

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

**Desenvolvimento e uso de marcadores SSR e DArT para estudos de
diversidade genética em macadâmia (*Macadamia integrifolia*)**

Graciela da Rocha Sobierajski

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

**Piracicaba
2012**

Graciela da Rocha Sobierajski
Bióloga

Desenvolvimento e uso de marcadores SSR e DArT para estudos de diversidade genética em
macadâmia (*Macadamia integrifolia*)
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

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

Piracicaba
2012

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

Sobierajski, Graciela da Rocha

Desenvolvimento e uso de marcadores SSR e DArT para estudos de diversidade genética em macadâmia (*Macadamia integrifolia*) / Graciela da Rocha Sobierajski. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2012.

94 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Cruzamento vegetal 2. Diversidade genética 3. Macadâmia 4. Marcador molecular 5. Melhoramento genético vegetal 6. Variação genética em plantas
I. Título

CDD 634.57
S677d

**Dedico este trabalho aos meus pais,
Leonardo e Ivone,
pelo grande amor e
por sempre me apoiarem em meus sonhos.**

Agradecimentos

Esta tese é resultado de um trabalho que tive o prazer de realizar em parceria com muitos colegas e amigos. Estes contribuíram para o meu crescimento como pesquisadora e como pessoa, muitos me dando apoio em momentos difíceis, fundamentais para a minha jornada. Assim, quero aproveitar a oportunidade para agradecer:

Ao Prof. Dr. Antonio Augusto Franco Garcia pela orientação e pelos ensinamentos transmitidos nestes quatro anos de convivência, assim como por sua amizade e de sua família;

À Profa. Dra. Anete Pereira de Sousa também pelos ensinamentos, por acreditar neste trabalho desde seu início, quando ainda era “apenas” uma ideia, e pela oportunidade de desenvolver os marcadores SSR;

Ao Dr. Andrzej Kilian por abrir as portas de seu laboratório e dedicar longas horas para que eu pudesse desenvolver os marcadores DArT;

Ao Instituto Agronômico e a minha chefia que autorizou a realização desta pós-graduação;

À EMBRAPA pela concessão da bolsa de estudo e pelo auxílio durante o desenvolvimento dos marcadores DArT na Austrália;

À FAPESP (processo nº 11/06141-0) pela concessão do auxílio que proporcionou a realização deste estudo;

Aos professores Roland Vencovsky, José Baldin Pinheiro e Giancarlo Conde Xavier Oliveira pelas contribuições dadas durante o exame de qualificação;

À Associação Brasileira de Macadâmia (ABM), pelo apoio e por incentivar o crescimento da cultura da macadâmia no Brasil, e em especial, ao Pedro Toledo-Piza por disponibilizar os pomares para coleta de material vegetativo e sementes;

Aos colegas do Centro de Fruticultura/IAC pelo auxílio no campo para manutenção do banco de germoplasma de macadâmia, em especial a Lucia, Mara, Marco, Marcos, Zé Luiz, Luiz e Diego;

Aos colegas do Laboratório de Genética Estatística: Rodrigo, Maria Marta, Gabriel, Marcelo, Renato, Edjane, Carina, João, Rodrigo, Luciano, Guilherme e Adriana pela convivência e amizade;

Aos professores e funcionários do Departamento de Genética;

À equipe do laboratório CEBEMEG/UNICAMP pelo auxílio no desenvolvimento dos marcadores SSR e na genotipagem das variedades, em especial à Letícia, Adna, Paty Juverlande, Danilo e Rafa sempre dispostos a me ajudar na bancada;

À equipe da empresa Diversity Arrays Technology pelos ensinamentos, amizade e carinho durante minha estadia na Austrália, em especial à Cina, Collen, Ling, Vanessa, (que saudade dos nossos cafés!), Claire, Sara, Kasia e Adriane, amigas inesquecíveis, que sempre me confortaram nos momentos em que a saudade “batia”;

Ao Renato e ao Guilherme pela valiosa ajuda nas análises dos dados e pelas sugestões que enriqueceram esta tese;

À Mara pela paciência de ler o protótipo deste trabalho;

Aos meus amigos que sempre me apoiaram, mas que principalmente nestes quatro anos, me deram muita força para continuar essa travessia, em especial a Lucia, Tais, Maurício, Mara, Nara, Marília, Andreia, Selmo, Elza, Carina, Renato, Larissa, Adriana, Rose, Roberto (e todos os amigos do CNM, por me apresentarem o mundo “sub”), Marco e Moreira;

À Andreia e ao Selmo também por sempre me receberam em sua casa tão gentilmente;

À Carina e à Eva pelas longas conversas e por também permitirem as minhas “invasões”;

À Carina pela amizade;

Ao Marcio, meu grande amigo de tantas horas;

E principalmente à minha família: minha mãe pelo apoio e incentivo em sempre buscar a melhor formação acadêmica e profissional e que desde o meu mestrado me ajuda com “a mão na massa”, me acompanhando nas coletas e nos trabalhos no viveiro, por estar sempre ao meu lado e por cuidar tão bem de mim; à minha irmã Rafaela e ao Júnior por terem me dado o melhor presente que recebi até hoje; ao meu sobrinho Tiago, que espero que no futuro entenda essa tia ausente nestes primeiros anos de sua linda vida; à minha madrinha Filomena (Nena) pelo apoio de todas as horas, por sempre me incentivar a continuar os meus estudos;

E a todos os colegas e amigos que passaram por minha vida, meu muito obrigado.

SUMÁRIO

RESUMO	09
ABSTRACT	11
1 INTRODUÇÃO	13
2 REVISÃO BIBLIOGRÁFICA	15
2.1 Macadâmia	15
2.1.1 Origem e domesticação	15
2.1.2 Citogenética	19
2.1.3 Variabilidade genética	19
2.1.4 Melhoramento genético	21
2.1.5 Parâmetros genéticos	25
2.1.6 Mapeamento genético	26
2.2 Marcadores moleculares	27
2.2.1 Marcadores Microssatélites – SSR	29
2.2.2 Marcadores <i>Diversity Array Technology</i> – DArT	32
3 MATERIAL E MÉTODO	37
3.1 Material Vegetal	37
3.2 Desenvolvimento dos marcadores microssatélites – SSR	39
3.3 Desenvolvimento dos marcadores <i>diversity array technology</i> – DArT	42
3.4 Análise estatística	43
3.4.1 Estrutura Genética	44
3.4.2 Distância Genética	45
3.4.3 Análise de Componentes Principais	47
4 RESULTADOS E DISCUSSÃO	49
4.1 Desenvolvimento dos marcadores microssatélites – SSR	49
4.2 Desenvolvimento dos marcadores <i>diversity array technology</i> – DArT	59
4.3 Caracterização molecular do germoplasma	60
4.3.1 Estrutura genética	60
4.3.2 Distância genética	72
5 CONCLUSÕES	85
REFERÊNCIAS	87

RESUMO

Desenvolvimento e uso de marcadores SSR e DArT para estudos de diversidade genética em macadâmia (*Macadamia integrifolia*)

A macadâmia (*Macadamia integrifolia* Maiden & Betche) é uma noqueira arbórea originária das florestas tropicais australianas. Sua distribuição natural ocorre por quase toda a costa leste da Austrália. No Brasil, foi introduzida em 1931, mas sua expansão aconteceu a partir da década de 70. As variedades comerciais são originadas de duas espécies, *Macadamia integrifolia* e *M. tetraphylla*, e seus híbridos. As informações sobre *pedigree* são geralmente incompletas e a identificação das variedades algumas vezes não é clara. O uso de técnicas moleculares auxilia em diversas fases dos programas de melhoramento, inclusive na caracterização do material genético e seleção de parentais envolvidos nos cruzamentos. A partir do uso de técnicas moleculares (microsatélites e *diversity array technology*), as 28 variedades que compõem o germoplasma brasileiro foram genotipadas com a finalidade de investigar a estrutura e a divergência genética entre estas. Foram utilizados 29 *primers* flanqueadores de locos microsatélites e 462 marcas de *diversity array technology*. A análise pelo programa *Structure* apresentou a formação de duas subpopulações: uma com a maioria das variedades desenvolvidas no Haváí/USA e Brasil, e outra composta pelas variedades desenvolvidas na Austrália e três variedades brasileiras. A análise da diversidade genética detectou valores medianos de heterozigosidade para o conjunto de variedades ($\hat{H}_E = 0,46$ e $\hat{H}_O = 0,43$), com indicação de um pequeno excesso de homozigoto em relação ao esperado sob Equilíbrio de Hardy-Weinberg. A estimativa do índice de fixação ($\hat{F} = 0,067$) para o conjunto de subpopulações apresentou grande variação, porém na média, foi baixo e não diferente de zero. Os resultados também mostram que a maior parte da diversidade genética do germoplasma está dentro de subpopulações ($\hat{\theta} = 0,041$). O índice de fixação devido ao sistema reprodutivo também variou entre os locos analisados, mas na média foi igualmente baixo e não diferente de zero ($\hat{f} = 0,027$). Os valores das Distâncias de Jaccard variaram entre 0,00 e 0,787 enquanto que as Distâncias de Roger Modificada variaram entre 0,227 e 0,671. O dendrograma UPGMA formado pelas Distâncias de Jaccard formou três grupos enquanto que pelas Distâncias de Roger Modificada formou dois grupos, sendo possível distinguir somente o grupo externo formado pelas duas espécies de grevílea. A falta de divergência entre alguns pares de variedades foi inesperada, assim como o não agrupamento entre variedades identificadas com o mesmo nome ou que apresentam histórico de parentesco. Da mesma forma, as duas subpopulações discriminadas pelo programa *Structure* não formaram agrupamento específico em ambos os dendrogramas. Os resultados das análises dos Componentes Principais foram coerentes aos obtidos nos dendrogramas formando os mesmos agrupamentos.

Palavras-chave: Distância genética; Germoplasma; Marcador molecular; Predição de cruzamentos; Variabilidade genética

ABSTRAT

Development and use of SSR and DArT genetic markers to study genetic diversity in macadamia (*Macadamia integrifolia*)

Macadamia (*Macadamia integrifolia* Maiden and Betche) is a nut crop tree widespread across eastern Australia. Although has been introduced in Brazil in 1931, just in the past 20 years its cultivated area has expanded more significantly. Commercial varieties are originated from *Macadamia integrifolia* and *M. tetraphylla*, as well as its hybrids. The information on pedigree is generally incomplete and the identification of varieties is sometimes not so clear. Molecular techniques could provide genetic information useful for breeding programs, such as, the kinship (relatedness) between individuals. Molecular techniques (microsatellites and diversity array technology) was used to genotype all 28 varieties that composes Brazil's germoplasma with the purpose to investigate the structure and the genetic divergence between these varieties. Twenty nine microsatellites locus and 462 markers of diversity array was used. Analysis using the Structure software showed evidence of two subpopulations: one with the majority of the varieties developed in the Hawaii/USA and Brazil, and another with varieties developed in Australia and three in Brazil. The analysis of genetic diversity detected medium values of heterozigosis ($\hat{H}_E = 0.46$ and $\hat{H}_O = 0.43$), with indication of a small excess of homozygotes in relation the expected value under Hardy-Weinberg equilibrium. The estimate of inbreeding coefficient showed great variation (from -0.522 to 0.558), however in the average ($\hat{F} = 0.067$) it was low and statistically not different from zero. The results also showed that most of the genetic diversity is within subpopulations ($\hat{\theta} = 0.041$). The inbreeding due to the reproductive system varied between analyzed loci, but on average it was low and statistically not different from zero ($\hat{f} = 0.027$). Values of Jaccard distances ranged from 0 to 0.787, whereas the modified Roger's distances ranged from 0.227 to 0.671. UPGMA Cluster analysis from Jaccard distances formed three groups, whereas modified Roger's distances of formed two groups. In the latter, it is possible to distinguish only the external group formed by the two species from *Grevilea* spp. For some varieties that we expected to be clustered together were unexpected assigned to different clusters. On the other hand, for some varieties that we expected to be placed in different clusters were unexpected clustered together. Unexpected, the two subpopulations discriminated in the analysis (Structure) did not form specific groups when using both Jaccard and modified Roger's distances. The results of Principal Coordinate Analysis were coherent with the UPGMA analyses.

Keywords: Genetic distance; Germplasm; Molecular marker; Genetic variability; Parental selection

1 INTRODUÇÃO

A Macadâmia (*Macadamia integrifolia* Maiden & Betche), pertencente à família botânica das Proteaceas, é uma noz australiana originária dos estados de Queensland (leste e sudeste) e Nova Gales do Sul (norte), ambos os estados na costa leste australiana. Seu habitat natural são as florestas tropicais e subtropicais, o que favoreceu a sua adaptação em outros países, tais como nos Estados Unidos/Havaí, África do Sul, América Central e Brasil. Atualmente apenas 1% da área original de floresta tropical Australiana está preservada e restrita à fragmentos (PEACE, 2002), o que torna indispensável uma ação para conservação desta espécie.

Trata-se de uma noz muito apreciada no mercado internacional e, apesar do Brasil estar entre os maiores produtores, ainda é pouco conhecida pelo consumidor brasileiro. A Austrália e o Havaí/USA detêm as maiores participações na produção mundial e juntos têm a participação de cerca de 70% do mercado. Em 2005 o Brasil ocupou o sexto lugar na classificação mundial com produção de 3.200 toneladas (ton.) de noz em casca, ou o correspondente a 740 ton. em amêndoas (TOLEDO-PIZA, 2005; SOBIERAJSKI *et al.*, 2006). O'Connor (2000) cita como sendo os principais estados produtores: São Paulo (33%), Espírito Santo (31%), Bahia (18%) e Rio de Janeiro (10%).

Assim como outras nozes, o consumo de macadâmia deve ser incentivado devido à presença de diversos minerais em sua constituição. As nozes são compostas por, aproximadamente, 70% de óleos monoinsaturados, 9,4% de proteína, 4,7% de carboidratos, além de minerais (potássio, fósforo, magnésio, cálcio, selênio, zinco, cobre e ferro), vitaminas (complexo B e E) e antioxidantes (incluindo flavonóides)(STEPHENSON, 2005).

O programa de domesticação e melhoramento genético de macadâmia no Brasil teve início no Instituto Agrônômico (IAC) na década de 40 com a introdução de sementes de variedades desenvolvidas no Havaí/EUA (DIERBERGER e MARINO NETO, 1981). Desde então foram selecionados 16 materiais, podendo ser destacadas as variedades IAC 4-12B, IAC 1-21 e IAC 9-20 pela sua expressiva produtividade e qualidade de nozes (PIZA *et al.*, 2006). Atualmente as variedades desenvolvidas no Brasil representam somente 30% dos pomares em produção (O'CONNOR, 2000). Essa baixa porcentagem pode ser considerada devido o último lançamento de variedades ter ocorrido em 1981, sendo a demanda por novos materiais uma prioridade citada pelos produtores durante as reuniões da Associação Brasileira de Noz-Macadâmia (ABM). Assim, o programa de melhoramento genético de macadâmia do IAC foi reativado e reestruturado visando atender às necessidades do setor.

Como as primeiras variedades desenvolvidas no IAC foram selecionadas a partir de sementes de poucas matrizes, acredita-se que exista baixa variabilidade genética no banco de germoplasma da Instituição. Até o presente, foram desenvolvidas 16 variedades pelo programa de melhoramento genético do IAC, originadas de sementes de polinização aberta. As informações sobre a genealogia não são completas, havendo somente em alguns casos, informações do genitor feminino.

A existência de variabilidade genética é o ponto de partida para o melhoramento genético uma vez que para haver a seleção, o melhorista necessita de identificar indivíduos superiores em uma população segregante. A população de seleção pode ser formada de diversas maneiras. Uma delas é o cruzamento entre genitores superiores. E a seleção de parentais divergentes para compor os cruzamentos é uma das formas de gerar a variabilidade. Conhecer a variabilidade destes materiais é, portanto, muito importante.

Uma forma eficaz de quantificar a variabilidade genética é através de estudos moleculares, e a utilização de marcadores microssatélites (SSR) aliados a marcadores *diversity array technology* (DArT) pode fornecer estimativas dos parâmetros genéticos populacionais. A falta de informações sobre o *pedigree* das atuais variedades é outro empecilho para o programa de melhoramento, pois não há informações suficientes sobre o grau de parentesco entre a maioria das variedades, o que dificulta a tomada de decisão sobre a escolha dos genótipos parentais.

Desta forma, objetivou-se no presente trabalho avaliar a estrutura genética de *Macadamia integrifolia*, caracterizando o material genético existente no Brasil, utilizando como ferramentas os marcadores microssatélites (SSR) e os *diversity array technology* (DArT).

2 REVISÃO BIBLIOGRÁFICA

2.1 Macadâmia

2.1.1 Origem e domesticação

Centro de Origem

A macadâmia é originada dos estados australianos de Queensland (leste e sudeste) e Nova Gales do Sul (norte), e suas populações naturais estão atualmente restringidas a fragmentos florestais, visto que se acredita que exista apenas 1% da área original da sua distribuição natural (BELL, 1993, citado por PEACE, 2002). Este cenário de rápida destruição de habitat fez com que espécies de macadâmia fossem classificadas como vulneráveis (HAMMER e KHOSHBAKHT, 2005) e sujeitas à extinção.

São poucas as informações sobre a distribuição natural original da macadâmia e, atualmente, as três espécies que ocorrem juntas ocupam uma região costal de apenas 350 km no sentido norte – sul e de 50 km no sentido leste – oeste (McCONACHIE, 1980). Existem relatos não confirmados por evidências que existiam outras populações de macadâmia distantes em até um raio de 100 km da zona de ocorrência atual, sugerindo que a distribuição natural da espécie fosse maior do que é observado (PEACE, 2002).

As populações de *M. integrifolia* são as que apresentam maior distribuição compreendendo quatro grandes regiões (Figura 1). A *M. tetraphylla* é a espécie que apresenta distribuição mais ao sul e em uma única região de ocorrência, porém de maior tamanho contínuo. As populações de *M. ternifolia* coexistem na mesma região das populações de *M. integrifolia*, mas também são encontradas separadamente (PEACE, 2002). A *M. jansanii* é a espécie mais isolada, sendo encontrada em apenas uma população composta por 33 indivíduos no município de Bundaberg (estado de Queensland), e afastada cerca de 150 km ao norte das demais espécies de macadâmia (GROSS e WESTON, 1992).

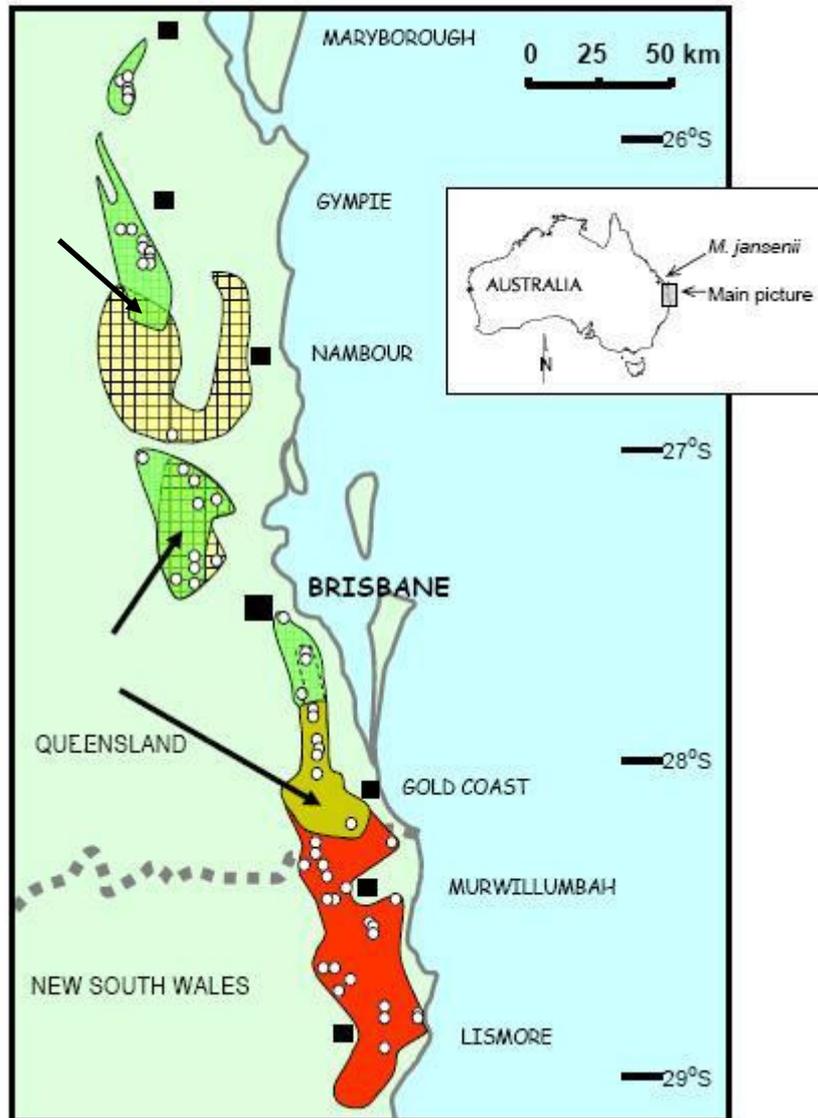


Figura 1 - Mapa da costa leste da Austrália demonstrando as áreas de ocorrência e hibridização das espécies *Macadamia integrifolia* (em verde), *M. tetraphylla* (em vermelho), *M. ternifolia* (em amarelo). *M. jansonii* apresenta-se isolada à 150 km na norte. Área sinalizada pelas setas indica locais de observação de híbridos naturais. As setas indicam as zonas de hibridação geográfica. Fonte: PEACE, 2002

A zona de hibridação natural entre *M. integrifolia* e *M. tetraphylla* ocorre próximo da cidade de Gold Coast, e Johnson (1954, citado por PEACE, 2002) citava ter observado indivíduos morfologicamente intermediários às duas espécies, indicando poder haver cruzamentos interespecíficos naturais. Estes apresentavam grande variabilidade, sugerindo que mais de uma geração de hibridação tivesse ocorrido, porém o indivíduo mais antigo com estas características foi datado como tendo cerca de 100 anos (JOHNSON, 1954, citado por PEACE, 2002), sugerindo que o contato entre estas duas espécies é recente e possivelmente causado por interferências humanas (WILLS, 1961, citado por PEACE, 2002). Já entre *M. integrifolia* e *M. ternifolia* ocorrem duas prováveis zonas de hibridação natural, e apesar desta

sobreposição, não foram observados híbridos naturais entre estas (McCONACHIE, 1980). Recentes expedições encontraram um único indivíduo que apresentava características comuns às duas espécies: folhas semelhantes à *M. ternifolia* e frutos grandes como os de *M. integrifolia* (PEACE, 2002).

Infelizmente, não existem estudos detalhados que contribuam com maiores informações para a corroboração do centro de origem da macadâmia.

Domesticação

A macadâmia é um rico alimento utilizado há muito tempo pelas comunidades aborígenes da Austrália. Relatos mostram que estas comunidades utilizavam as nozes das espécies *M. integrifolia* e *M. tetraphylla*, mas não de *M. ternifolia*, provavelmente por apresentar sabor amargo e tamanho pequeno (McCONACHIE, 1980). O mesmo raciocínio pode ser transposto à *M. janseni*, por ser esta espécie muito rara. Devido às suas características de coletores-caçadores, os pesquisadores não acreditam que estes povos tenham cultivado e ativamente trabalhado para a domesticação da espécie (PEACE, 2002). Porém, durante suas longas jornadas, as tribos carregavam consigo macadâmias coletadas para a alimentação, visto que era fonte rica em carboidratos e de difícil depreciação. Assim, os aborígenes tiveram uma importante contribuição para a dispersão da espécie (PEACE, 2002).

O processo de domesticação da macadâmia pode ser considerado recente, diferentemente da maioria das culturas cultivadas. A macadâmia foi descoberta pelos colonizadores europeus somente em 1828 e em alguns poucos anos começou a ser cultivada em sua forma selvagem (PEACE, 2002). Os primeiros cultivos iniciaram em 1860 na Austrália, a partir de sementes selvagens coletadas das duas espécies comestíveis, e eram realizados com poucas árvores (McCONACHIE, 1980), semelhantes aos pomares domésticos.

O primeiro cultivo comercial foi implantado entre 1878 e 1888 na cidade de Rous Mill, perto da cidade de Lismore, na Austrália. Em 1900 eram relatados cinco pomares comerciais de macadâmia no estado de Nova Gales do Sul, mas nenhum no estado de Queensland, onde a macadâmia era encontrada somente em parques e jardins públicos. O primeiro pomar comercial do estado de Queensland foi iniciado em 1910 com o plantio de 30 árvores de *M. tetraphylla*, e apenas em 1931 as primeiras árvores de *M. integrifolia* foram plantadas comercialmente (McCONACHIE, 1980). O cultivo da macadâmia teve um grande declínio na década de 40 quando problemas com pragas e doenças, além da falta de

conhecimento sobre tratos culturais, quase dizimaram os pomares australianos (McCONACHIE, 1980).

Apesar de ter origem na Austrália e os primeiros cultivos terem sido realizados neste país, o marco da domesticação da macadâmia ocorreu nos Estados Unidos, no estado do Havaí, entre 1880 e 1890 (PEACE, 2002). Foram realizadas três introduções independentes: 1881 – *M. integrifolia*; 1892 – *M. integrifolia*; e, 1892-1894 – *M. tetraphylla*. As sementes foram introduzidas na *Hawaii Agricultural Experiment Station* (HAES) e, semelhante ao ocorrido na Austrália, o interesse inicial era voltado para a *M. tetraphylla* devido sua maior produção de nozes (HAMILTON e FUKUNAGA, 1959; SHIGEURA e OOKA, 1984).

O potencial produtivo da macadâmia impulsionou a substituição dos pomares de café no Havaí e em 1918 os pesquisadores da HAES recomendavam plantios utilizando plantas originadas de sementes de *M. tetraphylla* (HAMILTON e FUKUNAGA, 1959). Porém, por volta de 1920 após avaliar diversas plantas de ambas as espécies, foi verificado que as plantas de *M. tetraphylla* eram mais suscetíveis ao ataque de doenças e que a *M. integrifolia* deveria ser utilizada com fonte de material genético (PEACE, 2002). Os principais avanços no cultivo de macadâmia foram realizados a partir de 1930 quando o primeiro programa de melhoramento genético foi estabelecido no Havaí.

Taxonomia

Para utilização dos recursos genéticos é necessário o conhecimento das relações de parentesco entre os diversos táxons, e a taxonomia tem central importância neste contexto. O primeiro exemplar observado de macadâmia foi coletado pelo botânico Allan Cunningham em 1828, 40 anos após os primeiros colonizadores europeus terem chegado à Austrália (PEACE, 2002). Tratava-se da *M. integrifolia*, porém a espécie somente foi descrita em 1857 quando o botânico Ferdinand Von Mueller realizou seu estudo de identificação de novas espécies baseado nas coletas realizadas em Brisbane (Queensland, Austrália), e homenageou seu colega John Macadam dedicando seu nome ao gênero (McCONACHIE, 1980).

Em 1843, o explorador Ludwig Leichhardt coletou o primeiro espécime de *M. ternifolia* (STOREY, 1959, citado por PEACE, 2002), e até 1956 os indivíduos de *M. tetraphylla* eram classificados como *M. integrifolia* ou *M. ternifolia*. Somente após a revisão do gênero realizado por Smith (1956, citado por PEACE, 2002) é que as três espécies foram descritas. A quarta espécie reconhecida como pertencente ao gênero, *M. janseni*, foi descoberta recentemente em uma única população isolada de outras populações naturais de macadâmia (GROSS e WESTON, 1992).

A controvérsia entre os pesquisadores da área de taxonomia sobre a classificação das espécies do gênero *Macadamia* ainda permanece atualmente. Geralmente são consideradas quatro espécies como compondo o *genepool* primário do gênero: *M. integrifolia*, *M. tetraphylla*, *M. ternifolia* e *M. jansanii* (PEACE, 2002). O *genepool* secundário é composto pelos híbridos formados pelo cruzamento entre as espécies do *genepool* primário, e por último, o terceiro *genepool* formado por espécies correlatas, mas que possuem distribuição geográfica distinta (PEACE, 2002).

2.1.2 Citogenética

Estudos gerais citogenéticos da família Proteacea (na qual o gênero *Macadamia* está inserido) sugeriam a hipótese de que a macadâmia seria um “paleo-poliplóide” de um ancestral $n = 5$ ou 7 (RAO, 1970; JOHNSON e BRIGGS, 1975; ambos citados por PEACE *et al.*, 2008). Porém, Stace, Douglas e Sampson (1998) descartam essa possibilidade devido a alguns gêneros estudados apresentarem cromossomos de tamanho pequeno, e sugerem que membros da família Proteacea sejam derivados de um genoma ancestral $n = 12$ ou 21 cromossomos.

Estima-se que o gênero *Macadamia* possua $n = 14$, porém nenhum estudo específico fora realizado até o momento e a necessidade de maior conhecimento sobre a organização do genoma do gênero *Macadamia* é consenso entre pesquisadores (PEACE *et al.*, 2008), o que mostra a necessidade de estudos básicos.

2.1.3 Variabilidade genética

Alguns estudos foram realizados para investigar a variabilidade genética do germoplasma de macadâmia. Um dos trabalhos pioneiros de avaliação molecular de variedades de macadâmia foi realizado por Vinthanage e Winks (1992), investigado a partir de marcadores isoenzimáticos. Foram avaliadas variedades provenientes de *M. integrifolia*, *M. tetraphylla*, seus híbridos e indivíduos de *M. ternifolia*. Os resultados apresentados mostraram a proximidade genética entre variedades relacionadas, tais como, Keauhou (HAES 246) conhecida como parental de Makai (HAES 800); e, Keaau (HAES 660) e Mauka (HAES 741) originárias do mesmo *seedling* formando grupos semelhantes (VINTHANAGE e WINKS, 1992). Porém, os autores encontraram resultados conflitantes entre as informações de *pedigree* e os resultados moleculares. Por exemplo, a variedade Own Venture relatada

como sendo desenvolvida a partir de sementes de polinização aberta de Own Choise em possível cruzamento com Renown, não foi alocada no mesmo grupo de seu parental feminino (VINTHANAGE e WINKS, 1992). Os autores citam o loco isoenzimático PGI como indicador de uma possível anotação equivocada, pois neste os alelos entre as duas variedades são completamente diferentes (VINTHANAGE e WINKS, 1992).

Também utilizando a técnica de isoenzimas, Aradhya *et al.* (1998) investigou a variabilidade genética em acessos de macadâmia existentes na Austrália, incluindo quatro variedades IAC, a saber: IAC Camp A, IAC Camp F, IAC-412-B e IAC Camp H. Com a análise dos dados foram obtidos quatro grupos, sendo o maior composto pelas variedades desenvolvidas a partir de *M. integrifolia* e *M. ternifolia* (ARADHYA *et al.*, 1998). O agrupamento 1 e 2 (composto por variedades de origem em *M. integrifolia*, *M. ternifolia* e *M. tetraphylla*) apresentaram cerca de 80% de similaridade genética dentro do grupo, enquanto que os grupos externos (formados por *M. hildebrandii* e *Hicksbeachia pinnatifolia*) apresentaram similaridade genética de 60% (ARADHYA *et al.*, 1998). O estudo mostrou que as variedades brasileiras ficaram dispersas entre variedades havaianas, californianas e australianas, porém todas no grupo de variedades ligadas predominantemente ao genoma da espécie *M. integrifolia* (ARADHYA *et al.*, 1998).

Peace *et al.* (2002) aplicaram a técnica *Randomly Amplified DNA Fingerprinting* (RAF) a um estudo sobre identificação de origem genética de 30 variedades de macadâmia. Neste estudo os autores consideraram como sendo de origem *M. integrifolia* as variedades que apresentassem entre 81 e 100% das marcas de *M. integrifolia*, híbridos as que apresentassem entre 43 e 60% das marcas de *M. integrifolia*, e como proveniente de *M. tetraphylla* as que não apresentassem marcas de *M. integrifolia*. Com esta classificação, os genótipos foram separados em três grupos, confirmando a existência de material comercial híbrido entre as duas espécies, além de indicar que variedades consideradas como sendo *M. integrifolia* continham até 19% do genoma *M. tetraphylla* (PEACE *et al.*, 2002).

Em outro estudo semelhante, Steiger *et al.* (2003) avaliaram a diversidade genética de 24 variedades (principalmente australianas e havaianas), e de três espécies correlacionadas a partir de marcadores *Amplified Fragment Length Polymorphism* (AFLP). O índice de similaridade genética entre todos os materiais foi de 0,742 e, entre somente variedades de origem em *M. integrifolia*, de 0,839 (STEIGER *et al.*, 2003). Os mesmos autores citam que apesar das variedades havaianas terem sido desenvolvidas a partir de apenas duas introduções, estas foram selecionadas de um grande número de árvores, e a polinização cruzada manteve a alta diversidade (STEIGER *et al.*, 2003).

Com o objetivo de avaliar o grau de parentesco do material existente na África do Sul, Peace *et al.* (2005) utilizaram o marcador RAF na genotipagem de 31 variedades de macadâmia. A partir dos dados de marcadores foi possível separar o material em sete grupos conforme a similaridade genética. A formação dos grupos seguiu o que era esperado por sua constituição de origem genética e, posteriormente, pelo local de seleção (PEACE *et al.*, 2005). Além destes resultados, os autores puderam afirmar a identidade de duas variedades já lançadas, as quais eram denominadas erroneamente naquele país.

Conforme Peace *et al.* (2008) as cultivares havaianas provém de um limitado conjunto gênico, derivado de um pequeno número de sementes introduzidas da Austrália, o que nos sugere que as variedades brasileiras compartilhem da mesma baixa heterozigosidade. Isto pode limitar seriamente o ganho com seleção nos programas de melhoramento genético.

2.1.4 Melhoramento genético

Programas de Melhoramento Genético

As características biológicas das espécies devem ser consideradas em um programa de melhoramento genético, tais como, duração da fase juvenil e tipo de sistema reprodutivo. A macadâmia possui um período juvenil de aproximadamente 5-7 anos em indivíduos originados por sementes, o que pode aumentar o tempo necessário para a obtenção de novos cultivares.

A espécie apresenta características de alogamia, embora não existam estudos a respeito do sistema reprodutivo. As variedades de macadâmia selecionadas até o momento são provavelmente provenientes de cruzamentos, ou polinização aberta, não havendo a possibilidade de formação de linhagens endogâmicas via autofecundação, devido à presença de genes deletérios, comuns em espécies com este sistema reprodutivo, e seu efetivo processo de melhoramento.

*Estados Unidos - Havaí

A macadâmia fora introduzida no Estado do Havaí entre 1881 e 1892 e, a partir da criação da *Hawaii Agricultural Experiment Station* (HAES), estudos de adaptação começaram a ser realizados visando a utilização comercial da espécie (HAMILTON e FUKUNAGA, 1959; SHIGEURA e OOKA, 1984). O primeiro programa de melhoramento genético de macadâmia foi instituído em 1930 no Havaí, e era fundamentado em avaliações do desempenho fenotípico individuais de inúmeras plantas em todas as localidades onde os

pesquisadores tinham conhecimento de plantios (HAMILTON e FUKUNAGA, 1959). Este trabalho sistemático de avaliação foi essencial para o desenvolvimento da cultura.

Em 1935-1936 a primeira seleção realizada pela HAES foi alcançado (HAMILTON, STOREY e FUKUNAGA, 1952). Este material era morfológicamente classificado como *M. integrifolia* e após o desenvolvimento de técnicas de propagação satisfatórias que garantiam a manutenção das características selecionadas, as primeiras variedades foram lançadas em 1948 (HAMILTON e FUKUNAGA, 1959; PEACE, 2002).

Este primeiro lote de variedades foi selecionado a partir de material selvagem e com apenas uma geração de seleção apresentou grandes melhorias na produção e qualidade das nozes, o que permitiu a rápida expansão da cultura e da indústria beneficiadora de nozes no Havaí (HAMILTON e FUKUNAGA, 1959). Em 1960 estas variedades já eram plantadas em diversos países, inclusive na Austrália. Os trabalhos de seleção genética continuaram na HAES com sucessivos lançamentos de variedades até recentemente, sendo este conjunto de material genético um dos mais plantados no mundo.

As principais variedades lançadas no Havaí são:

1935 – Primeira geração – Keauhou (HAES 246), HAES 294, HAES 344, HAES 508.

1948 – Segunda geração – Ikaika (HAES 333), Wailua (HAES 475) (HAMILTON, STOREY e FUKUNAGA, 1952); HAES 660, HAES 741, HAES 762, HAES 772, HAES 781, HAES 783, HAES 789, HAES 790, HAES 791, HAES 792, HAES 794, HAES 800, HAES 804, HAES 807, HAES 814, HAES 816, HAES 828, HAES 835, HAES 836, HAES 837, HAES 842, HAES 849, HAES 856.

1954 – Materiais selecionados da introdução, sem melhoramento, só adaptação – HAES 705.

*Estados Unidos – Califórnia

As duas espécies de macadâmia comestíveis também foram introduzidas na Califórnia (Estados Unidos) em 1880, no mesmo período da introdução realizada no Havaí, mas sem o mesmo trabalho de seleção (STOREY, 1963, citado por PEACE, 2002), o que não resultou no mesmo sucesso havaiano. Desta forma, poucos trabalhos foram desenvolvidos e não houve grande crescimento nos pomares desta região.

As principais variedades lançadas na Califórnia são: Faulkner, Queen Anne, Elimba (esta última desenvolvida a partir de *M. tetraphylla* é citada como implantada no Estado da Flórida/USA – MALO e CAMPBELL, 1972).

*Austrália

Seguindo o modelo de estudo havaiano, a Austrália iniciou as avaliações nos primeiros pomares estabelecidos. As primeiras variedades foram lançadas entre 1948 e 1952, porém estas foram na sua maioria perdidas devido à falta de técnicas de propagação que permitisse a manutenção das características selecionadas (PEACE, 2002). Porém, após alguns anos de estudos os pesquisadores adotaram métodos de seleção mais robustos, atualmente realizando a seleção com base no *pedigree* e na informação do desempenho entre progênies utilizando o método *Best Linear Unbiased Prediction* (BLUP)(HARDNER e McCONCHIE, 1999).

Boyton, Hardner e Kristiansen (1999) relatam que entre os anos de 1993 e 1996 foram implantados testes de progênies provenientes de polinização aberta e de cruzamentos entre variedades. No total foram implantadas 2.061 árvores, em três localidades (Bundaberg, Urbenville e Astonville). Em 1996 o *Commonwealth Scientific and Industrial Research Organisation* (CSIRO) estabeleceu um projeto de melhoramento genético para a espécie. Foram utilizados 48 parentais em cruzamentos controlados entre os anos de 1995 a 1999 (BOYTON, HARDNER e KRISTIENSEN, 1999). Entretanto, não foram encontradas informações a respeito do desenvolvimento de variedades comerciais oriundas destes experimentos.

As principais variedades lançadas na Austrália são:

Anterior à década de 60 – Daddow, Heilscher, Kopp, Mason 97, NG 13, NG 18, Own Choice, Own Venture, Release, Seedling 51, Renown, Teddington.

1954 – Beaumont, Fernleigh Special, Greber Hybrid, Hybrid Rankin, L35, L65, L68, NG3, NG4, NG7, NG8, NG29, NG35, NRG43.

1968 – Probert.

1980 – A4.

1981 – A16, A29, A38, A90, A104, A199.

1984 – A203, A232.

1986 – A268.

*África do Sul

A introdução na África do Sul ocorreu em duas fases: 1931 a partir de sementes obtidas no Havaí; e, 1935 com a importação de sementes australianas (PEACE, 2002).

Estudos de adaptação e seleção foram iniciados em 1966, porém os produtores continuavam a preferir o material importado. Somente a partir de década de 70 os materiais selecionados no país tornaram-se relevantes e continuam a ser desenvolvidos até hoje.

As principais variedades lançadas na África do Sul são: 741s, 741u, Reim'sInt1, Reim'sInt 2, UNP-F1, UNP-F4, UNP-F7.

*Brasil

As primeiras sementes de macadâmia foram introduzidas no Brasil na década de 30 pela família Dierberger, na cidade de Limeira/SP (DIERBERGER e MARINO NETO, 1981). Porém, o programa de domesticação e melhoramento genético de macadâmia teve início somente na década de 40 no Instituto Agrônomo (IAC) com a introdução de sementes de variedades desenvolvidas no Havaí. A seleção das primeiras variedades seguiu o padrão de avaliação do desempenho individual, assim como o método utilizado no Havaí e nas primeiras seleções australianas. Desde então foram selecionados 16 variedades por sua produtividade e qualidade de nozes (PIZA *et al.*, 2006).

Atualmente as variedades desenvolvidas no Brasil representam somente 30% dos pomares em produção (O'CONNOR, 2000). Essa baixa porcentagem possivelmente ocorre porque o último lançamento de variedades ocorreu em 1981, sendo a demanda por novos materiais uma prioridade dos produtores durante as reuniões da Associação Brasileira de Noz-Macadâmia (ABM). Assim, o programa de melhoramento genético de macadâmia do Instituto Agrônomo (IAC) foi reativado e reestruturado visando atender às necessidades do setor.

As principais variedades lançadas no Brasil são:

1970 – Primeira geração IAC – Waiasol (IAC 4-12) – progênie de Waialua (HAES 475 ou 36-475; HAMILTON, STOREY e FUKUNAGA, 1952); Kakere (IAC 5-10) e Kakedo (IAC 4-10) – progênies de Kakea (HAES 508); Keaufa (IAC 4-21), Keami (IAC 4-20) e Keaudo (IAC 2-23) – progênies de Keauhou (HAES 246) (OJIMA, CAMPO-DALL'ORTO e RIGITANO, 1989). IAC 1-21 e Keure (IAC 4-18) – progênies de Keauhou (HAES 246); IAC 1-11 – progênie de Kakea (HAES 508).

1981 – Segunda geração IAC – IAC Ti 4 - 12 B, IAC 4-12 B, IAC 3-4 C, IAC 2-23 C, IAC 2-23 A, IAC 1-21 A, Campinas – B, Campinas – A; selecionadas entre 935 *seedlings* provenientes de sementes de 21 variedades promissoras (VEIGA, OJIMA e RIGITANO, 1981).

Sem data – Selecionada pela família Dierberger – Aloha.

A fase atual do programa de melhoramento genético desenvolvido pelo IAC está estruturada com base em estudos realizados em outras espécies florestais, como por exemplo, o eucalipto. Em 2008 foi implantado no Centro APTA-Frutas (Jundiaí/SP) um teste de progênies, instalado em delineamento estatístico que proporcionará estimativas dos parâmetros genéticos, e assim reduzir os riscos de obtenção de progressos de seleção inferiores aos esperados. Porém, o grande gargalo do programa de melhoramento genético do IAC é o baixo número de parentais disponíveis.

Segundo Hardner e McConchie (1999) o uso de um grande número de parentais nas fases iniciais do programa de melhoramento fornece vantagens, tais como, inclusão de alelos favoráveis que se apresentam em baixa frequência, assim como reduz a possibilidade de perda destes alelos sob média e alta intensidade de seleção; melhora a estimativa da resposta à seleção, visto que a resposta à seleção em pequenas populações tem maior variância; reduz a taxa de parentesco reduzindo a relação de endogamia entre o material; e, aumenta a chance de identificação de parentais elites para compor os cruzamentos. Os mesmos autores estimam que um número de 50 parentais com 20 progênies por família seria necessário para manter a diversidade genética em programas de melhoramento genético.

Assim, fica evidente a necessidade de intercâmbio e importação de material genético de outros países, a fim de aumentar a variabilidade genética para seleção de variedades elite e a complementação do banco de germoplasma para a conservação destas variedades.

2.1.5 Parâmetros genéticos

A partir do estudo envolvendo 40 cultivares clonadas, Hardner *et al.* (2001) estimaram a herdabilidade e as correlações genéticas e fenotípicas de caracteres importantes na produção de nozes em macadâmia, tais como, peso de nozes, peso de amêndoas, taxa de recuperação de amêndoas (porcentagem do que é convertido de nozes em amêndoas), porcentagem de amêndoas com menos de $1,0 \text{ g/cm}^3$ (amêndoas com menos de $1,0 \text{ g/cm}^3$ geralmente possuem teor de óleo superior à 72% e consideradas como sendo de boa qualidade), e porcentagem de amêndoas inteiras.

Os resultados mostraram a partir dos valores estimados que os caracteres peso de nozes, peso de amêndoas e taxa de recuperação estão sob forte controle genético com herdabilidades de 0,63 para os três caracteres (HARDNER *et al.*, 2001). Os outros caracteres estudados apresentaram baixa herdabilidade (porcentagem de amêndoas com menos de

1,0g/cm³, $h^2 = 0,18$; porcentagem de amêndoas inteiras, $h^2 = 0,31$), sugerindo que estes estejam sob maior influência ambiental.

Os estudos de correlação genética entre caracteres mostram forte correlação entre peso de nozes e amêndoas ($r_g = 0,79$), sugerindo, segundo os autores, que genes similares estejam envolvidos no controle dos caracteres (HARDNER *et al.*, 2001). Para os demais caracteres foram obtidas correlações de magnitude intermediária, porém significativas (entre peso de amêndoa e taxa de recuperação: 0,48; peso de nozes e teor de óleo: 0,40; peso de amêndoa e teor de óleo: 0,41), e baixos e não significativos (peso de nozes e taxa de recuperação: -0,17; taxa de recuperação e teor de óleo: 0,13; e porcentagem de amêndoas inteiras com todos os demais caracteres).

A correlação entre a produção em diferentes idades em um mesmo indivíduo mostra que as avaliações realizadas aos seis anos tem grande correspondência às avaliações mais tardias ($r_g > 0,70$), porém inferior ao encontrado aos sete anos ($r_g > 0,90$) (HARDNER e McCONCHIE, 1999). Estes resultados mostram a possibilidade de realização de seleção precoce, sem que haja grande viés nas estimativas dos ganhos esperados na seleção.

2.1.6 Mapeamento genético

Um mapa de ligação genética de macadâmia disponível foi construído a partir de 56 indivíduos F₁ provenientes do cruzamento entre as variedades Keauhou e A16 (PEACE *et al.*, 2003). A variedade Keauhou tem origem genética em *M. integrifolia*, enquanto que a variedade A16 é um híbrido entre *M. integrifolia* e *M. tetraphylla* (PEACE *et al.*, 2002; PEACE *et al.*, 2005).

A estratégia utilizada foi a do pseudo-testcross (PEACE *et al.*, 2003), que consiste na construção, em separado, de dois mapas, um para cada parental e posterior detecção de grupos homólogos (GRATTAPAGLIA e SEDEROFF, 1994). Foram utilizadas 265 marcas moleculares, sendo: 259 *Randomly Amplified DNA Fingerprinting* (RAF), cinco *Random Amplified Polymorphic DNA* (RAPD), e uma *Sequence-Tagged Microsatellite Site* (STMS). Das 259 marcas RAF, 16 apresentaram herança codominante (PEACE *et al.*, 2003).

O mapa apresentou 11 grupos de ligação com marcadores comuns a ambos os genitores, seis grupos de ligação formados a partir de marcas que segregaram no parental Keauhou, e sete grupos de ligação com marcas que segregaram no parental A16 (PEACE *et al.*, 2003). O mapa referente ao genoma parental da variedade Keauhou formou 17 grupos de

ligação, com o comprimento de 570 cM, tendo a cobertura do genoma estimado entre 48 e 53%. Considerando as marcas que segregaram no parental A16 foi possível formar 18 grupos de ligação, com o comprimento de 920 cM, e cobertura do genoma estimada entre 61 e 69% (PEACE *et al.*, 2003). Provavelmente, melhores resultados seriam obtidos caso um maior número de progênies estivesse envolvido no estudo. Um número maior de indivíduos envolvidos no cruzamento aumentaria a probabilidade de observação de recombinantes, e desta forma melhor identificar marcas que segregam conjuntamente.

2.2 Marcadores moleculares

Marcadores genéticos são aqueles que mostram segregação mendeliana, ou seja, possuem mecanismos de herança simples que podem ter ação dominante ou codominante (BERED *et al.*, 1997; FUNGARO e VIEIRA, 2001; BORÉM e VIEIRA, 2005). Os primeiros estudos de genética e melhoramento genético utilizaram os marcadores morfológicos, baseados no polimorfismo da anatomia de determinados órgãos e estruturas vegetais (formato de folhas, ausência de espinhos, cor de flores e rugosidade da semente, entre outros), como marcadores genéticos (FERREIRA e GRATTAPAGLIA, 1998; FUNGARO e VIEIRA, 2001). Alguns desses polimorfismos apresentam variação discreta típica de características qualitativas, porém, muitas destas características possam estar sob influência ambiental (APPELS *et al.*, 1998). Outra restrição ao uso dos marcadores morfológicos está na não dispersão uniforme ao longo do genoma e sua baixa capacidade de detecção de polimorfismo. Também são utilizados os marcadores citológicos (constrições, bandas, satélites e knobs) como marcadores genéticos, e estes permitiram acompanhar as segregações de um cromossomo ao longo das gerações (FUNGARO e VIEIRA, 2001).

Os marcadores morfológicos e citológicos tiveram grande importância nos trabalhos em genética, contudo mostraram-se inapropriados na elucidação de diversos estudos genéticos, por exemplo, na probabilidade de se encontrar associações significativas entre estes marcadores e caracteres de importância econômica (FERREIRA e GRATTAPAGLIA, 1998; FUNGARO e VIEIRA, 2001). As desvantagens segundo Tanksley (1993), Bered *et al.* (1997) e Ferreira e Grattapaglia (1998) na utilização dos marcadores morfológicos incluem:

- * Baixo número de classes genéticas entre linhagens;
- * Efeitos deletérios das mutações que limitaram sua utilização para a geração de mapas genéticos e no melhoramento genético;

- * Grande efeito dos genes determinantes de marcadores morfológicos pode afetar a análise genética dos caracteres de importância econômica;
- * Poucos caracteres podem ser estudados simultaneamente devido aos efeitos das interações gênicas (como a epistasia, por exemplo);
- * Não são seletivamente neutros, ou seja, a influência que o ambiente pode causar na expressão dos genes que determinam os marcadores morfológicos;
- * Podem apresentar efeitos de dominância e epistasia.

Com os avanços tecnológicos, principalmente com o desenvolvimento da *Polymerase Chain Reaction* - PCR, por Mullis em 1987, diversas técnicas de análises em biologia molecular foram desenvolvidas. A principal diferença foi a direta avaliação da variação no DNA, sem mais haver a necessidade de avaliação do genótipo a partir do fenótipo (FERREIRA e GRATTAPAGLIA, 1998), ou seja, sem a influência ambiental (TANKSLEY, 1993). Estes marcadores, denominados moleculares, trouxeram grande avanço nas pesquisas e reduziram as limitações nos estudos genéticos (FUNGARO e VIEIRA, 2001).

Existe atualmente um grande número de marcadores moleculares, que incluem os protéicos (por exemplo, as isoenzimas) e os de DNA (por exemplo, *Random Amplified Polymorphic DNA* - RAPD, *Simple Sequence Repeats* - SSR, *Single Nucleotide Polymorphism* - SNP e *Diversity Arrays Technology* - DArT), que forneçam um alto grau de identificação do polimorfismo existente entre os organismos em estudo (BERED *et al.*, 1997; FUNGARO e VIEIRA, 2001).

Os avanços tecnológicos são constantes e o desenvolvimento de novos marcadores moleculares demonstra a melhoria contínua de como a variabilidade genética está sendo estimada. Pode-se inferir que as técnicas mais recentes de marcadores moleculares são aquelas mais utilizadas e que nos proporcionam um maior número de informações (Figura 2)(SCHLÖTTERER, 2004).

Os marcadores moleculares podem diferir entre si devido ao tipo de herança destes, sendo classificados como dominantes (quando a presença do genótipo heterozigoto não pode ser observado, sendo incluso nas classes dos genótipos dos homozigotos dominantes), ou como codominantes, (quando o genótipo do heterozigoto pode ser discriminado em uma classe a parte). Esta característica é extremamente importante na escolha de qual marcador utilizar, visto que, dependendo do tipo de estudo, pode ou não ser necessário identificar os genótipos heterozigotos.

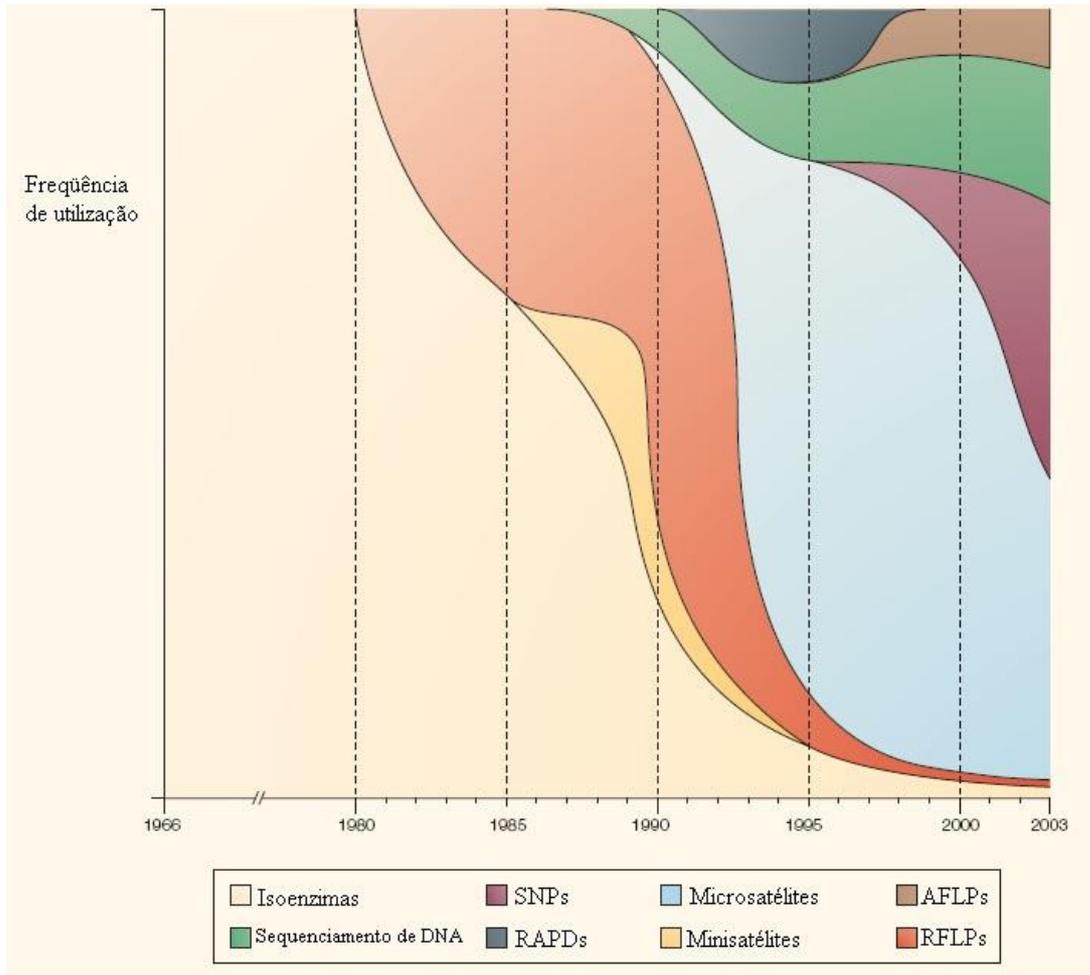


Figura 2 - Importância relativa dos diferentes marcadores moleculares. O eixo horizontal indica o tempo e o vertical corresponde à frequência de utilização de marcadores moleculares. Fonte: Schlötterer, 2004

2.2.1 Marcadores microsatélites - SSR

Os marcadores Microsatélite (WEBER e MAY, 1989), ou SSR, são sequências repetidas em tandem encontradas abundantemente no genoma de eucariotos, flanqueadas por sequências únicas denominadas *primers* (McCOUCH *et al.*, 1997). Os SSR estão distribuídos aleatoriamente no genoma, possibilitando uma ampla cobertura e são geralmente considerados neutros evolutivamente. Porém, nos últimos anos foram encontradas regiões contendo SSR em regiões de expressão (*Expressed Sequence Tag* - EST) no genoma, possibilitando a avaliação de locos próximos às regiões envolvidas com a expressão de caracteres. Esta nova classe de marcadores SSR localizados em regiões de expressão recebem o nome de SSR-EST (OLIVEIRA *et al.*, 2007). Estes marcadores, por estarem em regiões de alta pressão evolutiva, não podem ser considerados neutros e por isso não são indicados, por exemplo, em estudos

filogenéticos, porém são uma ótima opção em estudos de mapeamento de *Quantitative Trait Loci* - (QTL), entre outros.

Os SSR podem apresentar um padrão simples, quando não há a alteração do motivo (por exemplo, (CT)₁₉ - Figura 3), ou composto, quando o SSR é formado por dois ou mais conjuntos de motivos (por exemplo, (CG)₈(CT)₈). Outra classificação que pode ser realizada é quanto a interrupção do motivo, sendo os perfeitos formados sem interrupção e os imperfeitos apresentando uma porção não repetido entre duas regiões repetitivas (por exemplo, (CT)₈TA(CA)₁₀)(Figura 4). Estas sequências podem ser formadas por dois nucleotídeos, ou mais, sendo classificados como dinucleotídeos quando possuem dois nucleotídeos na sequência, trinucleotídeos na presença de três nucleotídeos, tetranucleotídeos quando o motivo é composto por quatro nucleotídeos, e assim, sucessivamente (por exemplo, (CT)₁₉, (TGC)₃, (TACC)₃, respectivamente).

A região do genoma contendo o SSR é amplificada utilizando-se um par de *primers* específico (ou testada a sua transferibilidade entre organismos cujo sítio de anelamento seja conservado – *primers* heterólogos) via PCR (FERREIRA e GRATTAPAGLIA, 1998). A visualização dos fragmentos pode ocorrer em gel de eletroforese de agarose ou acrilamida, assim como através de radiografias e *primers* marcados com radioisótopos (FERREIRA e GRATTAPAGLIA, 1998), ou diretamente utilizando sequenciadores e *primers* marcados com fluorescência.

Cada região SSR pode ser considerada um loco, altamente variável, o que fornece a partir do alto número de alelos, uma grande fonte de informação a respeito do genoma em estudo. Esta variação genética, o polimorfismo entre indivíduos, é detectada a partir de variações no número de repetições amplificadas de um determinado motivo SSR.

2.2.2 Marcadores diversity array technology - DArT

A *Diversity Array Technology* - DArT (KILIAN *et al.*, 2005) é uma técnica molecular que busca mensurar simultaneamente um grande número de variações genéticas utilizando plataformas de microarranjos para a análise do polimorfismo de DNA (JACCOUD *et al.*, 2001; KILIAN *et al.*, 2005). Segundo JACCOUD *et al.* (2001) a técnica detecta a alteração de pares de bases localizadas nos sítios das enzimas de restrição, embora essa tecnologia seja capaz de encontrar inserções, deleções e inversões no DNA, que são ocorrências mais frequentes em genomas humanos e de plantas do que as próprias substituições de base.

Os marcadores DArT utilizam a redução de complexidade do genoma, via enzimas de restrição, e a hibridização deste ao genoma em estudo para identificação do polimorfismo (KILIAN *et al.*, 2005). Esta técnica, inicialmente desenvolvida para a espécie *Oryza sativa* (KILIAN *et al.*, 2005), atualmente está sendo empregada em diversas espécies diplóides e poliplóides (AKBARI *et al.*, 2006), inclusive em estudos com espécies de fungos patogênicos (KILIAN *et al.*, 2005).

O desenvolvimento dos DArT requer a união de um *pool* gênico com a maior diversidade possível, para formação de um painel e a posterior clivagem do DNA utilizando-se enzimas de restrição de corte raro e restrito (KILIAN *et al.*, 2005; SANSALONI *et al.*, 2010). Com estes fragmentos, denominados DNA da representação, são construídas bibliotecas e posteriormente formam o microarranjo (Figura 5). O DNA da amostra é marcado com fluorescência e posto para hibridização sobre o DNA da representação (JACCOUD *et al.*, 2001; KILIAN *et al.*, 2005; SANSALONI *et al.*, 2010). As amostras marcadas são identificadas a partir da leitura em equipamentos apropriados.

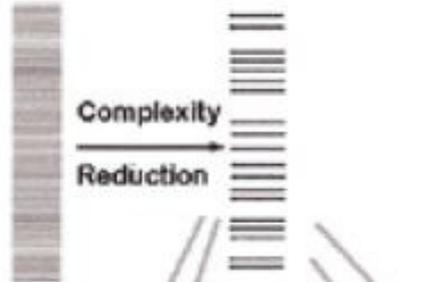
(A)

(B)

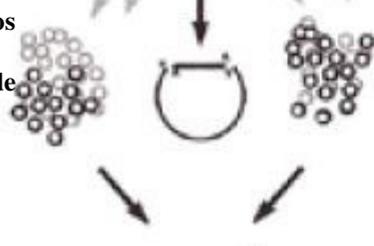
- Formação do painel de diversidade



- Redução da complexidade



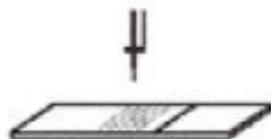
- Inserção dos fragmentos no vetor de clonagem



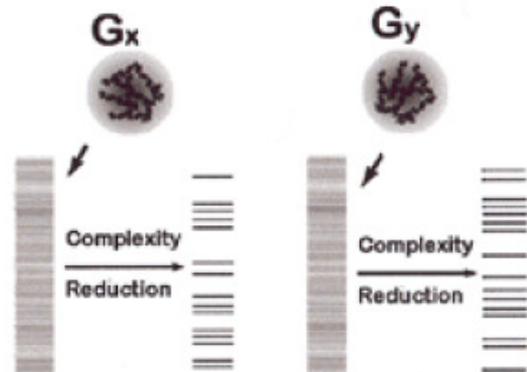
- Seleção das colônias transformadas



- Impressão do microarranjo



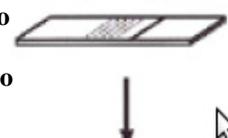
- DNA das variedades para avaliação



- Marcação com as respectivas fluorescências



- Hibridização sobre o microarranjo



- Visualização da hibridização



Figura 5 - Esquema representativo do desenvolvimento e genotipagem a partir dos marcadores *Diversity Array Technology* - DArT. A – formação do painel de diversidade, redução da complexidade a partir da clivagem do DNA pelas enzimas de restrição, inserção dos fragmentos no vetor de clonagem, seleção das colônias transformadas e impressão do microarranjo. B – genotipagem dos genótipos G_x e G_y , clivagem do DNA, marcação de cada DNA com a respectiva fluorescência, hibridização sobre o microarranjo previamente impresso e visualização da hibridização via *scanner*. Fonte: Jaccoud *et al.*, 2001

Cada ponto do microarranjo contendo um DNA da representação pode ser considerado um loco, por estes apresentarem baixa frequência de cópias no DNA (SANSALONI *et al.*, 2010). Estudos têm demonstrado que apenas 2% das marcas DArT possam possuir mais de uma localização no mesmo genoma (AKBARI *et al.*, 2006).

O polimorfismo é detectado diretamente do genoma e por não haver a necessidade de uso de *primers* para a amplificação dos fragmentos, que podem anelar em regiões homólogas erroneamente, a possibilidade de avaliação equivocada por duas regiões homólogas serem amplificadas e hibridizadas cruzadamente durante o arranjo é reduzida (AKBARI *et al.*, 2006). Outro grande avanço proporcionado por essa técnica é a facilidade de detecção de variações no genoma, sem a necessidade de conhecimento prévio das sequências em análise (WENZL *et al.*, 2004).

A detecção da variação é baseada na presença (ou ausência) de hibridização de um fragmento específico ao painel de representação (JACCOUD *et al.*, 2001), podendo ser assim caracterizado como um marcador dominante.

Uma das aplicações dos marcadores DArT é na avaliação da diversidade entre variedades. Em seu trabalho introdutório da técnica, Jaccoud *et al.* (2001) avaliaram nove cultivares de arroz, que a partir dos dados binários, calcularam as respectivas distâncias genéticas entre os materiais. As cultivares foram distintas quanto às classes *Indica* e *Japonica* e os materiais relacionados formaram grupos próximos. A *Oryza sativa* foi escolhida por tratar-se de uma espécie modelo, com um relativo pequeno genoma (430 Mpb) e com muitos estudos a respeito da origem genética de suas cultivares, fato que facilitou a comprovação dos resultados obtidos pelos marcadores DArT (KILIAN *et al.*, 2005).

Trinta e três cultivares e dois acessos selvagens de cevada foram avaliados por 383 marcas polimórficas, pelas quais foi possível identificar as relações de parentesco entre estas (WENZL *et al.*, 2004). Como esperado pelos autores, os dois acessos selvagens foram discriminados das cultivares, assim como duas cultivares desenvolvidas no leste asiático. As demais formaram um único grande agrupamento. Wenzl *et al.* (2004) incluíram diferentes amostras de mesmas cultivares, porém provenientes de diferentes instituições, e estas foram alocadas em grupos próximos, mas não identificados como idênticos. Suspeita-se que este fato ocorreu devido a alterações na metilação do DNA (que pode proporcionar pequenas alterações do sítio da enzima de restrição utilizada), que podem ocorrer devido às diferentes condições ambientais onde as cultivares foram implantadas (WENZL *et al.*, 2004). Os autores citam que estas mesmas diferenças foram encontradas quando utilizado o marcador *Restriction Fragment Length Polymorphism* (RFLP) em estudos com as mesmas cultivares.

Os marcadores DArT foram empregados por Akbari *et al.* (2006) em 62 cultivares hexaplóides de *Triticumaestivum* (trigo), incluindo três possíveis replicatas. A relação de genética entre as cultivares estudadas utilizando os marcadores DArT coincidiu com as informações obtidas por outros marcadores, inclusive com a distinção entre cultivares muito relacionadas (AKBARI *et al.*, 2006). Houve uma clara separação entre as cultivares desenvolvidas na Austrália e na Europa, sendo os materiais australianos mais divergentes do que os europeus. Este estudo deve ser considerado importante também por demonstrar a utilização destes marcadores em espécies e/ou variedades poliplóides.

A diversidade genética entre organismos não modelo também foi avaliada utilizando os marcadores DArT como ferramenta. Um exemplo foi o estudo desenvolvido por James *et al.* (2008) conduzido com 16 indivíduos do gênero *Asplenium* (14 de *A. viride*, um de *A. trichomanes* – todos diplóides; e, um *A. Adulterinum* – alotetraplóide; coletados em diversos locais da Europa e do norte da África), 15 indivíduos de *Garovaglia elegans* (coletados em diferentes locais da Austrália e Nova Guiné), e um de *G. powellii*. As análises realizadas com o material de *Asplenium* mostraram a formação de um grande grupo contendo os indivíduos de *A. viride* dividido em quatro subgrupos relativos às regiões de coleta do material, enquanto que o gênero *A. trichomanes* formou um grupo externo às espécies e o *A. adulterinum* ocupou uma posição intermediária, conforme esperado, visto que se trata de um alopólóide formado pela combinação dos genomas de *A. viridex A. trichomanes* (JAMES *et al.*, 2008).

Resultado semelhante fora encontrado com as espécies do gênero *Garovaglia*. Os espécimes foram agrupados conforme as respectivas espécies, sendo que o material de *G. elegans* coletados na Austrália e na Nova Guiné foi discriminado, enquanto que a espécie *G. powellii* formou um grupo externo às demais (JAMES *et al.*, 2008). Devido ao grande poder de discriminação dos marcadores DArT, os autores investiram no seu uso em estudo de reconstituição da história evolucionária destas espécies, com resultados que confirmaram outras técnicas e corroboraram para a elucidação de pontos anteriormente confusos.

3 MATERIAL E MÉTODOS

3.1 Material vegetal

As 35 variedades de macadâmia avaliadas neste estudo foram coletadas no Brasil (banco de germoplasma do Instituto Agrônomo (IAC), em Jundiaí/SP, e em produtores do município de Dois Córregos/SP), e na Austrália. Também foram acrescentadas espécies do terceiro *genepool* da macadâmia (*M. ternifolia* e *M. jansonii*) e dois indivíduos das espécies *Grevilea robusta* e *G. banksii* pertencentes à mesma família botânica (Proteacea).

A extração de DNA das variedades localizadas no Brasil e das duas espécies de grevilea foram conduzidas a partir de folhas jovens liofilizadas conforme protocolo estabelecido por PEACE (2002), enquanto que as variedades coletadas na Austrália tiveram seu DNA extraído de folhas frescas conforme protocolo estabelecido pela empresa DArT Technology (<http://www.diversityarrays.com/samplesub.html>).

As variedades avaliadas estão citadas na tabela 1.

Tabela 1 – Informações sobre o material genotipados no presente estudo com a designação do tipo botânico, marcador genético utilizado e local de coleta

(continua)

Espécie	Variedade	Tipo botânico	Marcador utilizado	Local de coleta
<i>M. integrifolia</i>	IAC 4-20	Genótipo típico	SSR e DArT	Dois Córregos/SP
	IAC 4-20	Fruto menor	SSR e DArT	Dois Córregos/SP
	IAC 4-20	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC Campinas B	Genótipo típico	SSR e DArT	Dois Córregos/SP
	IAC Campinas B	Arquitetura de copa aberta	SSR e DArT	Dois Córregos/SP
	IAC Campinas B	Arquitetura de copa fechada	SSR e DArT	Dois Córregos/SP
	IAC 4-10	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 9-20	Folha maior	SSR e DArT	Dois Córregos/SP
	IAC 9-20	Folha menor	SSR e DArT	Dois Córregos/SP
	IAC 9-20	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 11-8	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 1-21	Genótipo típico	SSR e DArT	Dois Córregos/SP

Tabela 1 – Informações sobre o material genotipados no presente estudo com a designação do tipo botânico, marcador genético utilizado e local de coleta

(conclusão)

Espécie	Variedade	Tipo botânico	Marcador utilizado	Local de coleta
	IAC 8-17	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 2-23	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 4-8	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 11-18	BAG IAC-Nozes	SSR e DArT	Jundiaí/SP
	IAC 4-12B	Genótipo típico	SSR e DArT	Dois Córregos/SP
	Aloha	Genótipo típico	SSR e DArT	Dois Córregos/SP
	Flor Roxa	Genótipo típico	SSR e DArT	Dois Córregos/SP
	HAES 8-16	Genótipo típico	SSR e DArT	Dois Córregos/SP
	HAES Ikaika	Genótipo típico	SSR e DArT	Dois Córregos/SP
	HAES 3-44	Genótipo típico	SSR e DArT	Dois Córregos/SP
	HAES 7-41	Genótipo típico	SSR e DArT	Dois Córregos/SP
	HAES 6-60	Genótipo típico	SSR e DArT	Dois Córregos/SP
	816	Sem informação	SSR e DArT	Austrália
	246	Sem informação	SSR e DArT	Austrália
	344	Sem informação	SSR e DArT	Austrália
	1A-4	Genótipo típico	SSR e DArT	Dois Córregos/SP
	2A-16	Genótipo típico	SSR e DArT	Dois Córregos/SP
	A16	Sem informação	DArT	Austrália
	3A-29	Genótipo típico	SSR e DArT	Dois Córregos/SP
	4A-38	Genótipo típico	SSR	Dois Córregos/SP
	Beaumont	Sem informação	DArT	Austrália
	Daddow	Sem informação	DArT	Austrália
	Heilscher	Sem informação	DArT	Austrália
<i>M. ternifolia</i>		Selvagem	DArT	Austrália
<i>M. janseni</i>		Selvagem	DArT	Austrália
<i>G. robusta</i>		Arborização de parques	SSR	São Paulo/SP
<i>G. banksii</i>		Arborização de parques	SSR	São Paulo/SP

Os diferentes tipos botânicos foram coletados em diferentes locais, porém todos com a mesma nomenclatura. Incluiu-se estes indivíduos para verificar possíveis erros de identificação por parte dos viveiristas que comercializam estas mudas enxertadas.

3.2 Desenvolvimento dos marcadores microssatélites - SSR

O desenvolvimento da biblioteca genômica enriquecida para dinucleotídeos (AG)₁₅ e (AC)₁₅, necessária para a obtenção dos SSR, foi realizado durante a disciplina NG 256 Construção de Bibliotecas Enriquecidas em Microssatélites de Eucariotos, ministrada pela Profa. Dra. Anete Pereira de Souza, na Universidade Estadual de Campinas (UNICAMP).

Para a construção desta biblioteca foi utilizado 6.500ng de DNA de um único genótipo (variedade IAC 9-20) e realizada seguindo as etapas de digestão do DNA genômico com a enzima *RsaI*; ligação dos adaptadores, amplificação, purificação e seleção dos fragmentos contendo microssatélites; amplificação dos fragmentos selecionados; ligação dos fragmentos a um vetor de clonagem; amplificação dos insertos clonados; e, isolamento do fragmento para posterior sequenciamento (BILLOTTE *et al.*, 1999).

Foram sequenciados 96 clones utilizando os *primers* SP₆ e T₇ *promoter*, obtendo-se assim 192 sequências. Com a obtenção das sequências, o próximo passo para o desenvolvimento dos *primers* foi a realização da edição e análise dessas para identificar quais apresentam regiões com microssatélites e obter os *primers* flanqueadores.

Inicialmente as sequências foram editadas pelo programa EditSeq (DNA STAR, 2000) para que fossem identificados possíveis erros de leitura nas sequências e realizadas as suas correções. Após análise e seleção das sequências por sua qualidade e presença de regiões microssatélites, estas foram alinhadas pelo programa SeqMan (DNA STAR, 2000) buscando a formação de clusters de sequências complementares. A remoção dos adaptadores foi realizada pelo programa Microsat (RISTERRUCCI *et al.*, 2005), assim como também a identificação dos sítios da enzima de restrição *Rsa I*. Para a busca por microssatélites nos clusters, foi utilizada a ferramenta *Simple Sequence Repeat Identification Tool* – SSRIT (disponível em www.gramene.org). Foram considerados como parâmetros de seleção dos microssatélites que estes apresentem no mínimo 10 nucleotídeos e/ou motivos com três repetições.

O desenho dos *primers* foi realizado a partir do programa *Primer Select* (DNA STAR, 2001). Estabeleceu-se como critérios que os *primers* devessem ter entre 18 e 22 pares de bases (pb), diferença de temperatura de anelamento entre os *primers forward* e *reverse* de no máximo 3,0 °C, tamanho do produto de PCR entre 200 e 250 pb e que não apresentem a formação de dímeros e *hairpins*.

Após a síntese dos *primers* selecionados, estes foram testados em cinco indivíduos da amostra (incluindo variedades brasileiras, havaianas e australianas) para constatar o

polimorfismo do loco e definir as condições ideais de amplificação e de eletroforese. Todos os pares de *primers* foram amplificados via PCR utilizando os programas *Touch Down* (DON *et al.*, 1991) e gradiente para definir a temperatura ótima de anelamento destes ao DNA da amostra. As reações de amplificação via PCR foram realizadas com volume final de 25 μL contendo 10 ng de DNA da amostra, 0,8 μM de cada *primer forward* e *reverse*, 100 μM de cada dNTP (MBI Fermentas, MD, USA), 1,5 mM de MgCl_2 , 10 mM de Tris-HCl, 50 mM de KCl, e 0,5 U Taq DNA Polimerase (Invitrogen, CA, USA). As reações de PCR foram realizadas conforme seguintes condições: 94°C por 1 minuto, seguido por 30 ciclos de 94°C por 1 minuto, T_a específica (entre 45 e 60 °C) por 1 minuto, 72°C por 1 minuto, e extensão final a 72°C por 5 minutos.

Quando necessário, realizou-se PCR com temperatura fixa (60 °C), também com o intuito de otimizar a temperatura de anelamento. O produto de PCR foi visualizado primeiramente em eletroforese horizontal com gel de agarose 3% e coloração com brometo de etídio, e posteriormente genotipados em gel de acrilamida 6% (eletroforese vertical). As eletroforeses e as colorações foram realizadas conforme protocolo estabelecido por Creste, Tulmann Neto e Figueira (2001).

Após a definição das condições de PCR para cada loco, estes foram utilizados para a genotipagem total da amostra. Os dados foram tabulados em uma planilha eletrônica para posterior análise, sendo observados os números de alelos contidos, assim como o peso em pares de base para cada alelo. Padronizou-se que o primeiro alelo do loco seria aquele que apresentasse o maior peso.

Os locos foram caracterizados conforme seu polimorfismo com base na porcentagem de locos polimórficos, no número de alelos por loco, na heterozigosidade esperada (H_E) e observada (H_O) e no *Polymorphic Index Content* (PIC), estimados conforme expressões abaixo.

A porcentagem de locos polimórficos (P) foi obtida por:

$$P = \frac{\text{n}^\circ \text{ de locos polimórficos}}{\text{n}^\circ \text{ total de locos}}$$

A estimativa das frequências alélicas (p_{ij}) foi obtida pela expressão:

$$p_{ij} = \frac{n_{ij}}{n_j}$$

Onde:

p_{ij} = frequência do alelo i na população j ;

n_{ij} = número de ocorrência do alelo i na população j ;

$n_{.j}$ = número total de alelos amostrados na população j .

A diversidade genética entre locos foi analisada pela heterozigidade observada (H_O) e esperada (H_E) segundo o Equilíbrio de Hardy-Weinberg conforme equações abaixo.

$$H_O = 1 - \sum_{i=1}^n p_{ii}$$

Em que, p_{ii} = frequência dos genótipos homozigotos.

$$H_E = 1 - \sum_{i=1}^n p_i^2$$

Em que, p_i^2 = frequência alélica estimada do i -ésimo alelo.

O *Polymorphic Index Content* (*PIC*), ou diversidade gênica, foi calculada conforme fórmula abaixo:

$$PIC = 1 - \sum_{i=1}^{n-1} f_i^2 - \sum_{j=i+1}^n 2 f_i^2 f_j^2$$

Onde: f_i^2 = frequência do i -ésimo alelo; e, f_j^2 = frequência do j -ésimo alelo.

Os genótipos analisados com este marcador foram: IAC 4-20, IAC 4-20 “a”, IAC 4-20 “c”, IAC Campinas B “a”, IAC Campinas B “b”, IAC Campinas B “f”, IAC 4-10, IAC 9-20, IAC 9-20 “d”, IAC 9-20 “e”, IAC 11-8, IAC 1-21, IAC 8-17, IAC 2-23, IAC 4-8, IAC 11-18, IAC 4-12B, Aloha, Flor Roxa, HAES 8-16, HAES Ikaika, HAES 3-44, HAES 7-41, HAES 6-60, 1A-4, 2A-16, 3A-29, 4A-38 e as espécies *Grevilea robusta* e *G. banksii*.

3.3 Desenvolvimento dos marcadores *diversity array technology*- DArT

O desenvolvimento dos marcadores DArT para macadâmia foi realizado em parceria com o Dr. Andrzej Kilian da empresa *Diversity Arrays Technology*, localizada na cidade de Canberra, Austrália.

O primeiro passo para o desenvolvimento dos marcadores DArT é a formação de um painel de diversidade com o maior número de genótipos divergentes para que sejam captadas as variações polimórficas e posterior redução da complexidade para gerar o genoma de representação (WENZL *et al.*, 2004). Para a formação do painel foi necessário 5ng de DNA de cada variedade, compondo no mínimo 100ng da mistura de DNAs (JACCOUD *et al.*, 2001; WENZL *et al.*, 2004). Para esta etapa do desenvolvimento foram utilizados os mesmos 30 genótipos da análise por SSR.

A redução da complexidade do genoma é baseada no uso de enzimas de restrição para a digestão do DNA total. Assim, à mistura de DNAs foram adicionadas as enzimas de restrição de corte raro (*Pst I*) e de corte frequente (*Taq I*) para clivagem do DNA e a formação de fragmentos que foram ligados a T₄ DNA ligase (WENZL *et al.*, 2004). Para verificar a adequada clivagem do DNA, estes foram visualizados em gel de agarose 1,2%, após eletroforese horizontal. Uma ideal digestão do DNA produz uma zona de fragmentos de diferentes tamanhos, sem a presença de bandas definidas (SANSALONI *et al.*, 2010).

O próximo passo para o desenvolvimento dos marcadores DArT é a construção de biblioteca genômica e o seu *screening* para comprovação do polimorfismo. Para isto, os fragmentos foram postos em contato com um vetor de clonagem (*Escherichia coli*), e selecionados para amplificação a partir dos plasmídeos que contiverem os fragmentos recombinantes, ou seja, as colônias brancas (JACCOUD *et al.*, 2001). Ressalta-se que cada placa de 384 poços contendo os fragmentos selecionados é considerada uma biblioteca.

O teste de polimorfismo das bibliotecas foi realizado em oito indivíduos. Além das bibliotecas desenvolvidas neste projeto, foram utilizadas outras seis construídas previamente pela equipe da empresa *Diversity Arrays Technology* para testes, totalizando 4.608 clones. Os fragmentos clonados, presentes nas bibliotecas, foram postos em microarranjo sobre uma lâmina de vidro. Todos os DNA receberam duas marcações fluorescentes (Cy-3 - de coloração verde e Cy-5 de coloração vermelha) para identificação da hibridização (JACCOUD *et al.*, 2001; SANSALONI *et al.*, 2010), e estes foram adicionados à

fluorescência azul (*blue fluorescently-labeled polylinker*), usado no controle da qualidade do microarranjo (SANSALONI *et al.*, 2010).

As placas contendo o material marcado foram postas para desnaturação a 96,0 °C por três minutos. O DNA marcado das variedades foi posto diretamente sobre o painel impresso nas lâminas de vidro e após hibridização *overnight* a 65 °C foi avaliado a partir do *scanner* Affymetrix 418 para análise da imagem. Os dados foram extraídos a partir do programa DArTSoft versão 7.44. Este programa verifica cada spots do microarranjo e compara as relativas intensidades das fluorescências, ou seja, azul *vs* verde e azul *vs* vermelho, e identifica pela intensidade do sinal a presença ou ausência de hibridização (SANSALONI *et al.*, 2010).

Após a validação da técnica de desenvolvimento dos marcadores DArT na amostra de oito indivíduos, o total da amostra fora genotipado seguindo os mesmos passos realizados nos testes. Para a genotipagem utilizando os marcadores DArT, além do material avaliado a partir dos marcadores SSR, foram somadas sete variedades (816, 246, 344, A16, Beaumont, Daddow e Heilscher) e dois indivíduos das espécies selvagens (*M. ternifolia* e *M. jansanii*). As duas espécies de grevilea utilizadas na genotipagem a partir dos marcadores SSR foram retiradas das análises com os marcadores DArT devido a má qualidade apresentada pelo DNA.

Os parâmetros selecionados para a avaliação da qualidade dos DArTs foram: correlação entre *spot* e *pixel*, interferência na leitura pelo sinal luminoso, correlação entre as duas fluorescências para uma mesma amostra, repetibilidade, polimorfismo.

Os genótipos avaliados pelos marcadores DArT foram: IAC 4-20, IAC 4-20 “a”, IAC 4-20 “c”, IAC Campinas B “a”, IAC Campinas B “b”, IAC Campinas B “f”, IAC 4-10, IAC 9-20, IAC 9-20 “d”, IAC 9-20 “e”, IAC 11-8, IAC 1-21, IAC 8-17, IAC 2-23, IAC 4-8, IAC 11-18, IAC 4-12B, Aloha, Flor Roxa, HAES 8-16, HAES Ikaika, HAES 3-44, HAES 7-41, HAES 6-60, 816, 246, 344, 1A-4, 2A-16, A16, 3A-29, Beaumont, Daddow e Heilscher, e as espécies *M. ternifolia* e *M. jansanii*.

3.4 Análise estatística

Os marcadores DArT, de herança dominante, foram utilizados para o cálculo da divergência genética pelo método de Jaccard, enquanto que os marcadores SSR, de herança codominante, foram utilizados para cálculo da divergência genética com base no método de

Roger Modificado. Ambos os marcadores foram utilizados na análise de estrutura de populações.

3.4.1 Estrutura genética

O primeiro passo para a o estudo da estrutura genética dos genótipos que compõem o germoplasma presente no Brasil foi a investigação da possível estruturação em subpopulações. A partir das frequências alélicas é possível estimar a probabilidade de cada genótipo ter sido originado de cada população e indivíduos sem informações de origem poder ser atribuídos às populações, ou subpopulações, conforme a probabilidade estimada (PRITCHARD, STEPHENS e DONNELLY, 2010).

O método escolhido neste trabalho foi o desenvolvido por Pritchard, Stephens e Donnelly (2010), implementado no programa *Structure* (PRITCHARD, WEN e FALUSH, 2009), baseado na locação dos indivíduos conforme o seu genótipo. Trata-se de método iterativo onde a probabilidade de um determinado indivíduo compor uma população é calculada simultaneamente à estimativa das frequências alélicas desta população (PRITCHARD, STEPHENS e DONNELLY, 2010).

Neste programa pode-se optar por classificar os indivíduos sob análise como sendo um modelo de não mistura (quando os genomas dos indivíduos têm origem em uma única população), ou modelo de mistura (quando os genomas dos indivíduos podem ter proporção de mais de uma população)(PRITCHARD, STEPHENS e DONNELLY, 2010). No presente estudo, por suspeitar-se da existência de híbridos, optou-se pelo modelo de misturas.

Inicialmente deve ser definido um número inicial de *sets* (K). Esta escolha geralmente é empírica, tomando como base o número de locais onde os indivíduos foram coletados, por exemplo. Embora a hipótese inicial fosse que os genótipos contidos no banco de germoplasma tivessem origem em duas ou três populações, utilizou-se como critério o número inicial de 10 subpopulações ($K = 10$), visando permitir ao programa *Structure* (PRITCHARD, WEN e FALUSH, 2009) a localização no número mais adequado. Assim, identificação da estrutura de populações foi realizada utilizando os marcadores codominantes SSR a partir do programa *Structure* (PRITCHARD, WEN e FALUSH, 2009).

A estrutura genética das variedades da macadâmia também foi caracterizada pelas frequências alélicas e pelas estatísticas F . As análises para as estimativas dos parâmetros de diversidade foram realizadas de acordo com o método de Weir e Cockerham (1984), e obtidas pelo programa *GDA* (LEWIS e ZAYKIN, 1999).

A estimativa das frequências alélicas (p_{ij}), heterozigozidade observada (H_O) e esperada (H_E) foram obtidas pelas expressões citadas anteriormente. A diversidade genética intrapopulacional também foi caracterizada pelo Índice de Fixação de Wright (f), obtida pela expressão:

$$f = 1 - \frac{H_O}{H_E}$$

A diversidade interpopulacional foi caracterizada pela estatística F conforme os parâmetros Índice de Fixação para o conjunto de subpopulações (F), Divergência Genética entre subpopulações (θ_p) e Índice de Fixação médio dentro de subpopulações (f). Estes parâmetros foram obtidos pelo método de Weir e Cockerham (1984), a partir do programa *Fstat* (GOUDET, 1995). Para verificar se as estimativas médias dos parâmetros F , θ_p e f eram estatisticamente diferentes de zero, estimou-se o intervalo de confiança a 95 e 99% de probabilidade pelo método de reamostragem *bootstrap* sobre locos, conforme procedimento realizado pelo programa *Fstat* (GOUDET, 1995).

3.4.2 Distância genética

As estimativas das distâncias genéticas entre as variedades e espécies estudadas foram estimadas conforme as características dos marcadores desenvolvidos, ou seja, obtidas segundo o método Jaccard a partir dos marcadores dominantes DArT, e pelo método de Rogers (1972), modificado por Wright (1978) para os marcadores codominantes SSR.

O coeficiente de Jaccard (desenvolvido por Paul Jaccard em 1901) é uma estatística usada para medir a similaridade entre uma diversidade de conjuntos em diversas áreas da ciência. Ele é definido pelo tamanho da interseção dividido pelo tamanho da união dos itens de dois conjuntos. A distância de Jaccard, que mede a dissimilaridade entre dois conjuntos, é complementar ao coeficiente de Jaccard e obtida pela subtração do coeficiente de 1. Este índice é muito utilizado em estudos por permitir mensurar a diversidade entre amostras

Assim, Distância de Jaccard foi obtida por:

$$J = 1 - \frac{c}{a + b + c}$$

Onde:

- * a é o número de alelos no grupo A;
- * b é o número de alelos no grupo B;
- * c é o número de alelos comuns em ambos os grupos.

A Distância de Roger Modificada obtida entre pares de genótipo estimada por:

$$MRD = \frac{1}{\sqrt{2m}} \sqrt{\sum_{i=1}^m \sum_{j=1}^{n_i} (p_{ijx} - p_{ijy})^2}$$

Onde:

- * m = número de locos avaliados;
- * p_{ijx} = frequência do alelo j , no loco i , no indivíduo x ;
- * p_{ijy} = frequência do alelo j , no loco i , no indivíduo y .

Os dendrogramas foram obtidos com base nas medidas de similaridade genética entre pares de indivíduos, para ambos os marcadores moleculares. Nesta análise são estabelecidos, de forma arbitrária, centros iniciais e cada indivíduo é alocado em relação ao grupo mais próximo geneticamente (MEYER, 2002). A cada inclusão, novos grupos são formados e os indivíduos são rearranjados até que haja estabilidade entre todos os grupos (MEYER, 2002).

Para verificar a acurácia das estimativas das distâncias genéticas em ambos os métodos realizou-se a reamostragem *bootstrap* sobre locos. Foram estipulados conjuntos de dados – *sets* – para cada marcador (SSR: 10, 15, 20, 25 e 29 locos; DArT: 100, 200, 300, 400 e 462 locos), e verificado a sua eficácia considerando um máximo de 10% de Coeficiente de Variação (CV). Assim, separadamente para cada marcador foi estimado um subconjunto de dados para cada *set* descrito acima, e calculado o CV entre os resultados obtidos de Distâncias Genéticas.

Os *bootstraps* foram conduzidos conforme Tivang, Nienhuis e Smith (1994) e Garcia *et al.* (2004), de forma possibilitar a seleção aleatória de cada loco para compor o conjunto de dados de cada *set*. Cada loco tinha a probabilidade de ser amostrado igual à $1/n$, sendo n = número total de locos, com a possibilidade de reamostragem de um mesmo loco para

composição de um conjunto de dados. Cada set teve uma reamostragem igual à 1.000 para cálculo do CV entre as Distâncias Genéticas.

As estimativas das distâncias de Jaccard e de Roger Modificada, assim como os *bootstraps* para estas estimativas e a obtenção dos dendrogramas foram conduzidas a partir do programa R (R DEVELOPMENT CORE TEAM, 2009).

3.4.3 Análise de componentes principais

A Análise de Componentes Principais (*Principal Component Analysis - PCA*) é um procedimento para determinação de componentes que expliquem a maior variabilidade possível em uma matriz de dados multidimensionais. Trata-se de um método estatístico de análise multivariada simples, onde se busca o agrupamento de variáveis discretas sobrepostas ou altamente correlacionados (combinações lineares das variáveis originais).

Segundo Anderson (2003) a matriz de covariância pode ter seus componentes decompostos em autovetores e autovalores. Os autovetores padronizados pela divisão da raiz quadrada do seu autovalor correspondente são alocados no eixo da coordenada principal (ANDERSON, 2003).

A partir deste método dispõe os indivíduos no espaço a fim de demonstrar a similaridades entre eles. A vantagem deste método é a redução do número de variáveis, uma vez que estas são agrupadas com base na relativa importância de cada variável em relação aos componentes principais.

Utilizou-se neste estudo o pacote PCO que calcula as coordenadas principais de qualquer matriz de distâncias simétrica conforme Gower (1966, citado por ANDERSON, 2003), a partir do programa R (R DEVELOPMENT CORE TEAM, 2009).

4 RESULTADOS E DISCUSSÃO

4.1 Desenvolvimento dos marcadores microssatélites - SSR

Dos 192 clones sequenciados, 190 foram considerados adequados para análise. O alinhamento realizado formou 131 *clusters* (47 *contigs* e 84 *singletons*), nos quais foram encontrados 82 microssatélites. Estes foram classificados como simples ou compostos, e posteriormente, como perfeitos ou imperfeitos. Desta forma, foram localizados 66 microssatélites perfeitos (dos quais 62 eram formados por dinucleotídeos, um por trinucleotídeos e três por tetranucleotídeos), e 16 microssatélites compostos por dois e três motivos.

Apesar de terem sido encontrados 82 regiões contendo microssatélites, algumas foram descartadas da análise, visto que estavam localizadas próximas as extremidades, impedindo o desenho do *primer*.

Para a caracterização dos locos foram utilizados 28 variedades de *M. integrifolia*, dentre variedades desenvolvidas na Austrália, Havaí e Brasil. Além das variedades de macadâmia foram genotipados um indivíduos da espécie *Grevilea robusta* e outro de *G. banksii*, ambas as espécies pertencentes à mesma família botânica da macadâmia (Proteacea), a fim de testar a transferibilidade dos *primers*.

Os produtos amplificados via PCR foram avaliados em gel de agarose 3% e posteriormente aplicado em gel de acrilamida 6% para genotipagem dos indivíduos, conforme exemplificados nas figuras 6 e 7. Dos 51 locos testados, 38 tiveram suas temperaturas ajustadas reproduzindo zonas de leitura de alelos identificáveis. Destes, 29 foram polimórficos e apresentaram entre dois e nove alelos por loco. No total foram observados 106 alelos.

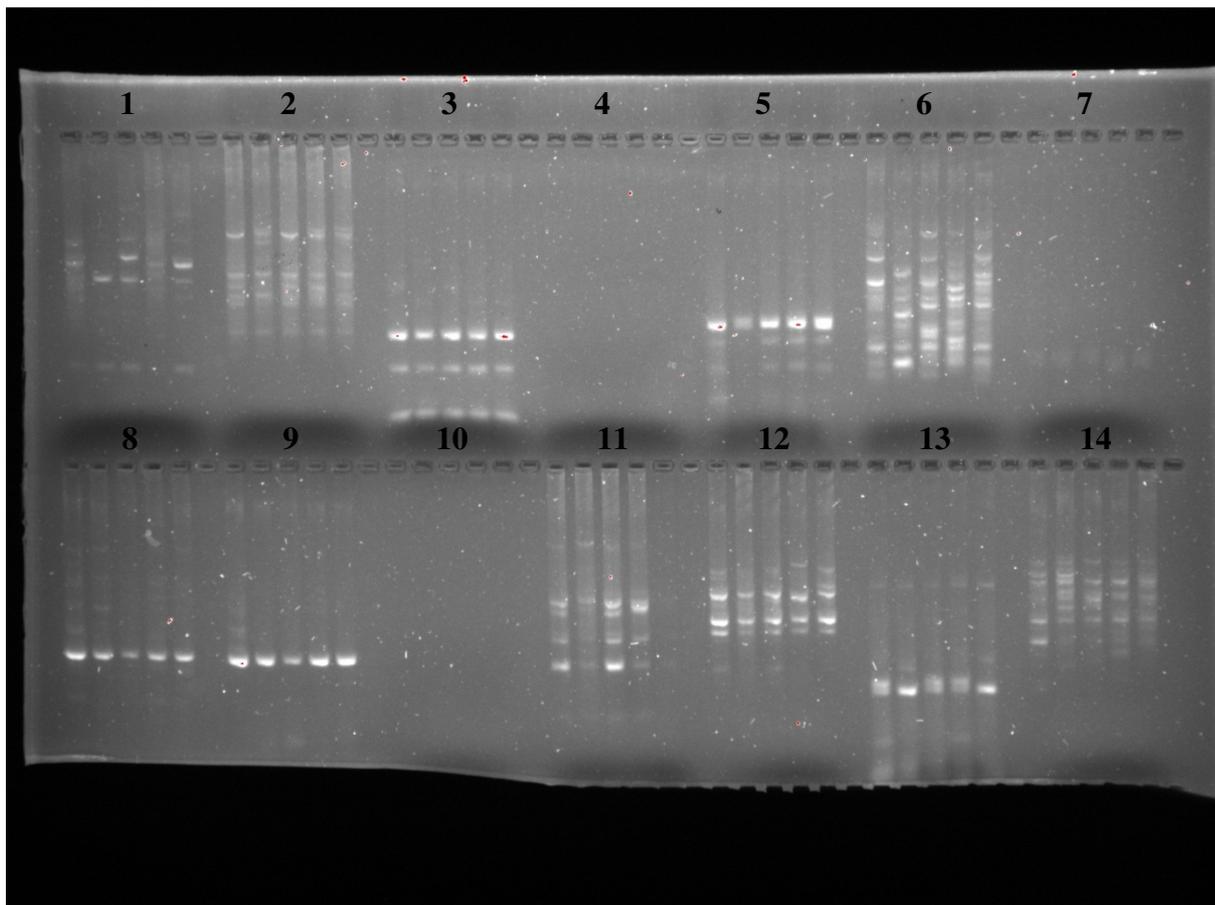


Figura 6 - Exemplo de gel de agarose (3%) onde foram testados cinco indivíduos para cada loco no programa *Touch Down*. Observa-se, por exemplo, nos locos 6, 11, 12 e 14 a presença de muitas bandas inespecíficas, demonstrando a necessidade de teste de ajuste de temperatura de anelamento dos *primers*

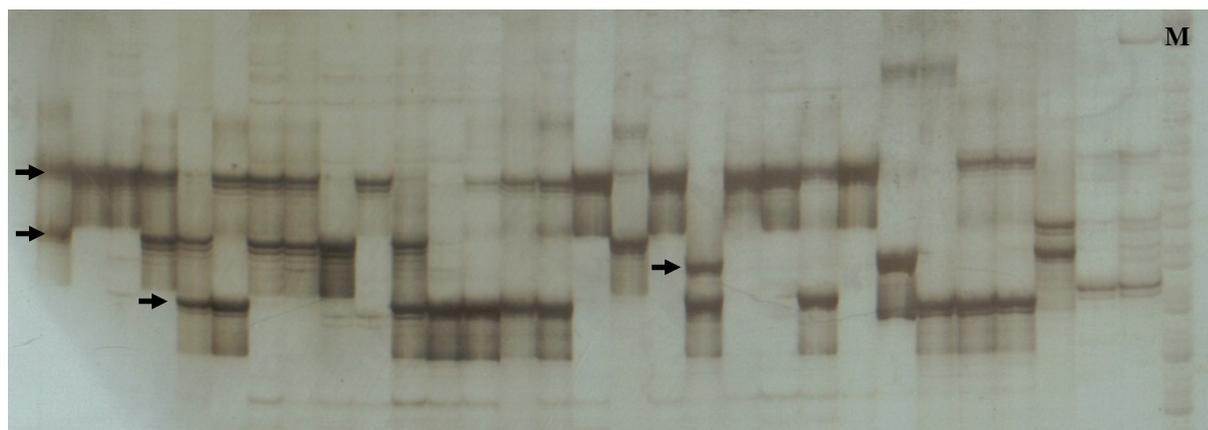


Figura 7 - Exemplo de gel de acrilamida (6%) onde foram testados os 30 indivíduos para o loco MI1C019A1-2 (5/6). Observa-se neste, a presença de quatro alelos. M = marcador *leader* 10 pb

A frequência dos alelos para o conjunto das 28 variedades e as duas espécies de grevilea variou entre 0,017 e 0,983 (Tabela 2). Foram observados 25 alelos raros representando 23,58% do total amostrado. Os locos MI1C025-1 (11/12) e MI3C004-B1e2 (65/66) foram os que apresentaram maior riqueza alélica com nove alelos, sendo cinco e três

destes raros, respectivamente (frequência entre 0,017 e 0,033). Em apenas um loco, MI4C022-A1 (93/94), um alelo apresentou frequência igual ou superior a 0,95 (frequência do alelo 1: 0,983), demonstrando grande possibilidade de fixação.

Tabela 2 - Frequência alélica de cada loco microssatélite na espécie macadâmia (*Macadamia integrifolia*) ($n =$ tamanho amostral)

(continua)			
Loco	Alelo	n	Frequência Alélica
MI1C002A1		29	
(1/2)	1		0,414
	2		0,121
	3		0,069
	4		0,396
MI1C019A1-2		30	
(5/6)	1		0,450
	2		0,183
	3		0,050
	4		0,317
MI1C019A3-4		28	
(7/8)	1		0,750
	2		0,214
	3		0,036
MI1C022-1		30	
(9/10)	1		0,183
	2		0,017
	3		0,017
	4		0,667
	5		0,117
MI1C025- 1		30	
(11/12)	1		0,083
	2		0,433
	3		0,017
	4		0,167
	5		0,200

Tabela 2 - Frequência alélica de cada loco microsatélite na espécie macadâmia (*Macadamia integrifolia*) ($n =$ tamanho amostral)

			(continuação)
Loco	Alelo	n	Frequência Alélica
	6		0,033
	7		0,017
	8		0,033
	9		0,017
MI1C25-4 (13/14)		29	
	1		0,207
	2		0,017
	3		0,776
MI1C028-1 (15/16)		27	
	1		0,037
	2		0,241
	3		0,204
	4		0,074
	5		0,019
	6		0,370
	7		0,056
MI1C043-1 (19/20)		30	
	1		0,183
	2		0,150
	3		0,550
	4		0,100
	5		0,017
MI1C038-1 (25/26)		30	
	1		0,700
	2		0,300
MI1C023-1e2 (29/30)		29	
	1		0,121
	2		0,034
	3		0,017
	4		0,103

Tabela 2 - Frequência alélica de cada loco microssatélite na espécie macadâmia (*Macadamia integrifolia*) ($n =$ tamanho amostral)

			(continuação)
Loco	Alelo	n	Frequência Alélica
	5		0,534
	6		0,172
	7		0,017
MI1C034-1e2 (31/32)		30	
	1		0,467
	2		0,583
MI1C36-3 (33/34)		28	
	1		0,679
	2		0,321
MI1C041-3 (35/36)		30	
	1		0,350
	2		0,650
MI1C041-4 (37/38)		28	
	1		0,214
	2		0,268
	3		0,518
MI1C045-A2e3 (39/40)		29	
	1		0,069
	2		0,759
	3		0,172
MI1C045-A4 (41/42)		30	
	1		0,067
	2		0,050
	3		0,816
	4		0,067
MI1C052-B1a3 (45/46)		30	
	1		0,933
	2		0,067
MI1C054-1		16	

Tabela 2 - Frequência alélica de cada loco microsatélite na espécie macadâmia (*Macadamia integrifolia*) ($n =$ tamanho amostral)

			(continuação)
Loco	Alelo	n	Frequência Alélica
(47/48)	1	30	0,875
	2		0,125
MI1C059-2e3		30	
(51/52)	1	30	0,900
	2		0,100
MI1C069-5		30	
(55/56)	1	30	0,033
	2		0,083
	3		0,884
MI1C080-A1		30	
(59/60)	1	30	0,450
	2		0,550
MI1C102-1e2		29	
(63/64)	1	29	0,172
	2		0,672
	3		0,086
	4		0,052
	5		0,017
MI3C004-B1e2		30	
(65/66)	1	30	0,167
	2		0,017
	3		0,250
	4		0,117
	5		0,033
	6		0,250
	7		0,067
	8		0,067
	9		0,033
MI4C006-1a3		29	

Tabela 2 - Frequência alélica de cada loco microssatélite na espécie macadâmia (*Macadamia integrifolia*) ($n =$ tamanho amostral)

Loco	Alelo	n	Frequência Alélica
(79/80)	1	30	0,534
	2		0,466
MI4C011-1		30	
(87/88)	1	28	0,067
	2		0,666
	3		0,267
MI4C011-2		28	
(89/90)	1	30	0,018
	2		0,616
	3		0,821
MI4C022-A1		30	
(93/94)	1	30	0,983
	2		0,017
MI4C025-2		30	
(95/96)	1	29	0,250
	2		0,750
MI4C026-1		29	
(99/100)	1	25	0,017
	2		0,086
	3		0,621
	4		0,276
Número total de alelos	106		
Número e proporção de alelos raros	25 (23,58%)		

A preocupação na conservação dos alelos em baixa frequência se dá devido a facilidade de extinção destes por deriva genética, não correspondendo que o alelo frequente fornecerá algum valor adaptativo. Além do que, se o ambiente sofrer alterações bruscas, o alelo predominante pode se revelar não adaptado e haver a necessidade do aumento da frequência do alelo raro, o que ocorre mais lentamente do que se este tivesse frequência intermediária (FUTUYMA, 1992). Este é um grande impasse nos programas de

melhoramento genético: promover o aumento da frequência dos alelos que contribuem favoravelmente para determinado carácter, sem que haja a perda de alelos na população que podem ser imprescindíveis para continuação do processo de seleção no futuro.

A heterozigosidade observada variou entre 0,034 e 0,967, enquanto que a heterozigosidade esperada variou entre 0,034 e 0,836, respectivamente nos locos MI4C022-A1 (93/94) e MI3C004-B1e2 (65/66)(Tabela 3). Esses valores podem ser considerados de baixos a altos e podem ter sido afetados pela análise ter ocorrido entre variedades possivelmente aparentadas.

Podemos observar que os índices de diversidade (Tabela 3), calculados a partir da frequência alélica, foram diretamente afetados pelo baixo número de alelos ou pela quase fixação de alelo. O loco MI4C022-A1 (93/94), com os menores índices de diversidade, foi o que apresentou o único valor de frequência superior a 0,95. Enquanto que o loco MI3C004-B1e2 (65/66), com os valores de diversidade mais altos, foi o que apresentou nove alelos, três destes raros, porém os demais com valores intermediários (entre 0,067 e 0,250).

O índice de fixação variou entre -0,526 e 0,557 demonstrando padrões deste o excesso de heterozigotos em alguns locos (valores com sinal negativo), como a sua deficiência em outros (Tabela 3). O índice *PIC* calculado foi baixo e variou de 0,03 a 0,80, mostrando alguns valores baixos, principalmente nos locos que apresentaram baixo número de alelos por loco. Esse baixo índice pode ser devido à proximidade genética entre as variedades.

Tabela 3 - Sequência dos *primers*^(*) e características dos microssatélites de *Macadamia integrifolia*, incluindo nome do loco, sequência do *primer*, tamanho amostral (*n*), número de alelos esperados por loco (*A*), heterozigosidade observada (*H_O*), heterozigosidade esperada (*H_E*), índice de fixação de Wright (*f*) e *polymorphic index content* (*PIC*)

Loco	Sequência do <i>primer</i> (5'→3')	<i>n</i>	<i>A</i>	<i>H_E</i>	<i>H_O</i>	<i>f</i>	<i>PIC</i>
MI1C002A1 (1/2)	F: CAGCCCATGAAATAACAAT R: CATAATGGTGCAGTGATAGTA	29	4	0,663	0,690	-0,040	0,58
MI1C019A1-2 (5/6)	F: CCTGGAAGAGCTGACCTAAAA R: CACAATCACGACCAGTAAACAA	30	4	0,672	0,533	0,210	0,60
MI1C019A3-4 (7/8)	F: TGGTTTATATCTCCTCCTTGTT R: ATCTCCACCCATTCCTTCT	28	3	0,397	0,286	0,285	0,34
MI1C022-1 (9/10)	F: ATCTGCCACTCCATTTTA R: GACTTCTGTTTCCCTTCTA	30	5	0,516	0,600	-0,165	0,46

(continua)

Tabela 3 - Sequência dos *primers*^(*) e características dos microssatélites de *Macadamia integrifolia*, incluindo nome do loco, sequência do *primer*, tamanho amostral (n), número de alelos esperados por loco (A), heterozigosidade observada (H_O), heterozigosidade esperada (H_E), índice de fixação de Wright (f) e *polymorphic index content* (PIC)

								(continuação)
Loco	Sequência do <i>primer</i> (5'→3')	n	A	H_E	H_O	f	PIC	
MI1C025- 1 (11/12)	F: GCCCAAATAGAGTCAAAGTC R: TTAAGCGGCAAATCAAGA	30	9	0,747	0,867	-0,163	0,70	
MI1C25-4 (13/14)	F: TAAGTTTGAAGACAGTGAGTGG R: TGAAAGTGTTAAGAGGCAGAA	29	3	0,361	0,448	-0,246	0,30	
MI1C028-1 (15/16)	F: AACCTGATGACACCGCCTTCTC R: CTCTCCTCCCCTTGTCTATTCA	27	7	0,767	0,852	-0,112	0,72	
MI1C043-1 (19/20)	F: ATTTTCGGTTGGGGTTCTCC R: CCCATTTGCTCTTTCATTTTCAT	30	5	0,642	0,600	0,066	0,59	
MI1C038-1 (25/26)	F: ACGTGCCTTAGATGTTTG R: TAATTTTGGTTGTAGGGGTAG	30	2	0,427	0,267	0,380	0,33	
MI1C023-1e2 (29/30)	F: CACTATTTGGGGCTCCTCAT R: ACCCATCCCTAACTCACTCTCA	29	7	0,669	0,552	0,177	0,62	
MI1C034-1e2 (31/32)	F: ATAATCAGTAAACAACGCTCAA R: TCCCACAATTCCAGACCA	30	2	0,506	0,733	-0,460	0,37	
MI1C36-3 (33/34)	F: CAACTTAACTTCAACGGGTAG R: AAAATTTTCATCCTCTTACAA	28	2	0,444	0,500	-0,128	0,34	
MI1C041-3 (35/36)	F: GTCACCGGTCCGATTCTGT R: ATGGCTTGGCTTCCTTGTTT	30	2	0,463	0,700	-0,526	0,35	
MI1C041-4 (37/38)	F: AATCGGATACTTCTTTCTAA R: ACCGTATGCTTCTGTTTT	28	3	0,625	0,286	0,548	0,54	
MI1C045-A2e3 (39/40)	F: GCTAACTACCATACGAACTGC R: TAAATTGAACTTGCTTCTT	29	3	0,397	0,276	0,309	0,35	
MI1C045-A4 (41/42)	F: TCCCAAGATGACCAAAGAAG R: AACTCAAAGAAAAGGCAAATC	30	4	0,327	0,167	0,495	0,31	
MI1C052-B1a3 (45/46)	F: TAGACCAGTAAAACATCAGG R: AAAGCTTTACCACCAACTAT	30	2	0,126	0,133	-0,054	0,14	
MI1C054-1 (47/48)	F: ACCCTGTCACCTTTGTCA R: GTTTTCGTGTCTTCTCTTTATG	16	2	0,226	0,250	-0,111	0,20	
MI1C059-2e3 (51/52)	F: TGCTCACCTCAAATCGTAG R: TTAAAGAGTGGGTGAGAAGAGT	30	2	0,183	0,133	0,275	0,16	
MI1C069-5 (55/56)	F: TCGGAAGAAAATGTCAC R: ACTCAAATGCTCTCAATCA	30	3	0,215	0,100	0,540	0,22	
MI1C080-A1 (59/60)	F: TATCAGGTAGTGTCTTTTT R: TGTGTTTTCTTATTTATTCTA	30	2	0,503	0,500	0,006	0,37	

Tabela 3 - Sequência dos *primers*^(*) e características dos microssatélites de *Macadamia integrifolia*, incluindo nome do loco, sequência do *primer*, tamanho amostral (n), número de alelos esperados por loco (A), heterozigosidade observada (H_o), heterozigosidade esperada (H_E), índice de fixação de Wright (f) e *polymorphic index content* (PIC)

Loco	Sequência do <i>primer</i> (5'→3')	n	A	H_E	H_o	f	(conclusão)
							PIC
MI1C102-1e2 (63/64)	F: GTAGGAACAAAGAAATGGTC R: TTACAAGTATGGAAGTGGAGT	29	5	0,517	0,379	0,269	0,47
MI3C004-B1e2 (65/66)	F: CGGTTTTTCTTCTTTCACTCAA R: TCATGGTTTCAGGATTTCACTA	30	9	0,836	0,967	-0,159	0,80
MI4C006-1a3 (79/80)	F: CAGACACTTGCCCCTACATAC R: GTCTCAAACCTCTCAACCACTCA	29	2	0,506	0,586	-0,161	0,37
MI4C011-1 (87/88)	F: CACCCTCTTCATCACCGTCGTT R: TTCCCCAGTCCATCCCAATCT	30	3	0,488	0,467	0,045	0,41
MI4C011-2 (89/90)	F: GATCTGACGCCTTACGCTGACT R: CTTTGGGGTTTCGATTGGAGAG	28	3	0,304	0,357	-0,176	0,26
MI4C022-A1 (93/94)	F: CAACCAAAGCCATCAAAGAGTG R: ACGCTAGACGGTGGGAGAAA	30	2	0,034	0,034	0,00	0,03
MI4C025-2 (95/96)	F: GTTCGCCATCCTTTTGACA R: CAGGGTGCTTTGAGAACATTAG	30	2	0,381	0,433	-0,139	0,30
MI4C026-1 (99/100)	F: AAAAGGGGGAAAAAGAGATTGT R: ACTTGTCGCTATGCACTGGATT	29	4	0,540	0,241	0,557	0,46

^(*) Sequências depositadas no GenBank – NCBI (<http://www.ncbi.nlm.nih.gov/WebSub/?tool=genbank>)

Utilizando a técnica de isoenzimas, Aradhya *et al.* (1998) estudaram 45 materiais, incluindo variedades provenientes de *M. integrifolia*, *M. tetraphylla*, híbridos e indivíduos de espécies correlacionadas. Foram obtidos valores entre 0,297 e 0,406 para o parâmetro H_o e, entre 0,262 e 0,307 para H_E (média de 0,354 e 0,297, respectivamente).

Schmidt, Scott e Lowe (2006) desenvolveram *primers* de SSR para 33 locos em macadâmia. Estes foram testados em 43 variedades comerciais, revelando monomorfismo em quatro locos. Os índices de diversidade avaliados variaram de baixos à altos entre locos. O número de alelos variou de um a 15 por loco. A heterozigosidade esperada (H_E) e observada (H_o) variou de 0,027 a 0,875 e 0,027 a 0,882, respectivamente. O índice de polimorfismo (PIC) geral foi de 0,480, variando entre 0,026 a 0,848 (SCHMIDT, SCOTT e LOWE, 2006). Desvios entre os valores de heterozigosidade esperada e observada foram encontrados em 15 locos no sentido de falta de heterozigotos e em somente um loco houve excesso de heterozigotos (SCHMIDT, SCOTT e LOWE, 2006).

Os valores apresentados por Schmidt, Scott e Lowe (2006) são levemente inferiores aos obtidos no presente estudo (Tabela 3), demonstrando que os resultados estão dentro do esperado em estudos de diversidade genética entre variedades de macadâmia.

Romão *et al.* (2011) avaliando 45 genótipos de seringueira (*Hevea brasiliensis*) e seis espécies selvagens obteve valores de heterozigosidade esperada entre 0,514 (para variedades desenvolvidas na Malásia) e 0,603 (para variedades desenvolvidas no Brasil); e heterozigosidade observada entre 0,271 e 0,259, respectivamente. Os resultados apresentados para a seringueira nos fornece a ideia da conservação da variabilidade genética em variedades desenvolvidas a partir de espécies arbóreas que, assim como a macadâmia, apresentam poucos ciclos de melhoramento e características semelhantes (tais como, ciclo de vida longo e dificuldade na formação de linhagens).

A transferibilidade intergenérica dos *primers* desenvolvidos ocorreu em 86% dos locos, mostrando a possível aplicação destes marcadores em estudos genéticos com as espécies *Grevilea robusta* e *G. banksii*.

4.2 Desenvolvimento dos marcadores *diversity array technology* - DArT

A combinação de enzimas de corte raro e frequente (*PstI/TaqI*) mostrou-se eficaz para a digestão do DNA da macadâmia, sendo estas as enzimas selecionadas para clivar o DNA utilizado no desenvolvimento dos marcadores. Foram geradas sete bibliotecas contendo cada uma 384 clones selecionados aleatoriamente das placas de crescimentos das colônias de *E. coli*.

No total foram obtidos 4.608 locos DArT, dos quais foram selecionados 462 para as análises genéticas. Apesar de representar apenas 10% do total de marcas produzidas, estas foram selecionadas de forma conservativa, ou seja, somente as marcas com os melhores parâmetros de qualidade foram selecionadas. Este aproveitamento é condizente com outros trabalhos sobre o desenvolvimento do marcador DArT. Por exemplo, Jaccoud *et al.* (2001) no desenvolvimento de DArT utilizando nove cultivares de arroz para compor o painel de diversidade, encontraram 14,5% de marcas polimórficas, e James *et al.* (2008) obtiveram 6 e 15% de polimorfismo no estudo com espécies de briófitas e pteridófitas (*Asplenium* sp. e *Garovaglia* sp., respectivamente), utilizando em ambos os painéis 16 indivíduos. Enquanto que no estudo com painel composto por 284 indivíduos de oito espécies de eucalipto realizado por Sansaloni *et al.* (2010) foram selecionadas 56% das marcas como polimórficas. A comparação entre estes resultados demonstra a importância da seleção dos indivíduos que irão

compor o painel de diversidade, pois dele dependerá a riqueza genética captada pela técnica (JACCOUD *et al.*, 2001; KILIAN *et al.*, 2005).

No presente estudo foram selecionadas somente as marcas com reprodutibilidade superior a 95%, discordância entre fluorescências igual à zero e correlação entre *spot* e *pixel* de no mínimo 90%. As 462 marcas apresentaram polimorfismo entre 53,79 e 97,24%.

4.3 Caracterização molecular do germoplasma

4.3.1 Estrutura genética

A análise a partir do programa *Structure* atribuiu a formação de duas subpopulações no germoplasma de macadâmia existente no Brasil (Figuras 8 e 9). Ressalta-se que nesta análise os dados das duas variedades de grevilea foram desconsiderados com o intuito de não induzir a alocação das variedades nos agrupamento, por se tratarem de indivíduos de gênero botânico diferente.

As variedades desenvolvidas nos Estados Unidos/Havaí e no Brasil (com exceção da IAC 4-20c, IAC Campinas Ba e Aloha) formaram a maior subpopulação com 70,5% dos genótipos, enquanto que a segunda subpopulação foi formada pelas variedades Australianas e pelas variedades brasileiras acima citadas (Figura 10). Podemos observar na figura 10 um gradiente de cores que, conforme grade de intensidade à esquerda, podemos verificar a contribuição de cada subpopulação na formação do genoma das variedades.

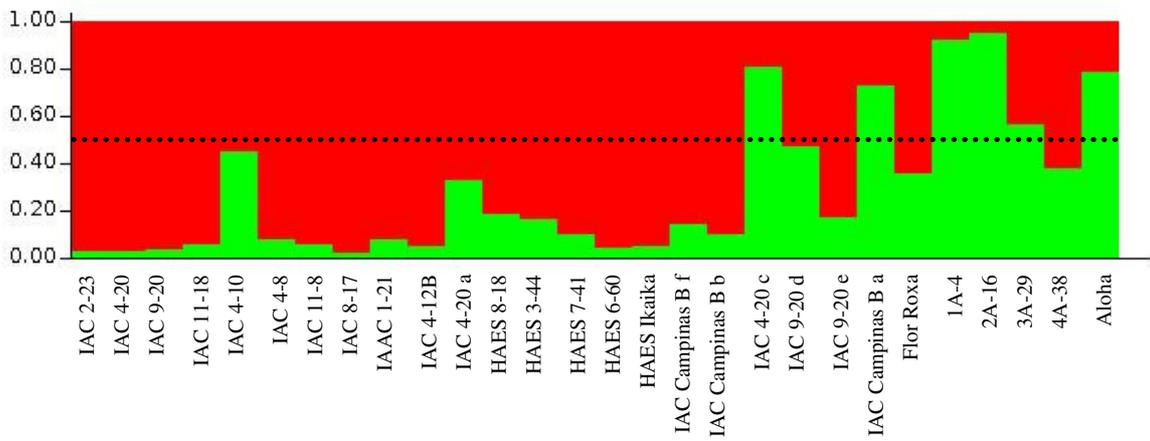


Figura 8 - Proporção do genoma das 28 variedades de macadâmia (*Macadamia integrifolia*) pertencentes às duas populações. Cada indivíduo é representado por uma coluna. Em vermelho está representada a proporção do genoma referente à subpopulação I e em verde à subpopulação II. Linha tracejada = limite entre as duas subpopulações

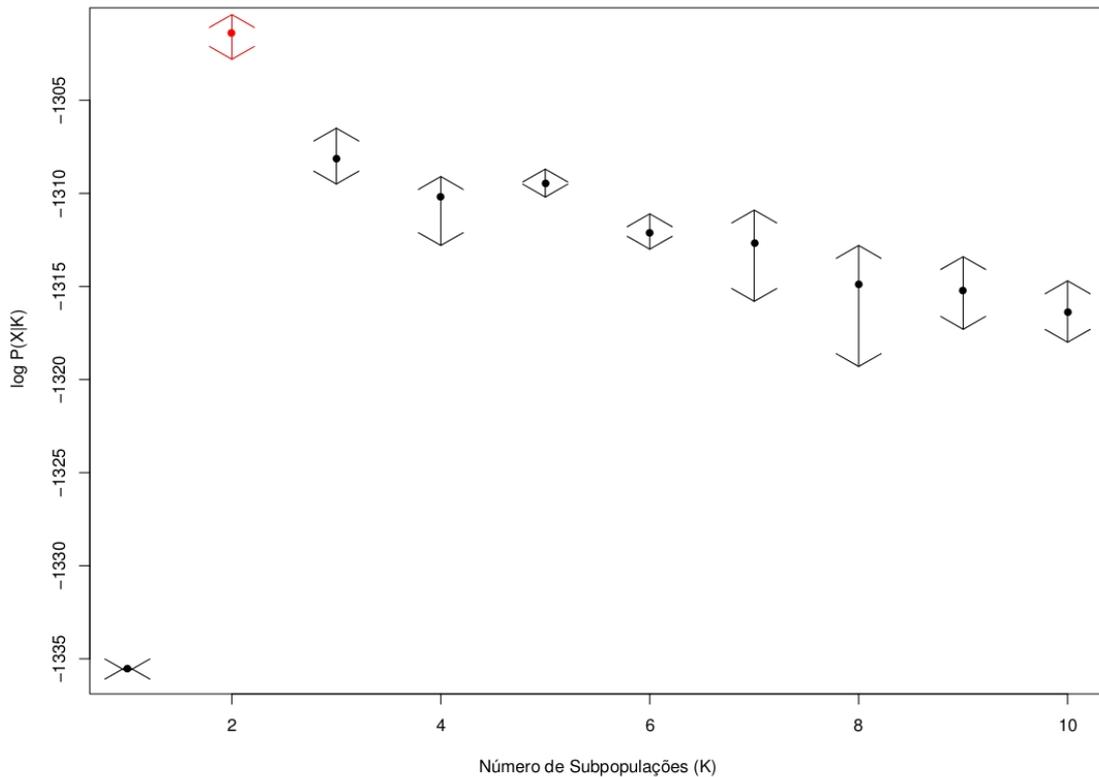


Figura 9 - Valores mínimos, médios e máximos dos logaritmos das probabilidades *a posteriori* referentes a 28 variedades da macadâmia (*Macadamia integrifolia*), de 10 análises realizadas para cada uma das 10 subpopulações (K) estabelecidas *a priori*. Os valores em vermelho indicam o número provável de subpopulações formando o germoplasma no Brasil

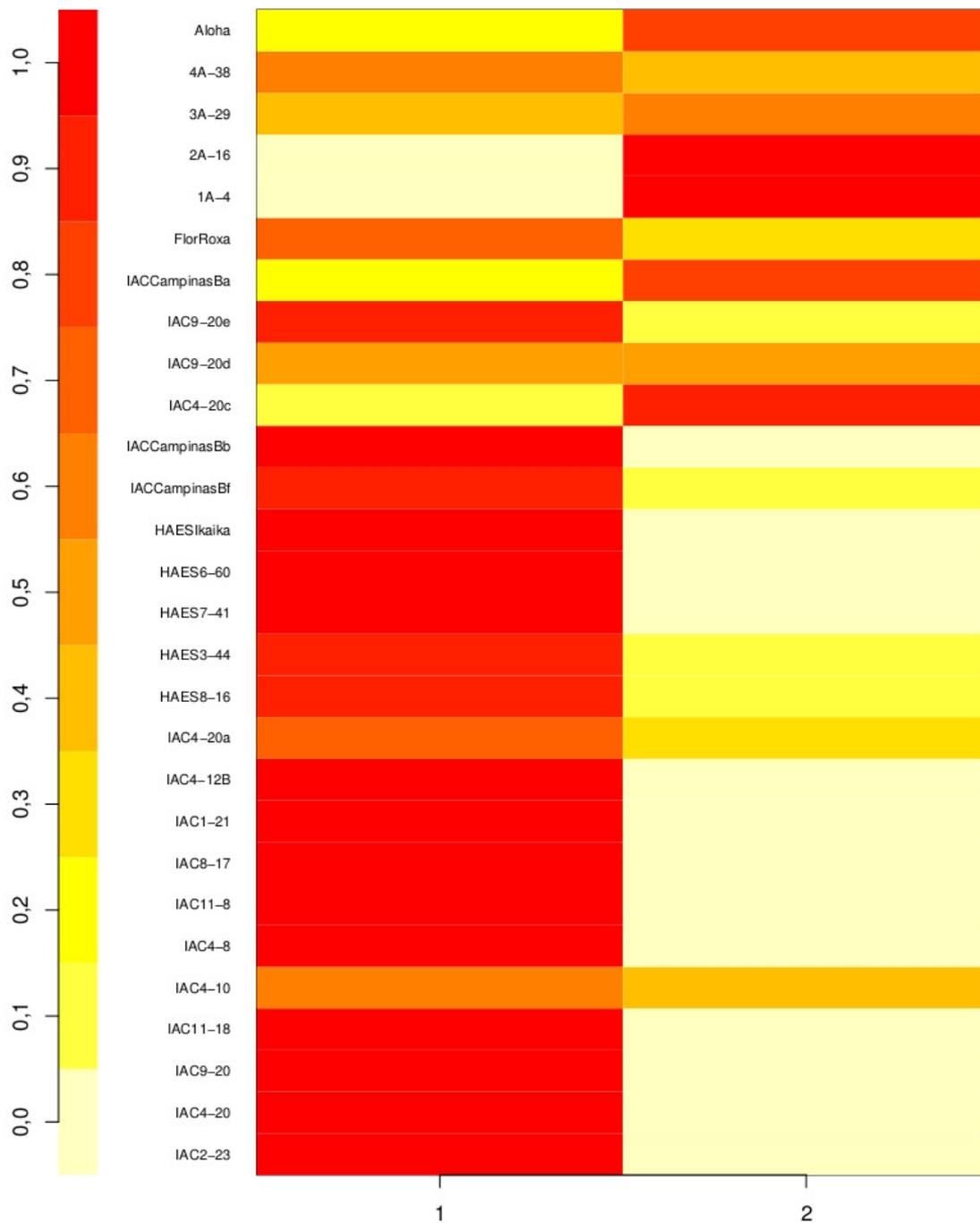


Figura 10 - Contribuição de cada subpopulação para formação do genoma das variedades de macadâmia (*Macadamia integrifolia*)

A diversidade genética foi estudada considerando o conjunto das variedades que compõem o germoplasma no Brasil como um todo e as duas subpopulações discriminadas pelo *Structure* (Tabela 4). Foram encontrados 32 alelos exclusivos entre as subpopulações,

sendo que a subpopulação I apresentou 25 alelos exclusivos. A frequência alélica dos alelos exclusivos variou entre 0,333 (loco 87/88, alelo 3, subpopulação I) e 0,021 (locos: 9/10, alelo 2; 11/12, alelo 9; 65/66, alelo 2; e, 93/94, alelo 2 – todos na subpopulação I).

A subpopulação I apresentou maior riqueza alélica com 18,18% de alelos raros, enquanto que a subpopulação II não apresentou alelos com frequência igual ou inferior a 0,05 (Tabela 4). Quando avaliadas em conjunto as variedades apresentaram 23,58% de alelos raros. Somente a subpopulação II apresentou alelos fixados (dois alelos com frequência igual a um – locos 47/48 e 93/94). A falta de detecção de alelos raros na subpopulação II pode ser atribuída ao baixo tamanho amostral (n entre 5 e 6), assim como à proximidade genética entre as variedades.

Tabela 4 - Estimativa das frequências alélicas em cada loco nas duas subpopulações de *Macadamia integrifolia* e no conjunto das variedades (n = tamanho amostral)

Loco	Alelo	Subpopulação I		Subpopulação II	
		n	Frequência	n	Frequência
MI1C002A1		24		5	
(1/2)	1		0,458		0,200
	2		0,125		0,100
	3		0,083		-
	4		0,333		0,700
MI1C019A1-2		24		6	
(5/6)	1		0,542		0,083
	2		0,145		0,333
	3		0,021		0,167
	4		0,292		0,417
MI1C019A3-4		23		5	
(7/8)	1		0,739		0,800
	2		0,239		0,100
	3		0,022		0,100
MI1C022-1		24		6	
(9/10)	1		0,167		0,250
	2		0,021		-
	3		-		0,083
	4		0,708		0,500

Tabela 4 - Estimativa das frequências alélicas em cada loco nas duas subpopulações de *Macadamia integrifolia* e no conjunto das variedades (n = tamanho amostral)

Loco	Alelo	Subpopulação I		Subpopulação II	
		n	Frequência	n	Frequência
MI1C025- 1 (11/12)	5		0,104		0,167
		24		6	
	1		0,104		-
	2		0,458		0,333
	3		0,021		-
	4		0,167		0,167
	5		0,208		0,167
	6		0,021		0,083
	7		-		0,083
MI1C25-4 (13/14)	8		-		0,167
	9		0,021		-
		23		6	
	1		0,196		0,250
MI1C028-1 (15/16)	2		0,022		-
	3		0,782		0,750
		21		6	
	1		0,048		-
	2		0,286		0,083
	3		0,119		0,500
	4		0,095		-
MI1C043-1 (19/20)	5		0,024		-
	6		0,405		0,250
	7		0,024		0,167
		24		6	
	1		0,208		0,083
2		0,125		0,250	
3		0,583		0,417	
4		0,083		0,167	
5		-		0,083	

Tabela 4 - Estimativa das frequências alélicas em cada loco nas duas subpopulações de *Macadamia integrifolia* e no conjunto das variedades (n = tamanho amostral)

Loco	Alelo	Subpopulação I		Subpopulação II	
		n	Frequência	n	Frequência
MI1C038-1		24		6	
(25/26)	1		0,667		0,833
	2		0,333		0,167
MI1C023-1e2		23		6	
(29/30)	1		0,065		0,333
	2		0,022		0,083
	3		0,022		-
	4		0,022		0,417
	5		0,652		0,083
	6		0,196		0,083
	7		0,022		-
MI1C034-1e2		24		6	
(31/32)	1		0,500		0,333
	2		0,500		0,667
MI1C36-3		22		6	
(33/34)	1		0,614		0,917
	2		0,386		0,083
MI1C041-3		24		6	
(35/36)	1		0,354		0,333
	2		0,646		0,667
MI1C041-4		23		5	
(37/38)	1		0,196		0,300
	2		0,261		0,300
	3		0,543		0,400
MI1C045-A2e3		23		6	
(39/40)	1		0,087		-
	2		0,739		0,833
	3		0,174		0,167
MI1C045-A4		24		6	
(41/42)	1		0,083		-

Tabela 4 - Estimativa das frequências alélicas em cada loco nas duas subpopulações de *Macadamia integrifolia* e no conjunto das variedades (n = tamanho amostral)

Loco	Alelo	(continuação)			
		Subpopulação I		Subpopulação II	
		n	Frequência	n	Frequência
	2		0,063		-
	3		0,771		1,000
	4		0,083		-
MI1C052-B1a3 (45/46)		24		6	
	1		0,958		0,833
	2		0,042		0,167
MI1C054-1 (47/48)		15		1	
	1		0,867		1,000
	2		0,133		-
MI1C059-2e3 (51/52)		24		6	
	1		0,917		0,833
	2		0,083		0,167
MI1C069-5 (55/56)		24		6	
	1		-		0,167
	2		0,104		-
	3		0,896		0,833
MI1C080-A1 (59/60)		24		6	
	1		0,479		0,333
	2		0,521		0,667
MI1C102-1e2 (63/64)		23		6	
	1		0,130		0,333
	2		0,717		0,500
	3		0,065		0,167
	4		0,065		-
	5		0,022		-
MI3C004-B1e2 (65/66)		24		6	
	1		0,125		0,333
	2		0,021		-
	3		0,271		0,167

Tabela 4 - Estimativa das frequências alélicas em cada loco nas duas subpopulações de *Macadamia integrifolia* e no conjunto das variedades (n = tamanho amostral)

Loco	Alelo	Subpopulação I		Subpopulação II	
		n	Frequência	n	Frequência
	4		0,125		0,083
	5		0,042		-
	6		0,271		0,167
	7		0,083		-
	8		0,063		0,083
	9		-		0,167
MI4C006-1a3 (79/80)		23		6	
	1		0,565		0,417
	2		0,435		0,583
MI4C011-1 (87/88)		24		6	
	1		0,063		0,083
	2		0,604		0,917
	3		0,333		-
MI4C011-2 (89/90)		22		6	
	1		0,023		-
	2		0,159		0,167
	3		0,818		0,833
MI4C022-A1 (93/94)		24		6	
	1		0,979		1,000
	2		0,021		-
MI4C025-2 (95/96)		24		6	
	1		0,229		0,333
	2		0,771		0,667
MI4C026-1 (99/100)		23		6	
	1		-		0,083
	2		0,087		0,083
	3		0,587		0,751
	4		0,326		0,083

Tabela 4 - Estimativa das frequências alélicas em cada loco nas duas subpopulações de *Macadamia integrifolia* e no conjunto das variedades (n = tamanho amostral)

Loco	Alelo	(conclusão)			
		Subpopulação I		Subpopulação II	
		n	Frequência	n	Frequência
N° total de alelos		99		81	
N° e proporção de alelos raros		18 (18,18%)		0	

Para o conjunto das 28 variedades de macadâmia e duas espécies de grevílea foram avaliados, em média, 3,08 alelos por loco. A H_E e H_O obtidas foram de 0,46 e 0,43, respectivamente, demonstrando uma pequena redução no número de heterozigotos esperados quando a população está sob Equilíbrio de Hard-Weinberg (Tabela 5).

Foram calculados os parâmetros de H_E , H_O e f para cada loco de cada subpopulação, assim como a média por subpopulação (Tabela 5). O loco *MI3C004-B1e2* obteve maiores valores para os parâmetros H_E e H_O (subpopulação I - H_E : 0,83 e H_O : 1,00; subpopulação II - H_E : 0,86 e H_O : 0,83), enquanto que o loco *MI4C002-A1*, obteve os menores (subpopulação I - H_E : 0,04 e H_O : 0,04; subpopulação II - H_E : 0,00 e H_O : 0,00), em ambas as subpopulações. O índice de fixação de Wright variou entre -0,32 e 0,78 na subpopulação I, e entre -0,43 e 1,00 na subpopulação II, revelando a ocorrência de excesso de homozigotos em grande número dos locos (45% dos locos da subpopulação I; 55% dos locos da subpopulação II), fato esperado, visto que se trata de amostras compostas por variedades que sofreram alguma intensidade de seleção artificial.

Tabela 5 - Estimativa dos parâmetros genéticos nas duas subpopulações de *Macadamia integrifolia*, em cada loco (n = número amostral; A = número de alelos; H_E = Heterozigosidade esperada; H_O = Heterozigosidade observada; f = Índice de Fixação de Wright)

Loco	Subpopulação I					Subpopulação II				
	n	A	H_E	H_O	f	n	A	H_E	H_O	f
	MI1C002A1 (1/2)	24	4	0,64	0,75	-0,17	5	3	0,60	0,40
MI1C019A1-2 (5/6)	24	3	0,61	0,50	0,18	6	4	0,76	0,67	0,13
MI1C019A3-4	23	3	0,41	0,26	0,36	5	3	0,38	0,40	-0,07

Tabela 5 - Estimativa dos parâmetros genéticos nas duas subpopulações de *Macadamia integrifolia*, em cada loco (n = número amostral; A = número de alelos; H_E = Heterozigosidade esperada; H_O = Heterozigosidade observada; f = Índice de Fixação de Wright)

Loco	Subpopulação I					Subpopulação II				
	n	A	H_E	H_O	f	n	A	H_E	H_O	f
(7/8)										
MI1C022-1	24	4	0,44	0,50	-0,13	6	4	0,75	1,00	-0,36
(9/10)										
MI1C025- 1	24	7	0,72	0,83	-0,16	6	6	0,85	1,00	-0,20
(11/12)										
MI1C25-4	23	3	0,36	0,43	-0,22	6	2	0,41	0,50	-0,25
(13/14)										
MI1C028-1	21	6	0,72	0,90	-0,25	6	1	0,71	0,67	0,07
(15/16)										
MI1C043-1	24	4	0,64	0,62	0,02	6	4	0,65	0,50	0,25
(19/20)										
MI1C038-1	24	2	0,45	0,25	0,45	6	2	0,30	0,33	-0,11
(25/26)										
MI1C023-1e2	23	7	0,57	0,56	0,01	6	5	0,80	0,50	0,40
(29/30)										
MI1C034-1e2	24	2	0,51	0,79	-0,57	6	2	0,41	0,50	-0,25
(31/32)										
MI1C36-3	22	2	0,48	0,59	-0,22	6	2	0,17	0,17	0,00
(33/34)										
MI1C041-3	24	2	0,47	0,71	-0,53	6	2	0,48	0,67	-0,43
(35/36)										
MI1C041-4	23	3	0,58	0,30	0,48	5	3	0,69	0,20	0,73
(37/38)										
MI1C045-A2e3	23	3	0,47	0,35	0,27	6	1	0,00	0,00	0,00
(39/40)										
MI1C045-A4	24	4	0,33	0,21	0,38	6	2	0,30	0,00	1,00
(41/42)										
MI1C052-B1a3	24	2	0,08	0,08	-0,02	6	2	0,30	0,33	-0,11
(45/46)										
MI1C054-1	16	2	0,22	0,25	-0,11	0	0	0,00	0,00	0,00
(47/48)										
MI1C059-2e3	24	2	0,16	0,17	-0,07	6	2	0,30	0,00	1,00
(51/52)										
MI1C069-5	24	2	0,19	0,04	0,78	6	2	0,30	0,33	-0,11

Tabela 5 - Estimativa dos parâmetros genéticos nas duas subpopulações de *Macadamia integrifolia*, em cada loco (n = número amostral; A = número de alelos; H_E = Heterozigosidade esperada; H_O = Heterozigosidade observada; f = Índice de Fixação de Wright)

Loco	Subpopulação I					Subpopulação II				
	n	A	H_E	H_O	f	n	A	H_E	H_O	f
(55/56)										
MI1C080-A1	24	2	0,51	0,54	-0,06	6	2	0,49	0,33	0,33
(59/60)										
MI1C102-1e2	25	5	0,41	0,30	0,26	6	4	0,74	0,66	0,11
(63/64)										
MI3C004-B1e2	24	8	0,83	1,00	-0,21	6	6	0,86	0,83	0,04
(65/66)										
MI4C006-1a3	23	2	0,49	0,65	-0,32	6	2	0,48	0,33	0,33
(79/80)										
MI4C011-1	24	3	0,55	0,58	-0,05	6	1	0,00	0,00	0,00
(87/88)										
MI4C011-2	22	3	0,31	0,36	-0,17	6	2	0,30	0,33	-0,11
(89/90)										
MI4C022-A1	24	2	0,04	0,04	0,00	6	1	0,00	0,00	0,00
(93/94)										
MI4C025-2	24	2	0,36	0,46	-0,28	6	2	0,48	0,33	0,33
(95/96)										
MI4C026-1	23	4	0,58	0,22	0,63	6	3	0,32	0,33	-0,05
(99/100)										
Média por subpopulação	23,2	3,38	0,45	0,46	-0,01	5,89	2,78	0,46	0,40	0,13

Comparando os resultados da tabela 5 com o número de alelos exclusivos e raros, podemos verificar que, apesar de ter apresentado maiores índices de alelos exclusivos e raros, a subpopulação I obteve, geralmente, valores de diversidade genética menores em relação à subpopulação II. Isto deve-se ao fato de que os alelos raros contribuem pouco para parâmetros genéticos que sejam derivados da frequência alélica, como são os parâmetros H_E , H_O e f .

A estimativa dos parâmetros interpopulacionais revelou que o índice de fixação para o conjunto de subpopulações teve grande variação (entre -0,522 e 0,558 – Tabela 6), porém, na média foi baixo e não diferente de zero. Este resultado sugere que, mesmo havendo a

pressão de seleção imposta pelo programa de melhoramento, o conjunto das variedades, na média dos locos, apresenta-se em Equilíbrio de Hardy-Weinberg.

Grande parte da diversidade genética do germoplasma presente no Brasil está dentro de subpopulações (95,9% - Tabela 6). A média de divergência genética entre subpopulações foi baixo ($\theta = 0,041$) e estatisticamente não pode ser diferido de zero, considerando o intervalo de confiança. Em estudos com espécies arbóreas, em populações naturais, foi verificado que esta tendência também ocorre, como por exemplo, em *Araucaria angustifolia* ($\theta = 0,044$, Auler *et al.*, 2002).

Dos 29 locos estudados, 23 apresentaram os valores para o índice de fixação dentro de subpopulações estatisticamente diferentes de zero (Tabela 6). Em apenas nove destes 23 locos, os índices de fixação foram positivos, sugerindo o excesso de homozigotos nestes locos. Estes resultados revelam baixa ocorrência do índice de fixação devido ao sistema reprodutivo, como resultado da autofecundação e de cruzamentos endogâmicos. Isto mostra o indicativo de que, apesar da semelhança genética entre as variedades, estas não foram desenvolvidas a partir de indivíduos aparentados.

Tabela 6 - Estimativa das estatísticas F com base no modelo de Weir e Cockerham (1984), entre locos de duas subpopulações de macadâmia (*Macadamia integrifolia*) (F = índice de fixação para o conjunto de subpopulações; θ = divergência genética entre procedências; f = índice de fixação médio dentro de subpopulações)

Loco	F	θ	f
MI1C002A1 (1/2)	0,030	0,096	-0,073
MI1C019A1-2 (5/6)	0,282	0,137	0,168
MI1C019A3-4 (7/8)	0,263	-0,045	0,294
MI1C022-1 (9/10)	-0,152	0,017	-0,172
MI1C025- 1 (11/12)	-0,162	0,002	-0,164
MI1C25-4 (13/14)	-0,276	-0,035	-0,232
MI1C028-1 (15/16)	-0,036	0,107	-0,161
MI1C043-1 (19/20)	0,067	0,001	0,066
MI1C038-1 (25/26)	0,377	-0,008	0,381
MI1C023-1e2 (29/30)	0,341	0,301	0,058
MI1C034-1e2 (31/32)	-0,433	0,028	-0,474
MI1C36-3 (33/34)	-0,018	0,151	-0,198
MI1C041-3 (35/36)	-0,552	-0,025	-0,514

(continua)

Tabela 6 - Estimativa das estatísticas F com base no modelo de Weir e Cockerham (1984), entre locos de duas subpopulações de macadâmia (*Macadamia integrifolia*) (F = índice de fixação para o conjunto de subpopulações; θ = divergência genética entre procedências; f = índice de fixação médio dentro de subpopulações)

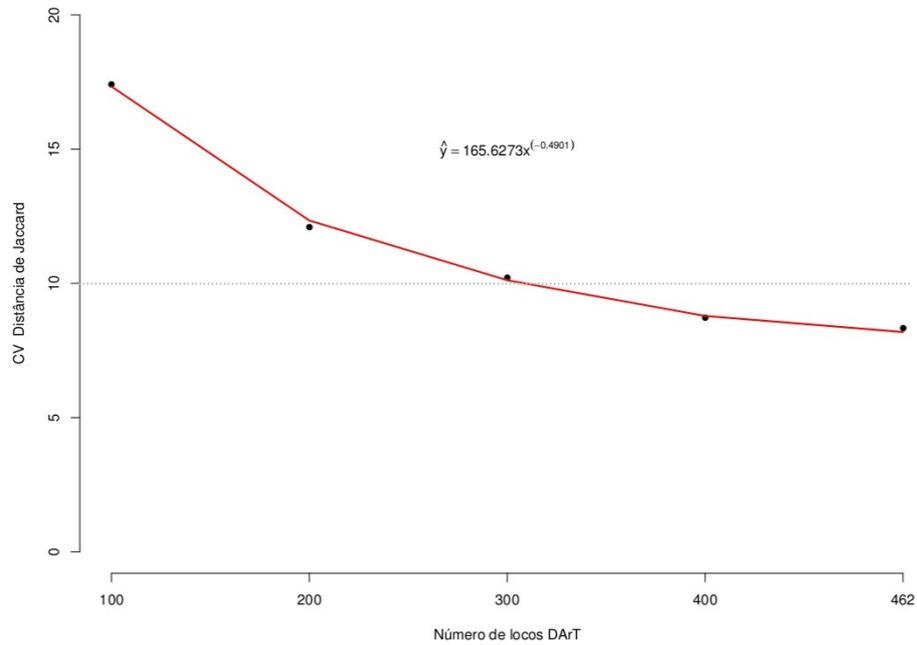
Loco	F	θ	f	(conclusão)
MI1C041-4 (37/38)	0,524	-0,074	0,557	
MI1C045-A2e3 (39/40)	0,285	-0,052	0,320	
MI1C045-A4 (41/42)	0,505	0,029	0,490	
MI1C052-B1a3 (45/46)	-0,003	0,073	-0,082	
MI1C054-1 (47/48)	-0,304	-0,199	-0,088	
MI1C059-2e3 (51/52)	0,260	-0,030	0,282	
MI1C069-5 (55/56)	0,545	0,018	0,537	
MI1C080-A1 (59/60)	0,000	-0,011	0,010	
MI1C102-1e2 (63/64)	0,285	0,033	0,261	
MI3C004-B1e2 (65/66)	-0,144	0,019	-0,167	
MI4C006-1a3 (79/80)	-0,161	0,000	-0,161	
MI4C011-1 (87/88)	0,140	0,149	-0,011	
MI4C011-2 (89/90)	-0,211	-0,045	-0,159	
MI4C022-A1 (93/94)	-0,028	-0,042	0,013	
MI4C025-2 (95/96)	-0,152	-0,017	-0,013	
MI4C026-1 (99/100)	0,558	0,003	0,557	
Média para todos os locos	0,067	0,041	0,027	
Intervalo de confiança (95%)	-0,041 – 0,177	0,004 – 0,083	-0,079 – 0,141	
Intervalo de confiança (99%)	-0,073 – 0,212	-0,006 – 0,097	-0,111 – 0,180	

4.3.2 Distância genética

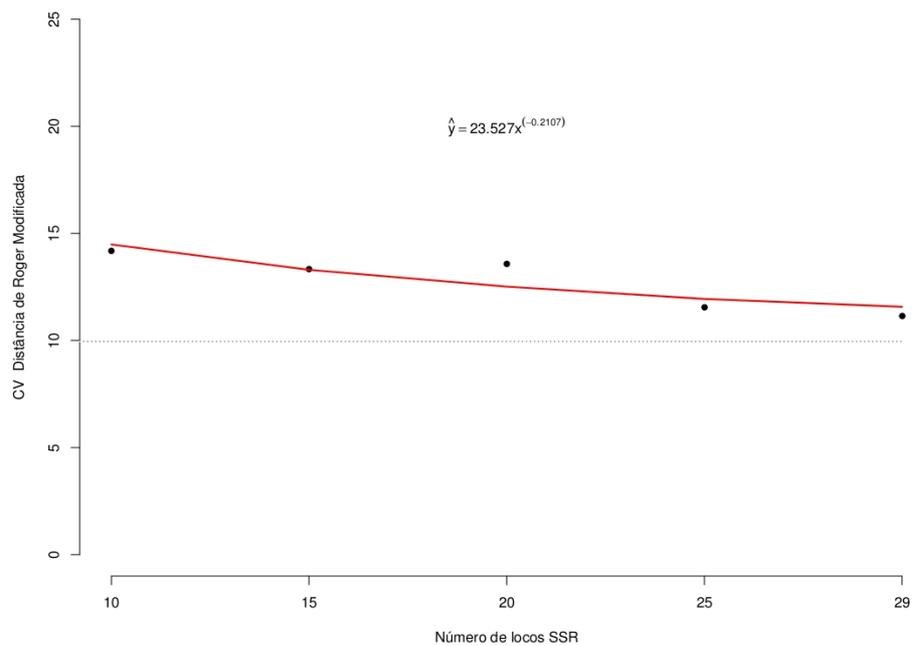
O número mínimo de locos DArT necessários para avaliar a Distância Genética destes acessos de macadâmia foi estimado observando a variação dos valores dos Coeficientes de Variação. Podemos observar que para alcançar um CV de 10%, seria necessário o uso de 307 marcas DArT (Figura 11). Com o uso de 462 marcas obtivemos um CV de 8,19%.

A mesma análise realizada com os marcadores SSR mostrou que o presente trabalho obteve 11,57% de CV, e que a amostragem de locos deveria ser de 58 para um máximo de

10% de CV, revelando a necessidade de aumento no número de locos para obtermos representatividade do genoma amostrado (Figura 12).



Figuras 11 - Relação entre o Coeficiente da Variação (CV) da Distância de Jaccard (J) e o número de locos DArT utilizando a análise *Bootstrap*. Y = ajuste da função exponencial em relação à mediana do CV



Figuras 12 - Relação entre o Coeficiente da Variação (CV) da Distância de Roger Modificada (MRD) e o número de locos SSR utilizando a análise *Bootstrap*. Y = ajuste da função exponencial em relação à mediana do CV

Os valores das Distâncias de Jaccard, estimados a partir dos marcadores DArT, entre pares de variedades variaram entre 0,0 (entre HAES Ikaika e IAC 9-20, HAES 7-41 e HAES 6-60, IAC 2-23 e IAC 8-17, IAC 4-20a e IAC 4-10, e A16 e 2A-16) e 0,787 (entre as variedades IAC 4-10 e 344) (Figura 13). A total falta de divergência entre estes materiais foi inesperada, exceto entre as variedades A16 e 2A-16, que acreditamos ser apenas divergência de nomenclatura. Uma hipótese é que as demais variedades citadas são geneticamente muito próximas, porém os locos amostrados não revelaram tal diferença. Esta hipótese pode ser reforçada pelos resultados dos marcadores SSR, onde, por exemplo, as variedades HAES 7-41 e HAES 6-60, apresentam maior semelhança no grupo. Outro exemplo são as variedades HAES Ikaika e IAC 9-20, que apresentaram diferenças em apenas um loco (*maPb-950215*) e a ocorrência de dados perdidos em um dos indivíduos em 16 locos.

Por outro lado, variedades que se acreditava serem as mesmas, a exemplo da HAES 344 (coletada no Brasil) e da 344 (coletada na Austrália), não apresentaram alta semelhança genética. O mesmo ocorreu entre os diferentes tipos botânicos das variedades IAC 4-20, IAC Campinas B e IAC 9-20, todas coletadas no Brasil. Estes fatos, apresentados em ambos os marcadores, podem ser atribuídos a problemas de identificação do material no campo.

O dendrograma mostrando a relação entre as variedades a partir da Distância de Jaccard (Figura 13) produziu três grupos: um externo formado pela espécie *M. ternifolia* e pela variedade 344; um segundo formado predominantemente pelas variedades desenvolvidas na Austrália, exceto pela variedade 246 desenvolvida nos Estados Unidos/Havaí; e um terceiro grande grupo formado pelas demais variedades.

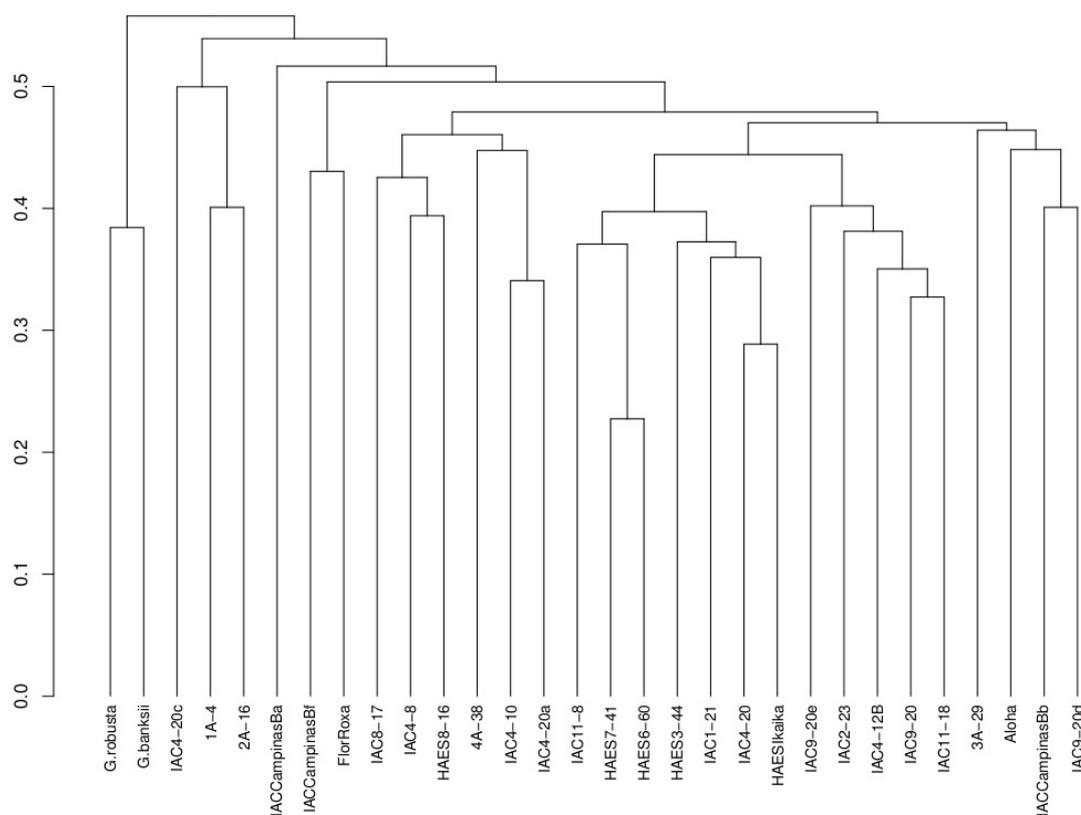


Figura 14 - Dendrograma multilocus baseada na Distância de Roger Modificada (MRD), estimados a partir dos marcadores SSR, apresentando a relação genética entre 28 variedades de macadâmia (*Macadamia integrifolia*) e as espécies *Grevilea robusta* e *G. banksii*. Eixo y = Dissimilaridade genética da Distância de Roger Modificada

As duas subpopulações discriminadas pelo programa *Structure* não formaram grupos específicos em ambos os dendrogramas. Na figura 13 podemos observar que as variedades da subpopulação II ficaram dispersas aos pares em três grupos (Aloha e 3A29; IAC 4-20c e IAC Campinas Ba; 1A4 e 2A16). Isto pode ter ocorrido devido à inclusão de outros indivíduos nos cálculos da Distância de Jaccard que não foram genotipados para os marcadores SSR e por isso não foram avaliados pelo *Structure*. Algo semelhante ocorreu comparando o dendrograma da figura 14 ao resultado do *Structure*, porém, com maior coerência entre os resultados. Podemos observar o agrupamento das variedades IAC 4-20c, IAC Campinas Ba, 1A4 e 2A16, enquanto que as variedades Aloha e 3A29 permaneceram juntas em um subgrupo separado.

A alocação da variedade 3-44 no grupo externo juntamente com a espécie *M. ternifolia* (Figura 13), não foi um resultado esperado, principalmente por a variedade HAES 3-44 ser caracterizada como proveniente de 100% de genoma de *M. integrifolia* (Peace *et al.*,

2005). Uma hipótese, considerando um erro de classificação, seria que a variedade 3-44 não é a mesma HAES 3-44, mas sim um híbrido tri-espécie, com uma considerável porção do genoma de *M. ternifolia*, assim como a variedade Sul-africana 7-91 (Peace *et al.*, 2005), o que a tornaria mais próxima da espécie selvagem do que das variedades reconhecidas como sendo originadas de genoma *M. integrifolia*. Esta contribuição de genomas de diferentes espécies foi mensurada por Peace *et al.* (2002), revelando que variedades anotadas como desenvolvidas a partir de *M. integrifolia* (por exemplo, HAES 2-46, HAES 8-16, HAES 3-44, Dadwon), pode conter até 19% de material genético de *M. tetraphylla*.

As variedades HAES 6-60 e HAES 7-41 são reportadas como sendo variedades originadas pelo mesmo conjunto de *seedling* (VITHANAGE e WINKS, 1992). Neste mesmo estudo estas duas variedades formaram um subgrupo com a variedade HAES 3-44. Resultado semelhante foi obtido no presente trabalho, com ambos os marcadores (Figuras 13 e 14), porém com a inclusão da variedade Beaumont no mesmo subgrupo (diferentemente do esperado, visto que a variedade Beaumont é um híbrido entre *M. integrifolia* x *M. tetraphylla*), e o observado por Vithanage e Winks (1992).

Aradhya *et al.* (1998) citam que dentre as variedades desenvolvidas no Havaí, com base genética em *M. integrifolia*, houve a formação de dois subgrupos, indicando que esta diferenciação pode ter sido devido às duas introduções realizadas em épocas e locais diferentes (1881 – em Kukuihaele; 1892 – em Honolulu). A outra subdivisão foi em relação às variedades desenvolvidas na Austrália, indicando que, apesar da distribuição geográfica restrita das populações naturais, estas possuíam divergência genética para a formação de diferentes *backgrounds* genéticos. Neste trabalho, as variedades HAES 6-60 e HAES 7-41 foram novamente alocadas em um mesmo grupo, também similares à HAES 3-44, o que reforça o estreito parentesco entre estas. As variedades IAC avaliadas ficaram dispersas em diferentes subgrupos, não formando um subgrupo específico, porém todas relacionadas às variedades de genoma predominante de *M. integrifolia* (ARADHYA *et al.*, 1998).

A semelhança entre as variedades HAES 6-60 e HAES 7-41 é relatada em diversos trabalhos, tais como, Vithanage e Winks (1992), Aradhya *et al.* (1998), Peace *et al.* (2002), Steiger *et al.* (2003) e Peace *et al.* (2005). A similaridade genética entre estas duas variedades foi mensurada em 0,985 (STEIGER *et al.*, 2003). Esta grande semelhança pode ser devido a uma mesma possível polinização, ou seja, as duas variedades podem ter sido originadas de um mesmo cruzamento, visto que foram desenvolvidas no mesmo local (Lawai Valley – Havaí/USA)(ARADHYA *et al.*, 1998).

As diferenças na formação dos agrupamentos encontradas entre os artigos publicados e os resultados do presente trabalho podem ser resultantes, em parte, por nenhum deste contemplar um grande número de variedades em comum, excluindo os casos de erro de identificação e nomenclatura. Isso por que as distâncias genéticas, usadas como base para a formação dos dendrogramas, são calculadas par a par, assim a inclusão de um genótipo pode influenciar o resultado final.

A análise dos Componentes Principais da Distância de Jaccard apresentou resultados consistentes ao dendrograma. Podemos verificar a presença de três grupos: dois correspondendo ao grupo externo do dendrograma da Distância de Jaccard (circulados em vermelho), e o outro formado pelas demais variedades (Figura 15). Observa-se que a variedade 344 e a espécie *M. ternifolia* não formaram um grupo próximo pela análise de PCA, porém trata-se de materiais bem distintos das demais variedades. O grupo apresentado no dendrograma, formado pelas variedades Australianas, não foi distinto na análise de PCA, porém os indivíduos compostos por ele ficaram próximos (salientados pelo círculo em amarelo).

Os resultados do dendrograma gerado a partir das Distâncias de Roger Modificada também foram coerentes à análise de PCA para os marcadores SSR. Assim como no dendrograma, as variedades ficaram dispersas, sendo possível somente a observação da separação das duas espécies de grevílea (destacadas em azul).

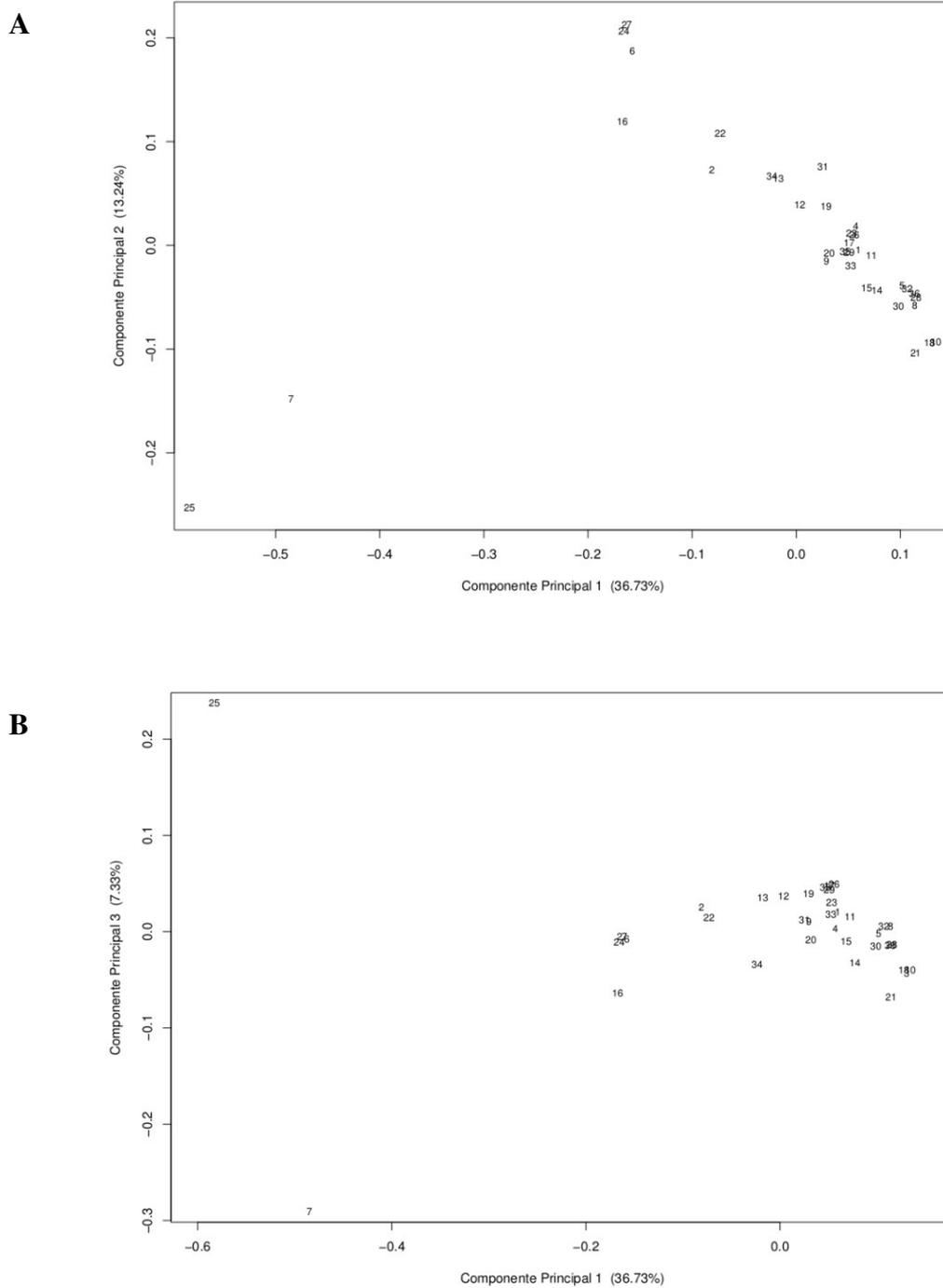
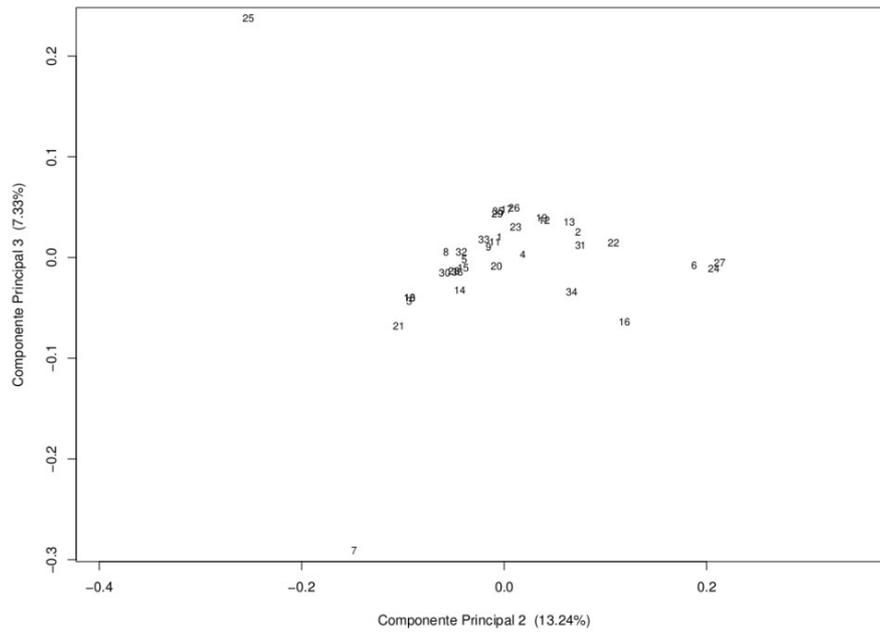


Figura 15 - Análise dos Componentes Principais (PCA) da Distância de Jaccard (J) com as respectivas proporções de resposta da variação. A – resposta do primeiro e segundo componentes; B – resposta do primeiro e terceiro componentes; C – resposta do segundo e terceiro componentes; D – resposta dos três componentes. (Variedades: 1 - HAES 8-16, 2 - HAES 246, 3 – HAES Ikaika, 4 - IAC Campinas B a, 5 - IAC 2-23, 6 - 1A-4, 7 – *Macadamia ternifolia*, 8 - AC 4-20, 9 – Beaumont, 10 - IAC 9-20 e, 11 - HAES 3-44, 12 - Flor Roxa, 13 – Aloha, 14 - IAC 9-20 d, 15 - IAC Campinas B b, 16 – Daddow, 17 - IAC 1-21, 18 - IAC 9-20, 19 - IAC 4-8, 20 - IAC Campinas B f, 21 – 816, 22 – Heilscher, 23 - IAC 4-12 B, 24 - A16, 25 – 344, 26 - IAC 11-8, 27 - 2A-16, 28 - IAC 4-20 a, 29 - HAES 7-41, 30 - IAC 11-18, 31 – IAC 4-20 c, 32 - IAC 8-17, 33 – *Macadamia jansenii*, 34 - 3A-29, 35 - HAES 6-60, 36 - IAC 4-10)(Círculo vermelho e amarelo: grupos formados pelo dendrograma)
(continua)

C



D

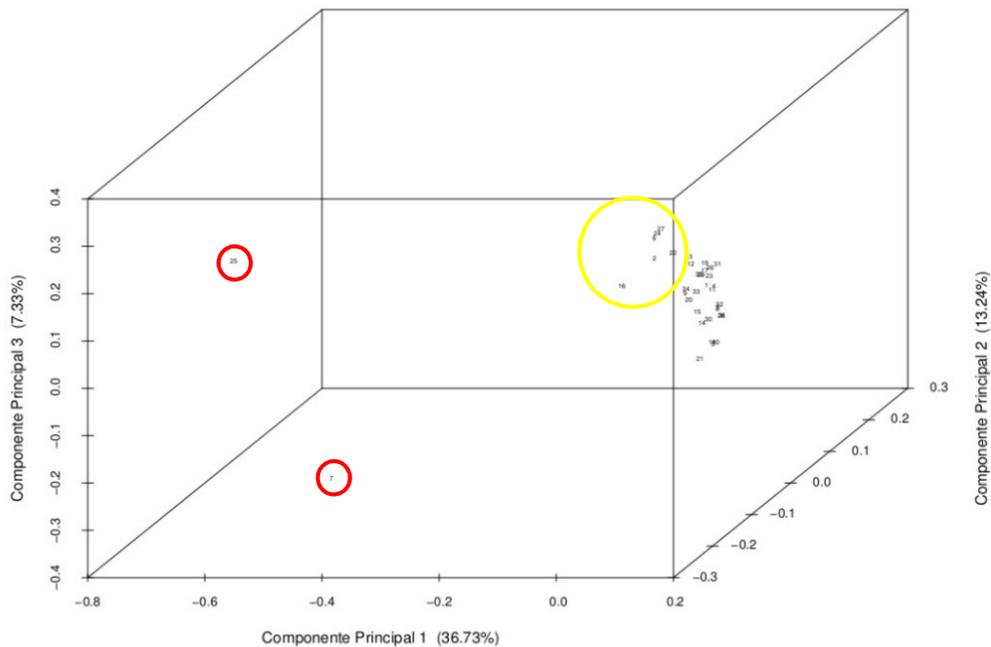


Figura 15 - Análise dos Componentes Principais (PCA) da Distância de Jaccard (J) com as respectivas proporções de resposta da variação. A – resposta do primeiro e segundo componentes; B – resposta do primeiro e terceiro componentes; C – resposta do segundo e terceiro componentes; D – resposta dos três componentes. (Variedades: 1 - HAES 8-16, 2 - HAES 246, 3 - HAES Ikaika, 4 - IAC Campinas B a, 5 - IAC 2-23, 6 - 1A-4, 7 - *Macadamia ternifolia*, 8 - AC 4-20, 9 - Beaumont, 10 - IAC 9-20 e, 11 - HAES 3-44, 12 - Flor Roxa, 13 - Aloha, 14 - IAC 9-20 d, 15 - IAC Campinas B b, 16 - Daddow, 17 - IAC 1-21, 18 - IAC 9-20, 19 - IAC 4-8, 20 - IAC Campinas B f, 21 - 816, 22 - Heilscher, 23 - IAC 4-12 B, 24 - A16, 25 - 344, 26 - IAC 11-8, 27 - 2A-16, 28 - IAC 4-20 a, 29 - HAES 7-41, 30 - IAC 11-18, 31 - IAC 4-20 c, 32 - IAC 8-17, 33 - *Macadamia janseni*, 34 - 3A-29, 35 - HAES 6-60, 36 - IAC 4-10) (Círculo vermelho e amarelo: grupos formados pelo dendrograma) (conclusão)

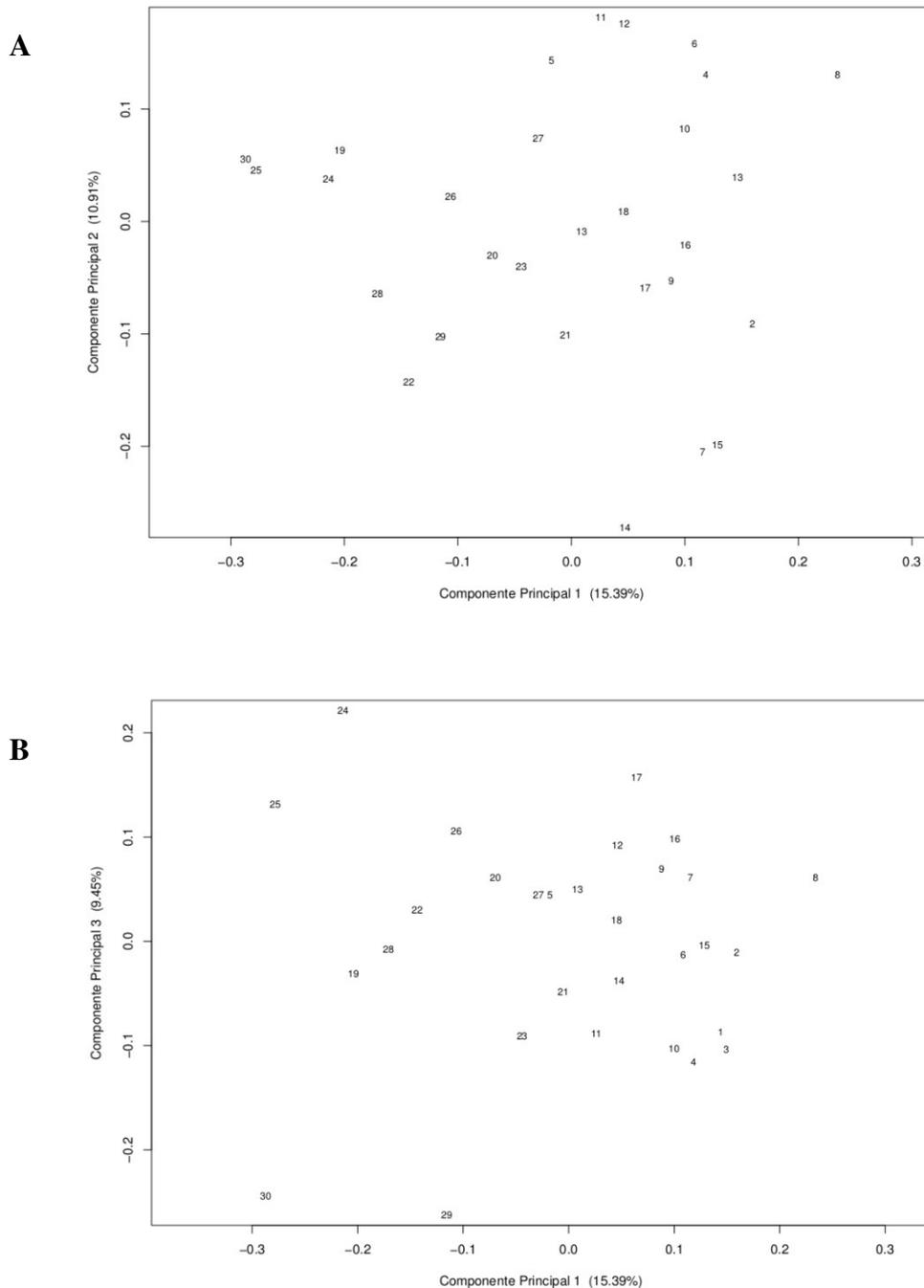


Figura 16 - Análise dos Componentes Principais (PCA) da Distância de Roger Modificada (MRD) com as respectivas proporções de resposta da variação. A – resposta do primeiro e segundo componentes; B – resposta do primeiro e terceiro componentes; C – resposta do segundo e terceiro componentes; D – resposta dos três componentes. (Variedades: 1 - IAC 2-23, 2 - IAC 4-20, 3 - IAC 9-20, 4 - IAC 11-18, 5 - IAC 4-10, 6 - IAC 4-8, 7 - IAC 11-8, 8 - IAC 8-17, 9 - IAC 1-21, 10 - IAC 4-12 B, 11 - IAC 4-20 a, 12 - HAES 8-16, 13 - HAES 3-44, 14 - HAES 7-41, 15 - HAES 6-60, 16 - HAES Ikaika, 17 - IAC Campinas B a, 18 - IAC Campinas B b, 19 - IAC 4-20 c, 20 - IAC 9-20 d, 21 - IAC 9-20 e, 22 - IAC Campinas B f, 23 - Flor Roxa, 24 - 1A-4, 25 - 2A-16, 26 - 3A-29, 27 - 4A-38, 28 - Aloha, 29 - *Grevilea robusta*, 30 - *Grevilea banksii*)(Círculo azul: grupo formado pelo dendrograma)

(continua)

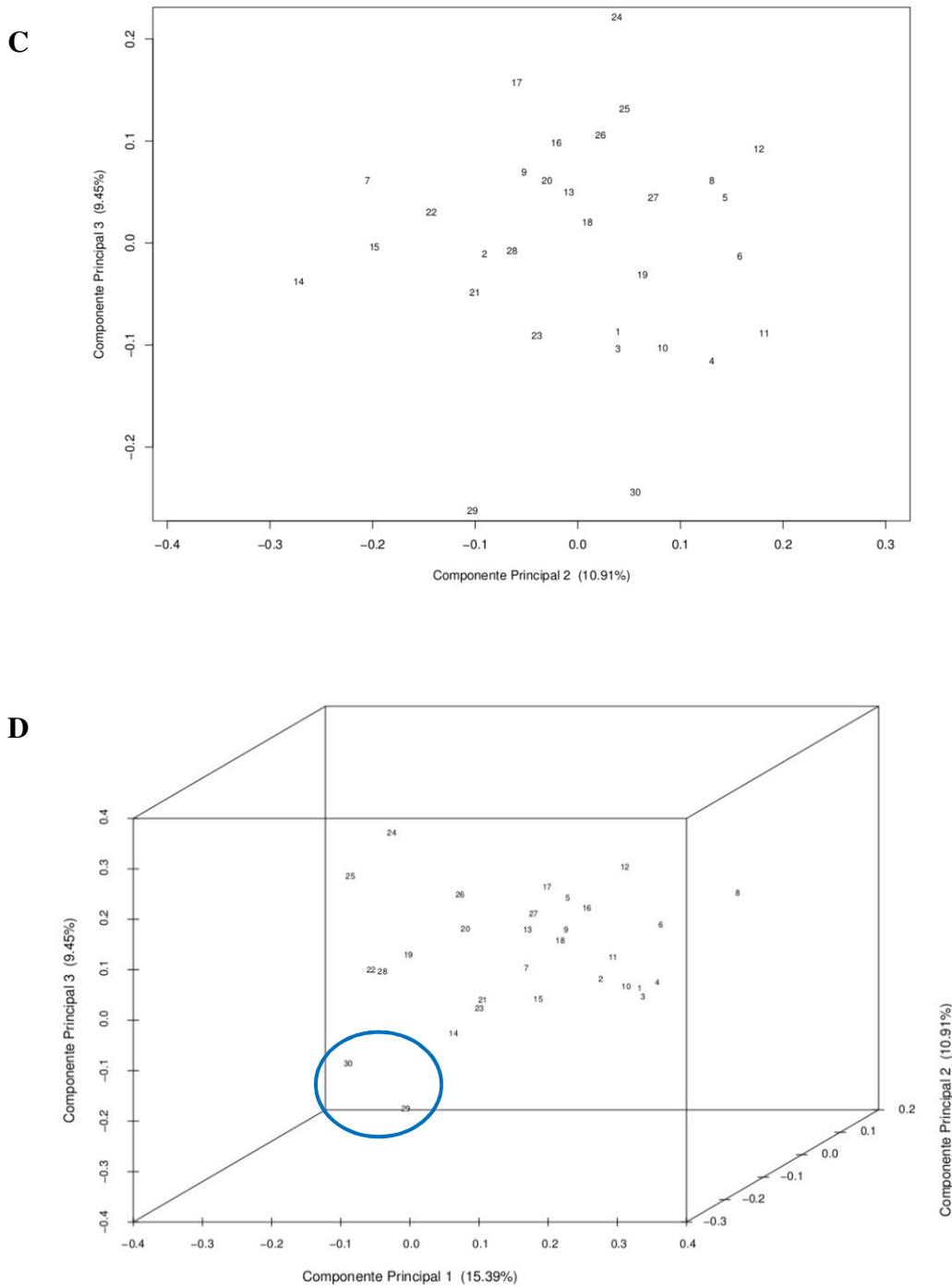


Figura 16 - Análise dos Componentes Principais (PCA) da Distância de Roger Modificada (MRD) com as respectivas proporções de resposta da variação. A – resposta do primeiro e segundo componentes; B – resposta do primeiro e terceiro componentes; C – resposta do segundo e terceiro componentes; D – resposta dos três componentes. (Variedades: 1 - IAC 2-23, 2 - IAC 4-20, 3 - IAC 9-20, 4 - IAC 11-18, 5 - IAC 4-10, 6 - IAC 4-8, 7 - IAC 11-8, 8 - IAC 8-17, 9 - IAC 1-21, 10 - IAC 4-12 B, 11 - IAC 4-20 a, 12 - HAES 8-16, 13 - HAES 3-44, 14 - HAES 7-41, 15 - HAES 6-60, 16 - HAES Ikaika, 17 - IAC Campinas B a, 18 - IAC Campinas B b, 19 - IAC 4-20 c, 20 - IAC 9-20 d, 21 - IAC 9-20 e, 22 - IAC Campinas B f, 23 - Flor Roxa, 24 - 1A-4, 25 - 2A-16, 26 - 3A-29, 27 - 4A-38, 28 - Aloha, 29 - *Grevilea robusta*, 30 - *Grevilea banksii*)(Círculo azul: grupo formado pelo dendrograma)

(conclusão)

Assim como o apresentado nos dendrogramas, as variedades não formaram agrupamentos claros conforme a origem de desenvolvimento destas. Podemos observar na figura 13 que a maioria das variedades formou um grande grupo, enquanto que a variedade 3-44 e a espécie *M. ternifolia* ficaram isoladas. Porém estas duas não formaram um agrupamento consistente, sugerindo que estas possam ter sido agrupadas no dendrograma não por sua semelhança, mas sim pela diferença com o restante do material. Peace *et al.* (2002) e Peace (2002) também alocaram as variedades em comum aos dois estudos, em um mesmo grupo pela análise de PCA. As variedades HAES 2-46, HAES 8-16, HAES 3-44, HAES 6-60 e HAES 7-41 ficaram muito próximas, estando as duas últimas quase sobrepostas (PEACE *et al.*, 2002). Peace (2002) observou pela mesma análise que a espécie *M. ternifolia* não formou agrupamento com outras variedades ou espécies, assim como no presente estudo.

Com base nos resultados apresentados podemos verificar que o desenvolvimento das variedades de macadâmia foi realizado por diferentes *backgrounds* genéticos, que aliado à escassez de informações sobre o *pedigree*, torna as relações de parentesco pouco claras.

Apesar da falta de informações precisas sobre os cruzamentos realizados no passado e dos possíveis erros de identificação das variedades, estas possuem potencial para a continuidade dos programas de melhoramento genético de macadâmia, visto que possuem poucas gerações de melhoramento.

5 CONCLUSÕES

- * O desenvolvimento dos marcadores microssatélites a partir da técnica de bibliotecas enriquecidas forneceu bons resultados, porém para a caracterização deste conjunto de variedades, seria necessária a inclusão de mais marcas para melhorar a cobertura do genoma;
- * Os marcadores *diversity array technology* desenvolvidos para o presente estudo forneceram resultados consistentes podendo ser utilizados em outros estudos genéticos com a espécie;
- * Foram identificados possíveis erros de identificação das variedades que compõem o germoplasma no Brasil, sendo necessária uma caracterização das árvores matrizes nos viveiristas para a correção;
- * Apesar da grande similaridade genética entre as variedades do germoplasma brasileiro, estas possuem potencial para a formação de novas populações de melhoramento, sendo necessário um cuidado especial na escolha dos parentais que estarão envolvidos nos cruzamentos;
- * É de extrema importância a introdução das variedades desenvolvidas em outros programas de melhoramento, tanto para a conservação *ex situ*, como para a sua caracterização e posterior uso no programa de melhoramento.

REFERÊNCIAS

AKBARI, M.; WENZEL, P.; CAIG, V.; CARLING, J.; XIA, L.; YANG, S.; USZYNSKI, G.; MOHLER, V.; LEHMENSIEK, A.; KUCHEL, H.; HAYDEN, M.J.; HOWES, N.; SHARP, P.; VAUGHAN, P.; RATHMELL, B.; HUTTNER, E.; KILIAN, A. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. **Theoretical and Applied Genetics**, New York, v. 113, p. 1409-1420, 2006.

ANDERSON, M.J. **PCO** (Principal Coordinate Analysis): a computer program. Auckland: Department of Statistics/University of Auckland, 2003. 8p.

APPELS, R.; MORRIS, R.; GILL, B. S.; MAY, C. E. **Chromosome Biology**. Boston: Kluwer Academic Publishers. 1998. 441p.

ARADHYA, M.K.; YEE, L.K.; ZEE, F.T.; MANSHARDT, R.M. Genetic variability in Macadamia. **Genetic Resources and Crop Evolution**, New York, v. 45, p. 19-32, 1998.

AULER, N.M.F.; REIS, M.S.; GUERRA, M.P.; NODARI, R.O. The genetics and conservation of *Araucaria angustifolia*: genetic structure and diversity of natural populations by means of non-adaptative variation in the state of Santa Catarina. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, n. 3, p. 239-338, 2002.

BERED, F.; BARBOSA NETO, J. F.; CARVALHO, F. I. F. Marcadores moleculares e sua aplicação no melhoramento genético de plantas. **Ciência Rural**, Santa Maria, v. 27, n. 3, p. 513-520, 1997.

BILLOTTE, N.; LAGODA, P.J.R.; RISTERUCCI, A.M.; BAURENS, F.C. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. **Fruits**, Les Ulis, v. 54, p. 277-288, 1999.

BOYTON, S.; HARDNER, C.; KRISTIANSSEN, P. New macadamia hybrids: production and testing. In: McCONCHIE, C. (Org.). **Advanced macadamia improvement program**. Gordon: Horticultural Research and Development Corporation, 1999. p. 54-65.

BORÉM, A.; VIEIRA, M. L. C. **Glossário de biotecnologia**. Viçosa: Folha de Viçosa, 2005. 119p.

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

DIERBERGER, J.E.; MARINO NETTO, L. **Noz Macadâmnia** – uma nova opção para a fruticultura brasileira. São Paulo: Nobel, 1981. 120p. il.

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.

DNA STAR (2000) **Laser Gene**: Expert Sequence Analysis Software. Version 5.01/5.02.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores em análise genética**. 3ª ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.

FUNGARO, M.H.P.; VIEIRA, M.L.C. Marcadores moleculares. In: SERAFINI, L.A.; BARROS, N.M.; AZEVEDO, J.L. **Biotecnologia na agricultura e na agroindústria**. Guaíba: Agropecuária, 2001. p.153-200.

FUTUYMA, D.J. **Biologia Evolutiva**. 2ª ed. Ribeirão Preto: Sociedade Brasileira de Genética/CNPq, 1992. 646p.

GARCIA, A.A.F.; BENCHIMOL, L.L.; BARBOSA, A.M.M.; GERALDI, I.O.; SOUZA, C.L.; SOUZA, A.P. Comparison of RAPD, AFLP and SSR markers for diversity studies in reopical maize inbred lines. **Genetics and Molecular Biology**, Ribeirão Preto, v. 27, n. 4, p. 579-588, 2004.

GOUDET, J. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. **Journal of Heredity**, Oxford, v. 86, n. 6, p. 485-486, 1995.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, Bethesda, v. 137, p. 1121-1137, 1994.

GROSS, C.L.; WESTON, P.H. *Macadamia janseni* (Proteaceae), a new species from central Queensland. **Australian Systematic Botanic**, Collingwood, v. 5, p. 725-728, 1992.

HARDNER, C.; McCONCHIE, C. A strategy for the improvement of macadamia in Australia by breeding. In.: McCONCHIE, C. (Org.). **Advanced macadamia improvement program**. Gordon: Horticultural Research and Development Corporation, 1999. 252p.

HAMILTON, R.A.; STOREY, W.B.; FUKUNAGA, E.T. **Two new macadamia nut varieties and an appraisal of the H.A.E.S. named varieties**. Hawaii: University of Hawaii, Agricultural Experiment Station, 1952. 1v. (Circular, 36).

HAMILTON, R.A.; FUKUNAGA, E.T. **Growing macadamia nuts in Hawaii**. Hawaii: University of Hawaii, Agricultural Experiment Station, 1959. 51p. (Bulletin, 121).

HAMMER, K.; KHOSHBAKHT, K. Towards a 'red list' for crop plant species. **Genetic Resources and Crop Evolution**, New York, v. 52, n. 3, p. 249-265, 2005.

HARDNER, C.; WINKS, C.; STEPHENSON, R.; GALLAGHER, E. Genetic parameters for nut and kernel traits in macadamia. **Euphytica**, Wageningen, v. 117, p. 151-161, 2001.

JACCOUD, D.; PENG, K.; FEINSTEIN, D.; KILIAN, A. Diversity Arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, Oxford, v. 29, n.4, p. e25, 2001.

JAMES, K.E.; SCHNEIDER, H.; ANSELL, S.W.; EVERS, M.; ROBBA, L.; USZYNSKI, G.; PEDERSEN, N.; NEWTON, A.E.; RUSSEL, S.J.; VOGEL, J.C.; KILIAN, A. Diversity Arrays Technology (DArT) for Pan-Genomic Evolutionary Studies of Non-Model Organisms. **PLoS ONE**, Cambridge, v. 3, n. 2, p. e1682, 2008.

KILIAN, A.; HUTTNER, E.; WENZL, P.; JACCOUD, D.; CARLING, J.; CAIG, V.; EVERS, M.; HELLER-USZYNSKA, C.; CAYLA, C.; PATARAPUWADOL, S.; XIA, L.; YANG, S.; THOMSON, B. The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In.: TUBEROSA, R.; PHILLIPS, R.L.; GALE, M. (Eds.). **Proceedings of the international congress “In the wake of the double helix: from the green revolution to the gene revolution”**. Bologna: Avenue Media. 2005. p. 443-461.

LEWIS, P.O.; ZAYKIN, D. **GDA - Genetic data analysis**: Version 1.0(d12) for Windows. The University of New Mexico, Albuquerque, NM, 87131-1051, 1999.

MALO, S.E.; CAMPBELL, C.W. The macadamia. **Fruits Crops Fact**, Hawaii, v 9, p. 1-2, 1972.

MEYER, A.S. **Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes**. 2002. 106 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, Brasil, 2002.

McCONACHIE, I. The macadamia story. **California Macadamia Yearbook**, Los Angeles, v. 26, p. 41-74, 1980.

McCOUCH, S.R.; CHEN, X.; PANAUD, O.; TEMNYKH, S.; XU, Y.; CHO, Y.G.; HUANG, N.; ISHII, T.; BLAIR, M. Microsatellite marker development, mapping and applications in rice genetics and breeding. **Plant Molecular Biology**, Netherlands, v. 35, p. 89-99, 1997.

O’CONNOR, L. **Macadamia Annual Report: Brazil**. [s.l.]: USDA. 2000. 5p.

OJIMA, M.; CAMPO-DALL'ORTO, F.A.; RIGITANO, O. **Germinação de sementes de noqueira-macadâmia**. Campinas: Instituto Agrônomo. 1989. 17p. (Boletim Técnico, 33).

OLIVEIRA, K.M.; PINTO, L.R.; MARCONI, T.G.; MARGARIDO, G.R.A.; PASTINA, M.M.; TEIXEIRA, L.H.M.; FIGUEIRA, A.V.; ULIAN, E.C.; GARCIA, A.A.F.; SOUZA, A.P. Functional integrated genetic linkage map based on EST-markers for sugarcane (*Saccharum* spp.) commercial cross. **Molecular Breeding**, Berlin, V. 20, p. 189-208, 2007.

PEACE, C.P.; VITHANAGE, V.; TURNBULL, C.N.; CARROLL, B.J. Characterizing macadamia germplasm with codominant radiolabelled DNA amplification fingerprinting (RAF) markers. **Acta Horticulture**, Leuven, v. 575, p. 481–490, 2002.

PEACE, C.P.; VITHANAGE, V.; TURNBULL, C.G. N.; CARROLL, B. J. A genetic map of macadamia based on randomly amplified DNA fingerprinting (RAF) markers. **Euphytica**, Wageningen, v. 134, p. 17–26, 2003.

PEACE, C.P.; ALLAN, P.; VITHANAGE, V.; TURNBULL, C.N.; CARROLL, B.J. Genetic relationships amongst macadamia varieties grown in South Africa as assessed by RAF markers. **South African Journal of Plant Soil**, Greenside, v. 22, p. 71–75, 2005.

PEACE, C.P.; MING, R.; SCHMIDT, A.; MANNERS, J.; VITHANAGE, V. Genomics of Macadamia, a recently domesticated tree nut crop. In.: MOORE, P.H. e MING, R. (Eds). **Plant Genetics/Genomics: Genomics of Tropical Crop Plants**. New York: Springer, 2008. v. 1, p. 313-332.

PEACE, C.P. **Genetic characterisation of macadamia with DNA markers**. 2002. 259 p. Tese (Doutorado) – School of Molecular and Microbial Sciences, School of Land and Food Sciences, University of Queensland, Brisbane, Australia, 2002.

PIZA, I.T.; TOLEDO PIZA, P.; VILHENA, S.; MORIYA, L.; RAMOS, O. (2006) Different aspects of Macadamia cultivars in Brazil. In.: SIMPÓSIO INTERNACIONAL DE NOZ MACADÂMIA, 3., 2006, Águas de São Pedro. **Resumos...** Água de São Pedro: Associação Brasileira de Noz Macadâmia (ABM), 2006. p. 47-51.

PRITCHARD, J.K.; WEN, X.; FALUSH, D. Documentation for structure software: Version 2.3. Chicago, Estados Unidos, 1999. Disponível em: <<http://pritch.bsd.uchicago.edu/structure.html>>. Acesso em: 21 jan. 2012.

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

R DEVELOPMENT CORE TEAM (2009). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2011. Disponível em: <<http://www.r-project.org>>. Acesso em: 21 jan. 2011.

RISTERUCCI, A.M.; DUVAL, M.F.; ROHDE, W.; BILLOTE, N. Isolation and characterization of microsatellite loci from *Psidium guajava* L. **Molecular Ecology Notes**, New York, v. 5, p. 745-748, 2005.

ROGERS, J.S. Measures of genetic similarity and genetic distance. In.: _____. **Studies in Genetics VII**. Austin: University of Texas. 1972. p.145-154.

ROMÃO, L.R.C.; PERSEGUINI, J.M.K.C; RUBIANO, L.B.; GONÇALVES, P.S.; MONDEGO, J.C.M. Divergência Genética de Seringueira Baseado em Marcadores EST-SSRs. In.: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA – CIIC, 5., 2011, Campinas. **Resumos...** Campinas: Instituto Agronômico, 2011. 1v.

SANSALONI, C.P.; PETROLI, C.D.; CARLING, J.; HUDSON, C.J.; STEANE, D.A.; MYBURG, A.A.; GRATAPAGLIA, D.; VAILLANCOURT, R.E.; KILIAN, A. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. **Plant Methods**, v. 6, p. 16, 2010. Disponível em: <<http://www.plantmethods.com/content/6/1/16>>. Acesso em: 20 jan. 2012.

SCHLÖTTERER, C. The evolution of molecular markers – just a matter of fashion? **Nature Reviews Genetics**, New York, v. 5, p. 63-69, 2004.

SCHMIDT, A.L., SCOTT, L.; LOWE, A.J. Isolation and characterization of microsatellite loci from Macadamia. **Molecular Ecology Notes**, Oxford, v. 6, n. 4, p. 1060-1063, 2006.

SHIGEURA, G.T.; OOKA, H. **Macadamia nuts in Hawaii**: history and production. Hawaii: [s.n.], 1984. 91p. (Research Extension Series, 39).

SOBIERAJSKI, G.R.; FRANCISCO, V.L.S.F.; ROCHA, P.; GHILARDI, A.A.; MAIA, M.L. Noz-macadâmnia: produção, mercado e situação no Estado de São Paulo. **Informações Econômicas**, São Paulo, v. 36, n. 5, p. 25-36, 2006.

STACE, H.M.; DOUGLAS, A.W.; SAMPSON, J.F. Did 'paleo-popolyploidy' rarely occur in Proteaceae? **Australian Systematic Botany**, Collingwood, v. 11, p. 613-629, 1998.

STEPHENSON, R. Macadamia: domestication and commercialisation. **Chronica Horticulturae**, Ghent, v. 45, n. 2, p. 11-15, 2005.

STEIGER, D.L.; MOORE, P.H.; ZEE, F.; LIU, Z.; MING, R. Genetic relationships of macadamia cultivars and species revealed by AFLP markers. **Euphytica**, Wageningen, v. 132, p. 269-277, 2003.

SCHMIDT, A.L., SCOTT, L.; LOWE, A.J. Isolation and characterization of microsatellite loci from Macadamia. **Molecular Ecology Notes**, New York, v. 6, n. 4, p. 1060-1063, 2006.

TANSKLEY, S. D. Mapping polygenes. **Annual Review of Genetics**, New York, v. 27, p. 205-234, 1993.

TIVANG, J.G.; NIENHUIS, J.; SMITH, O.S. (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. **Theoretical and Applied Genetic**, New York, v. 89, p. 259-264, 1994.

TOLEDO PIZA, P. **Brazilian macadamia export**. São Paulo: Associação Brasileira dos Produtores de Macadâmnia, 2005. 36 p.

VEIGA, A.A.; OJIMA, M.; RIGITANO, O. Novas seleções IAC de noqueira-macadâmia. In.: CONGRESSO BRASILEIRO DE FRUTICULTURA, 6., 1981, Recife. **Resumos...** Recife: Sociedade Brasileira de Fruticultura, 1981. p. 908-914.

VITHANAGE, V.; WINKS, C.W. Isozymes as genetic markers for Macadamia. **Scientia Horticulturae**, Leuven, v. 49, p. 103-115, 1992.

WEBER, Z.; MAY, P.E. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. **American Journal of Human Genetic**, New York, v. 44, p. 388-396, 1989.

WEIR, B.; COCKERHAM, C.C. Estimating F-statistics for the Analysis of Population Structure. **Evolution**, Lawrence, v. 38, p. 1358-1370, 1984.

WENZEL, P.; CARLING, J.; KUDRNA, D.; JACCOUD, D. HUTTNER, E.; KLEINHOF, A.; KILIAN, A. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. **PNAS**, New York, v. 101, n. 26, p. 9915-9920, 2004.

WRIGHT, S. **Evolution and the genetics of populations: variability within and among natural populations**. Chicago: University of Chicago Press, 1978. v. 4, 580p.