**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Desenvolvimento de locos de microssatélite para a caracterização da
diversidade genética de acessos de urucum (*Bixa orellana* L.)**

Gabriel Dequigiovanni

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Genética e
Melhoramento de Plantas

**Piracicaba
2012**

Gabriel Dequigiovanni
Engenheiro de Bioprocessos e Biotecnologia

Desenvolvimento de locos de microssatélite para a caracterização da diversidade genética de acessos de urucum (*Bixa orellana* L.)

Orientador:
Profa. Dra. **ELIZABETH ANN VEASEY**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Genética e
Melhoramento de Plantas

Piracicaba
2012

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Dequigiovanni, Gabriel

Desenvolvimento de locos de microsatélite para a caracterização da diversidade genética de acessos de urucum (*Bixa orellana* L.) / Gabriel Dequigiovanni.- - Piracicaba, 2012.

71 p: il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Banco de germoplasma 2. Diversidade genética 3. Marcadores microsatélites
4. Urucum I. Título

CDD 633.86
D426d

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICO,

Às pessoas que amo sem medidas ou proporções, apenas as amo com toda força e intensidade. Á minha fortaleza, meus pais Alda e Enio pelo amor e apoio incondicional. Ao meu irmão Diego por, mesmo distante, sempre estar presente em minha vida. A minha noiva Patricia, pelo amor, carinho, respeito e companheirismo, e por permitir que todos os nossos sonhos sejam realizados.

De coração agradeço

A Deus, pelo dom da vida, por sempre me iluminar e me guiar;

À Universidade de São Paulo, por meio da Escola Superior de Agricultura “Luiz de Queiroz”, pela oportunidade de realização do curso de Mestrado;

À Dr^a Elizabeth Ann Veasey, do Laboratório de Ecologia Evolutiva e Genética Aplicada, minha orientadora, pela confiança, pelas oportunidades, pelos esclarecimentos e pela grande ajuda na redação deste trabalho e dos projetos a ele relacionados. Por nunca medir esforços na realização deste projeto, pelas excelentes correções, e pelo total apoio em todas as minhas decisões, a você Betty, meu mais sincero muito obrigado;

Aos meus estimados pais, Enio e Alda, pelos valores, princípios e amor incondicional, com toda certeza devo tudo a vocês;

Ao meu irmão Diego e minha cunhada Carla pela amizade e incentivo em todos os momentos, que mesmo distante sempre mandaram muitas mensagens de incentivo;

À minha noiva Patrícia, pelo amor, paciência, compreensão, apoio, incentivo e carinho em todos os momentos. Sem você nada disso seria possível, muito obrigado por tudo;

A família Schaker, Jamil, Soili, Letícia e Felipe, pela confiança, e por todo apoio durante todos estes anos;

A pesquisadora Dra. Eliane Gomes Fabri (IAC) pela ajuda na escolha e fornecimento do material vegetal, viagens e coletas, e também pela disponibilidade em solucionar dúvidas, e a todo seu empenho que me permitiu a realização deste trabalho;

Aos amigos e colegas do Laboratório de Ecologia Evolutiva e Genética Aplicada, Lidinalva, Wellington, Berenice, Thiago, Marcos, Danielle, Carol, Jucelene, Mariana, Sueme, Ana, Santiago, e Nancy pela amizade, paciência e pela ajuda durante o desenvolvimento deste projeto;

A todos os professores do Departamento de Genética da ESALQ/USP, pelos grandes ensinamentos durante o curso de mestrado;

Aos funcionários do Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz”, em especial a Léia, por sempre ter auxiliado em tudo que foi necessário;

Ao amigo MSc. Marcos Vinícius Bohrer Monteiro Siqueira, por todo o apoio na minha primeira vinda a Piracicaba;

Ao técnico do laboratório, Marcos Cella (Marcão), pelo auxílio na realização dos experimentos, e pela grande amizade;

A Ana Claudia Pereira Cuenca pelo auxílio no desenvolvimento do trabalho;

Aos amigos do Café, Mariana, Sueme, Alessandro, Santiago e Glaucia, por tornarem meus dias muito mais felizes e harmoniosos;

Ao meu amigo Alessandro Alves pelo grande auxílio nas análises estatísticas, além das discussões que renderam resultados neste trabalho;

Ao Prof. Dr. José Baldin pelo uso do laboratório durante o desenvolvimento dos iniciadores, e pelas boas sugestões no desenvolvimento do projeto;

A Prof. Dr. Maria Imaculada Zucchi, pelo auxílio no desenvolvimento da biblioteca enriquecida e pelos grandes ensinamentos;

Aos amigos Carlos e Miklos, do Laboratório de Diversidade Genética e Melhoramento do Departamento de Genética da ESALQ/USP, pelo auxílio no desenvolvimento da biblioteca enriquecida;

Ao Dr. Eduardo de Andrade Bressan pelas sugestões durante o desenvolvimento do projeto, na escrita e também nas rotinas do laboratório;

A todos aqueles que de forma direta ou indireta tenham contribuído para a realização deste trabalho.

*And please remember people, that no matter who you are, and what you do to
live, thrive and survive, there are still some things that make us all the same.
You, me them, everybody, everybody!*

(Elwood Blues)

"A mente que se abre a uma nova ideia jamais volta ao seu tamanho original"

(Albert Einstein)

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO.....	19
1.1 Objetivos.....	20
1.2 Hipóteses	20
2 REVISÃO BIBLIOGRÁFICA.....	21
2.1 Descrição geral	21
2.2 Origem, evolução e aspectos nutricionais	23
2.3 Diversidade genética e conservação	25
2.3.1 Marcadores moleculares	27
2.3.1.1 Marcadores microssatélites	28
2.3.1.2 Diversidade genética em urucum	30
3 MATERIAL E MÉTODOS.....	33
3.1 Material vegetal.....	33
3.2 Desenvolvimento de iniciadores de microssatélites	36
3.2.1 Construção de biblioteca genômica enriquecida com SSR	36
3.2.1.1 Extração de DNA	36
3.2.1.2 Restrição do DNA	36
3.2.1.3 Ligação dos adaptadores.....	36
3.2.1.4 Pré-Amplificação via PCR	37
3.2.1.5 Seleção dos fragmentos contendo SSR	37
3.2.1.6 Amplificação dos fragmentos selecionados via PCR.....	37
3.2.1.7 Clonagem dos fragmentos selecionados.....	37
3.2.1.8 Sequenciamento dos clones positivos para microssatélites.....	38
3.2.1.9 Construção dos microssatélites.....	38
3.3 Avaliação dos acessos de urucum com marcadores microssatélites	39
3.3.1 Extração de DNA	39
3.3.2 Análises de PCR e eletroforese para obtenção das regiões microssatélites	40
3.4 Análise estatística dos resultados	41

4 RESULTADOS E DISCUSSÃO	45
4.1 Desenho dos iniciadores	45
4.2 Condições de amplificação dos locos microssatélites	48
4.3 Avaliação dos acessos do Banco de Germoplasma com os iniciadores desenvolvidos	48
5 CONCLUSÕES	61
REFERÊNCIAS	63

RESUMO

Desenvolvimento de locos de microssatélite para a caracterização da diversidade genética de acessos de urucum (*Bixa orellana* L.)

O objetivo do presente projeto foi caracterizar a diversidade genética de 63 acessos de urucum (*Bixa orellana*), mantidos no Banco de Germoplasma do Instituto Agronômico (IAC), a partir da utilização de marcadores microssatélites. Para tanto foram desenvolvidos locos microssatélites para a espécie por meio de uma biblioteca genômica enriquecida. A partir da biblioteca genômica desenvolvida foram obtidas e sequenciadas um total de 84 colônias. Deste total, foram encontrados microssatélites em 57 colônias, o que representa 67,9% de enriquecimento. Foram identificados 70 microssatélites, sendo que destes, foram desenhados e selecionados um total de 31 iniciadores. Dentre estes, 25 iniciadores apresentaram produtos de amplificação, sendo 15 (60%) monomórficos para o grupo de indivíduos estudados. A caracterização dos acessos presentes no germoplasma de urucum, com os 10 iniciadores polimórficos permitiu identificar 38 alelos polimórficos, variando de 2 a 6 alelos por loco, obtendo-se uma média de 3,8 alelos por loco. Os valores de heterozigosidade esperada (H_e) variaram de 0,464 a 0,765, com média de 0,572. Por outro lado, a heterosigozidade observada (H_o) variou de 0 a 0,744, com média de 0,362. A partir dos índices de fixação de Wright foi possível inferir a taxa aparente de cruzamento, que neste caso foi de 0,39. Os valores de PIC observados neste estudo variaram de 0,359 a 0,725, com média de 0,497. Os marcadores BorB4, BorG3, BorC12, BorG4, BorA2 e BorF5_2 demonstraram ser moderadamente informativos enquanto os iniciadores BorB10, BorB12, BorF1 e BorE7 foram considerados marcadores altamente informativos. A análise de agrupamento para os genótipos avaliados apresentou estruturação em dois grandes grupos, um com acessos da região Norte e outro com os acessos da região Sudeste. Os valores das distâncias de Rogers modificada estimados a partir dos marcadores SSR, variaram de 0 a 0,968, podendo ser identificadas apenas duas duplicatas. O coeficiente de correlação de Pearson (r) entre as distancias genéticas e geográficas obtido com o teste de Mantel foi baixo ($r = 0,367$), embora significativo ($P < 0,01$). A partir da análise bayesiana realizada pelo software Structure, o conjunto de acessos foi dividido em dois grupos, que coincidiram com os grupos obtidos na análise de agrupamento, demonstrando a consistência dos dados obtidos. Foi possível identificar que o grupo de acessos do Grupo 2 (região Norte) apresenta índices de diversidade superiores ao Grupo 1 (região Sudeste). Os valores médios obtidos de F_{ST} (0,183) e G_{ST} (0,194) indicam a existência de estruturação na amostra total e corrobora os resultados obtidos pela análise de agrupamento, e também com os dados obtidos pela análise de estruturação com o software Structure. Os valores médios de F_{IT} (0,483) e H_T (0,632) demonstraram a elevada diversidade genética no conjunto de acessos estudado. O fato de que os valores de F_{IS} e G_{IS} sejam maiores do que os valores de F_{ST} e G_{ST} para praticamente todos os locos, indicam a existência de maior diversidade dentro de cada grupo de acessos do que entre os grupos. A manutenção do banco ativo de germoplasma de urucum possibilita a preservação do patrimônio genético e pode fornecer matéria prima para as atividades de melhoramento e fomento da cadeia produtiva da cultura.

Palavras chave: Urucum; *Bixa orellana*; Marcadores microssatélites; Diversidade genética; Banco de Germoplasma

ABSTRACT

Development of microsatellite loci to characterize the genetic diversity of annatto (*Bixa orellana* L.) accessions

The aim of this project was to characterize the genetic diversity of 63 accessions of annatto (*Bixa orellana*), maintained at the Germplasm Bank of the Agronomic Institute (IAC), using microsatellite markers. For that purpose, microsatellite loci have been developed for this species through an enriched DNA genomic library. From the genomic library developed a total of 84 colonies were obtained and sequenced. Microsatellites were found in 57 colonies, representing 67.9% of enrichment. Seventy microsatellites were identified and, from those, a total of 31 primers were selected. Among these, 25 primers showed amplification products and 15 (60%) were monomorphic for the group of accessions studied. The germplasm accessions characterization with the 10 polymorphic primers showed 38 polymorphic alleles, ranging from 2 to 6 alleles per locus, yielding an average of 3.8 alleles per locus. Values of expected heterozygosity (H_e) ranged from 0.464 to 0.765, averaging 0.572. Moreover, the observed heterozygosity (H_o) ranged from 0 to 0.744, averaging 0.362. From the Wright's fixation index, it was possible to infer the apparent outcrossing rate, which in this case was 0.39. The PIC values observed in this study ranged from 0.359 to 0.725, averaging 0.497. The markers BorB4, BorG3, BorC12, BorG4, and BorA2 BorF5_2 were shown to be moderately informative, and the primers BorB10, BorB12, BorF1 BorE7 were considered highly informative markers. The cluster analysis for the genotypes showed structuring into two major groups, one with the accessions from the North and another with the accessions from the Southeast. The values of the modified Rogers distances estimated from the SSR markers ranged from 0 to 0.968, and only two duplicates were identified. The Pearson correlation coefficient (r) between genetic and geographic distances obtained with the Mantel test was low ($r = 0.367$), although significant ($P < 0.01$). From the Bayesian analysis performed by the software Structure, the accessions were divided into two groups, which coincided with the groups obtained from the cluster analysis, demonstrating the consistency of the results obtained. It was possible to identify that the accessions in Group 2 (Northern region) present diversity indexes higher than Group 1 (Southeast). The average values of F_{ST} (0.183) and G_{ST} (0.194) indicate the existence of genetic structure in the total sample and confirms the results obtained in the cluster analysis, and also with the data obtained in the structure analysis using the software Structure. The average values of F_{IT} (0.483) and H_T (0.632) demonstrated the high genetic diversity among the accessions studied. The fact that the F_{IS} and G_{IS} values are higher than the values of F_{ST} and G_{ST} for almost all loci indicates that diversity is higher within each group than between groups of accessions. Maintaining the annatto germplasm bank enables the preservation of genetic resources and can provide material for breeding activities and improvements in the crop production chain.

Keywords: Annatto; *Bixa orellana*; Microsatellite markers; Genetic diversity; Germplasm Bank

LISTA DE FIGURAS

- Figura 1 – Aspectos morfológicos da planta de Urucum, com detalhe para as folhas jovens e adultas, flor, sementes e fruto. PROSEA Network Office, Herbarium Bogoriense, Indonesia..... 22
- Figura 2 – Distribuição dos registros de ocorrência de urucum de acordo com herbários. Disponível em <http://www.discoverlife.org>. PROSEA Network Office, Herbarium Bogoriense, Indonesia..... 24
- Figura 3 – Plantas de *Bixa orellana* na coleção de germoplasma do Instituto Agromômico em Pindorama – SP 35
- Figura 4 – Mapa do Brasil com a identificação dos acessos brasileiros mantidos pelo Banco de Germoplasma do IAC. Os estados marcados em verde são da Região Sudeste, em azul da Região Centro-Oeste e em vermelho da Região Norte..... 35
- Figura 5 – Géis de poliacrilamida resultante da amplificação dos locos BorC12 (A), BorF1 (B), BorG3 (C), BorB10 (D), e BorG4 (E) em 36 indivíduos de urucum 49
- Figura 6 – Frequências alélicas dos 63 acessos de urucum (*Bixa orellana*) estimadas com 10 locos SSR 51
- Figura 7 – Dendrograma obtido com o método de UPGMA e distância de Rogers modificada. Os acessos marcados em vermelho são oriundos da região Norte, os marcados em verde são da região Sudeste, enquanto os marcados em azul são da região Centro-Oeste. O acesso marcado em amarelo é oriundo do Peru e em preto encontram-se os acessos sem identificação (SI), de origem desconhecida 55
- Figura 8 – Teste de atribuição para os 63 acessos do banco de germoplasma de urucum a partir de 10 locos microssatélites. No primeiro gráfico os genótipos foram

classificados em dois grupos ($K=2$). No segundo gráfico, os acessos foram classificados em três grupos ($K=3$) e no terceiro gráfico, os acessos foram agrupados em quatro grupos ($K=4$). Os acessos estão representados pelas barras verticais coloridas. A mesma cor para genótipos diferentes indica que eles pertencem ao mesmo grupo. Cores diferentes no mesmo indivíduo indicam a porcentagem de alelos compartilhados com cada grupo. O eixo das ordenadas mostra o coeficiente de participação do indivíduo em cada grupo56

Figura 9 – Valores de ΔK para cada valor de K . O maior valor de ΔK representa o valor ideal de K de acordo com o proposto por Evanno et al. (2005)57

Figura 10 – Análise da estrutura genética dos 63 acessos do banco de germoplasma de urucum a partir de 10 locos microssatélites. Os genótipos foram classificados em dois grupos ($K = 2$). Acessos de cor verde são originados da região norte e condizentes com o primeiro grupo da análise de agrupamento e de cor vermelha são, em sua maioria, originados da região Sudeste e condizentes com o segundo grupo na análise de agrupamento58

LISTA DE TABELAS

Tabela 1 - Relação dos acessos de urucum, número de identificação dos acessos mantidos no banco de germoplasma do Instituto Agronômico (IAC), origem e Estado de coleta	33
Tabela 2 - Sequências dos iniciadores incluindo nome, motivo, porcentagem de GC, temperaturas de anelamento, tamanhos esperados dos produtos em pares de bases (pb), e classificação de cada iniciador desenvolvido	46
Tabela 3 - Parâmetros de diversidade genética avaliados na coleção de urucum (<i>Bixa orellana</i>) do Banco de Germoplasma do IAC, incluindo número de alelos (A), heterozigosidade esperada (H_e), heterozigosidade observada (H_o), PIC e índice de fixação de Wright (f)	50
Tabela 4 - Teste exato de aderência ao Equilíbrio de Hardy-Weinberg para cada loco	53
Tabela 5 - Estimativas dos índices de fixação de Nei (H_S , D_{ST} , H_T , G_{ST} , G_{IS}) e Wright (F_{IT} , F_{ST} e F_{IS}) avaliados na coleção de urucum (<i>Bixa orellana</i>) do Banco de Germoplasma do IAC, considerando dois grupos gerados pela análise de estruturação.....	58
Tabela 6 - Parâmetros de diversidade genética avaliados na coleção de urucum (<i>Bixa orellana</i>) do Banco de Germoplasma do IAC, considerando dois grupos gerados pela análise de estruturação, incluindo número de alelos (A), heterozigosidade esperada (H_e), heterozigosidade observada (H_o), e índice de fixação de Wright (f)	59

1 INTRODUÇÃO

O urucum (*Bixa orellana* L.) é uma cultura tropical originária da América provavelmente da região Amazônica. A espécie pertence à família Bixaceae. Dentro do seu gênero é a única espécie amplamente cultivada em todos os países tropicais. Sendo assim, os principais países produtores de urucum são o Brasil, Peru e Quênia, sendo também encontrado na República Dominicana, Colômbia, Jamaica, Costa Rica, Suriname e outros países da Ásia (AKSHATHA et al., 2011). No Brasil, o urucum é cultivado em diversos estados, sendo São Paulo a principal região produtora (FABRI et al., 2011). O urucum apresenta grande importância econômica, por ser a única fonte do pigmento natural bixina. Estes corantes extraídos do urucum encontram-se entre os poucos permitidos pela Organização Mundial da Saúde (OMS). Além de não serem tóxicos, não alteram o sabor dos alimentos, tendo sido muito utilizados em diversos ramos da indústria, desde a farmacêutica até a alimentícia.

A conservação e manipulação de recursos genéticos do urucum representam as principais e mais difíceis metas almejadas nessa cultura. Por ser considerado o centro de origem do Urucum, o Brasil apresenta grande diversidade para esta espécie, sendo de grande importância a caracterização dessa diversidade visando a conservação de genótipos, além de fornecer informações aos programas de melhoramento genético.

Bancos de germoplasma têm como objetivo principal evitar a perda de alelos que diminui a variabilidade genética disponível de determinada espécie. A perda da variabilidade leva a diminuição da eficiência do processo de seleção e aumenta a vulnerabilidade a pragas e condições abióticas adversas. Neste sentido, são considerados repositórios de material genético e representam uma alternativa interessante visando à manutenção da variabilidade genética parcial ou total de uma determinada espécie, contribuindo para a manutenção da biodiversidade global.

Esta preocupação com a conservação dos recursos genéticos tem estimulado o estabelecimento de bancos de germoplasma *ex situ* de urucum no Brasil. A variabilidade genética disponível em bancos de germoplasma ainda é pouco explorada, em grande parte devido à carência na caracterização desses recursos genéticos. A caracterização destes bancos pode fornecer informações fundamentais para que a variabilidade presente seja utilizada em programas de melhoramento genético e de conservação.

1.1 Objetivos

Pretende-se com esse trabalho atingir dois principais objetivos:

- Desenvolver *primers* de microssatélites para a espécie *B. orellana*, por meio da tecnologia de biblioteca genômica enriquecida;
- Com os locos de microssatélites desenvolvidos neste estudo, caracterizar a diversidade genética dos acessos de urucum (*B. orellana*) mantidos no Banco de Germoplasma do Instituto Agrônomo (IAC), verificando-se o nível de diversidade genética da coleção, bem como a distribuição da variabilidade entre e dentro de regiões geográficas.

1.2. Hipóteses

- Os acessos do banco de germoplasma de urucum do IAC apresentam elevada diversidade genética por possuírem materiais de diferentes procedências;
- Os acessos provenientes da região Norte apresentam maior diversidade por serem procedentes de uma região que pode estar representando o centro de diversidade da espécie.

2 REVISÃO BIBLIOGRÁFICA

2.1 Descrição geral

O urucum (*Bixa orellana* L.) pertence à família Bixaceae. Esta família é representada por apenas um gênero, *Bixa* L. Deste gênero, *B. orellana* é a única espécie amplamente cultivada em todos os países tropicais (BAER, 1976; PAIVA NETO et al., 2003).

O urucuzeiro é uma planta arbustiva que em condições normais atinge entre 3 e 4 metros de altura, podendo chegar a até 10 metros dependendo do manejo, da idade e das condições de clima e solo em que é cultivada. Possui folhas completas, com longos pecíolos, cordiformes, acuminadas, dispostas alternadamente em relação aos ramos, glabras (quando adultas), medindo em média 8 cm de comprimento e 4 cm de largura (FRANCO et al., 2008).

As flores são grandes e a cor varia de branca a várias tonalidades de rosa até lilás, apresentando um ovário contendo uma série de óvulos em seu interior. São hermafroditas, com cinco sépalas, surgindo nas extremidades dos ramos, formando fascículos onde nascem cápsulas com cinco pétalas orbiculares, glandulosas na base, decíduas, obovais e inteiras (SANTOS, 1958; FRANCO et al., 2008).

O fruto é em forma de cápsula deiscente, ovóide, com dois ou três carpelos, coberto de espinhos flexíveis quando juvenis. Pode ser avermelhado, esverdeado ou pardo, medindo de três a cinco centímetros de comprimento, contendo muitas sementes pretas cobertas por um arilo ceroso de cor avermelhada e odor característico. Os frutos são dispostos em cachos com até 17 unidades. Um fruto bem desenvolvido pode fornecer 40 a 60 sementes (LORENZI, 2002; FRANCO et al., 2008). No Estado de Yucatan, México, Rivera-Madrid et al. (2006) observaram uma média de 32 sementes por fruto. Akshatha et al. (2011), avaliando acessos de urucum em casa-de-vegetação, na Índia, observaram que o maior número de sementes estava associado aos frutos de coloração avermelhada, com 46 sementes por fruto, em média, tendo estes também apresentado as maiores produções de sementes (1.5 – 2.0 kg de sementes por planta de quatro anos de idade).

Figura 1 – Aspectos morfológicos da planta de Urucum, com detalhe para as folhas jovens e adultas, flor, sementes e fruto. PROSEA Network Office, Herbarium Bogoriense, Indonesia

É uma espécie diplóide, com um número básico de cromossomos de $x=7$ ou 8 . O número de cromossomos de $2n=14$ tem sido relatado para a espécie (OHRI et al., 2004), embora $2n=16$ tenha sido também relatado (KRISHNAN; AYYANGAR, 1987).

Almeida e Pinheiro (1992) descreveram uma série de características da biologia floral do urucuzeiro do tipo verde piloso, realizando inclusive polinizações controladas. Concluíram que a fecundação cruzada é o método natural de reprodução no tipo estudado. Vilares et al. (1992) realizaram o mesmo tipo de estudo e concluíram pela existência de autofecundação natural em urucuzeiro, sem a presença de insetos polinizadores. Já Rivera-Madrid et al. (2006), realizando polinizações controladas em acessos de urucum do México, obtiveram 57% de polinização cruzada e 31,4% de autofecundação. Valdez-Ojeda et al. (2010), analisando oito progênies de autopolinização com 10 indivíduos com o marcador dominante SRAP (Sequence-Related Amplified Polymorphism), também no México, concluíram que a espécie se reproduz por polinização cruzada, estimando uma taxa de cruzamento multilocos de 0,748. Joseph et al. (2011) realizaram estudos bem detalhados sobre a biologia reprodutiva

do urucum na Índia, também concluindo que a espécie apresenta polinização cruzada, podendo também se reproduzir por autofecundação. Miller e Gross (2011) sugerem que a reprodução por alogamia é comum em plantas perenes e que esta pode mudar durante o processo de domesticação, chegando até a existir populações em que a autogamia é comum. Esta mudança é facilitada em plantas perenes cultivadas onde é comum haver meios irmãos próximos um dos outros, o que aumenta a probabilidade de cruzamentos entre parentes, o que explica os resultados obtidos por Rivera-Madrid et al. (2006) e Valdez-Ojeda et al. (2010).

2.2 Origem, evolução e aspectos nutricionais

O urucum é uma cultura tropical originária da América (ARCE, 1999), provavelmente da região amazônica (SANDY-CUEN; BECERRA, 2003; CLEMENT et al., 2010). O urucum encontra-se disperso em todas as terras baixas tropicais e subtropicais da América do Sul e até a Mesoamérica e México (SCHULTES, 1984), onde ganhou seu primeiro nome na língua Maia a pelo menos 2.400 anos antes do presente (BROWN, 2010). É também encontrado em alguns locais na Oceania, África e Ásia (Figura 2).

É conhecido por “achiote” na língua espanhola, “annatto” na língua inglesa, e “urucú” ou “urucum” na língua portuguesa (AKSHATHA et al., 2011). Na língua dos índios Ka'apor, no Maranhão, é possível distinguir tipos domesticados dos tipos silvestres, uma vez que “uruku” se refere ao urucum domesticado, enquanto que “urukuran” se refere ao urucum em estado silvestre (BALÉE, 1994).

Os principais países produtores de urucum são Brasil, Peru e Quênia, embora seja também cultivado na República Dominicana, Colômbia, Jamaica, Costa Rica e Suriname, tendo sido cultivado também, em menor escala, em alguns países da Ásia (AKSHATHA et al., 2011). No Brasil, o urucum é cultivado em diversos estados, sendo São Paulo a principal região produtora (FABRI et al., 2011). A produção brasileira de urucum situa-se em torno de 12.000 a 13.000 t/ano, sendo que desse total, 60% são destinados à fabricação de corante/colorífico, 30% à fabricação de corantes e 10% à exportação (IBGE, 2009). Essa disseminação está relacionada com a crescente utilização dos corantes naturais vermelhos e amarelos, nas indústrias farmacêuticas, têxteis, de laticínios, de alimentos, de bebidas, de tintas e de cosméticos (MELO; LIMA, 1990; CARVALHO, 1992). Sua grande importância econômica reside no fato de ser a única fonte do pigmento natural, a bixina, um apocarotenóide extraído do pericarpo das sementes, que representa mais de 80% de todos os carotenóides presentes (ANGELUCI et al., 1980; GIULIANO et al., 2003). A partir dela são

obtidos os demais pigmentos, como a norbixina, que é um corante lipossolúvel, o sal de norbixina, hidrossolúvel, e os produtos de degradação térmica, que têm como características a lipossolubilidade e uma coloração amarela mais estável (CARVALHO, 1992).

Figura 2 – Distribuição dos registros de ocorrência de urucum de acordo com herbários. Disponível em <http://www.discoverlife.org>

A utilização desses corantes naturais tem sido impulsionada devido à crescente proibição da utilização de corantes sintéticos nas indústrias de alimentos e cosméticos (FERREIRA; FALESI, 1989). Os corantes extraídos do urucum encontram-se entre os poucos permitidos pela Organização Mundial da Saúde (OMS). Além de não serem tóxicos, não alteram o sabor dos alimentos. A partir da restrição do uso de corantes sintéticos na indústria alimentícia, os extratos de urucum passaram a ganhar interesse e estímulo a pesquisas (PRENTICE-HERNANDEZ; RUSIG, 1992).

Além do uso para produção de corantes naturais, o urucum é também considerado planta medicinal (LORENZI, 2002). Há referências do uso das sementes na literatura etnofarmacológica como medicação estomáquica, tonificante do aparelho gastrointestinal, antidiarréica, antifebril, bem como para o tratamento caseiro das palpitações do coração, crises de asma, coqueluche e gripe (FRANCO et al., 2008). Além disso, os corantes de urucum apresentam atividade antimicrobiana (COELHO et al., 2003) e também auxiliam na redução de glicose e colesterol do sangue (MCGONIGLE; MCCRACKEN, 2002; RUSSEL et al., 2005). Outras atividades farmacológicas são verificadas no urucum, tais como antioxidante (LIMA et al., 2001), anti-virais (CÁCERES et al., 1995), hipotensor (RODRIGUES, 1988) e hiperglicemiante (MORRISON et al., 1987; THOMPSON, 1990).

Os corantes de urucum também tem grande importância na história do Brasil, uma vez que ainda são utilizados pelas tribos indígenas na pintura do corpo e tingimento de vestes (PLOTKIN, 1993).

A conservação e manipulação de recursos genéticos do urucum representam as principais e mais difíceis metas almejadas nessa cultura, face à grande heterogeneidade de tipos existentes nos quintais do Brasil e outros países de América Latina, e à necessidade de se desenvolver variedades mais homogêneas e produtivas, adaptadas às características edafoclimáticas de cada região do país (RODRIGUES, 1995). Além da grande abundância de variedades locais de urucum mantidas em comunidades tradicionais da Amazônia, bem como em outras regiões do país, há indícios da ocorrência de populações silvestres em Rondônia, especialmente ao longo do Rio Mamoré, tanto em vegetações ripárias como em ambientes mais secos longe dos rios, bem como em Roraima e no Pará. Algumas populações estão representadas por estandes quase monoespecíficos longe das margens dos rios. E o exame em herbários do INPA e New York Botanical Garden, detectou urucum silvestre coletado também em Oriximiná, PA, no Parque Nacional do Jaú, AM, e em Brasília, AC (informação verbal)¹.

2.3 Diversidade genética e conservação

Com o crescente aumento da busca por produtos com alta produtividade, qualidade nutricional, e demais características de interesse para os produtores e consumidores, esta seleção tem levado a um estreitamento da base genética em determinadas culturas, com isso há uma grande preocupação da comunidade científica no que diz respeito à diminuição ou perda de variabilidade genética de espécies cultivadas e de seus parentes silvestres, bem como de variedades (HALLAUER; MIRANDA, 1988). A condição de vulnerabilidade resultante do estreitamento da base genética vegetal só pode ser evitada com a existência de variabilidade, a qual depende da conservação dos recursos genéticos disponíveis (MUNOZ, 2003).

Bancos de germoplasma são considerados um método de conservação *ex situ*, em que amostras da variabilidade genética da espécie em interesse é conservada em condições controladas, fora do habitat natural da espécie (NASS, 2001). Bancos de germoplasma têm como objetivo principal evitar a perda de alelos que diminui a variabilidade genética disponível de determinada espécie. A perda da variabilidade leva a diminuição da eficiência

¹ Juliana Lins. Instituto Nacional de Pesquisas da Amazônia (INPA).

do processo de seleção e aumenta a vulnerabilidade a pragas e condições abióticas adversas (YORINORI et al., 2001). Neste sentido, são considerados repositórios de material genético, como plantas inteiras e/ou sementes, e representam uma alternativa interessante visando à manutenção da variabilidade genética parcial ou total de uma determinada espécie, contribuindo para a manutenção da biodiversidade global (BORÉM, 1999).

Esta preocupação com a conservação dos recursos genéticos tem estimulado o estabelecimento de bancos de germoplasma *ex situ* de urucum no Brasil. Coleções de germoplasma são mantidas pela Embrapa e por outras instituições (MORAIS et al., 1999). No Estado de São Paulo, o Instituto Agronômico de Campinas (IAC) tem trabalhado com recursos genéticos de urucum. O Banco de germoplasma do IAC contém 63 acessos coletados em regiões brasileiras, e também inclui um acesso do Peru (informação verbal)².

A variabilidade genética disponível em bancos de germoplasma ainda é pouco explorada pelos melhoristas, em grande parte devido à carência na caracterização desses recursos genéticos (NASS, 2001). A caracterização destes bancos pode fornecer informações fundamentais para que a variabilidade presente seja utilizada em programas de melhoramento genético e de conservação. Programas de melhoramento genético de urucum ainda estão em fases primárias, tendo sido iniciados na década de 80, e consistem basicamente na comparação fenotípica entre acessos provenientes de diferentes regiões (VILLARES et al., 1992).

Em bancos de germoplasma, o estudo da diversidade genética pode ser realizado através da utilização de vários descritores, desde os fitotécnicos (morfológicos) até os moleculares (JATASRA; PARODA, 1983).

Os marcadores moleculares são ferramentas úteis para detectar variações no genoma, aumentando o poder da análise genética das plantas. Estes marcadores estão substituindo rapidamente outros marcadores em vários tipos de estudos genéticos, principalmente devido à sua reprodutibilidade e simplicidade técnica, à pequena quantidade de DNA requerida, ao baixo custo, ao grande poder de resolução e aos altos níveis de polimorfismo (CAIXETA et al., 2006).

² Eliane Gomes Fabri. Instituto Agronômico (IAC).

2.3.1 Marcadores moleculares

Até a década de 60, os marcadores utilizados em estudos de diversidade genética eram controlados por genes associados a caracteres morfológicos, com fenótipo de fácil identificação visual (FERREIRA; GRATTAPAGLIA, 1998). Contudo, com o baixo número de marcadores morfológicos existentes, diminuía a probabilidade de encontrar associações significativas entre estes marcadores e a variação entre os indivíduos estudados (BORÉM, 1999).

A evolução deste processo deu-se através do desenvolvimento de marcadores isoenzimáticos, aumentando significativamente o número de marcadores genéticos disponíveis para análise. Com isso, os marcadores isoenzimáticos passaram a ser amplamente utilizados em inúmeras espécies vegetais (BORÉM, 1999). Posteriormente, surgiram diversos métodos de detecção de polimorfismo genético diretamente ao nível de DNA. Inicialmente, foi desenvolvida a técnica de RFLP (Restriction Fragment Length Polymorphism). O método baseia-se na digestão de moléculas de DNA via utilização de enzimas de restrição e posterior análise do polimorfismo no comprimento dos fragmentos gerados. Em seguida, o desenvolvimento do processo de amplificação de DNA em cadeia (PCR), utilizando a enzima DNA polimerase, conduzindo à descrição de outras classes de marcadores moleculares, tais como o RAPD (Random Amplified Polymorphic DNA), que procede da detecção de polimorfismo de DNA amplificado ao acaso (os iniciadores tem sequência não definida, portanto sua sequência alvo no DNA é desconhecida), os marcadores SSR (Simple Sequence Repeats), também conhecidos como microssatélites (que utilizam iniciadores específicos, com sequências de anelamento definidas, para regiões do DNA que apresentam sequências simples repetidas) e, os marcadores AFLP (Amplified Fragment Length Polymorphism), os quais combinam as técnicas de restrição com o potencial de amplificação do DNA via PCR (FERREIRA et al., 2007). A PCR é um método *in vitro* para a síntese enzimática de sequências específicas de DNA usando dois iniciadores de oligonucleotídeos que hibridizam em fitas opostas de uma molécula molde, flanqueando a região-alvo no DNA total (SAMBROOK; RUSSEL, 2001). Mais recentemente surgiram os marcadores denominados SNP (Single Nucleotide Polymorphism), que tem grande potencial para aplicação em estudos de genética (VARSHNEY et al. 2007; SEHGAL; RAINA, 2008).

A disponibilidade de uma variedade de marcadores moleculares e suas modificações levaram a diversos estudos comparativos entre eles em muitas culturas incluindo soja, trigo e cevada (POWELL et al. 1996; RUSSELL et al. 1997; BOHN et al. 1999). Entre todos esses

marcadores, SSR ganharam considerável importância em genética de plantas devido ao fato de apresentarem diversas características que os tornam amplamente utilizados (KALIA et al., 2012).

2.3.1.1 Marcadores microssatélites

Microssatélites são regiões de sequência de DNA intergênicas que consistem em arranjos de unidades de até seis bases repetidas em *tandem* presentes em todos os organismos eucarióticos (TAUTZ; RENZ, 1984). Podem ser designados como SSR (Simple Sequence Repeats), STR (Short Tandem Repeats) ou microssatélites (FERREIRA; GRATTAPAGLIA, 1998).

A geração de polimorfismo em cada loco microssatélite é resultado da diferença no número de repetições de cada unidade de repetição. Podemos, então, distinguir vários alelos, de um mesmo loco, pelo tamanho. As sequências microssatélites em um loco variam de um indivíduo a outro, contudo, as sequências de DNA que flanqueiam os microssatélites são bastante conservadas entre indivíduos da mesma espécie. Desse modo, conhecendo-se estas regiões adjacentes às sequências microssatélites, é possível o desenho de pares de iniciadores específicos, possibilitando a amplificação, via PCR, dos fragmentos que contêm o DNA repetitivo em diferentes genótipos (ZANE et al., 2002).

Esses marcadores baseados em reações de amplificação por meio de PCR revelam polimorfismo em um loco devido a diferenças no tamanho dos alelos que decorre do número de vezes em que uma determinada repetição é encontrada (FERREIRA; GRATTAPAGLIA, 1998; BUSO et al., 2003). Os SSR podem ser classificados em perfeitos, imperfeitos e compostos, em função do tipo de repetição que eles apresentam (WANG et al., 2009). São ditos perfeitos os microssatélites onde a repetição não apresenta interrupção (por exemplo, CACACACACA). São ditos imperfeitos quando a sequência é interrompida por alguma base que não pertença ao motivo (por exemplo, CACACAgCACA). E são considerados compostos os microssatélites que apresentam dois ou mais tipos de repetição conjuntamente (por exemplo, CACACAGCGCGC) (CAIXETA, 2006).

Os SSR são marcadores muito populares devido à herança codominante, sendo muito frequentes e distribuídos ao acaso no genoma, permitindo uma boa cobertura de qualquer genoma eucarioto, mostrando elevada diversidade genética, além de apresentar alta reprodutibilidade (AGARWAL et al., 2008; KALIA et al., 2012).

Os marcadores microssatélites fornecem relevantes informações para a construção de mapas de ligação, na análise gênica e de locos quantitativos (OLIVEIRA et al., 2006), na identificação de acessos em unidades de conservação de germoplasma (MASONA et al., 2005), e para investigar processos genéticos, tais como padrão de fluxo gênico, ocorrência de deriva, proteção de variedades (PRIOLLI et al., 2002) e análise de *pedigrees* (DREISIGACKER et al., 2004), constituindo marcadores amplamente utilizados em estudos envolvendo inúmeras espécies vegetais (OLIVEIRA et al., 2006; PARIDA et al., 2009; HALE et al., 2012).

Para iniciar estudos com marcadores microssatélites, há a necessidade de se ter disponível iniciadores desenvolvidos que irão flanquear as regiões microssatélites. Caso não existam para a espécie de interesse, há a possibilidade de procurar por sequências desta natureza em bancos de dados, ou o desenvolvimento de uma biblioteca genômica enriquecida para a obtenção de sequências candidatas a apresentar regiões microssatélites (KALIA et al., 2012). Em urucum a construção de uma biblioteca genômica enriquecida é bastante interessante e necessária já que existem poucas sequências desta espécie depositadas em bancos de dados, e nenhuma referente a regiões microssatélites. O método de bibliotecas genômicas enriquecidas tem sido o mais utilizado para isolamento de microssatélites e apresenta em geral uma alta eficiência (BILLOTE et al., 1999). Este método leva à obtenção de grandes quantidades de clones contendo microssatélites e apropriados para o desenho de primers (HAMILTON, 1999).

A metodologia desenvolvida por Billote et al. (1999) para o desenvolvimento de bibliotecas enriquecidas consiste da fragmentação do DNA com enzimas de restrição, sendo posteriormente ligado cada fragmento a adaptadores específicos. Os fragmentos são amplificados e purificados, e posteriormente hibridizados com sondas biotiniladas contendo repetições de microssatélites. As sondas que estão ligadas aos fragmentos que tem afinidade pela biotina aderem a contas magnéticas que estão aderidas à estreptavidina. Os fragmentos que possuem microssatélites são então selecionados e, são submetidos a lavagem para remoção do DNA ligado. O DNA é então suspenso em tampão e amplificado por PCR. Em seguida o DNA é clonado em vetores e posteriormente sequenciado. A partir das sequências, são identificadas as regiões microssatélites, bem como desenvolvidos os iniciadores, que irão flanquear estas regiões e serão uteis para a posterior amplificação das regiões microssatélites (BILLOTE et al., 1999). O desenvolvimento de bibliotecas enriquecidas tem se tornado mais acessível devido principalmente às tecnologias de sequenciamento automático, e ao desenvolvimento de metodologias mais eficientes e baratas (KALIA et al., 2012). Em função

disto, já foram desenvolvidos iniciadores de microsatélites para diversos grupos de plantas utilizando a tecnologia de bibliotecas enriquecidas, como por exemplo, Bajay et al. (2009) em mamona (*Ricinus communis* L.), Decroocq et al. (2003) em uva (*Vitis vinifera*), Parida et al. (2009) em cana de açúcar (*Saccharum* sp.), Mori et al. (2010) em uma espécie arbórea dos manguezais (*Avicennia germinans*), Siqueira et al. (2011) em inhame (*Dioscorea alata* L.), Bressan et al. (2012) para pinhão-manso (*Jatropha curcas* L.), entre inúmeros outros.

2.3.1.2 Diversidade genética em urucum

A importância da variabilidade genética para o melhoramento genético já é bem conhecida, e o uso de marcadores moleculares na determinação desta variabilidade é uma estratégia bastante utilizada (SEHGAL; RAINA, 2008). Apesar disso, e de toda importância que o urucum representa como fonte de corantes naturais, verifica-se que são poucas as pesquisas direcionadas à avaliação da variabilidade genética presente dentro desta espécie e à forma como esta variabilidade se distribui no território brasileiro.

Medina et al. (2001) avaliaram 36 genótipos de urucum coletados na Venezuela e no Brasil com marcadores isoenzimáticos, observando divergência genética de 40 a 60% entre os acessos. Os autores não observaram correlação entre as distâncias geográficas e genéticas, possivelmente devido à influência antrópica na distribuição desta cultura. Carvalho et al. (2005), avaliando 60 genótipos mantidos em coleções de germoplasma, de diversas regiões do Brasil, concluíram que apenas três dos 21 locos isoenzimáticos foram polimórficos (14,3%). A análise de agrupamento diferenciou claramente as procedências de urucum do Norte, daquelas oriundas do Brasil Central, com significativa variabilidade na procedência do Maranhão. Tezotto et al. (2006) avaliaram 27 acessos do Banco de Germoplasma do IAC, observando 25,4% de locos isoenzimáticos polimórficos, semelhante ao encontrado por Carvalho et al. (2005), e puderam separar os acessos originários de Mato Grosso dos demais. Os autores observaram, além disso, que a maior parte da diversidade genética observada encontrava-se entre grupos de acessos.

Estudos com marcadores de DNA para a caracterização da variabilidade genética de urucum são escassos, podendo-se citar o uso de marcadores AFLP (FAY et al., 2005) e SRAP (Sequence-Related Amplified Polymorphism) (VALDEZ-OJEDA et al., 2008). Fay et al. (2005) avaliaram apenas quatro acessos de urucum, obtendo 25 a 27,5% de bandas polimórficas com AFLPs. Já Valdez-Oheda et al. (2008) avaliaram 87 acessos coletados em duas áreas agrícolas no sul do México, observando elevada variabilidade utilizando SRAPs,

tanto entre como dentro dos locais de coleta. Outros estudos moleculares têm sido conduzidos, como os de Soares et al. (2011), na tentativa de se descobrir novos genes associados ao desenvolvimento das sementes e metabolismo em *B. orellana*. Os autores caracterizaram um conjunto de Etiquetas de Sequências Expressas (ESTs) a partir de uma biblioteca de cDNA.

3 MATERIAL E MÉTODOS

3.1 Material vegetal

Foram avaliados neste estudo 63 acessos do Banco de Germoplasma de urucum (*B. orellana* L.) do IAC, mantidos no Pólo Regional Centro Norte em Pindorama-SP, oriundos de municípios dos Estados de Mato Grosso (dois acessos), compondo a região Centro-Oeste, São Paulo, Minas Gerais e Rio de Janeiro (14 acessos), compondo a região Sudeste, e Rondônia (26 acessos), compondo a região Norte. Também estão contidos na coleção de germoplasma um acesso do Peru e 20 acessos com origem desconhecida (Tabela 1; Figuras 3 e 4).

Tabela 1 - Relação dos acessos de urucum, número de identificação dos acessos mantidos no banco de germoplasma do Instituto Agronômico (IAC), origem e Estado de coleta

(continua)

Número do acesso	Local de coleta - Município	Estado
IAC-01	Cuiabá	MT
IAC-02	Cuiabá	MT
IAC-03	Vale do Ribeira	SP
IAC-04	Viçosa	MG
IAC-05	Rezende	RJ
IAC-06	Sorocaba	SP
IAC-07	Igaratá	SP
IAC-08	Tupi Paulista	SP
IAC-09	Desconhecida	-
IAC-10	Guaraci	SP
IAC-11	São José do Rio Preto	SP
IAC-12	Desconhecida	-
IAC-13	Desconhecida	-
IAC-14	Icem	SP
IAC-15	Desconhecida	-
IAC-16	Minas Gerais	MG
IAC-17	Barretos	SP
IAC-18	Desconhecida	-
IAC-19	Desconhecida	-
IAC-20	Icem	SP
IAC-21	Olímpia	SP
IAC-22	Peru	PERU
IAC-23	Desconhecida	-
IAC-24	Desconhecida	-
IAC-25	Desconhecida	-
IAC-26	Desconhecida	-
IAC-27	Desconhecida	-
IAC-28	Desconhecida	-
IAC-29	Desconhecida	-

Tabela 1 - Relação dos acessos de urucum, número de identificação dos acessos mantidos no banco de germoplasma do Instituto Agrônômico (IAC), origem e Estado de coleta

(conclusão)

Número do acesso	Local de coleta - Município	Estado
IAC-30	Desconhecida	-
IAC-31	Desconhecida	-
IAC-32	Desconhecida	-
IAC-33	São João do Pau D' alho	SP
IAC-34	Desconhecido	-
IAC-35	Desconhecido	-
IAC-36	Desconhecido	-
IAC-37	Desconhecido	-
IAC-38	Corumbiara	RO
IAC-39	Nativo-Corumbiara	RO
IAC-40	Colorado do Oeste	RO
IAC-41	Colorado do Oeste	RO
IAC-42	Colorado do Oeste	RO
IAC-43	Corumbiara	RO
IAC-44	Corumbiara	RO
IAC-45	Corumbiara	RO
IAC-46	Corumbiara	RO
IAC-47	Corumbiara	RO
IAC-49	Colorado do oeste	RO
IAC-50	Colorado do oeste	RO
IAC-51	Colorado do oeste	RO
IAC-52	Corumbiara	RO
IAC-53	Colorado do oeste	RO
IAC-54	Colorado do oeste	RO
IAC-55	Colorado do oeste	RO
IAC-56	Colorado do oeste	RO
IAC-57	Corumbiara	RO
IAC-58	Colorado do oeste	RO
IAC-59	Colorado do oeste	RO
IAC-60	Corumbiara	RO
IAC-61	Colorado do oeste	RO
IAC-62	Colorado do oeste	RO
IAC-63	Colorado do oeste	RO

Figura 3 - Plantas de *Bixa orellana* na coleção de germoplasma do Instituto Agrônomo em Pindorama – SP

Figura 4 - Mapa do Brasil com a identificação dos acessos brasileiros mantidos pelo Banco de Germoplasma do IAC. Os estados marcados em verde são da Região Sudeste, em azul da Região Centro-Oeste e em vermelho da Região Norte

3.2 Desenvolvimento de iniciadores de microsatélites

A etapa de desenvolvimento de iniciadores de microsatélites de urucum foi realizada no Laboratório de Diversidade Genética e Melhoramento do Departamento de Genética da ESALQ/USP, em Piracicaba, SP.

3.2.1 Construção de biblioteca genômica enriquecida com SSR

A construção da biblioteca genômica enriquecida foi realizada de acordo com a metodologia descrita por Billote et al. (1999).

3.2.1.1 Extração de DNA

Foram selecionados ao acaso 3 acessos do banco de germoplasma e o DNA extraído utilizando-se a metodologia de Doyle & Doyle (1990) com modificações. A metodologia esta descrita no item 3.3.1.

3.2.1.2 Restrição do DNA

Cerca de 10µg de DNA genômico foi digerido com 60U da enzima *Rsa* I, adicionando na reação 10µl de tampão da enzima, 400 mM de espermidina e água ultrapura a fim de completar o volume final de 100µl. O material foi incubado a 37°C *overnight*. O resultado da restrição foi visualizado em gel de agarose 0,8% em tampão 1X TBE corado com *SYBRgold* sob eletroforese a 3V.cm⁻¹.

3.2.1.3 Ligação dos adaptadores

Os adaptadores *Rsa*21 e *Rsa*25 foram ligados aos fragmentos digeridos através das extremidades específicas por meio da T4 DNA *ligase*. Para tanto foi utilizado 0,2 µM do adaptador *Rsa*21 e 0,2 µM do adaptador *Rsa*25, e 1µg do DNA genômico digerido na etapa anterior, tampão – 7,5mM Tris-HCl (pH 7,8); 2,5mM de MgCl₂; 2,5mM de dNTP, 0,2 µM de cada adaptador, 5U de T4 DNA *ligase* e água Milli-Q para completar 200µL. A reação foi incubada *overnight* a 4°C.

3.2.1.4 Pré-Amplificação via PCR

Os fragmentos digeridos foram submetidos a reação de amplificação (PCR) utilizando o adaptador *Rsa21*, visando a obtenção de uma maior quantidade de DNA para seleção. Após a seleção, os fragmentos foram pré-amplificados. A amplificação procedeu conforme os seguintes ciclos: 95°C por 4 min, seguidos de 20 ciclos de 94°C por 30 s, 60°C por 1 min, 72°C por 1 min. e extensão final de 72°C por 8 min. Uma amostra da pré-amplificação foi analisada em gel de agarose 1,2%, em tampão 1X TBE, corado com *SYBRgold*, sob eletroforese a 3V.cm⁻¹.

3.2.1.5 Seleção dos fragmentos contendo SSR

O processo de enriquecimento foi realizado por meio de hibridização de oligos biotinilados [biotIII(CTT)₁₀, biotIII(TA)₁₀ e biotIII(GT)₁₀]. Os fragmentos contendo as repetições foram recuperados com estreptavidinas ligadas a contas magnéticas utilizando o Kit *Streptavidin Magesphere Paramagnetic Particle*.

3.2.1.6 Amplificação dos fragmentos selecionados via PCR

Após a seleção dos fragmentos enriquecidos, estes foram submetidos a amplificação utilizando o iniciador *Rsa21*. A amplificação foi realizada adicionando-se 10μL do produto dos fragmentos selecionados, 0,4μM do iniciador, 10μL de 10X *Taq Buffer*; 2mM de dNTP, 0,75mM de MgCl₂, 2U de *Taq polymerase* em um volume final de 100μL.

A amplificação procedeu conforme os seguintes ciclos: 95°C por 1 min., seguidos de 25 ciclos de 94°C por 40 s, 60°C por 1 min., 72°C por 1 min. e extensão final de 72°C por 5 min. O produto da amplificação foi visualizado em gel de agarose 1,2%, preparado em tampão 1X TBE, corado com *SYBRgold*.

3.2.1.7 Clonagem dos fragmentos selecionados

Os fragmentos selecionados foram clonados no vetor *pGEM-T Easy* e a reação incubada *overnight* a 4°C. A transformação bacteriana foi realizada em células eletrocompetentes de *E. coli* e plaqueadas em meio seletivo.

3.2.1.8 Sequenciamento dos clones positivos para microssatélites

As reações de sequenciamento foram realizadas utilizando o Kit de sequenciamento *DYEnamic ET Terminator Cycle Sequencing*, seguindo as seguintes concentrações para a amplificação dos fragmentos por PCR: 5 μ M do iniciador T7, 150 ng de DNA plasmidial, 2 μ L de *Sequencing reagent premix* e água Milli-Q para um volume final de 10 μ L. Para tanto foram utilizadas as seguintes condições em termociclador: 30 ciclos 95°C por 40 s, 60°C por 1 min e 72°C por 1 min. Os produtos das amplificações foram purificados adicionando 2 μ L de Acetato de Sódio/ Ácido Etilenodiamino Tetra-Acético (EDTA) em cada poço, seguido da adição de 60 μ L de etanol absoluto e centrifugação por 45 min à 4000 rpm à 4°C. O sobrenadante foi então descartado e foi adicionado 150 μ L de etanol 70% e em seguida centrifugado por 15 min a 4000 rpm à 4°C. O sobrenadante foi descartado e os *pellets* secos a 40°C por 10 min.

3.2.1.9 Construção dos microssatélites

Após a seleção os fragmentos contendo clones positivos para SSR foram extraídos e os plasmídeos sequenciados em sequenciador automático. Foi realizada uma reação com 1 μ g do plasmídeo, 2 μ L do Mix Bid Dye (Applied Biosystem, Foster City, CA, EUA) e água ultrapura para completar 20 μ L.

Os pares de iniciadores, complementares às sequências que flanqueiam as ilhas de repetição SSRs, foram desenhados utilizando o programa *Primer3* (ROZEN; SKALETSKY, 2009). O software foi configurado para a obtenção de iniciadores cujo produto final da amplificação esteja no intervalo de 100 a 250 pares de base e o tamanho dos iniciadores deveria conter de 18 a 22 pb. Além disso, os seguintes critérios foram definidos: conteúdo mínimo de 40% e máximo de 60% de CG; diferença de temperatura de anelamento dos iniciadores variando entre 55°C e 70°C; e diferença na temperatura de anelamento entre os pares de iniciadores variando no máximo em 3°C. Para avaliar a qualidade dos iniciadores quanto à formação de estruturas secundárias e para se evitar o desenho de iniciadores redundantes foram utilizados os programas *Gene Runner* (<http://www.generunner.net/>) e *Chromas 2* (<http://technelysium.com.au/>).

De posse dos primers sintetizados foi realizada a otimização dos locos, ou seja, a escolha da melhor temperatura de anelamento e melhor quantidade de reagentes para cada par de iniciadores.

3.3 Avaliação dos acessos de urucum com marcadores microssatélites

3.3.1 Extração de DNA

As análises moleculares foram realizadas no Laboratório de Ecologia Evolutiva e Genética Aplicada do Departamento de Genética, ESALQ/USP, em Piracicaba, SP.

Foram coletadas folhas jovens recém-expandidas dos acessos do banco de germoplasma, as quais foram armazenadas em gel de CTAB [30 mg CTAB, 350 mg NaCl, 0,7 mL de água destilada] contido em microtubos de 1,5 mL. Os microtubos contendo gel e as folhas jovens coletadas foram armazenados em geladeira por sete dias. Em seguida as folhas foram submetidas à maceração com tampão STE [0,13 mg de Sacarose; 45 µL de Tris – HCL (1M); 150 µL de EDTA (0.5M), completando-se com H₂O destilada para o volume final de 1,5 mL] (Bhattacharjee et al., 2009). Foi utilizada a metodologia de Doyle & Doyle (1990) com modificações para a extração de DNA. Os tubos contendo 0,8 mL do tampão de extração [30mM EDTA pH 8,0; 0,1M Tris-HCL pH 8,0; 1,2M NaCl; 3% CTAB; adicionando-se 2-mercaptoetanol a 1% imediatamente antes do uso] mais 50 mg do material macerado foram encubados por uma hora em banho-maria à 65°C, e posteriormente homogeneizados. Foi adicionado 0,5 mL de uma mistura que consiste em 24 partes de clorofórmio em uma parte de álcool isoamílico (24:1). O material foi centrifugado a 12000 g por 10 minutos sendo posteriormente retirado 0,5 mL do sobrenadante e transferindo o mesmo para um novo tubo. Novamente foi adicionado 0,5 mL da solução 24:1, agitando-se o tubo, posteriormente centrifugado a 12260 g durante 10 min. 0,4 mL do sobrenadante foi transferido para novos tubos, adicionando-se 0,35 mL de isopropanol pré-esfriado à -20°C. Após este período o material foi submetido a centrifugação a 12260 g por 10 min. O sobrenadante retirado e o pellet foram secos a temperatura ambiente por aproximadamente 12 horas. O pellet foi suspenso em solução tampão TE contendo RNase (10 mg/µL). Os tubos foram submetidos a banho-maria por 30 min à 37°C, para posteriormente serem armazenados no freezer.

A concentração de DNA da solução estoque de cada indivíduo foi estimada através da corrida eletroforética das amostras em gel de agarose 0,8%, sendo que para a coloração das bandas foi utilizado brometo de etídio, sendo possível sua visualização sob luz UV com

auxílio do sistema de fotodocumentação Syngene (Synoptics Ltda, Cambridge, United Kingdom). Para a quantificação das amostras de DNA foram utilizadas soluções padrões de DNA de fago *lambda* (5, 20, 40, 80 e 160 ng). A corrida eletroforética procedeu em cuba contendo tampão TBE 10X a uma voltagem de $3,75 \text{ V.cm}^{-1}$.

3.3.2 Análises de PCR e eletroforese para obtenção das regiões microssatélites

Os iniciadores desenvolvidos na etapa 3.2 foram utilizados para amplificação das regiões microssatélites. Foram amostrados ao acaso seis indivíduos da coleção de urucum para as reações de otimização que foram realizadas num volume final de $16 \mu\text{L}$, composta por 1U de *Taq*-Polimerase; 1X de tampão (10 mM Tris-HCl; pH 8,3; 50 mM KCl; 1,5 mM MgCl_2); 2,4mM de MgCl_2 ; 2,4 pmoles do iniciador *Forward*; 2,4 pmoles do iniciador *Reverse*; 3,2 mM de cada dNTP e 20 ng de DNA. Para proceder as reações de amplificação foi utilizado o termociclador modelo MyCycler Thermal Cycler (BioRad, Hercules, Califórnia, EUA).

Inicialmente testou-se o programa de amplificação *TouchDown* onde não é necessário determinar a melhor temperatura de anelamento, pois é feita uma pré-amplificação dos fragmentos com a temperatura diminuindo 1°C a cada ciclo (DON et al., 2001). As etapas do programa *TouchDown* estão descritas a seguir: uma etapa inicial de desnaturação por 5 min a 94°C , seguida de 10 ciclos de préamplificação (30 s a 95°C , 30 s na temperatura de anelamento diminuindo 1°C a cada ciclo, iniciando em 60°C para o programa *TouchDown60* e em 55°C para o programa *TouchDown55* e 50 s a 72°C), seguidos de 30 ciclos de desnaturação (30 s a 95°C , 30 s a 50°C e 50 s a 72°C) e uma fase final de extensão a 72°C por 5 min.

Para os pares de iniciadores para os quais não foi possível realizar a amplificação por esta metodologia, por *TouchDown*, aplicou-se diferentes temperaturas de anelamento para verificar a mais adequada. As temperaturas de anelamento de cada iniciador foram avaliadas para o termociclador acima citado, testando-se as mesmas em sistema de gradiente, avaliando-se quatro temperaturas para cada iniciador nas seguintes condições: desnaturação inicial a 94°C durante 5 min, seguida por 35 ciclos de desnaturação a 94°C por 30 s, 1 min na temperatura de anelamento (variando de 54°C a 60°C), e 72°C durante 1 min para extensão da cadeia de DNA, com uma extensão ao final dos 35 ciclos a 72°C por 8 min.

O material amplificado foi separado em gel desnaturante de poliacrilamida a 7% (233,4 mL de acrilamida líquida 29:1, 100 mL de TBE 10X, 420 g de uréia), 400 µL de persulfato de amônio a 10% e 168 µL de TEMED.

Os géis foram pré-aquecidos por 40 min a 70 W de potência constante. Foram aplicados 4 µL das amostras previamente desnaturadas acrescidas de 2 µL de tampão de carregamento (95% formamida, 0,05% de xylenicynol; 0,05% de azul de bromofenol e 12,5% de sacarose) a 95°C por 5 min. A eletroforese foi realizada a potência constante de 60 W durante por 120 minutos, utilizando-se como marcador de peso molecular DNA *ladder* de 10pb (Life Technology, Carlsbad, Ca, USA). Para revelação das bandas de microssatélites foi utilizada a metodologia de nitrato de prata (CRESTE et al., 2001). Os géis foram visualizados em transluminador para a interpretação das bandas de microssatélites e posteriormente fotografados.

3.4 Análise estatística dos resultados

Em função da natureza co-dominante dos marcadores, foram estimados para os locos de microssatélites avaliados, para o conjunto dos acessos do banco de germoplasma, as frequências alélicas e genotípicas, índices de diversidade genética como número de alelos por loco, porcentagem de locos polimórficos, heterozigosidade média observada e esperada (NEI, 1978) e índice de fixação de Wright. Esses índices foram obtidos a partir dos programas GDA - Genetic Data Analysis, versão 1.0 (LEWIS; ZAYKIN, 2001) e TFPGA (MILLER et al., 1997). Foi também estimado o coeficiente de Conteúdo de Informação de Polimorfismo (PIC), segundo Botstein et al. (1980) de acordo com a seguinte expressão matemática:

$$PIC = 1 - \sum_{i=1}^n (P_i)^2 - 2 \sum_{i=1}^{n-1} \sum_{j=1}^n (P_i)^2 (P_j)^2$$

onde:

p_i = frequência do i – ésimo alelo;

p_j = frequência do j – ésimo alelo; e

n = número total de alelos observados.

Para cada loco foi estimada também a taxa de cruzamento aparente a partir da fórmula:

$$Ta = (1 - f)/(1 + f)$$

Onde f é o coeficiente de endogamia de Wright, determinado para cada loco em função das heterozigosidades esperadas e observadas.

As distâncias genéticas entre os indivíduos foram calculadas utilizando a distância de Rogers modificada por Wright (WRIGHT, 1978), resultando em uma matriz de dissimilaridade usada na construção de um dendrograma por acessos individuais, incluindo os de origem desconhecida e o acesso do Peru, utilizando o método de agrupamento UPGMA (Unweighted Pair-Group Method Using Arithmetic Averages). Para a estimativa das distâncias genéticas foi utilizado o software TFPGA (MILLER et al., 1997), enquanto que para a construção do dendrograma foi utilizado o software DARwin5 (PERRIER et al., 2003). A estabilidade dos agrupamentos foi verificada através de 10.000 reamostragens bootstrap, utilizando o software TFPGA (MILLER et al., 1997).

Foi verificado, para cada loco, a aderência ao equilíbrio de Hardy-Weinberg, utilizando o software FSTAT (GOUDET, 2001).

Foi verificada a existência ou não de correlação entre as distâncias genéticas obtidas pela análise de microssatélites e as distâncias geográficas. Para tal análise foram considerados apenas os acessos com origem definida. Foi construída uma matriz de distâncias geográficas utilizando o software DIVA GIS (HIJMANS, 2004). A análise de correlação foi realizada estimando-se o coeficiente de correlação de Pearson (r), cuja significância foi verificada pelo teste de Mantel (1967) com 10000 permutações. Para tanto, foi utilizado o programa NTSYS-pc (ROHLF, 1992).

Além destas análises, foi realizada uma análise bayesiana com o programa STRUCTURE 2.3 (PRITCHARD et al., 2000) para definir grupos de indivíduos geneticamente relacionados com o objetivo de visualizar a estrutura genética do banco de germoplasma. Nesta análise existe a possibilidade de se optar por um modelo de ancestralidade com mistura, quando os genomas dos indivíduos podem apresentar proporção de mais uma população, e modelo de não mistura, quando os genomas dos indivíduos tem origem em populações distintas. Optou-se pelo modelo com mistura devido ao fato de o banco de germoplasma de urucum já ter sido replantado, e este plantio ter sido realizado por via sexual, existindo portando a possibilidade de polinizações cruzadas.

Para essas análises foram realizadas 20 simulações independentes feitas para cada valor de número de agrupamento K (K de 1 a 10). Em cada simulação, foram feitas 500.000 simulações de Monte Carlo de cadeias de Markov (MCMC) com um descarte inicial (*burn-in*) de 250.000. O número mais provável de agrupamentos foi estimado pelo método *ad hoc* ΔK de Evanno et al. (2005) e também pelo método descrito por Pritchard et al. (2000).

Após a definição do número de grupos presentes entre os acessos estudados, pelo software STRUCTURE 2.3, o conjunto de acessos foi dividido em tantos grupos quantos foram determinados pelo melhor K, e a partir destes grupos foram calculados os parâmetros de diversidade genética como número de alelos por loco, porcentagem de locos polimórficos, heterozigosidade média observada e esperada (NEI, 1978) e índice de fixação de Wright. Para esses grupos de acessos foram também estimadas as estatísticas F de Wright (WRIGHT, 1978), inferindo os valores de F_{ST} , F_{IT} e F_{IS} . Foram também calculados os índices de diversidade de Nei (1973). Estes índices permitem estimar a caracterização da diversidade entre os grupos e dentro dos grupos. Os parâmetros de diversidade genética foram estimados pelos programas GDA (LEWIS; ZAYKIN, 2001) e TFPGA (MILLER et al., 1997). Os índices de fixação (F) foram calculados utilizando-se o software TFPGA (MILLER et al., 1997). Para os índices de diversidade de Nei (1973), onde se observa a partição da diversidade genética entre grupos D_{ST} , e dentro de grupos, H_S , bem como a diversidade genética total da espécie, H_T , o índice G_{IS} que é um estimador do índice de fixação F_{IS} e a proporção da diversidade genética entre grupos, G_{ST} (onde $G_{ST} = D_{ST}/H_T$), foi utilizado o software FSTAT (GOUDET, 2001).

4 RESULTADOS E DISCUSSÃO

4.1 Desenho dos iniciadores

A partir da biblioteca genômica desenvolvida foram obtidas e sequenciadas um total de 84 colônias. Deste total, foram encontrados microssatélites em 57 colônias, o que representa um total de 67,9% de enriquecimento. Resultados similares foram obtidos por Morgante et al. (2012), que obtiveram 65,9% de enriquecimento em Cataia (*Pimenta pseudocaryophyllus*), uma lenhosa pertencente a família Myrtaceae. Os resultados do presente trabalho foram superiores aos encontrados em algumas espécies, como, por exemplo, 40,3% de enriquecimento obtidos em *Mussaenda pubescens*, um arbusto da família Rubiaceae (DUAN et al., 2012), 35,9% em *Protium subseratum* uma espécie, assim como urucum, também nativa da Amazônia (MISIEWICZ et al., 2012), 50,4% em *Aquilaria malaccensis*, uma espécie também de clima tropical (TNAH et al., 2012), 39,2% obtido em cacau (*Theobroma cacao*) (ARAÚJO et al., 2007) e 28% em melão (*Cucumis melo*) (RISTCHEL et al., 2004). Contudo, em mamona foi observado um valor bastante superior, de 92,9% (BAJAY et al., 2009). Portanto, este percentual de enriquecimento é muito variável entre espécies, sendo influenciado também pelos parâmetros definidos na busca por microssatélites dentro das sequências.

Foram identificados 70 microssatélites, sendo que destes, 57 foram dinucleotídeos (81,5%), quatro trinucleotídeos (5,7%), sete tetranucleotídeos (10%), um pentanucleotídeo (1,4%) e um hexanucleotídeo (1,4%). Dentre os 70 microssatélites obtidos, alguns se apresentavam no início das sequências, ou muito ao final, impossibilitando o desenho de iniciadores. Esses iniciadores foram retirados das análises, e ao final foram desenhados e selecionados um total de 31 iniciadores (Tabela 2).

Os marcadores SSR sintetizados a partir da biblioteca genômica desenvolvida apresentam porcentagem média de GC igual a 47,3%, variando de 40% a 57,9%, amplificando produtos com tamanho médio de 186 pb, variando de 122 pb a 228 pb, e temperatura de anelamento média de 59,1°C, variando de 56°C a 62°C. Dos 31 SSR sintetizados, 28 (90,4%) foram classificados como perfeitos, 1 (3,2%) como composto, 1 (3,2%) como interrompido e 1 (3,2%) composto-interrompido.

Tabela 2 - Sequências dos iniciadores incluindo nome, motivo, porcentagem de GC, temperaturas de anelamento, tamanhos esperados dos produtos em pares de bases (pb), e classificação de cada iniciador desenvolvido

(continua)

Nome	Motivo	Sequência	%GC	Temp (°C)	Produto (pb)	Classificação
BorA2	(AC) ₇	F: CGAGGGGATGTGAAATTAGG	50,0	60	228	Perfeito
		R: CTGTTAGGGGACATTTTCATTG	40,9	62		
BorA4	(TG) ₁₆ (AG) ₁₈	F: GCGTGGACTAGCTTAGCATT	50,0	60	224	Composto
		R: GGTGGGAGAAATGGACAAGA	50,0	60		
BorA7	(TCT) ₅	F: CTTTCCGCTTTTGGTCTTTG	45,0	58	184	Perfeito
		R: TGCAACCTTAATCCCTAGCA	45,0	58		
BorA9	(CAA) ₅	F: CTCCTTCCGCCAAAATCTC	52,6	58	228	Perfeito
		R: GTCCAAGATGCAGCAAATCA	45,0	58		
BorB4	(GA) ₁₆	F: GCATCATCTCTGTGGGATCA	50,0	60	218	Perfeito
		R: ACGAGAAAATGAACGAGCTG	42,9	60		
BorB6	(AC) ₈	F: TTCCTCAGTTCTTCTGAACC	45,4	64	180	Perfeito
		R: CGGAGCACTTTTCTTTTGAG	47,6	62		
BorB10	(AG) ₁₄	F: CATGCCTCCTCTCGATCTCT	55,0	62	178	Perfeito
		R: ATCATTTCTTGCCCAACTG	45,0	58		
BorB12	(TG) ₁₄ (AG) ₁₉	F: CACATGCCCTTTGATGGTTA	45,0	58	172	Composto
		R: TCATGTTCCGATCCTATGTGA	43,3	60		
BorC2	(AC) ₈	F: AGCAGAAGCGTACCACGTC	57,8	60	201	Perfeito
		R: TGTGGAAGTTGAGTGGAGGA	42,8	60		
BorC8	(CTT) ₃	F: CATTGTCAAAACCAGGAGGA	42,8	58	214	Perfeito
		R: GCTAGATTGGTGGCAAAAGG	50,0	60		
BorC11	(GAAA) ₃	F: TGCTCCATGAACGTCATT	45,0	58	181	Perfeito
		R: TGTGTGGGAGTTGGAAAACA	45,0	58		
BorC12	(CAT) ₄	F: GAGATTGTGTCGTCATGGGTA	50,0	60	194	Perfeito
		R: AGGAGGAAGGAAAGGAGGTG	55,0	62		
BorD4	(AC) ₇	F: TCAAACATCGGCACATTGAT	40,0	56	167	Perfeito
		R: CCCCCTTTTGAATTGCTGT	47,4	56		
BorD7	(GT) ₉	F: GTTCGTGGTTGGAGCATACA	50,0	60	209	Perfeito
		R: CAGCAGGAAGTGGAGGTGA	57,9	60		
BorD8	(AC) ₈	F: TGATGTGAAGAACCGAAATCC	42,9	60	169	Perfeito
		R: CAAAGCAGCAACCTCTTTTTC	42,9	60		
BorD12	(TG) ₇	F: GGATTCCATACATTTCAAGTCG	40,9	62	207	Perfeito
		R: CAGCATGATTGGTCAAAACTG	42,9	60		
BorE7	(AG) ₅	F: TCTCTTATCCCAGCGAGGAA	50,0	60	206	Perfeito
		R: CACATTTCTCTCTCCATGCA	42,9	58		

Tabela 2 - Sequências dos iniciadores incluindo nome, motivo, porcentagem de GC, temperaturas de anelamento, tamanhos esperados dos produtos em pares de bases (pb), e classificação de cada iniciador desenvolvido

(conclusão)

Nome	Motivo	Sequência	%GC	Temp (°C)	Produto (pb)	Classificação
BorE9	(CA) ₁₀	F: TCCCTTCCTTAAGTCAAGC	50,0	60	198	Perfeito
		R: CCTTGGTTTCATTGTGTGGA	45,0	58		
BorE12	(ATC) ₃	F: GTCATGCCAAGTTGTTTTCC	42,8	60	191	Perfeito
		R: AGGTGTTTAGGCGAAAGTGG	50,0	60		
BorF1	(TG) ₆ (CT) ₁₇	F: CGTTCGTCTGAAAATCTG	45,0	58	154	Imperfeito Composto
		R: CATTCCCAGTGCAAGACC	50,0	60		
BorF3	(CA) ₉	F: TCATTCGCTTTCACGCTATT	40,0	56	174	Perfeito
		R: GCCTTTATTCCTACGTTTCG	50,0	56		
BorF4	(CTT) ₃	F: CTTCTTGATTGCCTTTATTG	40,9	58	142	Perfeito
		R: TGCAGACCATTGGAGTGA	47,3	56		
BorF5	(CA) ₆	F: GCGTGGACTAGCACTTTCTC	55,0	58	210	Perfeito
		R: GGGTGTGTTAGCCGAATTGTA	47,6	62		
BorF5_2	(AC) ₈	F: GACATACAATTCGGCTAACA	45,0	56	168	Perfeito
		R: TGAAGGGTGAAGAGGACTGT	47,62=	60		
BorF8	(AG) ₁₅	F: CACGGGGACAATTAAAGGAA	45,0	58	182	Perfeito
		R: TCGTTGATTGATGAGGATGC	45,0	58		
BorF9	(TATT) ₃	F: TCAACACACCACCACAATGA	45,0	58	122	Perfeito
		R: AGCTATGCCAGAAAAAGGA	45,0	58		
BorF12	(ATA) ₃	F: CCTCTCATGGGGCTAAAAGA	50,0	60	214	Perfeito
		R: GGGCATTTTCAACCTGAAC	45,0	58		
BorG3	(GT) ₇	F: GTGTCACCCACAGAACACAA	50,0	60	175	Perfeito
		R: TGTTCCTCAAGAAGCTAATGCAG	41,0	62		
BorG4	(CA) ₁₆	F: TCCCTTTGTTTTCCATTGCT	40,0	56	209	Perfeito
		R: TATCCACCAGCAACATTGAC	42,8	58		
BorG6	(GT) ₈	F: CGTGGACTAACAAGGACTGT	52,3	60	130	Perfeito
		R: TAATTTCTGCTTGCACTCC	42,8	58		
BorH6	(AC) ₁₀	F: CAAGCACCTTCTCTTTTCC	50,0	60	160	Perfeito
		R: GCTAAGGTTGATGGCAATG	45,0	58		

4.2 Condições de amplificação dos locos microssatélites

Dentre os 31 iniciadores sintetizados, 13 amplificaram produtos utilizando-se o programa de *TouchDown* com a temperatura variando de 50°C a 60°C e as condições de amplificação determinadas no item 3.3.2. Oito iniciadores amplificaram produtos utilizando-se o programa *TouchDown* com a temperatura variando de 45°C a 55°C e dois com o programa *TouchDown* variando de 55°C a 65°C. Para os iniciadores BorA4, BoB10, BoB12, BorC2, BorE9, BorF3, BorF5 e F12 não foi possível a amplificação com a utilização deste programa. Para tanto foram testadas quatro temperaturas de anelamento para cada iniciador, sendo elas, 54°C, 56°C, 58°C e 60°C. As condições de amplificação também foram mantidas conforme as indicadas no item 3.3.2. Os iniciadores BoB12 e BoB10 amplificaram produtos com maior resolução sob a temperatura de anelamento de 56°C, enquanto os iniciadores BorA4, BorC2, BorE9, BorF3, BorF5 e F12 não apresentaram produtos de reação, sendo então eliminados do estudo.

Dentre os 25 iniciadores que apresentaram amplicons, ou seja, produtos de amplificação, 15 (60%) foram monomórficos para o grupo de indivíduos estudados. Contudo, isto não impede que em futuros estudos estes iniciadores passem a ser utilizados em outros grupos de indivíduos e apresentem polimorfismo. Este baixo índice de polimorfismo (40%) foi obtido também por outros autores trabalhando também com *Bixa orellana*. Carvalho et al. (2005) avaliando 60 genótipos de urucum mantidos em coleções de germoplasma, de diversas regiões do Brasil, concluíram que apenas três dos 21 locos isoenzimáticos foram polimórficos (14,3%). Tezotto et al. (2006), avaliando 27 acessos deste mesmo Banco de Germoplasma do IAC, observaram apenas 25,4% de locos isoenzimáticos polimórficos.

4.3 Avaliação dos acessos do Banco de Germoplasma com os iniciadores desenvolvidos

Os 10 iniciadores polimórficos desenvolvidos neste estudo foram utilizados para avaliar o conjunto de 63 acessos de urucum do banco de germoplasma mantido pelo IAC (Figura 5).

Foram identificados 38 alelos polimórficos, variando de 2 (loco BorG3) a 6 alelos por loco (BorB10 e BorB12), obtendo-se uma média de 3,8 alelos por loco (Tabela 3). As frequências alélicas variaram entre 0,016 e 0,68 (Figura 6). Dentre os 38 alelos, 7 (18,4%) apresentaram frequência inferior a 5% sendo considerados alelos raros. A preocupação na conservação destes alelos em baixa frequência se deve ao fato de que eles podem ser

facilmente levados a extinção pela deriva genética, o que não significa que os alelos mais frequentes possuam algum valor adaptativo.

Figura 5 - Geis de poliacrilamida resultante da amplificação dos locos BorC12 (A), BorF1 (B), BorG3 (C), BorB10 (D), e BorG4 (E) em 36 indivíduos de urucum

Tabela 3 – Parâmetros de diversidade genética avaliados na coleção de urucum (*Bixa orellana*) do Banco de Germoplasma do IAC, incluindo número de alelos (A), heterozigosidade esperada (H_e), heterozigosidade observada (H_o), PIC e índice de fixação de Wright (f)

Locos	Nº Genbank	A	H_e	H_o	PIC	f
BorB4	KC152853	3	0,525	0,204	0,406	0,614
BorG3	KC152871	2	0,474	0,274	0,359	0,424
BorC12	KC152859	3	0,464	0,229	0,404	0,508
BorG4	KC152872	4	0,514	0,456	0,471	0,116
BorA2	KC152850	3	0,514	0,158	0,464	0,665
BorF5_2	KC152868	5	0,563	0,744	0,493	-0,326
BorB10	KC152855	6	0,765	0,491	0,725	0,359
BorB12	KC152856	6	0,505	0,196	0,531	0,671
BorF1	KC152866	4	0,645	0,428	0,583	0,338
BorE7	KC152864	3	0,614	0,000	0,533	1,000
Média		3,8	0,572	0,362	0,497	0,439

A heterozigosidade é uma medida da variabilidade genética que estima o quanto de variação existe numa população e como esta variação está distribuída em função dos alelos presentes em um determinado loco. Os valores de heterozigosidade esperada (H_e) variaram de 0,464 (BorC12) a 0,765 (BorB10), com média de 0,572 (Tabela 3). Por outro lado, a heterozigosidade observada (H_o) variou de 0 (BorE7) a 0,744 (BorF5_2), com média de 0,362 (Tabela 3).

Quando plantas provenientes de populações diferentes são agrupadas e analisadas em um único grupo, como ocorre com acessos em bancos de germoplasma, a heterozigosidade esperada é geralmente maior do que a heterozigosidade observada, conhecido como efeito Wahlund (HARTL; CLARK, 1997). Por essa razão, as heterozigosidades esperadas excederam as obtidas para heterozigosidade observada em nove dos dez locos.

Figura 6 – Frequências alélicas dos 63 acessos de urucum (*Bixa orellana*) estimadas com 10 locos SSR

Valores altos de f observados na maioria dos locos analisados permitem a interpretação de que os genótipos estudados estão submetidos a forte endogamia. A partir destes valores foi calculada a taxa de cruzamento aparente (Ta), obtendo-se um valor de 0,39. A partir deste valor, é possível interpretar que a taxa aparente de autofecundação, neste caso, foi de 0,61.

A existência de cruzamentos entre aparentados gera autofecundação e, consequentemente, aumento da homoziguidade (RITLAND, 2002). Estes resultados estão de acordo com os obtidos por Rivera-Madrid et al. (2006), realizando polinizações controladas em acessos de urucum em campo experimental no México, sugerindo que urucum pode tolerar ambos tipos de polinização, tendo registrado valores de polinização cruzada de 57% e auto-fecundação de 31,4%. Vilares et al. (1992) também concluíram pela existência de autofecundação natural em urucuzeiro.

Bixa orellana é normalmente considerada, pela sua variabilidade em caracteres morfológicos, como uma planta de polinização cruzada (CARVALHO et al., 2005). Contudo, em plantas que normalmente são classificadas como alógamas, uma porcentagem de autofecundação geralmente ocorre, devido a inúmeros fatores, como a morfologia da inflorescência (SLEPER; POEHLMAN, 2006). A biologia reprodutiva de urucum ainda não está bem explicada, contudo, Rivera e Flores (1988), indicaram que a posição das flores masculinas e femininas de urucum tendem a promover a autofecundação.

Um dado interessante foi obtido por Valdez-Ojeda et al. (2010) a respeito do coeficiente de Wright F. Os autores concluíram que para plantas com flores brancas foram observados altos valores de f (0,395), demonstrando excesso de homozigotos, enquanto para as variantes com flores rosas este coeficiente foi bem menor e negativo (-0,006), indicando excesso de heterozigotos. No experimento dos mesmos, utilizando cruzamentos não controlados, o valor do índice de fixação foi de 0,332, indicando alta homozigosidade, corroborando aos dados do atual estudo.

De acordo com Ritland (1983), diversos fatores podem influenciar neste desvio, como seleção contra genótipos homozigotos, frequência desbalanceada de grãos de pólen na população. Resultados similares foram obtidos por Muluvi et al. (2004) em *Moringa oleifera*, uma espécie com sistema reprodutivo misto, utilizando marcadores AFLP, obtendo uma taxa de cruzamento de 0,74. Assim como urucum, esta espécie tolera a autofecundação.

Cabe salientar que as taxas de cruzamento e autofecundação aparente são estimativas, e não substituem análises com progênies para determinação do sistema reprodutivo, sendo assim, necessários mais estudos para confirmar estes dados.

O índice PIC refere-se ao potencial de um marcador em detectar polimorfismo dentro de uma população ou grupo de indivíduos. Este índice depende do número de alelos detectáveis e da distribuição das frequências alélicas. Tem sido muito utilizado para inferir quão informativo é um marcador. Os valores de PIC observados neste estudo variaram de 0,359 (BorG3) a 0,725 (BorB10), com média de 0,497 (Tabela 3). De acordo com Botstein et al. (1980), os marcadores BorB4, BorG3, BorC12, BorG4, BorA2 e BorF5_2 demonstraram ser moderadamente informativos devido aos valores de PIC estarem situados entre 0,25 e 0,50. Os iniciadores BorB10, BorB12, BorF1 e BorE7 apresentaram valores de PIC superiores a 0,5 sendo considerados marcadores altamente informativos.

O teste de aderência ao Equilíbrio de Hardy-Weinberg (Tabela 4) demonstra a ocorrência de desvios na proporção do Equilíbrio de Hardy-Weinberg em oito dos dez locos estudados a partir do valor p de probabilidade. Os locos que apresentam valores de “ p ” abaixo de 0,05 não estão em equilíbrio. Apenas os locos BorG4 (0,3977) e BorF5_2 (1,000) aparentaram estar sob equilíbrio de Hardy-Weinberg. Tal resultado já era esperado, uma vez que as premissas do Equilíbrio de Hardy-Weinberg determinam que as populações devem ser panmíticas, ou seja, com cruzamentos ao acaso, não ocorrendo autofecundação, o que não acontece com *Bixa orellana*. Além disso, trata-se de acessos de um banco de germoplasma e não populações naturais.

Tabela 4 - Teste exato de aderência ao Equilíbrio de Hardy-Weinberg para cada loco

Locos	Valor p de probabilidade
BorB4	0,0000
BorG3	0,0009
BorC12	0,0000
BorG4	0,3977
BorA2	0,0000
BorF5_2	1,0000
BorB10	0,0000
BorB12	0,0000
BorF1	0,0000
BorE7	0,0000

Os valores das distâncias de Rogers modificada estimados a partir dos marcadores SSR, variaram de 0 a 0,968, o que indica grande variabilidade para o conjunto de acessos avaliado. A análise de agrupamento para os genótipos avaliados apresentou estruturação em dois grandes grupos (Figura 7). O primeiro grupo foi formado pelos acessos da região Norte do Brasil, do estado de Rondônia, e também um acesso da região Centro-Oeste, da cidade de Cuiabá – MT. O único acesso oriundo de fora do país, do Peru, ficou agrupado com os acessos da região Norte, o que faz sentido em função da maior proximidade deste país com a região norte do Brasil e pertence também ao bioma Amazônia, local definido como centro de origem do Urucum. O segundo grupo foi formado pelos acessos da região Sudeste (São Paulo, Minas Gerais e Rio de Janeiro) bem como de um acesso da região Centro-Oeste, da cidade de Cuiabá - MT. A separação entre os acessos de Rondônia dos acessos da região Sudeste ficou bem clara. A análise de Bootstraps revelou um valor de 74,8% suportando esse nó, dando consistência ao agrupamento formado. Os acessos de Cuiabá, por estarem em região intermediária (Centro-Oeste) ficaram separados, um em cada um dos dois grupos.

Foi possível identificar duas duplicatas, uma em cada um dos dois grupos. No grupo 1 dois acessos de Colorado do Oeste – RO apresentaram distancia genética igual a zero. No grupo dois, um acesso de Igaratá – SP não apresentou diferença genética frente a um acesso de origem desconhecida. Sugere-se, baseado nesta clara separação dos acessos nos dois grupos, que os acessos de origem desconhecida do primeiro grupo tenham se originado no

norte do país, enquanto que os acessos de origem desconhecida do segundo grupo sejam originados da região sudeste.

Estes resultados vão de acordo com os obtidos por Carvalho et al. (2005), que apresentaram, em análise de agrupamento, clara diferenciação dos acessos de procedências do Norte, daqueles oriundos de outras regiões do país. Tezotto et al. (2006) avaliando 27 acessos deste mesmo Banco de Germoplasma do IAC, utilizando marcadores isoenzimáticos, puderam separar os acessos originários de Mato Grosso dos demais.

Os valores de correlação obtidos com o teste de Mantel foram significativos, apresentando coeficiente de correlação de Pearson (r) igual a 0,367, indicando baixa correlação, embora significativa ($P < 0,01$), entre as distâncias genéticas e geográficas. Portela de Carvalho et al. (2005) utilizando marcadores isoenzimáticos também detectou que as diferenças genéticas correspondem a diferenças geográficas. No entanto, Medina et al. (2001), avaliando 36 genótipos de urucum coletados na Venezuela e no Brasil utilizando também marcadores isoenzimáticos, não observaram correlação entre as distâncias geográficas e genéticas. No presente estudo, pode-se inferir que a correlação positiva e significativa se deva à estruturação dos acessos em dois grupos, sendo que acessos da região Norte (RO) encontram-se mais distantes tanto geograficamente como geneticamente dos acessos do Sudeste (SP, RJ e MG).

Figura 7 – Dendrograma obtido com o método de UPGMA e distância de Rogers modificada. Os acessos marcados em vermelho são oriundos da região Norte, os marcados em verde são da região Sudeste, enquanto os marcados em azul são da região Centro-Oeste. O acesso marcado em amarelo é oriundo do Peru e em preto encontram-se os acessos sem identificação (SI), de origem desconhecida

Para a análise da estruturação dos genótipos utilizando a abordagem bayesiana realizada pelo programa STRUCTURE 2.3, foram testados os valores de K variando entre 1 e 10. Na Figura 8 estão representados os resultados das análises considerando os valores de K=2, K=3 e K=4. A partir do K=5 os agrupamentos se encontravam totalmente inconsistentes, por isso não estão representados. O melhor K foi estimado a partir das metodologias de Pritchard et al. (2000) e também de Evanno et al. (2005), sendo definido como K= 2 (Figura 9).

Figura 8 – Teste de atribuição para os 63 acessos do banco de germoplasma de urucum a partir de 10 locos microssatélites. No primeiro gráfico os genótipos foram classificados em dois grupos (K=2). No segundo gráfico, os acessos foram classificados em três grupos (K=3) e no terceiro gráfico, os acessos foram agrupados em quatro grupos (K=4). Os acessos estão representados pelas barras verticais coloridas. A mesma cor para genótipos diferentes indica que eles pertencem ao mesmo grupo. Cores diferentes no mesmo indivíduo indicam a porcentagem de alelos compartilhados com cada grupo. O eixo das ordenadas mostra o coeficiente de participação do indivíduo em cada grupo

Quando o número de K é aumentado de 2 para 3, verificamos maior consistência no grupo de acessos do sudeste, enquanto o grupo de acessos da região norte é subdividido. Os acessos 22 (Peru), 38, 44, 45, 46, 57 e 60 (Corumbiara-RO), 49, 51, 55, 56 e 59 (Colorado d'Oeste - RO), além de dois acessos de origem desconhecida e os acessos 2 (Cuiabá-MT), 8 (Tupi Paulista - SP) e 14 (Icem - SP) formaram um novo agrupamento (verde). Porém, quando aumentamos o número de k para 4, a maior parte dos acessos da região sudeste se subdivide, enquanto os acessos da região norte permanecem mais consistentes com relação ao valor de k=3.

Aparentemente não há explicação biológica para as subdivisões em k=3 e k=4, confirmando que o valor de k=2, definido pelas metodologias de Evanno et al. (2005) e Pritchard et al. (2002), é o mais prudente para o grupo de acessos em estudo (Figura 9).

Figura 9 - Valores de ΔK para cada valor de K. O maior valor de ΔK representa o valor ideal de K de acordo com o proposto por Evanno et al. (2005)

O gráfico de barras resultante da análise bayesiana realizada com o programa STRUCTURE, com K=2 está representado na Figura 10. Nesta análise também foi verificada a existência de híbridos entre os acessos, caracterizando a migração de alelos entre os dois grupos. Os acessos 2 e 3 que foram classificados como pertencentes à região 1 apresentaram 70,3% e 51,9% de probabilidade de eles, ou seus ancestrais, serem migrantes da região 2. Este dado ajuda a explicar o fato do acesso 2 (Cuiabá-MT) ter sido agrupado distante do acesso 1 também de Cuiabá-MT.

Figura 10 – Análise da estrutura genética dos 63 acessos do banco de germoplasma de urucum a partir de 10 locos microssatélites. Os genótipos foram classificados em dois grupos ($K = 2$). Acessos de cor verde são originados da região norte e condizentes com o primeiro grupo da análise de agrupamento e de cor vermelha são, em sua maioria, originados da região Sudeste e condizentes com o segundo grupo na análise de agrupamento

A partir do número de grupos definidos na análise bayesiana realizada pelo software Structure, o conjunto de acessos foi dividido em dois grupos, que coincidiram com os grupos obtidos na análise de agrupamento (Figura 7), demonstrando a consistência dos dados obtidos. A Tabela 5 apresenta os parâmetros que analisam a estrutura de diversidade genética como as estatísticas F de Wright e a estatística de Nei (1973) para cada loco.

Tabela 5 – Estimativas dos índices de fixação de Nei (H_S , D_{ST} , H_T , G_{ST} , G_{IS}) e Wright (F_{IT} , F_{ST} e F_{IS}) avaliados na coleção de urucum (*Bixa orellana*) do Banco de Germoplasma do IAC, considerando dois grupos gerados pela análise de estruturação

Locos	H_S	D_{ST}	H_T	G_{ST}	G_{IS}	F_{IT}	F_{ST}	F_{IS}
BorB4	0,315	0,399	0,714	0,559	0,323	0,697	0,510	0,382
BorG3	0,462	0,038	0,5	0,076	0,39	0,446	0,077	0,399
BorC12	0,461	0,043	0,504	0,085	0,504	0,546	0,086	0,503
BorG4	0,507	0,019	0,526	0,036	0,062	0,116	0,035	0,083
BorA2	0,518	0,09	0,608	0,148	0,525	0,605	0,149	0,536
BorF5_2	0,551	0,003	0,554	0,006	-0,31	-0,315	0,006	-0,325
BorB10	0,683	0,163	0,846	0,193	0,297	0,431	0,191	0,296
BorB12	0,371	0,407	0,778	0,523	0,486	0,751	0,492	0,509
BorF1	0,605	0,078	0,683	0,114	0,306	0,372	0,104	0,299
BorE7	0,623	-0,014	0,609	-0,024	1,000	1,000	-0,026	1,000
Média	0,509	0,123	0,632	0,194	0,361	0,483	0,183	0,368

Tabela 6 - Parâmetros de diversidade genética avaliados na coleção de urucum (*Bixa orellana*) do Banco de Germoplasma do IAC, considerando dois grupos gerados pela análise de estruturação, incluindo número de alelos (A), heterozigosidade esperada (H_e), heterozigosidade observada (H_o), e índice de fixação de Wright (f)

Grupos	A	H_e	H_o	f
Grupo 1	34	0,488	0,355	0,272
Grupo 2	38	0,517	0,299	0,387
Total sem divisão	38	0,572	0,632	0,439

¹Grupo 1: acessos da região Sudeste; Grupo 2: acessos da região Norte.

Os valores médios obtidos de F_{ST} (0,183) e G_{ST} (0,194) indicam a existência de estruturação na amostra total e corrobora os resultados obtidos pela análise de agrupamento, e também com os dados obtidos pela análise de estruturação com o software STRUCTURE. Os valores médios de F_{IT} (0,483) e H_{IT} (0,632) demonstraram a elevada diversidade genética no conjunto de acessos estudado. O fato de que os valores de F_{IS} e (HS) G_{IS} sejam maiores do que os valores de F_{ST} e G_{ST} para praticamente todos os locos, assim como os valores H_S (0,509) demonstraram-se superiores aos valores de D_{ST} (0,123), indica a existência de maior diversidade dentro de cada grupo de acessos do que entre os grupos. Valdez-Ojeda et al. (2008) utilizando marcadores SRAP para avaliar acessos originários do México, também identificaram a existência de maior diversidade dentro dos grupos estudados (61,55%) do que entre eles (38,45%). É importante mencionar que os índices de Nei são os mais adequados para estudos com bancos de germoplasma, pois não tem os pressupostos de ancestralidade comum.

É possível identificar que o grupo de acessos do Grupo 2 (região Norte) apresentou índices de diversidade superiores ao agrupamento 1 (região Sudeste). Os valores de heterozigosidade esperada e número de alelos, que são representativos da diversidade presente na amostra, apresentaram leve incremento quando comparados aos valores obtidos pelo Grupo 1 (Tabela 6). Outro dado interessante é que o Grupo 2 apresentou todos os alelos presentes no estudo, enquanto no Grupo 1 alguns alelos não estão presentes. Tal explicação pode ser devida ao fato dos acessos da região norte estarem mais próximos ao centro de origem da espécie, que é considerado como sendo na região Amazônica (SANDY-CUEN; BECERRA, 2003; CLEMENT et al., 2010).

A manutenção do banco ativo de germoplasma de urucum possibilita a preservação do patrimônio genético e pode fornecer matéria prima para as atividades de melhoramento e fomento da cadeia produtiva da cultura. Para que ocorra um manejo mais eficiente desses

bancos é necessário um profundo conhecimento sobre a variabilidade genética contida nos mesmos. Neste sentido, os dados obtidos neste estudo foram muito importantes no sentido de que puderam separar os genótipos em dois grupos geneticamente distintos, o que pode ser utilizado na seleção de parentais divergentes para obtenção de híbridos em programas de melhoramento desta cultura. Além disso, os dados forneceram informações extras sobre a possível origem dos acessos sem identificação, pelo menos pode-se inferir a que região os mesmos pertencem.

5 CONCLUSÕES

A partir das análises com marcadores moleculares conclui-se que existe significativa variabilidade entre os genótipos avaliados, com diferenciação dos genótipos da região Norte e da região Sudeste.

Foi possível identificar que os acessos provenientes da região norte apresentam maior diversidade em comparação com os acessos da região Sudeste. Também foi possível identificar que a maior diversidade está contida dentro dos grupos de acessos e não entre eles, sendo essa informação útil no planejamento de coletas para incremento do Banco de Germoplasma.

Os iniciadores de microssatélites sintetizados neste estudo permitiram a detecção de alto grau de polimorfismo entre todo o conjunto de acessos estudados, indicando que os mesmos são adequados para emprego em estudos de diversidade genética em *Bixa orellana*.

Os marcadores SSR sintetizados nesse trabalho poderão ser utilizados para a análise de genótipos de urucum, em estudos de diversidade, mapeamento genético e outras áreas da genética.

Os genótipos analisados apresentaram um excesso de homozigose na maioria dos locos estudados, sugerindo forte endogamia, provavelmente resultante da ocorrência de autofecundação em urucum.

Foi possível agrupar os acessos com origem desconhecida apenas pela região de provável origem destes acessos, permanecendo os mesmos ainda com suas informações de origem incompletas.

Os marcadores SSR revelaram-se importantes ferramentas para a construção de perfis genéticos e de dispersão dos acessos avaliados, apresentando potencial para aplicação em programas de melhoramento genético e de conservação *in situ* e *ex situ* de urucum.

REFERÊNCIAS

- AGARWAL; M.; SHRIVASTAVA, N.; PADH, H. Advances in molecular marker techniques and their applications in plant sciences. **Plant Cell Report**, Berlin, v. 27, p. 617–631, 2008.
- AKSHATHA, P. G.; RAVISHANKAR, G.A. Morphological diversity in *Bixa orellana* L. and variations in annatto pigment yield. **The Journal of Horticultural Science & Biotechnology**, Ashford, v.86, p. 319-324, 2011.
- ALMEIDA, E.C.; PINHEIRO, A.L. **Biologia floral e mecanismo de reprodução em urucuzeiro (*Bixa orellana* L.) I. tipo “fruto verde piloso”**. In: REUNIÃO TÉCNICO CIENTÍFICA SOBRE O MELHORAMENTO GENÉTICO DO URUCUZEIRO, Belém, PA. **Anais...** Belém: EMBRAPA – CPATU, 1992. v.1, p.72–81. 1991.
- ANGELUCI, E.; ARIMA, H.K.; KUMAGAI, E.A. **Urucum I. Dados preliminares sobre a composição química**. Campinas: ITAL, p.86-95, 1980.
- ARAÚJO, I.S.; INTORNE, A.C.; PEREIRA, M.G.; LOPES, U.V.; SOUZA-FILHO, G.A. Development and characterization of novel tetra-, tri- and di-nucleotide microsatellite markers in cacao (*Theobroma cacao* L.). **Molecular Breeding**, New York, v. 20, p. 73-81, 2007.
- ARCE, J. **El achiote *Bixa orellana* L. cultivo promisorio para el tropic**. Turrialba: Earth, 1999. 149p.
- BAER, D.F. **Systematics of the genus *Bixa* and geography of the cultivated annatto**. 1976. 252 p. Tese (Doutorado) - University of California, Los Angeles , 1976.
- BAJAY, M. M.; PINHEIRO J. P.; BATISTA, C. E. A.; NÓBREGA, M. B. M.; ZUCCHI M. I. Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. **Conservation Genetic Resources**, Durham, v.1, p. 237-239, 2009.
- BALÉE, W. **Footprints of the forest. Ka'apor Ethnobotany - The historical ecology of plant utilization by an Amazonian people**. New York: Columbia University Press, 1994.
- BHATTACHARJEE, R.; FERGUSON, M.; GEDIL, M.; DOMINIQUE, D.; INGELBRECHT, I. Field collection, preservation and large scale DNA extraction procedures for cassava (*Manihot esculenta* Crantz). **African Journal of Biotechnology**, Nigéria, v. 8, p. 3424-3430, 2009.
- BILLOTTE, N.; LAGODA, P.J.L.; RISTERUCCI, A.M.; BAURENS, F.C. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops, **Fruits**, Cambridge, v.54, n.4, p. 277-288, 1999.
- BOHN, M.; UTZ, H.F.; MELCHINGER, A.E. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. **Crop Science**, Madison, v. 39, p. 228–237, 1999.

BOREM, A. **Melhoramento de plantas**. Viçosa: UFV, 1999. 547p.

BOTSTEIN, D.; WHITE, R.L.; SKOLNIEK, M.; DAVIS, R.V. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetic**, Boston, n. 32, p. 314-331, 1980.

BRESSAN, E.A.; SCOTTON, D.C.; FERREIRA, R.R.; JORGE, E.C.; SEBBENN, A.M.; GERALD, L.S.T.; FIGUEIRA, A. Development of microsatellite primers for *Jatropha curcas* (Euphorbiaceae) and transferability to congeners. **American Journal of Botany**, Baltimore, v.99, p.e237–e239, 2012.

BROWN, C.H. Development of agriculture in prehistoric Mesoamerica: the linguistic evidence. In: STALLER, J.E.; CARRASCO, M.D. (Eds.) **Pre-Columbian foodways: interdisciplinary approaches to food, culture, and markets in ancient Mesoamerica**. New York: Springer Science, 2010. p. 71-107.

BUSO, G.S.C.; CIAMPI, A.Y.; MORETZOHN, M.C.; AMARAL, Z.P.S.; BRONDANI, R.V. Marcadores microssatélites em espécies vegetais. **Revista Biotecnologia e Ciência e Desenvolvimento**, Brasília, v. 30, p. 46-50, 2003.

CÁCERES A.; MENÉNDEZ H.; MÉNDEZ E.; COHOBÓN E.; SAMAYOA B.E.; JAUREGUI E.; PERALTA E.; CARRILLO G. Antigonorrhoeal activity of plants used in Guatemala for the treatment of sexually transmitted diseases. **Journal of Ethnopharmacology**, Limerick, n. 48, p.85-88, 1995

CAIXETA, E.T.; OLIVEIRA, A.C.B. de; BRITO, G.G. de; SAKIYAMA, N.S. Tipos de marcadores moleculares. In: BOREM, A.; CAIXETA, E.T. (Ed.). **Marcadores moleculares**. UFV, Viçosa, 2006.

CARVALHO, P.R.N. Corantes de urucum hidrossolúveis. **Revista Brasileira de Corantes Naturais**, Viçosa, v.1, n.1, p. 242-243, 1992.

CARVALHO, J.F.R.P.; ROBINSON, I.P.; ALFENAS, A.C. Isozymic variability in a Brazilian collection of annatto (*Bixa orellana* L.). **Pesquisa Agropecuária Brasileira**, Brasília, v.40, n.7, p.653-660, 2005.

CLEMENT, C.R.; CRISTO-ARAÚJO, M.; D'EECKENBRUGGE, G.C.; ALVES PEREIRA, A.; PICANÇO-RODRIGUES, D. Origin and domestication of native Amazonian crops. **Diversity**, Basel, v.2, p.72-106, 2010.

COELHO, A. M. S. P.; SILVA, G. A.; VIEIRA, O. M. C.; CHAVASCO, J. K. Atividade antimicrobiana de *Bixa orellana* L. (Urucum). **Lecta**, Bragança Paulista, v. 21, n. 1, p. 47-54, 2003.

CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, New York, v.19, p. 299-306, 2001.

DECROOCQ, V.; FAVE, M.G.; HAGEN, L.; BORDENAVE, L.; DECROOCQ, S. Development and transferability of apricot and grape EST microsatellite markers across taxa. **Theoretical and applied genetics**, Berlin, v. 106, p. 912–922, 2003.

DON, R.H.; COX, P.T.; WAINWRIGHT, B.J.; BAKER, K.; MATTICK, J.S. “Touchdown” PCR to circumvent spurious priming during gene amplification. **Nucleic Acids Research**, Oxford, v. 19, p. 4008, 1991.

DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, Ithaca, v.12, p. 13-15, 1990.

DREISIGACKER, S.; ZHANG, P.; WARBURTON, M.L.; van GINKEL, M.; HOISINGTON, D.; BOHN, M.; MELLCHINGER, A.E. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments. **Crop Science**, Madison, v. 44, p. 381-388, 2004.

DUAN, T.; GONG, W.; ZHANG, D. Development of microsatellite markers from *Mussaenda pubescens* (Rubiaceae). **American Journal of Botany**, Baltimore, v. 99, n.11, p. 437–439, 2012.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software Structure: a simulation study. **Molecular Ecology**, Vancouver, v. 14, p.2611–2620, 2005.

FABRI, E. G.; IATESTA, C.S.; GOLLA, A.R.; TERAMOTO, J.R.S.; PANTANO, A.P. Ocorrência de urucum no Estado de São Paulo. Campinas, 2011. Disponível em: <<http://www.iac.sp.gov.br/Tecnologias/Urucumsp/urucumsp.htm>>. Acesso em: 31 de agosto de 2011.

FAY, M.F.; COWAN, R.S.; LEITCH, I.J. The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. **Annals of botany**, Londres, v.95, p.237-246, 2005.

FERREIRA, T.G.T.; OLIVEIRA, C.A.; JUNQUEIRA, N.; FIDELES, M.; ZUCCHI, M.I.; VIEIRA, M.L.C. Estimativa da taxa de cruzamento de *Passiflora alata* var. curtis utilizando marcadores moleculares. IN: Congresso Brasileiro de Genética, 53, 2007, Águas de Lindóia. **RESUMOS...** Ribeirão Preto:USP, 2007. Resumo GP070.

FERREIRA, W.A.; FALESI, I.C. **Características nutricionais do fruto e teor de bixina em urucum (*Bixa orellana* L.)**. Belém: EMBRAPA-CPATU, 1989. 31p.

FERREIRA, M.E.; GRATAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: EMBRAPA-CENARGEN, 1998. 220p.

FRANCO, C.F. de O.; FABRI, E.G.; NETO, M.B.; MANFIOLLI, M.H.; HARDER, M.N.C.; RUCKER, N.G. de A. **Urucum: sistemas de produção para o Brasil**. João Pessoa: Emepa, 2008. 112p.

GIULIANO, G.; AL-BABILI, S.; VON LINTIG, J. Carotenoid oxygenases: cleave it or leave it. **Trends in Plant Science**, Oxford, v.8, p. 145–149, 2003.

GOUDET, J. **FSTAT, a program to estimate and test gene diversities and fixation indices** (version 2.9.3). 2001. Disponível em: <http://www.unil.ch/izea/software/fstat.html>. Acesso em: 14 setembro 2012.

HALE, M.L.; BURG, T.M.; STEEVES, T.E. Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. **Plosone**, San Francisco, v. 7, n. 9, p. 1 – 7, 2012.

HALLAUER, A.R.; MIRANDA FILHO, J.B. **Quantitative genetics in maize breeding**. 2.ed. Ames: Iowa State University, 1988. 468p.

HAMILTON, M.B.; PINCUS, E.L.; Di-FIORE, A.; FLEISHER, R.C. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. **Biotechniques**, Natick, v. 27, p. 500-507, 1999.

HARTL, D. L.; CLARK, A.C. **Principles of Population Genetics**. 3ª ed. Sinauer Associates, Sunderland, MA. 1997. 542 p.

HIJMANS, R. J.; GUARINO, L.; BUSSINK, C.; MATHUR, P.; CRUZ, M.;BARRANTES, I.; ROJAS, E. **Sistema de infomacion geografica para el análisis de datos de distribucion en especies**. Roma: IPGRI-CIP, 2004. 83p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Disponível em: <<http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=1612>>. Acesso em: 12 Set. 2012.

JATASRA, D.S.; PARODA, R.S. Genetic divergence in wheat. **Indian Journal of Genetic**, India, v. 43, p. 63-67, 1983.

JOSEPH, N.; SIRIL, E.A.; NAIR, G.M. Reproductive characterization and preliminary studies on controlled breeding of Annatto (*Bixa orellana* L.). **Plant Systematics and Evolution**, Liede, v. 298, p 239-250, 2011.

KALIA, R.K.; RAI, M.K.; KALIA, S.; SINGH, R.; DHAWAN, A.K. Microsatellite markers: an overview of the recent progress in plants. **Euphytica**, Wageningen, v. 177, p. 309–334, 2011.

KRISHNAN, N.; AYYANGAR, K.R. Taxonomical relationships of *Bixa* L. In: Indian science congress association, 7., 1987, Mumbai. Resumos... Mumbai: 1987. Resumo 105.

LEWIS, P.O.; ZAYKIN, D. **Genetic Data Analysis**: computer program for the analysis of allelic data. *Version 1.0 (d15)*. Disponível em: <<http://alleyn.eeb.uconn.edu/gda>>. Acesso em: 14 Jun. 2011.

LIMA L.R.P.; OLIVEIRA T.T.; NAGEM T.J.; PINTO A.S.; STRINGHETA P.C.; TINOCO A.L.A.; SILVA J.F. Bixina, norbixina e quercetina e seus efeitos no metabolismo lipídico de coelhos. **Brazilian Journal of Veterinarian Research and Animal Sciences**, São Paulo, v. 38, p.196-200, 2001.

LORENZI, H. **Plantas medicinais no Brasil: nativas e exóticas cultivadas**. Nova Odessa: Instituto Plantarum, 2002. 512 p.

MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research**, Philadelphia, v. 27, p.209–220, 1967.

MASONA, S.L.; STEVENSA, M.R.; JELLENA, E.N.; BONIFACIOB, A.; FAIRBANKSA, D.J.; COLEMANA, C.E.; McCARTYA, R.R.; RASMUSSENA, A.G.; MAUGHANA, P.J. Development and use of microsatellite markers for germplasm characterization in quinoa (*Chenopodium quinoa* Willd.). **Crop Science**, Madison, v. 45, p. 1618-1630, 2005.

MCGONIGLE, K.E.M.; MCCRACKEN, K.J. Study on the effects of annatto (*Bixa orellana*) and *Yucca schidegera* on cholesterol content and yolk colour. **British Poultry Science**, Edinburgh, v.43, p.S41-S42, 2002.

MEDINA, A.; CLARET, M.; RAMIS, C., DÍAZ, A. Caracterización morfológica de frutos de onoto (*Bixa orellana* L.) y su correspondencia con patrones de proteínas e isoenzimas. **Acta Científica Venezolana**, Caracas, v.52, p.14-23, 2001.

MELO, A.A.A.; LIMA, L.C.F. A situação do urucum no Brasil e perspectivas. In: SÃO JOSÉ, A.R. & REBOUÇAS, T.N.H. **A cultura do urucum no Brasil**. Vitória da Conquista: UESB, BA, p.9-19, 1990.

MILLER, A.J.; GROSS, B.L. From forest to field: perennial fruit crop domestication. **American Journal of Botany**, Baltimore, v.98, p.1389–1414, 2011.

MILLER, M. P. **Tools for Population Genetic Analyses (TFPGA) 1.3**: a windows program for the analysis of allozyme and molecular population genetic data. Disponível em: <<http://www.public.asu.edu>>. Acesso em: 20 jun. 2011.

MISIEWICZ, T.M.; BARBOSA, C.E.A.; FINE, P.V.A. Microsatellite primers for an amazonian lowland tropical tree, *Protium subseratum* (Burseraceae). **American Journal of Botany**, Baltimore, v. 99, n.11, p. 465–467, 2012.

MORAIS, O.M.; SÃO JOSÉ, A.R.; REBOUÇAS, T.N.H.; ATAÍDE, E.M. Mejoramiento genético del achiote en Brasil. **Revista Brasileira de Corantes Naturais**, Viçosa, v.3, p.109-111, 1999.

MORGANTE, P.G.; SEBASTIÃO, I.; SILVEIRA, L.E.D.; MORI, G.M.; CONTE, M.; COFFANI-NUNES, J.V. Development of microsatellite markers for *Pimenta pseudocaryophyllus* (Myrtaceae), a wild south american species. **American Journal of Botany**, Baltimore, v. 99, n.11, p. 434–436, 2012.

MORI, G.M.; ZUCCHI, M.I.; SAMPAIO, I.; SOUZA, A.P. MICROSATELLITES FOR THE MANGROVE TREE *AVICENNIA GERMINANS* (ACANTHACEAE): TOOLS FOR HYBRIDIZATION AND MATING SYSTEM STUDIES. **American Journal of Botany**, Baltimore, v.97, p.e79-e81, 2010.

MORRISON E.Y.; SMITH-RICHARDSON S.; WEST M.; BROOKS S.H.E; PASCOE K.; FLECTCHER C. Toxicity of the hyperglycaemic-inducing extract of the annatto (*Bixa orellana*) in the dog. **West Indian Medical Journal**, Kingston, v. 36: 99-103, 1987.

MULUVI, G.M.; SPRENT, J.I.; ODEE, D.; POWELL, W. Estimates of outcrossing rates in *Moringa oleifera* using Amplified Fragment Length Polymorphism (AFLP). **African journal of biotechnology**, Nairobi, v.3, p. 146–151, 2004.

MUNOZ, Y.C.G. **Caracterización morfológica y molecular de genotipos de *Dioscorea alata* y *D. trifida* del Instituto de Investigación Agropecuaria de Panamá, IDIAP y CATIE, Costa Rica**. 2003. 93p. Dissertação (Mestrado em Ciências) – Escuela de Postgrado, Centro Agronômico Tropical de Investigación y Enseñanza - CATIE. 2003.

NASS, L.L. Utilização de recursos genéticos vegetais no melhoramento. In NASS, L.L.; MELO, I.S.; VALADARES-INGLIS, M.C. (Eds.). **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, 2001. p.29-56.

NEI, M. Analysis of gene diversity in subdivided populations. **Proceedings of the National Academy of Science**, Washington, v. 70, p. 3321-3323, 1973.

NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**, Pittsburgh, U.S.A., v. 89, p. 583-590, 1978.

OHRI, D.; BHARGAVA, A.; CATTERJEE, A. Nuclear DNA amounts in 112 species of tropical hardwoods—new estimates. **Plant Biology**, Londres, v.6, p. 555–561, 2004.

OLIVEIRA, E.J.; PÁDUA, J.G.; ZUCCHI, M.I.; VENCovsky, R.; VIEIRA, M.L.C. Origin, evolution and genome distribution of microsatellites. **Genetics and Molecular Biology**, Ribeirão Preto, v.29, n.2, p.294-307, 2006.

PAIVA NETO, V. B.; CARVALHO, C. R.; OTONI, W. C. Mannose - a potential selection system for genetic transformation of annatto (*Bixa orellana* L.). **Biologia Plantarum**, Prague, v. 47, n. 3, p. 441-444, 2003.

PARIDA, S.K.; KALIA, S.K.; SUNITA, K.; DALAL, V.; HEMAPRABHA, G.; SELVI, A.; PANDIT, A.; SINGH, A.; GAIKWAD, K.; SHARMA, T.R.; SRIVASTAVA, P.S.; SINGH, N.K.; MOHAPATRA, T. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. **Theoretical and applied genetics**, Berlin, v. 118, p. 327–338, 2009.

PERRIER, X.; FLORI, A.; BONNOT, F. Data analysis methods. In: HAMON, P.; SEGUIN, M.; PERRIER, X.; GLASZMANN, J. C. (Eds.), **Genetic diversity of cultivated tropical plants**. Enfield, Montpellier: Science Publishers, 2003. p. 43 - 76.

PLOTKIN, M.J. **Tales of a Shaman's apprentice**. New York: Penguin Books, 1993. 233p

POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in plant science**, Oxford, v.1, p. 215–222, 1996.

PRENTICE-HERNANDEZ, C.; RUSIG, O. Extrato de urucum (*Bixa orellana* L.) obtido utilizando álcool etílico como solvente. **Arquivos de Biologia e Tecnologia**, Curitiba, v.35, p.63-74, 1992.

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; ARANTES, N.E.; CONTEL, E.P.B. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, p. 185-193, 2002.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, Pittsburgh, v. 155, p. 945-959, 2000.

RISTCHEL, P.S.; LINS, T.C.L.; TRISTAN, R.L.; BUSO, G.S.C.; FERREIRA, M.E. Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). **BMC Plant Biology**, London, v. 4, p. 1-14, 2004.

RITLAND, K. Estimation of mating systems. In TANKSLEY, S.D.; ORTON, T.J. (Ed.) **Isozymes in plant genetics and breeding, part A**. Amsterdam: Elsevier Science Publisher, 1983, p. 289-302.

RITLAND, K. Extensions of models for the estimation of mating systems using *n* independent loci. **Heredity**, Edinburgh, v. 88, p. 221-228, 2002.

Rivera, D.I., and E.M. Flores. Morfología floral del achiote, *Bixa orellana* L (Bixaceae). **Revista de Biología Tropical**. San Jose, v. 36, p. 499-509, 1988.

RIVERA-MADRID, R.; ESCOBEDO-GM, R.M.; BALAM-GALERA, E.; VERA-KU, M.; HUGES HARRIES, H. Preliminary studies toward genetic improvement of annatto (*Bixa orellana* L.). **Scientia Horticulturae**, Amsterdam, v. 109, p.165-172, 2006.

RODRIGUES, J. P. F. **Análise de isoenzimas em progênies de meio-irmãos de urucum (*Bixa orellana* L.)**. Viçosa: UFV, 1995. 76p.

RODRIGUES L.A.; FRACASSO J.F.; YASHUDA Y. The hypotensive action of the extracts from seeds of *Bixa orellana* L. **Revista de Ciências Farmacêuticas**, São Paulo, v.10, p. 41-44, 1998

ROHLF, F.J. **NTSYS-PC: numerical taxonomy and multivariate analysis system**. New York: Exeter Publisher, 1989. 210p.

ROZEN, S.; SKALETISKY, H.J. **PRIMER3**. Disponível em: <http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi>. Acesso em: 26 set. 2011.

RUSSEL, K.R.M.; MORRISON, E.Y.S.A.; RAGOOBIRSINGH, D. The effect of annatto on insulin binding properties in the dog. **Phytotherapy Research**, Londres, v.19, p.433-436, 2005.

RUSSELL, J.R.; FULLER, J.D.; MACAULAY, M.; HATZ, B.G.; JAHOR, A.; POWELL, W.; WAUGH, R. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. **Theoretical and applied genetics**, Berlin, v. 95, p. 714-722, 1997.

SAMBROOK, J.; RUSSEL, D.W. **Molecular Cloning: A Laboratory Manual**. 3ª Edição. New York: Cold Spring Harbour Laboratory Press, 2001.

SANDY-CUEN, P.M.; BECERRA, R. Manejo campesino de recursos naturales. El achiote. **BioDiversitas**, Belo Horizonte, v.7, p. 7–11, 2003.

SANTOS, E. **O urucum**. Ministério da Agricultura. Serviço de Informação agrícola. Rio de Janeiro:MAPA, 1958. 14 p.

SCHULTES, R.E. Amazonian cultigens and their northward and westward migrations in pre-Columbian times. In: Stone, D. (Ed.) **Pre-Columbian Plant Migration**. Papers of the Peabody Museum of Archaeology and Ethnology. Cambridge: Harvard University, 1984. p. 19-38.

SEHGAL, D.; RAINA, S.N. DNA markers and germplasm resource diagnostics: new perspectives in crop improvement and conservation strategies. In: ARYA, I.D.; ARYA, S. (Ed.) **Utilization of biotechnology in plant sciences**. Dehradun: Microsoft Printech, 2008. p. 39–54.

SIQUEIRA, M.V.B.M.; MARCONI, T.G.; BONATELLI, M.L.; ZUCCHI, M.I.; VEASEY, E.A. Development and characterization of microsatellite markers in water yam (*Dioscorea alata*). **American Journal of Botany**, Baltimore, v.98, p.144-146, 2011.

SLEPER, D.A.; POEHLMAN, J.M. Methods in plant breeding.. In: SLEPER, D.A.; POEHLMAN, J.M. (Ed.) **Breeding field crops**. Ames: Blackwell Publ, 2006. p. 155–182

SOARES, V.L.F.; RODRIGUES, S.M.; OLIVEIRA, T.M.; QUEIROZ, T.O.; LIMA, L.S.; HORA-JÚNIOR, B.T.; GRAMACHO, K.P.; MICHELI, F.; CASCARDO, J.C.M.; OTONI, W.C.; GESTEIRA, A.S.; COSTA, M.G.C. Unraveling new genes associated with seed development and metabolism in *Bixa orellana* L. by expressed sequence tag (EST) analysis. **Molecular biology reports**, Basel, v. 38, p. 1329–1340, 2011.

TAUTZ, D.; RENZ, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. **Nucleic Acids Research**, Londres, v.12, p. 4127–4138, 1984.

TEZOTTO, T.; BORGES, A.; RECCHIA, G.H.; MAY, A.; VEASEY, E.A. Caracterização da variabilidade genética de urucum (*Bixa orellana* L.) utilizando análise isoenzimática. In: Congresso brasileiro de genética, 52, 2007, Foz do Iguaçu: **RESUMOS...** Foz do Iguaçu, 2007, Resumo GP086.

THOMPSON H.J. Isolation, purification and identification of the hyperglycaemic principle of the annatto (*Bixa orellana*). **West Indian medical journal**., Kingston, v. 15, p. 287, 1990.

TNAH, L.H.; LEE, C.T.; LEE, S.L.; NG, K.K.S; NG, C.H.; NURUL-FARHANAH, Z.; LAU, K.H.; CHUA, L.S.L. Isolation and characterization of microsatellite markers for an important tropical tree, *Aquilaria malaccensis* (Thymelaeaceae). **American Journal of Botany**, Baltimore, v.99, p. 431–433, 2012.

VALDEZ-OJEDA, R.; HERNANDEZ-STEFANONI, J.L.; AGUILAR-ESPINOSA, M.; RIVERA-MADRID, R. Assessing Morphological and Genetic Variation in Annatto (*Bixa orellana* L.) by Sequence-related Amplified Polymorphism and Cluster Analysis. **Horticultural Science**, Alexandria, v.43, n.7, p.2013-2017, 2008.

VALDEZ-OJEDA, R.; QUIROS, C.F.; AGUILAR-ESPINOSA, M.L.; RIVERA-MADRID, R. Outcrossing rates in annatto determined by sequence-related amplified polymorphism. **Agronomy Journal**, Madison, v.102, n.5, p.1340-1345, 2010.

VARSHNEY, R.K.; THUDI, M.; AGGARWAL, R.; BORNER, A. Genic molecular markers in plants: development and applications. In: VARSHNEY, R.K.; TUBEROSA, R. (Ed.) **Genomics-assisted crop improvement: genomics approaches and platforms**. Dordrecht: Springer, 2007. p. 13–29.

VILARES, A.S. et al. Estudo da biologia floral de urucuzeiro (*Bixa orellana* L.). **Revista Brasileira de Corantes Naturais**, Viçosa, v.1, n.1, p.101-105, 1992.

WANG, M.L.; BARKLEY, N.A.; JENKINS, T.M. Microsatellite markers in plants and insects. Part I. Applications of biotechnology. **Genes Genomes Genomics**, v. 3, p. 54–67, 2009.

WRIGHT, S. **Evolution and the genetics of populations: variability within and among natural populations**. Chicago: University of Chicago Press, 1978. p. 573.

YORINORI, J.T; KIIHL, R.A.S. Melhoramento de plantas visando resistência a doenças. In NASS, L.L.; MELO, I.S.; VALADARES, INGLIS, M.C. (Ed.). **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT,2001. p. 715-735.

ZANE, L.; BARGELLONI, L.; PATARNELLO, T. Strategies for microsatellite isolation: a review. **Molecular Ecology**, Oxford, v.11, p.1-16, 2002.