Pádua et al., 2005), and studies related to the development of microsatellites have been published for *P. cincinnata*

(Cerqueira-Silva et al., 2012) and *P. contracta* Vitta (Cazé et al., 2012) only recently. The markers available are still insufficient for performing consistent genetic studies of most *Passiflora* species because the evaluated populations exhibit low variability and percentages of polymorphic loci (between 0% and 26%) (Pereira, 2010; Ortiz et al., 2012; Cerqueira-Silva et al., 2012). Thus, considering the difficulty in obtaining informative microsatellites for *Passiflora* and to enhance the genetic investigation of both wild and commercial populations, we isolated, characterized, and evaluated the cross-amplifications of microsatellites for *P. edulis*, *P. setacea*, and *P. cincinnata*.

## METHODS AND RESULTS

Two microsatellite-enriched genomic libraries were developed using genotypes from the germplasm collection of *P. edulis* (Pe-UESB01) and *P. setacea* (Ps-UESB01) from the Universidade Estadual do Sudoeste da Bahia (UESB; Itapetinga, Bahia, Brazil). Genomic DNA was isolated from fresh leaves using the cetyltrimethylammonium bromide (CTAB) method, and libraries were constructed following Billote et al. (1999). DNA samples (5 µg) were digested with *AfaI* and ligated to the double-stranded adapters 5'-CTCTTGCTTA-CGCGTGGACTA-3' and 5'-TAGTCCACGCGTAAGCAAGAGCACA-3'. Enrichment was performed using a hybridization-based capture with (GT)<sub>8</sub> and (CT)<sub>8</sub> biotin-linked probes and streptavidin-coated magnetic beads (Streptavidin Magnosphere Paramagnetic Particles; Promega Corporation, Madison, Wisconsin, USA). The selected fragments were cloned into a pGEM-T Easy Vector (Promega Corporation) and used to transform *Escherichia coli* x11-blue competent

<sup>1</sup>Manuscript received 19 July 2013; revision accepted 18 October 2013.

The authors thank the Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, PROCAD-NF2008) for financial support, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a research fellowship to A.P.S., and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and CNPq for graduate fellowships to G.M.M. and J.G.P.V., respectively.

<sup>7</sup>Author for correspondence: anete@unicamp.br

doi:10.3732/apps.1300061

TABLE 1. Characteristics of the 69 new microsatellite markers developed for passion fruit species (17 markers for *Passiflora edulis* and 52 markers for *P. setacea*) and cross-amplification assays.

| Locus         | Primer sequences (5'–3')                              | Allele size (bp) | Repeat motif                                          | PCR amplification conditions <sup>a</sup> | GenBank accession no. | Cross-amplification |         |         |
|---------------|-------------------------------------------------------|------------------|-------------------------------------------------------|-------------------------------------------|-----------------------|---------------------|---------|---------|
|               |                                                       |                  |                                                       |                                           |                       | Pe                  | Ps      | Pc      |
| mPe-UNICAMP01 | F: CCTGTGCGGAAGACTTCTGC<br>R: GGATCGTTGTGGAGTGTGGT    | 230–232          | (AC) <sub>4</sub>                                     | TD58                                      | KF142650              |                     | 232     | 232     |
| mPe-UNICAMP02 | F: TCGAGTGAGATTGGCAGTG<br>R: TTGGCTTCGAGGAGAAGAA      | 165–171          | (GT) <sub>8</sub>                                     | TD58                                      | KF142651              |                     | 161–163 | 163     |
| mPe-UNICAMP03 | F: ATAGGCATTTCACACAGCAC<br>R: AAGCATCCGTGAGACAGGT     | 261              | (AC) <sub>8</sub>                                     | TD58                                      | KF142652              |                     | 261     | 261     |
| mPe-UNICAMP04 | F: GCTAACAAGCCCAATCAAC<br>R: CAGACCATGAGACGGCAGTA     | 296              | (CA) <sub>5</sub>                                     | TD65                                      | KF142653              |                     | 296     | 296     |
| mPe-UNICAMP05 | F: CGGGGTTATGCAAGGTAACA<br>R: ACTGGGTGGACTAGGAAACG    | 121              | (TG) <sub>8</sub>                                     | TD65                                      | KF142654              |                     | —       | —       |
| mPe-UNICAMP06 | F: GTTCGAACCTTGGTTCTCTTG<br>R: AATCCTCTCCCGGTATCCAC   | 292              | (TG) <sub>4</sub>                                     | TD65                                      | KF142655              |                     | —       | 290–320 |
| mPe-UNICAMP07 | F: GGAACCGTGTGATGGGATAC<br>R: ACCGATTGACAGCTCTGCC     | 255              | (AG) <sub>8</sub>                                     | TD65                                      | KF142656              |                     | 255     | 255     |
| mPe-UNICAMP08 | F: GCTGAGAACCCCGTGACTTA<br>R: CGAGTATGGCACATCCCTG     | 196              | (CA) <sub>4</sub>                                     | TD65                                      | KF142657              |                     | 196     | 196     |
| mPe-UNICAMP09 | F: TGCCCTCTCGGATATTTACAGC<br>R: CGCATGTCCCCATACGAC    | 212              | (AC) <sub>5</sub>                                     | TD58                                      | KF142658              |                     | 212     | 248–261 |
| mPe-UNICAMP10 | F: GTCACCTGACGCTGGTATAGTT<br>R: GAACATATTCGGCAGATGGA  | 251              | (CT) <sub>5</sub>                                     | TD58                                      | KF142659              |                     | 251     | 251     |
| mPe-UNICAMP11 | F: GCAGCAATCAATGCAATCAG<br>R: GCCATTCTCCTCTCACCGTA    | 180              | (CA) <sub>9</sub> (AT) <sub>5</sub>                   | TD58                                      | KF142660              |                     | 172     | 176     |
| mPe-UNICAMP12 | F: CACACAAGCGTTTCTTACG<br>R: TGATATGAACGATACGGTAGGC   | 214              | (CA) <sub>7</sub>                                     | TD65                                      | KF142661              |                     | —       | —       |
| mPe-UNICAMP13 | F: TTCGTGCATTGTTCTATTACC<br>R: GCCTTCTTGTCTATGTTGGA   | 202              | (TC) <sub>5</sub>                                     | TD58                                      | KF142662              |                     | 202     | 166–168 |
| mPe-UNICAMP14 | F: GACTTCGTATGACGCCAGGT<br>R: TGCAAGAATCCGAAGACTCA    | 263              | (CA) <sub>8</sub>                                     | TD65                                      | KF142663              |                     | 263     | 260     |
| mPe-UNICAMP15 | F: CATTCTCACCTCACGAA<br>R: TGGTTGTGTGGTTTGTGCTT       | 253              | (AC) <sub>5</sub>                                     | TD58                                      | KF142664              |                     | 253     | 253     |
| mPe-UNICAMP16 | F: CGTGGGTGAGTGTGAATGAG<br>R: TGATGTGAGCATGGTTGGTT    | 195              | (AT) <sub>4</sub> (TG) <sub>11</sub>                  | TD65                                      | KF142665              |                     | —       | —       |
| mPe-UNICAMP17 | F: GCCACGTGCAATGTCACT<br>R: CGTGTGTGACCAAGGAG         | 300              | (AC) <sub>9</sub>                                     | TD65                                      | KF142666              |                     | —       | —       |
| mPs-UNICAMP01 | F: TAGCTTAACACAATGCAACAGA<br>R: CAACGGAGAACGATGTCAG   | 153–154          | (TG) <sub>5</sub> (TG) <sub>5</sub>                   | TD58                                      | KF171014              | 158–168             |         | 154     |
| mPs-UNICAMP02 | F: TAGCTTAACACAATGCAACAGA<br>R: CAACGGAGAACGATGTCAG   | 154–156          | (TG) <sub>5</sub> (TG) <sub>5</sub>                   | TD58                                      | KF171015              | 160–170             |         | 156     |
| mPs-UNICAMP03 | F: GTAGCGTCTCGGCAGGTC<br>R: ACTCTAAGTCGGCCACTCTTG     | 176–177          | (CT) <sub>4</sub>                                     | TD65                                      | KF171016              | 176                 |         | 176     |
| mPs-UNICAMP04 | F: CAACAGGAGGTGAGGTGTGA<br>R: GACAGTGCAACTTTAGGCGAC   | 156–157          | (TG) <sub>4</sub>                                     | TD65                                      | KF171017              | 156                 |         | 156     |
| mPs-UNICAMP05 | F: TCGGTCTTCGTATTCAACTCTG<br>R: GAGGAACTGGCATCGCAT    | 194–218          | (CT) <sub>8</sub>                                     | 61.5°C                                    | KF171018              | 210–220             |         | 213–216 |
| mPs-UNICAMP06 | F: GTTGGATCAAAGGGTCACA<br>R: CAACACTGGATCGAACTGGTA    | 218–224          | (CGTG) <sub>3</sub> (ATGA) <sub>3</sub>               | TD65                                      | KF171019              | 194–224             |         | 215     |
| mPs-UNICAMP07 | F: ACAGGGGTGAGGCACATTC<br>R: TCTGTTATTATCATCGGCAGGA   | 143–145          | (CA) <sub>4</sub>                                     | TD58                                      | KF171020              | —                   |         | —       |
| mPs-UNICAMP08 | F: AGTGCCAGTGGCTTCGTATT<br>R: GATCGTCATGGCTGTTGCTA    | 207–211          | (TGCAA) <sub>3</sub>                                  | TD65                                      | KF171021              | 174                 |         | 176     |
| mPs-UNICAMP09 | F: GGGCCGTTGTCAAAGTAGT<br>R: GAGGTTAAGGCAAGCACTG      | 250–268          | (AC) <sub>4</sub>                                     | 61.5°C                                    | KF171022              | 258–260             |         | 260     |
| mPs-UNICAMP10 | F: ACTCTCACTCAATCGACC<br>R: AATTGTTACTCCGTTTCTCTGA    | 256–260          | (AG) <sub>4</sub> (GT) <sub>5</sub> (GT) <sub>4</sub> | 60°C                                      | KF171023              | 264–268             |         | 260–268 |
| mPs-UNICAMP11 | F: CAGACGTTGTGTTTGGTAAT<br>R: TCAGGTTAGGAAGCTGCATC    | 232–270          | (CA) <sub>4</sub> (CA) <sub>4</sub> (AT) <sub>4</sub> | 60°C                                      | KF171024              | 262                 |         | —       |
| mPs-UNICAMP12 | F: ACAGGGGTGAGGCACATACA<br>R: GTAGTGCGTGGCTTGGGTAG    | 201–204          | (CA) <sub>4</sub>                                     | TD65                                      | KF171025              | 208                 |         | 208     |
| mPs-UNICAMP13 | F: CCTATACCTGCCAGTCAGC<br>R: ACTTAAGCACCCCAATCGTT     | 146–148          | (CA) <sub>4</sub>                                     | TD65                                      | KF171026              | 144                 |         | —       |
| mPs-UNICAMP14 | F: CGTTCATAAGTGAATCAGTCAA<br>R: GGATCGACAAACAAAGGTAGA | 112–116          | (CA) <sub>4</sub>                                     | TD65                                      | KF171027              | 114                 |         | 114     |
| mPs-UNICAMP15 | F: TATGGAGTTGCGAGGCTTAG<br>R: CGGGCAACGAACACTTTATT    | 145–148          | (GT) <sub>4</sub>                                     | 60°C                                      | KF171028              | 143–145             |         | 146     |
| mPs-UNICAMP16 | F: GAGAAAGCGAGTCAGCGAGA<br>R: GACTCCAATATCGGCACTTCA   | 157–165          | (GAG) <sub>6</sub> (CAA) <sub>4</sub>                 | TD65                                      | KF171029              | 163–167             |         | 159–170 |

TABLE 1. Continued.

| Locus         | Primer sequences (5'–3')                               | Allele size (bp) | Repeat motif                                           | PCR amplification conditions <sup>a</sup> | GenBank accession no. | Cross-amplification |    |         |
|---------------|--------------------------------------------------------|------------------|--------------------------------------------------------|-------------------------------------------|-----------------------|---------------------|----|---------|
|               |                                                        |                  |                                                        |                                           |                       | Pe                  | Ps | Pc      |
| mPs-UNICAMP17 | F: CATCCAACTCCGAACCTTA<br>R: TACCCAGTCCGGTCCATTAG      | 142–148          | (AC) <sub>5</sub>                                      | 60°C                                      | KF171030              | 147                 |    | 146     |
| mPs-UNICAMP18 | F: GGGGTTCTTCACTCATCCAC<br>R: TGACGACTAGGGGATTACAGG    | 262–278          | (CA) <sub>10</sub> (AT) <sub>6</sub>                   | TD65                                      | KF171031              | —                   |    | —       |
| mPs-UNICAMP19 | F: CTGTGGCAAGTGGCTAACAA<br>R: CCACCCCTACTCGACCACTC     | 290–294          | (TG) <sub>4</sub>                                      | 60°C                                      | KF171032              | 290                 |    | 290     |
| mPs-UNICAMP20 | F: GCTGGCTCTAGCTCAACTCG<br>R: GCCAGCATAGGATGTCAGGT     | 200              | (CT) <sub>5</sub>                                      | TD65                                      | KF171033              | 200                 |    | 200     |
| mPs-UNICAMP21 | F: CCCAATCGCTGAGAGGAGT<br>R: CGGTAGGCTCATTCGTGTCA      | 228              | (TG) <sub>4</sub>                                      | TD58                                      | KF171034              | —                   |    | —       |
| mPs-UNICAMP22 | F: AGGCATGCCCCATCAATG<br>R: CACTAAACCTGCAAAGCGAA       | 131              | (GT) <sub>5</sub> (GT) <sub>4</sub>                    | TD58                                      | KF171035              | —                   |    | —       |
| mPs-UNICAMP23 | F: GAGCAGCTAAAAGAACCTAC<br>R: TAGAGGTTGTGCTGGAGTC      | 298              | (AC) <sub>5</sub> (CA) <sub>4</sub>                    | TD58                                      | KF171036              | 298                 |    | 298     |
| mPs-UNICAMP24 | F: GAGGTCCACCACTGTCAGT<br>R: CTAGCGTCACCTCCAGAG        | 254              | (AG) <sub>4</sub>                                      | TD58                                      | KF171037              | 254                 |    | 258–260 |
| mPs-UNICAMP25 | F: GTGTTTGTGGCGATGTGATTA<br>R: GACAAACGTTGTTCCCGTC     | 162              | (AAG) <sub>5</sub>                                     | TD58                                      | KF171038              | 162                 |    | 162     |
| mPs-UNICAMP26 | F: TGTGGCATGTGTATGACTTGAT<br>R: CATAGATATGGGATGAGCGACA | 166              | (TG) <sub>4</sub>                                      | TD58                                      | KF171039              | 166                 |    | 174     |
| mPs-UNICAMP27 | F: AGATGGAACAGGTGGGTGAG<br>R: TAGGCTTGTCTGGCTCTGG      | 151              | (CCA) <sub>5</sub>                                     | TD58                                      | KF171040              | 151                 |    | 151     |
| mPs-UNICAMP28 | F: AATTGTATCGGTAAACCTGC<br>R: TGCCATTGCGAGTGAATAAG     | 274              | (AC) <sub>6</sub>                                      | TD58                                      | KF171041              | 274                 |    | 274     |
| mPs-UNICAMP29 | F: GAGAAATCTCAGCACACGCA<br>R: CGGTTCTTGGTTTGTGGAT      | 204              | (CA) <sub>5</sub>                                      | TD58                                      | KF171042              | —                   |    | —       |
| mPs-UNICAMP30 | F: CGGCTGAAGGAGGAGGTAG<br>R: TGAAAAACAAGTCAGCCAACA     | 118              | (GT) <sub>6</sub>                                      | TD58                                      | KF171043              | —                   |    | —       |
| mPs-UNICAMP31 | F: GGTGTGGTAGCCTGTTTGTC<br>R: CCGCATCTCTTACATCGTTA     | 211              | (TG) <sub>4</sub> (GT) <sub>5</sub>                    | TD65                                      | KF171044              | 215                 |    | 215–219 |
| mPs-UNICAMP32 | F: CAGACGTTGCATCTTGGTAAT<br>R: CATCGGAGGAGTTTACACATT   | 172              | (CA) <sub>4</sub> (AC) <sub>9</sub> (AT) <sub>6</sub>  | TD65                                      | KF171045              | 172                 |    | 172     |
| mPs-UNICAMP33 | F: GCAGCAATCAATGCAATCAG<br>R: GCCATTCTCCTCACCCTGA      | 184              | (AT) <sub>4</sub> (CA) <sub>10</sub> (AT) <sub>6</sub> | TD65                                      | KF171046              | 184                 |    | 184     |
| mPs-UNICAMP34 | F: GGCAGGATATGCTTTGGTT<br>R: GCTGTGCGACACATGGAC        | 162              | (TC) <sub>10</sub>                                     | TD65                                      | KF171047              | 160                 |    | 158–161 |
| mPs-UNICAMP35 | F: TCGAGAGTTGCGTGTGTTTC<br>R: CATTCTCCTGCCACCTGAGT     | 183              | (TG) <sub>4</sub>                                      | TD65                                      | KF171048              | 183                 |    | 183     |
| mPs-UNICAMP36 | F: GGGAGTCGGGTTGAGTTA<br>R: AGTCGAGGACAGTCAAAG         | 228              | (TG) <sub>4</sub> (TG) <sub>7</sub>                    | TD65                                      | KF171049              | 228                 |    | 228     |
| mPs-UNICAMP37 | F: TTGTTTGGGTTAGCGTGTGAG<br>R: CCCTGCCACCTGAGTAATCA    | 172              | (TG) <sub>6</sub>                                      | TD65                                      | KF171050              | 172                 |    | 172     |
| mPs-UNICAMP38 | F: CCTGACCTCTGGCACTACC<br>R: GAGGCGTATCAGGCTTTGA       | 112              | (TGC) <sub>6</sub>                                     | TD65                                      | KF171051              | 112                 |    | 112     |
| mPs-UNICAMP39 | F: GGAGGGTTGTTGTGTGAGTG<br>R: CTCCTGTGCGAAAGACTTCTG    | 230              | (GT) <sub>4</sub>                                      | TD65                                      | KF171052              | 230                 |    | —       |
| mPs-UNICAMP40 | F: GAATCAATGGAACACAAGCA<br>R: CCAGCCCCACTAGACCACCT     | 224              | (AC) <sub>5</sub>                                      | TD65                                      | KF171053              | 234                 |    | 230     |
| mPs-UNICAMP41 | F: CTTCACTGCAGCCTTCCAT<br>R: ATACCGATACTCGCCTTGATAG    | 168              | (GT) <sub>4</sub>                                      | 60°C                                      | KF171054              | 168                 |    | 170     |
| mPs-UNICAMP42 | F: AGTGCCAGTGGCTTCGTATT<br>R: GATCGTCATGGCTGTTGCTA     | 174              | (TGCAA) <sub>3</sub>                                   | 61.5°C                                    | KF171055              | 174                 |    | 174     |
| mPs-UNICAMP43 | F: CTCAGTGAGGAATAAGCAATCA<br>R: ATTTGGCATGCTGTTACGC    | 192              | (CA) <sub>4</sub>                                      | 61.5°C                                    | KF171056              | 198                 |    | 198     |
| mPs-UNICAMP44 | F: AGTCGTGCTTGTGTTGTTGAG<br>R: CCACTGTTGAGGTCCAGATG    | 275              | (GATT) <sub>3</sub>                                    | TD65                                      | KF171057              | 280                 |    | 275     |
| mPs-UNICAMP45 | F: CCTATACCTGCCCACTCAG<br>R: GTATGTGTGTGCCGTGGATT      | 110              | (AT) <sub>4</sub> (CA) <sub>4</sub>                    | TD65                                      | KF171058              | 110                 |    | 110     |
| mPs-UNICAMP46 | F: TGGCTGTTGTCCACCAT<br>R: GACTGAGCGGACTCACATCA        | 138              | (CT) <sub>8</sub>                                      | TD65                                      | KF171059              | 138                 |    | 138–139 |
| mPs-UNICAMP47 | F: AAATTCGGCATGGTTTATG<br>R: CCGAGATCGTTGGAGCTTA       | 298              | (AC) <sub>5</sub> (CA) <sub>4</sub>                    | 60°C                                      | KF171060              | 294                 |    | 298     |
| mPs-UNICAMP48 | F: AGCTTACCGGCTCACTCTTG<br>R: GACAGGCTTGGAAGTGAAT      | 144              | (AC) <sub>6</sub>                                      | 60°C                                      | KF171061              | 143                 |    | 142     |
| mPs-UNICAMP49 | F: TGTATGAGTGAGAAATGAGCCCA<br>R: CAATCAACATGAGACAAGCGG | 118              | (TA) <sub>4</sub>                                      | TD65                                      | KF171062              | 126                 |    | 126     |
| mPs-UNICAMP50 | F: TTCTGCGAAACTGGTGAGTG<br>R: CGCCCGTATTTGTCATGA       | 202              | (TA) <sub>6</sub>                                      | 60°C                                      | KF171063              | 202                 |    | 202     |



TABLE 1. Continued.

| Locus         | Primer sequences (5'–3')                           | Allele size (bp) | Repeat motif      | PCR amplification conditions <sup>a</sup> | GenBank accession no. | Cross-amplification |    |         |
|---------------|----------------------------------------------------|------------------|-------------------|-------------------------------------------|-----------------------|---------------------|----|---------|
|               |                                                    |                  |                   |                                           |                       | Pe                  | Ps | Pc      |
| mPs-UNICAMP51 | F: CTTGCACACTCAGGCTAAA<br>R: CAACCTACTGGATCGAAGTGA | 152              | (GT) <sub>5</sub> | 60°C                                      | KF171064              | 152                 |    | 150–152 |
| mPs-UNICAMP52 | F: GTCGGTTGAGAACCCGTA<br>R: ACCAATCGTTGAGAGTTCGTG  | 118              | (AT) <sub>5</sub> | 60°C                                      | KF171065              | 118                 |    | —       |

Note: — = unsuccessful amplification; Pe = *Passiflora cincinnata*; Ps = *Passiflora edulis*; Pc = *Passiflora setacea*.

<sup>a</sup>TD65 and TD58 indicate touchdown PCR programs with temperatures ranging from 65°C to 55°C and 58°C to 48°C, respectively.

cells (Stratagene, La Jolla, California, USA). The recombinant colonies were selected using blue/white screening. In total, 480 positive clones (192 for *P. edulis* and 288 for *P. setacea*) were randomly selected and double-sequenced using an ABI PRISM 377 automated DNA sequencer (Applied Biosystems, Foster City, California, USA). Every sequence was aligned and edited using SeqMan software (DNASTAR, Madison, Wisconsin, USA). We used the MICROSAT software developed by A. M. Risterucci at the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, France; unpublished) to identify and eliminate the adapters and restriction sites from the edited sequences.

Sequences containing microsatellites (134 for *P. edulis* and 114 for *P. setacea*) were identified using the SSR Identification Tool (SSRIT; Temnykh et al., 2001). Approximately 85% of the microsatellite motifs observed for both of the species were dinucleotides. We designed a total of 30 (*P. edulis*) and 75 (*P. setacea*) primer pairs using PrimerSelect (DNASTAR) and Primer3Plus

(Untergasser et al., 2007). The 105 primer pairs exhibited the following characteristics: annealing temperatures ranging from 45°C to 65°C (with a maximum difference of 3°C between the forward and reverse primers), CG concentrations ranging from 40% to 70%, and amplified product sizes varying from 100 to 300 bp. We used 16 genotypes of passion fruit (eight for each species) for the amplification tests. PCRs were conducted using a final volume of 15 µL (containing 15 ng of template DNA) with the reagents and concentrations described by Cerqueira-Silva et al. (2012). Every marker was evaluated by PCR amplification as follows: 94°C for 5 min; 34 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min; and a final extension at 72°C for 10 min. The loci that showed unsatisfactory amplification with an annealing temperature of 60°C were subjected to two different touchdown PCR protocols (TD 65–55°C and TD 58–48°C) as follows: an initial denaturation (94°C for 5 min); 10 cycles of 94°C for 1 min and an annealing temperature decreasing by 1°C from 65–55°C or 58–48°C every cycle for 1 min; 14 cycles of 94°C for 1 min, 55°C or 48°C

TABLE 2. Results of the initial screening of polymorphic microsatellite markers in populations of *Passiflora edulis*, *P. setacea*, and *P. cincinnata*.

| Locus                      | <i>P. edulis</i> (N = 42) |                |                |                    | <i>P. setacea</i> (N = 42) |                |                |                      | <i>P. cincinnata</i> (N = 31) |                |                |                     |
|----------------------------|---------------------------|----------------|----------------|--------------------|----------------------------|----------------|----------------|----------------------|-------------------------------|----------------|----------------|---------------------|
|                            | A                         | H <sub>o</sub> | H <sub>e</sub> | PIC                | A                          | H <sub>o</sub> | H <sub>e</sub> | PIC                  | A                             | H <sub>o</sub> | H <sub>e</sub> | PIC                 |
| mPe-UNICAMP01              | 2                         | 0.051          | 0.047          | 0.476              | 1                          | —              | —              | —                    | 1                             | —              | —              | —                   |
| mPe-UNICAMP02              | 4                         | 0.651          | 0.515          | 0.404              | 2                          | 0.261          | 0.497          | 0.371 <sup>a</sup>   | 1                             | —              | —              | —                   |
| mPe-UNICAMP06              | 1                         | —              | —              | —                  | 0                          | —              | —              | —                    | 2                             | 0.083          | 0.079          | 0.077               |
| mPe-UNICAMP09              | 1                         | —              | —              | —                  | 1                          | —              | —              | —                    | 6                             | 0.458          | 0.679          | 0.628 <sup>a</sup>  |
| mPe-UNICAMP13              | 1                         | —              | —              | —                  | 1                          | —              | —              | —                    | 3                             | 0.055          | 0.205          | 0.191 <sup>ab</sup> |
| mPs-UNICAMP01              | 5                         | 0.578          | 0.723          | 0.642 <sup>a</sup> | 2                          | 0.333          | 0.512          | 0.393 <sup>a</sup>   | 1                             | —              | —              | —                   |
| mPs-UNICAMP02              | 5                         | 0.631          | 0.768          | 0.704              | 2                          | 0.333          | 0.511          | 0.389 <sup>a</sup>   | 1                             | —              | —              | —                   |
| mPs-UNICAMP03              | 1                         | —              | —              | —                  | 2                          | 0.311          | 0.266          | 0.225                | 1                             | —              | —              | —                   |
| mPs-UNICAMP04              | 1                         | —              | —              | —                  | 2                          | 0.142          | 0.133          | 0.123                | 1                             | —              | —              | —                   |
| mPs-UNICAMP05              | 4                         | 0.381          | 0.471          | 0.476 <sup>b</sup> | 4                          | 0.261          | 0.593          | 0.468 <sup>a,b</sup> | 3                             | 0.401          | 0.513          | 0.392               |
| mPs-UNICAMP06              | 4                         | 0.191          | 0.176          | 0.157              | 2                          | 0.424          | 0.401          | 0.322                | 1                             | —              | —              | —                   |
| mPs-UNICAMP07              | 0                         | —              | —              | —                  | 2                          | 0.251          | 0.221          | 0.194                | 0                             | —              | —              | —                   |
| mPs-UNICAMP08              | 1                         | —              | —              | —                  | 2                          | 0.102          | 0.097          | 0.093                | 1                             | —              | —              | —                   |
| mPs-UNICAMP09              | 2                         | 0.024          | 0.024          | 0.023              | 4                          | 0.761          | 0.614          | 0.551                | 1                             | —              | —              | —                   |
| mPs-UNICAMP10              | 3                         | 0.119          | 0.197          | 0.186 <sup>a</sup> | 3                          | 0.357          | 0.583          | 0.493 <sup>ab</sup>  | 4                             | 0.448          | 0.637          | 0.577 <sup>a</sup>  |
| mPs-UNICAMP11              | 1                         | —              | —              | —                  | 3                          | 0.166          | 0.157          | 0.149                | 0                             | —              | —              | —                   |
| mPs-UNICAMP12              | 1                         | —              | —              | —                  | 2                          | 0.208          | 0.187          | 0.371                | 1                             | —              | —              | —                   |
| mPs-UNICAMP13              | 1                         | —              | —              | —                  | 2                          | 0.282          | 0.456          | 0.351 <sup>a</sup>   | 0                             | —              | —              | —                   |
| mPs-UNICAMP14              | 1                         | —              | —              | —                  | 4                          | 0.589          | 0.674          | 0.678 <sup>b</sup>   | 1                             | —              | —              | —                   |
| mPs-UNICAMP15              | 2                         | 0.024          | 0.024          | 0.023              | 3                          | 0.101          | 0.531          | 0.411 <sup>ab</sup>  | 1                             | —              | —              | —                   |
| mPs-UNICAMP16              | 3                         | 0.476          | 0.585          | 0.499              | 3                          | 0.391          | 0.485          | 0.395                | 4                             | 0.561          | 0.541          | 0.464               |
| mPs-UNICAMP17              | 1                         | —              | —              | —                  | 4                          | 0.833          | 0.714          | 0.656 <sup>b</sup>   | 1                             | —              | —              | —                   |
| mPs-UNICAMP18              | 0                         | —              | —              | —                  | 4                          | 0.524          | 0.454          | 0.412                | 0                             | —              | —              | —                   |
| mPs-UNICAMP19              | 1                         | —              | —              | —                  | 2                          | 0.189          | 0.173          | 0.566                | 1                             | —              | —              | —                   |
| mPs-UNICAMP24              | 1                         | —              | —              | —                  | 1                          | —              | —              | —                    | 2                             | 0.125          | 0.187          | 0.169               |
| mPs-UNICAMP31              | 1                         | —              | —              | —                  | 1                          | —              | —              | —                    | 3                             | 0.217          | 0.326          | 0.282               |
| mPs-UNICAMP34              | 0                         | —              | —              | —                  | 1                          | —              | —              | —                    | 5                             | 0.401          | 0.671          | 0.592 <sup>ab</sup> |
| mPs-UNICAMP46              | 1                         | —              | —              | —                  | 1                          | —              | —              | —                    | 2                             | 0.041          | 0.041          | 0.041               |
| mPs-UNICAMP51              | 1                         | —              | —              | —                  | 1                          | —              | —              | —                    | 2                             | 0.033          | 0.033          | 0.038               |
| mPe-UNICAMP11 <sup>c</sup> | 4                         | 0.237          | 0.447          | 0.424 <sup>b</sup> | 0                          | —              | —              | —                    | 1                             | —              | —              | —                   |
| mPe-UNICAMP19 <sup>c</sup> | 0                         | —              | —              | —                  | 4                          | 0.418          | 0.411          | 0.367                | 1                             | —              | —              | —                   |

Note: — = information not available; A = number of alleles per locus; H<sub>e</sub> = expected heterozygosity; H<sub>o</sub> = observed heterozygosity; PIC = polymorphism information content.

<sup>a</sup>Markers with the probability of null allele occurrence after a Bonferroni correction.

<sup>b</sup>Markers deviating from Hardy-Weinberg equilibrium after a Bonferroni correction ( $P < 0.004$  [*P. edulis* and *P. cincinnata*];  $P < 0.002$  [*P. setacea*]).

<sup>c</sup>Microsatellite markers published by Cerqueira-Silva et al. (2012).

for 1 min, and 72°C for 1 min; and a final extension at 72°C for 10 min. For markers that showed inconsistent amplification after the touchdown protocols, we tested reactions with an annealing temperature gradient ranging from 65°C to 50°C. The products were visualized using vertical electrophoresis on 6% denaturing polyacrylamide gels run in 1× TBE and stained with silver nitrate. The product sizes were determined using a 10-bp DNA ladder (Invitrogen, Carlsbad, California, USA). In total, 17 and 52 markers generated consistent patterns of amplification that matched the expected sizes based on the sequenced fragments from *P. edulis* and *P. setacea*, respectively (Table 1). Cross-amplification assays were performed according to previously described protocols, with all 69 primer pairs showing a high percentage of amplification (88% [*P. edulis*], 70% [*P. setacea*], and 80% [*P. cincinnata*]) (Table 1). Cross-amplification assays were also performed with the 25 loci previously characterized for *P. cincinnata* (Cerqueira-Silva et al., 2012), presenting a percentage of amplification of 48% in *P. edulis* (mPc-UNICAMP02, -04, -06, -10, -11, -14, -15, -17, -18, -20, -21, and -24) and 28% in *P. setacea* (mPc-UNICAMP02, -04, -06, -10, -15, -19, and -20).

To characterize all the loci, we used genotypes from the germplasm collection of the Embrapa Mandioca Fruticultura Center (Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]), Cruz das Almas, Bahia, Brazil, and of the UESB, Itapetinga, Bahia, totaling 114 genotypes. For each species, 42, 42, and 30 genotypes from *P. edulis* (all from EMBRAPA), *P. setacea* (30 from EMBRAPA and 12 from UESB), and *P. cincinnata* (all from EMBRAPA), respectively, were used (Appendix S1). We performed a descriptive statistical analysis for all the polymorphic loci using GENEPOP software (Raymond and Rousset, 1995; Table 2). The polymorphism information content was calculated using PIC Calculator software (Kemp, 2002), and the probability of null alleles was estimated using MICRO-CHECKER software (van Oosterhout et al., 2004), with significant probabilities between two and six loci observed for the three species evaluated (Table 2).

The percentage of polymorphic microsatellites observed was 15% in *P. edulis*, 29% in *P. setacea*, and 20% in *P. cincinnata*, totaling 11, 21, and 11 polymorphic loci, respectively (Table 2). This low number of polymorphic loci was expected because low variability appears to be a characteristic of the genus *Passiflora*, as suggested by Cerqueira-Silva et al. (2012). The number of alleles per locus ranged from two to six, with a mean of 3.1 for the three species evaluated; overall, the observed heterozygosity was lower than expected heterozygosity. Of the 31 polymorphic microsatellites, only one (*P. edulis*), six (*P. setacea*), and two (*P. cincinnata*) showed significant deviation from Hardy–Weinberg equilibrium (HWE) after a Bonferroni correction. Deviations from HWE can be explained by linkage disequilibrium (LD) or the occurrence of null alleles. Among the 320 possible pairs of microsatellites, we observed significant LD for two pairs (in *P. edulis*;  $P < 0.004$ ), 49 pairs (in *P. setacea*;  $P < 0.002$ ), and one pair (in *P. cincinnata*;  $P < 0.004$ ) after a Bonferroni correction. However, with no additional information, LD should not be attributed solely to physical linkages among loci because of the possibility of population processes such as nonrandom mating (Hedrick, 2005).

## CONCLUSIONS

We present the first set of microsatellites developed for *P. setacea* and characterize new markers for *P. edulis* and *P. cincinnata*, thereby increasing the number of available markers for these species. This effort potentiates the use of microsatellites in genetic studies of wild and commercial populations of *Passiflora* species, enabling the development of more efficient conservation and genetic breeding strategies.

## LITERATURE CITED

- BILLOTE, N., P. J. L. LAGODA, A. M. RISTERUCCI, AND F. C. BAURENS. 1999. Microsatellite-enriched libraries: Applied methodology for the development of SSR markers in tropical crops. *Fruits* 54: 277–288.
- CAZE, A. L. R., R. A. KRIEDT, L. B. BEHEREGARAY, S. L. BONATTO, AND L. B. FREITAS. 2012. Isolation and characterization of microsatellite markers for *Passiflora contracta*. *International Journal of Molecular Sciences* 13: 11343–11348.
- CERQUEIRA-SILVA, C. B. M., E. S. L. SANTOS, A. M. SOUZA, G. M. MORI, E. J. OLIVEIRA, R. X. CORRÊA, AND A. P. SOUZA. 2012. Development and characterization of microsatellite markers for the wild South American *Passiflora cincinnata* (Passifloraceae). *American Journal of Botany* 99: e170–e172.
- CERVI, A. C., M. A. MILWARD-DE-AZEVEDO, AND L. C. BERNACCI. 2010. *Passifloraceae*. In *Lista de Espécies da Flora do Brasil* [online]. Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil. Website <http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB182> [accessed 19 January 2014].
- FALEIRO, F. G. F., N. T. V. JUNQUEIRA, AND M. F. BRAGA. 2005. Germoplasma e melhoramento genético do maracujazeiro: Desafios da pesquisa, 55–78. In F. G. Faleiro, N. T. V. Junqueira, and M. F. Braga [eds.], *Maracujá: Germoplasma e melhoramento genético*. Embrapa Cerrados, Planaltina, Brazil.
- HEDRICK, P. W. 2005. *Genetics of populations*, 3rd ed. Jones Bartlett Publishers, Boston, Massachusetts, USA.
- KEMP, S. 2002. PIC Calculator Extra. Website <http://www.genomics.liv.ac.uk/animal/pic.html> [accessed 25 April 2012].
- OLIVEIRA, E. J., J. G. PADUA, M. I. ZUCCHI, L. E. A. CAMARGO, M. H. P. FUNGARO, AND M. L. C. VIEIRA. 2005. Development and characterization of microsatellite markers from the yellow passion fruit (*Passiflora edulis* f. *flavicarpa*). *Molecular Ecology Notes* 5: 331–333.
- ORTIZ, D. C., A. BOHÓRQUEZ, M. C. DUQUE, J. TOHME, D. CUÉLLAR, AND T. M. VÁSQUEZ. 2012. Evaluating purple passion fruit (*Passiflora edulis* Sims f. *edulis*) genetic variability in individuals from commercial plantations in Colombia. *Genetic Resources and Crop Evolution* 59: 1089–1099.
- PÁDUA, J. G., E. J. OLIVEIRA, M. I. ZUCCHI, G. C. X. OLIVEIRA, L. E. A. CAMARGO, AND M. L. C. VIEIRA. 2005. Isolation and characterization of microsatellite markers from the sweet passion fruit (*Passiflora alata* Curtis Passifloraceae). *Molecular Ecology Notes* 5: 863–865.
- PEREIRA, G. S. 2010. Desenvolvimento de Marcadores SSR, M-AFLP e SNP visando à integração de mapas genético-moleculares de *Passiflora alata* Curtis. M.Sc. dissertation, Universidade Estadual de São Paulo, Escola Superior de Agricultura ‘Luiz de Queiroz’, Piracicaba, São Paulo, Brazil.
- RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Heredity* 86: 248–249.
- TEMNYKH, S., G. CLERCK, A. LUKASHOVA, L. LIPOVICH, S. CARTINHO, AND S. MCCOUCH. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* 11: 1441–1452.
- ÜNTERGASSER, A., H. NIJVEEN, X. RAO, T. BISSELING, R. GEURTS, AND J. A. M. LEUNISSEN. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35 (Supplement 2): W71–W74.
- VAN OOSTERHOUT, C. V., W. F. HUTCHINSON, D. P. M. WILLS, AND P. SHIPLEY. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.





Anexo

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação de Mestrado/tese de Doutorado intitulada "**Diversidade genético-molecular de cacauzeiros descendentes das primeiras introduções ocorridas na Bahia**":

( ) não se enquadra no § 4º do Artigo 1º da Informação CCPG 002/13, referente a bioética e biossegurança.

Tem autorização da(s) seguinte(s) Comissão(ões):

(X) CIBio – Comissão Interna de Biossegurança, projeto No. 13/2003, Instituição:

( ) CEUA – Comissão de Ética no Uso de Animais, projeto No. \_\_\_\_\_, Instituição:

( ) CEP - Comissão de Ética em Pesquisa, protocolo No. \_\_\_\_\_, Instituição:

*\* Caso a Comissão seja externa ao IB/UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*

*Elisa Susilene Lisboa dos Santos*  
Aluno: Elisa Susilene Lisboa dos Santos

*Anete Pereira de Souza*  
Orientador: Anete Pereira de Souza

Para uso da Comissão ou Comitê pertinente:

(X) Deferido ( ) Indeferido

*Edi Lúcia Sartorato*  
Profª. Dra. EDI LÚCIA SARTORATO  
Lab. Genética Humana  
CBMEG - UNICAMP

Carimbo e assinatura

Para uso da Comissão ou Comitê pertinente:

( ) Deferido ( ) Indeferido

Carimbo e assinatura

CCPG / IB / UNICAMP - 21-Mai-2014-10:58-001019-2/3