



VALQUIRIA MARTINS PEREIRA

**DIVERSIDADE GENETICA E ESTRUTURA
POPULACIONAL DE *Elaeis oleifera* POR MEIO
DE DArTSeq**

LAVRAS - MG

2015

VALQUIRIA MARTINS PEREIRA

DIVERSIDADE GENETICA E ESTRUTURA POPULACIONAL DE
***Elaeis oleifera* POR MEIO DE DArTSeq**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Agronomia, para a obtenção do título de Doutor.

Orientador

Dr. Manoel Teixeira Souza Junior

Coorientadores

Dr. Alexandre Alonso Alves

Dr. Eduardo Fernandes Formighieri

LAVRAS - MG

2015

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Pereira, Valquiria Martins.

Diversidade genética e estrutura populacional de *Elaeis oleifera*
por meio de DArTSeq / Valquiria Martins Pereira. – Lavras: UFLA,
2015.

128 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2015.

Orientador: Manoel Teixeira Souza Junior.

Bibliografia.

1. Palma de Óleo. 2. *E. oleifera*. 3. Diversidade Genética. I.
Universidade Federal de Lavras. II. Título.

VALQUIRIA MARTINS PEREIRA

**DIVERSIDADE GENETICA E ESTRUTURA POPULACIONAL DE
Elaeis oleifera POR MEIO DE DArTSeq**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Agronomia, para a obtenção do título de Doutor.

APROVADA em 27 de fevereiro de 2015.

Dra. Julceia Camillo	Universidade de Brasília
Dr. Eduardo Fernandes Formighieri	Embrapa Agroenergia
Dr. Alexandre Alonso Alves	Embrapa Agroenergia
Dr. Márcio de Carvalho Moretzsohn	Embrapa Recursos Genéticos e Biotecnologia

Dr. Manoel Teixeira Souza Junior
Orientador

**LAVRAS – MG
2015**

AGRADECIMENTOS

A Deus, primeiramente e principalmente...

A minha família, especialmente meus pais Sr. Miguel e Sra. Maria de Lourdes, por acreditarem sempre em mim, me apoiar, pelo amor e confiança. Amo vocês...

Aos meus irmãos José Pascoal e Liama, pelo apoio incondicional, sempre. Esse sucesso também pertence a vocês!

Ao meu Orientador, Dr. Manoel Teixeira, pela orientação, paciência, apoio e oportunidade para realização da Tese.

A Universidade Federal de Lavras e ao Programa de Biotecnologia Vegetal, pela oportunidade concedida para realização do doutorado. Assim como também agradecer a todos os professores que fazem parte do Programa, obrigada pelos ensinamentos.

Ao Dr. Alexandre Alonso e Dr. Eduardo Fernandes pela Co-Orientação, paciência, ajuda e dedicação no desenvolvimento da Tese.

A Embrapa Agroenergia, pela disponibilização do que foi necessário para a realização desta pesquisa.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudo concedida.

A FINEP (Financiadora de Estudos e Projetos do Ministério da Ciência e Tecnologia) pelo financiamento do projeto ao qual foi parte da Tese.

Ao LGB/AGROENERGIA (Laboratório de Genética e Biotecnologia) e todos que fazem parte e caminharam junto nestes cinco anos.

Aos membros da Banca que se disponibilizaram a contribuir com o manuscrito da Tese: Dra. Julceia Camillo, Márcio de C. Moretzsohn, Dr. Eduardo F. Formighieri, Dr. Alexandre A. Alves e os suplentes Dra. Sara de A. Rios, Dr. Guy de Capdeville e Dr. Bruno Laviola.

A equipe da Embrapa Amazônia Ocidental pela colaboração com as pesquisas e coletas de materiais.

A colaboração nesta longa jornada de Julcéia, André Leão, Marcelo Picanso, Luiz Galli, Rodrigo Furtado, Lucas, Marcelo de Paula e demais colegas.

Aos amigos pela força sempre..., Anne Costa, Julcéia Camillo, Jaire Alves, Tatiana Costa, Eduardo Fernandes, Gisele Domciano, Renata Duarte, Lucimeire Costa, Conceição da Silva, Marília Burle, Ana Paula Ribeiro, Gláucia Emy, Paula Jaramillo, Glaycon, Jomar, Claudia Reis, Camila Marques e Tatiane Balzon, Luciane Favarin e Paloma Reis. Meu muito obrigado.

Os amigos (as) e colegas que mesmo distante sempre estiveram presente: Alberto Melo, Suely Xavier, Janivan Suassuna, Livia Campos e família, Sandra Guimarães, Marcelo Helmich, Patrick Callegari, Flávio Costa e família, Whellyson Cordão, Erica Almeida, Jean Carlo, Ylana Claudia, Paulo Henrique, José Zinho, Rafaela Reis e família, Valdeci França, Rayanna Campos, Priscila Targino, Débora Sá, Amanda França, Livia Pinto, Tamires Amorim, Juliana Alves, Antônio Suassuna e Aldo Muniz.

A todos que estava na expectativa para que essa etapa fosse concluída!

Obrigada!

Princípios da eficiência: não temer o futuro e nem idolatrar o passado. O insucesso é apenas uma oportunidade de começar um novo com mais inteligência. O passado só nos serve para mostrar nossas falhas e fornecer indicações para o progresso no futuro. O fracasso é a oportunidade de começar de novo inteligentemente. (Henry Ford)

RESUMO

O objetivo deste estudo foi caracterizar a estrutura genética de 19 populações nativas de *Elaeis oleifera* (caiaué) e avaliar a diversidade genética entre subamostras que compõem o Banco Ativo de Germoplasma (BAG) da Embrapa, por meio de marcadores SNPs (*Single Nucleotide Polymorphism*) e PAVs (*Presence/Absence Variants*). Um total de 552 plantas mantidas pela Embrapa foi utilizado para extração de DNA utilizando o protocolo CTAB. A genotipagem foi realizada em plataforma DArTSeq. A filtragem dos marcadores foi realizada com base em parâmetros de *Call Rate* ($>0,90$) e *MAF* ($>0,05$). Após estas análises, o número de marcadores SNPs foi reduzido a 1.667, e acrescentando-se o parâmetro de *Q-value* (> 2) o número de PAVs, foi reduzido a 3.187. Foram realizadas análises de diversidade genética para verificar a relação entre subamostras de *E. oleifera* com marcadores PAVs, assim como também para caracterizar a estruturação populacional por meio de marcadores SNPs. Verificou-se a existência de moderada diversidade genética entre subamostras/populações nativas de *E. oleifera* (0,301), podendo ser observada a relação genética inter e intra-populacional. As subamostras mais relacionadas entre si são (Autazes e Careiro) e as mais distantes geneticamente são (BR 174 e Tonantins). O índice F, que mede o coeficiente de endogamia, foi relativamente baixo para todas as subpopulações (0,073 em média). Por outro lado, a heterozigosidade observada (H_o) foi moderada (0,185 em média), o que pode ser um indicativo de que as subpopulações se reproduzem principalmente por meio de cruzamentos entre indivíduos não aparentados. O índice F_{ST} foi, no entanto, elevado (0,315 em média), o que demonstra a existência de estruturação, ou a diferenciação genética entre as subpopulações *E. oleifera*. Análise da estrutura de fato revelou a existência de 02 a 03 maiores aglomerados e a AMOVA indicou que apenas 54% da variância molecular ocorrem dentro das populações. Por conseguinte, para a conservação *ex situ*, as coleções devem incluir como muitos sites que possível. Esses resultados reforçam ainda a hipótese que a Amazônia Brasileira deve ser um centro de diversificação secundário da espécie. A existência de variabilidade genética moderada entre e dentro de populações de *Elaeis oleifera*, indica possibilidade de exploração comercial da espécie. Para fins de melhoramento, diferentes subamostras/subpopulações devem ser testadas em cruzamentos a fim de maximizar a diversidade na população base.

Palavras-Chave: Palma de Óleo. *E. oleifera*. Diversidade Genética. PAVs. SNPs. DArTSeq.

ABSTRACT

The objective of this study was to characterize the genetic structure of 19 native populations of *Elaeis oleifera* (caiaué) and assess the genetic diversity among samples that make up the Active Germplasm Bank (BAG) of Embrapa, through SNP (*Single Nucleotide Polymorphism*) and PAVs (*Presence/Absence Variants*) markers. A total of 552 plants maintained by Embrapa was used for DNA extraction using the CTAB protocol. Genotyping was performed on DArTSeq platform. The filtering of the markers was performed based on Call Rate (> 0.90), and MAF (> 0.05). After these analyzes, the number of SNP markers was reduced to 1,667, and after adding the Q-value parameter (> 2) the number of PAVs, was reduced to 3,187. Genetic diversity analyzes were performed to verify the relationship between *E. oleifera* subsamples with PAVs markers, as well as to characterize the population structure by means of SNPs markers. We verified the existence of moderate genetic diversity among samples/native populations of *E. oleifera* (0,301) while also observing the existence of inter and intra-population genetic relationship. The subsamples which are more related to each other are (Autazes and Careiro) and the most distant genetically are (BR 174 and Tonantins). The F index, which measures the inbreeding coefficient, was relatively low for all subpopulations (0.073 on average). Conversely, the observed heterozygosity (H_o) was moderate (0.185 on average) which may be indicative that the subpopulations reproduce mainly by crosses between unrelated individuals. The F_{ST} index was, however, high (0.315 on average), demonstrating the existence of structuration, or genetic differentiation among the *E. oleifera* subpopulations. Structure analysis indeed revealed the existence of 02 to 03 major clusters and the AMOVA indicated that only 54% of the molecular variance occurs within populations. Consequently, for ex situ conservation, the collections should include as many sites as possible. These results further support the hypothesis that the Brazilian Amazon should be a secondary center of diversification of species. The existence of moderate genetic variability among and within populations of *Elaeis oleifera*, points for commercial exploitation of the species. For breeding, different subsamples/subpopulations should be tested in crosses to maximize diversity.

Keywords: Oil Palm. *E. oleifera*. Genetic Diversity. PAVs. SNPs. DArTSeq.

LISTA DE FIGURAS

CAPÍTULO 2

- Figura 1 Fluxograma do procedimento DArT-Seq baseado em NGS (*Next Generation Sequencing*).....54
- Figura 2 Representação Gráfica da Análise de Coordenadas Principais (PCoA) para as 19 populações nativas de *E. oleifera* em seis regiões amazônicas. Análise realizada pelo programa GenAIEx65
- Figura 3 Dendrograma gerado a partir de distâncias genéticas pelo Índice de Coincidência Simples e pelo método não ponderado de *Neighbour-joining*, representando o agrupamento das 552 amostras de *E. oleifera*. Análise realizada pelos programas Darwin5 e FigTree67
- Figura 4 Média dos valores de: a) LnP(D) de Pritchard et al. (2000) e b) Evanno et al. (2005) para 10 repetições de simulações no programa Structure com $k = 1$ a 10 para 552 indivíduos de *E. oleifera*. Análise realizada pelo programa Structure Harvester74
- Figura 5 Estrutura genética de populações de *E. oleifera* quando avaliadas por $K = 2$ pelo Programa Structure76
- Figura 6 Estrutura genética de populações de *E. oleifera* pertencentes a seis regiões amazônicas, quando avaliadas de 2 a 4 níveis de estruturação genética pelo Programa Structure.....77
- Figura 7 Análise da frequência de marcadores SNPs no genoma de referência *Elaeis guineensis*. Os 16 cromossomos foram divididos em 130 intervalos de 5Mb.....78

CAPÍTULO 3

- Figura 1 Distribuição de frequência das estimativas de distância genética, utilizando o coeficiente de dissimilaridade de Jaccard, obtidas entre os pares de 206 subamostras de *E. oleifera*. Análise realizada através do Programa Darwin v5.5..... 101
- Figura 2 Dispersão Gráfica de 206 subamostras de *E. oleifera* obtida pelo método de Análise das Coordenadas Principais (PCoA), utilizando o coeficiente de dissimilaridade de Jaccard. Análise realizada através do Programa Darwin v5.5 103
- Figura 3 Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para 206 subamostras de *E. oleifera* que compõe o BAG da Embrapa. Análise realizada através do Programa Darwin v5.5..... 105
- Figura 4 Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para 29 subamostras (representadas por 2 indivíduos cada) de *E. oleifera* que compõe o BAG da Embrapa, evidenciando possíveis erros de alocação de indivíduos em subamostra..... 110
- Figura 5 Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para 159 subamostras (representadas por 3 indivíduos cada) de *E. oleifera* que compõe o BAG da Embrapa, evidenciando possíveis erros de alocação de indivíduos em subamostra..... 111

LISTA DE TABELAS

CAPÍTULO 2

- Tabela 1 Dissimilaridades baseadas no coeficiente de Coincidência Simples entre indivíduos e dentro de populações e regiões, para 552 indivíduos de *E. oleifera*, por meio de 1.667 marcadores SNPs. N = número da amostra. Análise realizada pelo programa Darwin560
- Tabela 2 Estatística descritiva dos 1.667 SNPs caracterizados em 552 subamostras de *E. oleifera*: tamanho da amostra (N), número de alelos (Na), número efetivo de alelos por locos (Ae), heterozigosidade esperada (He), heterozigosidade observada (Ho), e índice de fixação (*f*) para as 19 populações. Valores entre parênteses são estimativas de desvio padrão. Análise realizada pelo programa GenAIEx.....62
- Tabela 3 Estimativas de diversidade genética. Tamanho da amostra (N), número de alelos privados (AP), conteúdo de informação de polimorfismo (PIC), estimativas de frequência alélica máxima (Fmáx) e tamanho efetivo populacional (Ne) em 19 populações de *E. oleifera*, por meio de 1.667 marcadores SNPs. Análise realizada pelo programa GenAIEx.....64
- Tabela 4 Diversidade genética dentro de populações e regiões de *E. oleifera*. Número de Amostras (NA), frequências alélicas baseadas no Índice de Shannon (SHa), Percentagem de locos polimórficos por população (%LP), diversidade média dentro de populações (P), diversidade média dentro de regiões (R). Análise realizada pelos programas Genes e GenAIEX69

Tabela 5	Distância Genética entre 19 populações de <i>E. oleifera</i> pelo método de Nei 1972, a partir de 1.667 marcadores SNPs. Análise realizada pelo programa GenAIEX.....	71
Tabela 6	Análise de Variância Molecular (AMOVA) para 552 subamostras de <i>E. oleifera</i> representando 19 famílias agrupadas em seis regiões geográficas da Floresta Amazônica Brasileira, por meio de 1.667 marcadores SNPs. Análise realizada pelo programa Genes.....	72
Tabela 7	Estatísticas F - Wright e fluxo gênico (Nm) para 19 populações de <i>E. oleifera</i> . Análise realizada pelo programa GenAIEx.....	74

CAPÍTULO 3

Tabela 1	Agrupamento baseado pelo método de Tocher Modificado, utilizando o coeficiente de dissimilaridade de Jaccard como medida de distância genética, para 206 subamostras de <i>E. oleifera</i>	107
Tabela 2	Análise de Variância Molecular (AMOVA) para 206 subamostras de <i>E. oleifera</i> representadas por 19 famílias distribuídas em seis regiões geográficas da Floresta Amazônica Brasileira.....	108
Tabela 3	Porcentagem de erro de alocação em subamostras de <i>E. oleifera</i>	109

SUMÁRIO

CAPÍTULO 1 INTRODUÇÃO GERAL	15
1 INTRODUÇÃO	15
2 REFERENCIAL TEÓRICO	19
2.1 Introdução	19
2.2 O caiaué - <i>Elaeis oleifera</i> (Kunth) Cortés	21
2.3 <i>Elaeis oleifera</i> como fonte de variabilidade	23
2.4 Recursos genéticos de <i>E. oleifera</i>	24
2.5 Recursos genômicos de <i>E. oleifera</i>	28
2.6 Desafios e perspectivas para pesquisa em <i>E. oleifera</i>	36
3 CONSIDERAÇÕES FINAIS	39
REFERÊNCIAS	40
CAPITULO 2 Diversidade e estrutura genética de populações nativas de <i>Elaeis oleifera</i> inferidas por meio de marcadores SNPs ...	47
1 INTRODUÇÃO	49
2 MATERIAL E MÉTODOS	53
2.1 Material Biológico	53
2.2 Extração de DNA, purificação e quantificação	53
2.3 Genotipagem-por-sequenciamento em plataforma DArTSeq	53
2.4 Mapeamento dos marcadores SNPs	55
2.5 Análise de Variabilidade Genética	55
2.6 Análise de Estruturação genética das populações	57
3 RESULTADOS	59
3.1 Parâmetros Genéticos Populacionais - Índice de Diversidade	59
3.2 Parâmetros Genéticos Populacionais - Análise de Estruturação Genética	72
3.3 Cobertura do genoma de <i>Elaeis oleifera</i> por marcadores SNPs	75
4 DISCUSSÃO	79
4.1 Diversidade genética entre populações	79
4.2 Estrutura Populacional de <i>Elaeis oleifera</i>	83
5 CONCLUSÕES	86
REFERÊNCIAS	87
CAPÍTULO 3 Genotipagem de alta densidade via plataforma DArTSeq fornece novas informações sobre a extensão e a estrutura da diversidade genética da coleção de germoplasma brasileira de <i>Elaeis oleifera</i>	92
1 INTRODUÇÃO	94
2 MATERIAL E MÉTODOS	97
2.1 Material Biológico	97

2.2	Genotipagem-por-sequenciamento em plataforma DArTSeq	97
2.3	Análise de dados	98
3	RESULTADOS	100
3.1	Diversidade Genética	100
3.2	Organização/Estrutura da diversidade amostrada no BAG	101
3.3	Partição da variância genética	107
3.4	Erros de alocação de subamostras e de indivíduos dentro de subamostras	108
4	DISCUSSÃO	112
4.1	A extensão da diversidade genética de <i>E. oleifera</i> é menos restrita que se supunha	112
4.2	O uso de uma bateria de alta densidade de marcadores moleculares proporcionou uma maior discriminação de subamostras por meio de técnicas dispersivas, aglomerativas e de otimização	114
4.3	O conhecimento detalhado da organização da diversidade genética da coleção permite se traçar estratégias de conservação, coleta e melhoramento mais eficientes	116
5	CONCLUSÕES	119
	REFERÊNCIAS	120
	ANEXOS	123

CAPÍTULO 1 Introdução Geral

1 INTRODUÇÃO

O caiaué (*Elaeis oleifera*) é uma palmeira nativa da Amazônia Brasileira que tem despertado interesse como fonte de diversidade genética para os programas de melhoramento genético de palma de óleo (*Elaeis guineensis*) no continente Americano. A espécie tem se destacado por suas características de interesse agrônomo, como porte reduzido, qualidade do óleo e por apresentar tolerância às principais doenças que afetam a cultura da palma de óleo no continente.

A palma de óleo é responsável pela maior parte do fornecimento de óleo vegetal consumido no mundo. Sua alta capacidade de produção por unidade de área faz com que seu rendimento médio seja de sete a dez vezes superior à produtividade da soja (RIOS et al., 2012), espécie que ocupa posição de destaque no Brasil para produção de óleo. Embora ocupe atualmente apenas 5% da área total utilizada para o cultivo de espécies oleaginosas no mundo, a palma de óleo responde por 33% da produção mundial de óleo vegetal (SINGH et al., 2013a). Suas aplicações são extensas, em setores como a indústria de tintas e vernizes, saboaria, medicamentos e cosméticos fazem uso rotineiro do óleo de palma. Mais recentemente, o óleo extraído da palma têm sido utilizado também para a produção de biocombustíveis como o biodiesel. Entretanto, seu principal uso é na indústria alimentícia (BARCELOS, 1995).

A demanda mundial por óleo de palma vem crescendo progressivamente, tendo passado de 18 milhões de toneladas em 1998 para aproximadamente 56 milhões de toneladas em 2013 (FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, 2015). Estimativas indicam que em 2015 a demanda possa chegar a 63 milhões de

toneladas e considerando-se que o mercado mundial de biocombustíveis está em plena expansão, esta poderá ser bem maior. No Brasil, a palma de óleo é considerada a principal espécie capaz de complementar a soja no abastecimento das usinas produtoras de biodiesel (BECKER, 2010), principalmente, após o lançamento, pelo Governo Federal, dos programas de Produção Sustentável de Palma de Óleo e Produção e Uso de Biodiesel.

Como forma de responder a esta crescente demanda, além do aumento da área plantada (esforço orientado no Brasil por meio de um extenso estudo de mapeamento edafoclimático), vem-se buscando também desenvolver cultivares mais produtiva e principalmente, tolerante a doenças, sendo um dos principais limitadores da expansão dos cultivos na América. Nesse sentido, os programas de melhoramento têm buscado a produção de híbridos interespecíficos entre a palma de óleo e o caiaué. Esta ação visa introduzir novas características aos materiais comercialmente disponíveis, principalmente a resistência ao Amarelecimento Fatal – AF, uma vez que a principal área de expansão da cultura (estado do Pará) coincide com a área de maior ocorrência de AF, cujo agente etiológico ainda é desconhecido (SANTOS, 2010).

O caiaué também representa uma importante fonte de variabilidade para resistência à Fusariose (*Fusarium oxysporum* f.sp. *elaeidis*), além de características relacionado à arquitetura da planta. A espécie americana apresenta porte reduzido, o que pode propiciar um aumento considerável no período de exploração econômica da cultura além de elevado teor de ácidos graxos insaturados, que resulta em um óleo de melhor qualidade industrial. Sua incorporação aos programas de melhoramento genético de palma de óleo poderá resultar assim em incremento da produção, melhora na qualidade de óleo, e conseqüentemente, colaborar com a sustentabilidade da cultura da palma de óleo no país (RIOS et al., 2012).

O caiaué é uma espécie amplamente distribuída nas Américas Central e do Sul, principalmente na região Amazônica. Extensas expedições de coleta de germoplasma têm sido realizadas desde as décadas de 1970/80. No Brasil, uma grande coleção de germoplasma de *Elaeis* é mantida pela Embrapa e foi estabelecida com base em uma série de expedições organizadas pela Instituição em parceria com o CIRAD (*Centre de Coopération Internationale en Recherches Agronomique pour le Développement*), nas áreas amazônicas. Assim como no Brasil, outras coleções de germoplasma de *E. oleifera* são mantidas na Malásia, Costa do Marfim e Costa Rica (CORLEY; TINKER, 2003).

A coleção brasileira de palma de óleo tem sido estudada com a finalidade de se verificar a extensão e a organização da diversidade genética contida, uma vez que o conhecimento do modelo de variação entre os materiais é importante na definição de estratégias a serem utilizadas nos programas de melhoramento da espécie (OOI et al., 1981). Esses estudos têm sido desenvolvidos, na maior parte, com base em marcadores moleculares e um dos primeiros resultados publicados com base nos materiais do banco de germoplasma da Embrapa, foi descrito por Barcelos et al. (2002). Estes autores identificaram a existência de quatro grupos diferentes e geograficamente distintos: 1. Brasil; 2. Peru; 3. América Central/Norte da Colômbia e 4. Suriname/Guiana Francesa. Posteriormente, Moretzsohn et al. (2002) observaram a existência de variabilidade genética intra e interpopulacionais em *E. oleifera*. Estes trabalhos, no entanto, foram baseados em uma quantidade limitada de marcadores e, considerando-se que *E. oleifera* é uma espécie alógama, o nível de diversidade existente talvez não possa ser adequadamente verificado com uma quantidade reduzida de marcadores.

Tecnologias como a genotipagem-por-sequenciamento, entretanto, têm tornado possível a realização de análises genética em grande escala em um curto prazo de tempo ao permitir a genotipagem de milhares de marcadores de

maneira rápida e com custo acessível. Assim, tendo em vista a importância que *E. oleifera* hoje representa para o programa de melhoramento genético de palma de óleo e, dada à necessidade de se as possibilidades que novas ferramentas genômicas oferecem, o presente estudo teve por objetivos: (i) avaliar a estrutura populacional de populações nativas de *E. oleifera* por meio de marcadores SNPs (*Single Nucleotide Polymorphisms*) e (ii) avaliar a diversidade genética entre subamostras de *E. oleifera* que compõem o Banco Ativo de Germoplasma (BAG) da Embrapa, por meio de marcadores PAVs (*presence/absence variants*).

2 REFERENCIAL TEÓRICO

2.1 Introdução

No cenário global, o óleo de palma responde por 33% do total de óleos vegetais produzidos e por 45% de todo o óleo vegetal utilizado na indústria alimentícia, embora ocupe uma área equivalente a apenas 5% do total cultivado com espécies oleaginosas. Com mais de oito milhões de hectares plantados no mundo, a palma de óleo é responsável pela produção de mais de 40 milhões de toneladas de óleo; tendo superado já em 2008 a produção de óleo de soja (FARIA et al., 2010). No Brasil, a produção local é insuficiente para atender ao mercado interno, sendo necessária a complementação desta por meio de importações. Na possibilidade de uso de palma de óleo para a produção de biocombustíveis, em especial o biodiesel, há a necessidade de aumentar a oferta interna de óleo, para evitar aumento no déficit da balança comercial devido a um possível aumento da importação deste produto. Para produção de biodiesel no Brasil, o óleo de soja é a principal matéria prima utilizada, tendo sido no último ano responsável por aproximadamente 73% da produção, seguido de gordura animal com 22%, e o óleo de algodão, com aproximadamente 3%. O uso de óleo de dendê para a produção de biodiesel no Brasil ainda é residual, isto é, menor que 1% (AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E COMBUSTÍVEIS, 2015).

A alta demanda por energia no mundo globalizado tem resultado em elevados níveis de poluição decorrentes, em grande parte, do uso dos combustíveis fósseis. Alinhados com a questão da poluição, têm-se ainda os efeitos decorrentes do aquecimento global e o fato das reservas de combustíveis fósseis serem finitas. Em decorrência disto, torna-se necessário o desenvolvimento de fontes alternativas de energia, de preferência não poluidoras

e renováveis. Uma dessas alternativas, no que tange aos combustíveis líquidos, que movem quase que a totalidade do setor de transportes no mundo, é a produção de biodiesel (produzido a partir de óleos vegetais ou gordura animal). O biodiesel produzido a partir de óleo vegetal é um combustível que emite menos gases nocivos do que as fontes convencionais (ISSARIYAKUL et al., 2008).

Em decorrência da promoção da diversificação de matérias-primas para a produção de biodiesel no Brasil, combinada com possibilidades de promoção de uma maior inserção regional e social nesta cadeia de produção de biocombustíveis, a palma de óleo africana (*E. guineensis*) apresenta-se como uma das principais fontes alternativas de matéria-prima disponível. Isto se deve, entre outros fatores, ao elevado rendimento de óleo, entre 5 e 7 toneladas ha⁻¹ ano⁻¹ e ao alto poder calorífico do mesmo. Porém, o cultivo desta palmeira no Brasil, especificamente no estado do Pará – maior produtor nacional, com > 90% do óleo produzido – apresenta uma grande limitação, que é a ocorrência do Amarelecimento Fatal – AF. O AF é uma doença que acomete as plantas de palma de óleo, levando-as invariavelmente à morte. Não se identificou até o momento fonte de resistência ao AF em coleções de germoplasma de *E. guineensis* e, após mais de 30 anos de pesquisa, a única alternativa economicamente viável encontrada para contornar o problema é a exploração da palma de óleo americana ou caiaué (*E. oleifera*), mediante desenvolvimento de híbridos interespecíficos.

O germoplasma de *E. oleifera* tem sido explorado em duas linhas: melhoramento do híbrido F1, buscando-se identificar combinações de caiaué com dendezeiro para produção de híbridos interespecíficos F1 com potencial de cultivo comercial, e introdução de características do caiaué no dendezeiro através de retrocruzamentos, utilizando o dendezeiro como genitor recorrente para alcançar altas produtividades e o caiaué como genitor doador de

características de interesse e que não ocorrem naturalmente em dendê (BARCELOS; NUNES; CUNHA, 2000).

Um dos grandes problemas enfrentados pela cultura da palma de óleo no Brasil é a reduzida capacidade de fornecimento e baixa diversidade de sementes produzidas. A capacidade produtiva da Embrapa, maior produtora nacional, está entre 4,5 a 5 milhões de sementes por ano, número insuficiente para atender a um aumento em larga escala da área plantada com palma de óleo no país. O programa de Pesquisa, Desenvolvimento e Inovação (PD&I) de Palma de Óleo da Embrapa, que é o maior atualmente em curso no Brasil e que tem o melhoramento genético de cultivares desta espécie como prioridade, vem sendo apoiado fortemente pelo governo federal – FINEP/MCTI - desde 2009, no sentido de atender às demandas internas decorrentes de políticas governamentais de apoio a empresas de produção de óleo de palma no país.

A existência de variabilidade genética caracterizada tanto para dendê quanto para caiaué traz expectativas promissoras para os programas de melhoramento genético existentes, com possibilidade de ganho expressivo com as sucessivas gerações de seleção, retrocruzamentos e recombinação, ou seja, permitindo a expansão da cultura da palma de óleo no país de modo mais sustentável.

2.2 O caiaué - *Elaeis oleifera* (Kunth) Cortés

A espécie *E. oleifera* é nativa, mas não endêmica da flora do Brasil, amplamente distribuída na região Amazônica. Tem despertado o interesse de pesquisadores por ser uma importante fonte de variabilidade (e.g. resistência a doenças, arquitetura de plantas e etc) para o melhoramento genético do dendê (*E. guineensis*). Por esta razão, ao longo das últimas décadas, têm sido empreendidos esforços na coleta e formação de bancos de germoplasma da

espécie, sobretudo na América Latina, com vistas a atender às demandas dos programas de melhoramento genético (RIOS et al., 2012). O gênero *Elaeis*, pertencente à classe *Liliopsida* (Monocotiledônea), ordem *Arecales* (*Palmales*), família *Areaceae* (*Palmaceae* ou *Palmae*), subfamília *Arecoideae*, tribo *Cocoseae* (*Cocoinaea*) e subtribo *Elaeidinae*. O gênero *Elaeis* é composto por duas espécies: *E. guineensis* (dendê ou palma de óleo) e *E. oleifera* (caiaué), ambas diploides e com 16 pares de cromossomos ($2n = 32$) (CRONQUIST, 1981). Segundo Rajanaidu (1986a), o dendê é originário do Oeste Africano, com alguma dispersão no Centro-Oeste do continente; enquanto que o caiaué é encontrado nas Américas Central e do Sul, disperso em vários países, inclusive no Brasil (RAJANAIDU, 1986b). O caiaué não tem expressão comercial devido ao fato de sua produtividade de óleo ser inferior à do dendezeiro. Segundo Clement (1999), é considerada uma espécie incipiente, isso porque vem sendo modificada ou melhorada pela intervenção humana.

O interesse no germoplasma do caiaué se deve ao fato de que algumas características desta espécie podem ser incorporadas ao programa de melhoramento de *E. guineensis*. Dentre as principais características observadas destacam-se: baixo porte das plantas, qualidade do óleo (maioria dos ácidos graxos apresenta elevada insaturação), bem como tolerância ao AF e murcha de *Fusarium*, algumas das doenças que afetam os cultivos da palma de óleo (BERGAMIN-FILHO et al., 1998; HARDON; RAO; RAJANAIDU, 1985; RENARD; NOIRET; MEUNIER, 1980).

Entretanto, parâmetros genéticos de *E. oleifera* ainda são pouco conhecidos. Conhece-se, por exemplo, muito pouco sobre a extensão da variabilidade genética existente entre os materiais coletados e mantidos em bancos de germoplasma, ou sobre a estruturação genética das populações nativas, entre outros fatores importantes para a definição de estratégias em programa de melhoramento (OOI et al., 1981). Entender a diversidade genética

entre e dentro de espécies cultivadas e silvestres é ainda essencial para desenvolver estratégias de conservação, coleta e utilização do germoplasma no desenvolvimento de novas variedades.

2.3 *Elaeis oleifera* como fonte de variabilidade

Uma das principais linhas de programas de pesquisa com palma de óleo na América Latina é a busca por fontes de resistência a doenças (RIVAS; BARBIERI; MAIA, 2012), pois acredita-se que um dos principais fatores que inibem os investimentos na expansão da palma de óleo na região seja a ausência de conhecimentos sobre as doenças como o AF, que dizimam os plantios (RAMOS; VEIGA; FURLAN JUNIOR, 2006). A estratégia atual dos programas de melhoramento genético da palma de óleo no Brasil visa à produção e avaliação de híbridos interespecíficos entre palma de óleo africana e palma de óleo americana, a seleção dos melhores para reprodução e uso em plantios comerciais ou para obtenção de gerações avançadas de retrocruzamentos (CHIA et al., 2009). Assim, o conhecimento acerca da variabilidade contida em coleções de germoplasma torna-se informação privilegiada uma vez que permite a melhoristas delimitar um conjunto de genótipos (com base em padrões de diversidade e estruturação) a ser triado em bioensaios para resistência a doenças. Pouco, no entanto, se sabe sobre a variabilidade genética do caiaué, espécie que apresenta resistência a várias doenças importantes da cultura, em nível molecular.

O entendimento da diversidade genética de espécies cultivadas e silvestres é essencial para desenvolver estratégias de conservação, coleta e utilização do germoplasma no desenvolvimento de novas variedades. Especificamente no caso da palma de óleo, o emprego das técnicas convencionais de melhoramento é dificultado por diversos fatores, destacando-

se o longo período da fase juvenil, a demora em estabilizar a produção, a alta heterozigosidade dos melhores indivíduos, e a ineficiência das técnicas de melhoramento em minimizar o mascaramento imposto pelas influências do ambiente. Por ser uma cultura perene com longo ciclo de produção e devido ao alto custo de manutenção e avaliação dos experimentos, é necessário definir o período de tempo mínimo para uma avaliação estável de produção dos híbridos interespecíficos (TURNER; YONG, 1969).

Apesar de esforços na coleta e conservação, a caracterização do germoplasma de *E. oleifera* ainda se encontra em fases iniciais quando comparado a *E. guineensis*. Os principais desafios de pesquisa para o caiaué são, portanto: conhecer profundamente a diversidade genética disponível no Banco Ativo de Germoplasma (BAG); desvendar a estrutura e organização populacionais; e identificar genótipos ou populações superiores para futuros cruzamentos e produção de sementes. Devido ao fato das coleções existentes atualmente serem, em geral, pouco avaliadas, a avaliação dos recursos genéticos da cultura deve ser considerada assim como de alta prioridade, a fim de se efetivamente aproveitar a variabilidade existente em programas de melhoramento genético (BARCELOS et al., 2002; MORETZSOHN et al., 2002).

2.4 Recursos genéticos de *E. oleifera*

Um dos primeiros bancos de germoplasma da espécie foi estabelecido na Costa Rica, a partir de 1967 pela *United Brands Company*. A coleção foi formada por indivíduos coletados inicialmente na Costa Rica, Panamá e Colômbia e posteriormente, nos anos 1975/76 foi enriquecida com a realização de coletas em diferentes populações no Suriname, Honduras, Nicarágua e Brasil. A coleção contava com 4974 plantas divididas em 326 acessos, coletados em

sete países e 36 zonas de coleta (ESCOBAR, 1981) e atualmente conta com 370 acessos representativos de todas as principais regiões produtoras, além de materiais silvestres originados de ambientes específicos (RAJANAIDU; AINUL, 2013).

Além da coleção mantida na Costa Rica, atualmente estão disponíveis no mundo mais três coleções de germoplasma de *E. oleifera* distribuídas da seguinte forma: (1) na Malásia, a *Malaysian Palm Oil Board* (MPOB) mantém uma coleção, com acessos coletadas entre os anos de 1981 a 2004 na Colômbia, Panamá, Costa Rica, Honduras, Brasil e Suriname; (2) no Equador é mantida uma coleção, cujas plantas foram selecionadas com base na espessura de mesocarpo; e (3) no Brasil, a Embrapa mantém a segunda maior coleção de *E. oleifera*, contendo 226 acessos coletados durante a década de 1980, em parceria com o instituto francês *Coopération Internationale en Recherche Agronomique Pour le Développement* (CIRAD), ao longo da bacia do Rio Amazonas (RAJANAIDU; AINUL, 2013), os materiais foram coletados em seis regiões distintas: Região de Manaus, Rio Amazonas, Rio Negro, Região de Caracará, Rio Solimões e Rio Madeira e mantidos no BAG de caiaué da Embrapa.

Alguns estudos da diversidade genética foram realizados nos últimos anos, com base em características morfológicas de diferentes acessos de *E. oleifera* mantidos no banco de germoplasma no Brasil, pertencente à Embrapa e em outros países da América Latina. Rios et al. (2012) avaliaram subamostras de caiaué coletadas na região de Coari, Amazonas, e observaram a existência de correlação positiva e significativa entre algumas variáveis morfológicas e o número de frutos por cacho. Camillo (2012), com base na biometria de frutos de 18 acessos distintos de *E. oleifera* da bacia amazônica, observou que estes tendem a formar grupos genéticos, de acordo com a distribuição geográfica da espécie, porém com significativas diferenças entre acessos de uma mesma

origem e, algumas características agronômicas importantes podem ser associadas a caracteres morfológicos.

Escobar (1981) observou diferenças significativas entre acessos mantidos em banco de germoplasma na Costa Rica e foi uns dos primeiros autores a quantificar esta variabilidade através de características vegetativas como a conformação de plantas, pecíolo, coloração dos frutos e composição do óleo. Rajanaidu, Tan e Rao (1983) realizaram um estudo sobre diferentes aspectos morfológicos em populações coletadas nas Américas do Sul e Central. Dentre os parâmetros avaliados, a biometria de frutos mostrou diferenças significativas entre as populações dos diferentes países e principalmente, entre populações de um mesmo país. Mohd-Din, Rajanaidu e Jalani (2000), estudaram os aspectos morfológicos de progênies coletadas no Panamá, Costa Rica, Colômbia e Honduras e observaram diferenças significativas entre progênies de diferentes países quanto ao peso médio de frutos e peso médio de diásporos e amêndoas.

Em estudos realizados por Barcelos (1986), foi evidenciado que o caiaué é encontrado ao longo do Rio Madeira, em solo conhecido como terra preta do índio, como é comum também no baixo Rio Solimões e alto Rio Amazonas. Sua raridade em ambientes naturais ao longo do Rio Madeira é um indicativo de que o caiaué foi introduzido intencionalmente pelos povos indígenas, muito provavelmente na época pré-colombiana. Além disso, Wessels Boer (1965) sugere que o caiaué tenha sido introduzido na Amazônia, vindo da América Central, embora as análises moleculares de Barcelos, Cunha e Nouy (2001) não ofereçam apoio claro para essa proposta.

Já Martins, Miranda e Nunes (2003), quando compararam a morfologia polínica entre populações de *E. oleifera* coletadas na região amazônica, verificaram que, na comparação com outros métodos de análise de diversidade genética (ecótipos, isoenzimas ou DNA nuclear), a divisão dos grupos pode

coincidir ou não, dependendo do estudo realizado. Os autores relatam ainda, a divisão das populações em três grupos distintos, coincidentes com a localização geográfica, em populações do norte do Amazonas, baixo Rio Madeira e um terceiro grupo com populações espalhadas ao longo do Rio Solimões, Amazonas e afluentes menores. Resultados semelhantes são descritos por Barcelos et al. (2002), para populações da espécie situadas nas Américas Central e do Sul. O estudo, com marcadores RFLP e AFLP, mostrou que a diversidade genética observada nas populações brasileiras (0,157) foi maior do que aquela observada em populações coletadas no Peru (0,042), América Central (0,047), Guiana Francesa (0,042) e Suriname (0,035), agrupadas de acordo com a origem geográfica.

Em outro estudo sobre a diversidade genética de *E. oleifera* no Brasil, verificou-se que os acessos apresentam níveis moderados de diversidade genética (0,675), quando comparados com aqueles observados em *E. guineensis* (0,472), apesar das diferenças de tamanho de amostragem; os grupos de similaridade formados estavam correlacionados com a dispersão geográfica da espécie ao longo da bacia amazônica e não, com a distância geográfica entre as populações e que, a maior diversidade genética encontrada foi intrapopulacional (MORETZSOHN et al., 2002).

A espécie *E. oleifera* vem sendo utilizada como fonte de recursos genéticos nos programas de melhoramento de palma de óleo e, apesar do isolamento geográfico entre as espécies, estas apresentam compatibilidade de cruzamento com produção de híbridos férteis, a exemplo temos o híbrido interespecífico (HIE) BRS Manicoré, que é um dos três únicos HIE lançados no mundo e foi desenvolvido e registrado no MAPA (Ministério da Agricultura, Pecuária e Abastecimento) pela EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) (RIOS et al., 2012). Apesar de não ser utilizado comercialmente, o caiaué apresenta características favoráveis a serem introduzidas no genoma de

palma de óleo, e por isso seu cultivo tem sido intensificado nas últimas décadas (BARCELOS; CUNHA; NOUY, 2001).

2.5 Recursos genômicos de *E. oleifera*

Os desafios são constantes na agricultura moderna, levando os programas de melhoramento genético, ao longo dos anos, a buscar estratégias eficientes para a caracterização dos recursos genéticos e populações de melhoramento assim como para a efetiva seleção de genótipos que apresentem desempenho superior. Assim, como ferramenta-suporte aos programas de melhoramento genético, as técnicas baseadas na utilização de marcadores moleculares surgiram para agregar informações às etapas de caracterização e seleção de genótipos superiores, permitindo estudar em nível genômico espécies e indivíduos. Com o avanço da tecnologia têm-se usado cada vez mais o auxílio de ferramentas biotecnológicas como suporte a programas de melhoramento genético. Para a espécie *Elaeis oleifera*, este panorama não é diferente e alguns estudos já foram realizados na área da genômica.

Shah e Cha (2000) isolaram e caracterizaram cDNA específico, MO1 (434 pb) em *E. oleifera* e demonstraram que sua maior expressão ocorre no mesocarpo entre a quinta e vigésima semana após a antese. O método de apresentação diferencial foi usado para isolar cDNAs correspondentes aos transcritos que se acumulam durante o período da síntese de lípidos, entre 12 e 20 semanas após a antese (WAA) no mesocarpo de duas palmeiras de óleo, *Elaeis guineensis* e *Elaeis oleifera*. Foi extraído RNA total do mesocarpo e da amêndoa de frutos tenera (15 WAA), foram utilizados para obter os padrões de expressão diferencial de genes entre estes tecidos nas duas espécies. Os autores descrevem o isolamento e a caracterização de um clone de cDNA específico, MO1 (434 pb), que foi demonstrado ser específico no mesocarpo e

exclusivamente para a espécie de *E. oleifera*. O sequenciamento deste fragmento mostrou homologia com a enzima sesquiterpeno-sintase. O seu maior clone de cDNA, pMO1 (1072 pb), isolado a partir de uma biblioteca de cDNA de *E. oleifera*, confirmou que codifica para sesquiterpeno-sintase. Sesquiterpeno é um composto defensivo (fitoalexinas) secretados em planta adultas contra fungos ou infecções microbianas (STEELE et al., 1998; BACK; CHAPPEL, 1995). Daí o potencial de *E. oleifera* a resistência a algumas doenças causadas pela infecção fúngica (RAO; CHIN; RAJANAIDU, 1989). Portanto, o clone isolado que codifica para uma sintase de sesquiterpeno, a qual está envolvida na síntese de um sesquiterpeno específico podendo atuar como um composto defensivo nesta espécie.

A sequência completa de 1976 pb foi obtida utilizando o método de RACE de 5%. A hibridação por Northern mostrou que Mo1 e pMO1 em mRNA transcritos são altamente expressos apenas no mesocarpo de *E. oleifera* de 5 a 20 WAA. Não sendo detectada expressão na amêndoa (WAA 12-17) e em tecidos vegetativos de ambas as espécies nem no mesocarpo de *E. guineensis*. Este resultado foi o primeiro a ser divulgado sobre o isolamento e a caracterização de expressão no mesocarpo em palma de óleo. O isolamento de genes específicos no mesocarpo e na amêndoa é importante para o isolamento de promotores específicos para a expressão de transgenes alvo. Os autores usaram a exposição diferencial de mRNA para procurar diferenças de expressão gênica em amêndoas e no mesocarpo de ambas as espécies de palma de óleo, em 15 (WEE) período em que ocorre a síntese de óleo ativa. Estes resultados relatam o uso do método de exibição diferencial para isolar um gene que é expresso especificamente no mesocarpo de *E. oleifera* assim como também os primeiros ensaios sobre o isolamento de um gene de tecidos específicos da espécie de óleo de palma (SHAH; CHA, 2000).

Asemota e Shah (2004) verificaram que o gene oleoyl-thioesterase- ACP está presente no caiaué, à sequência foi comparada com outros genes thioesterases no Centro Nacional de Informações sobre Biotecnologia (NCBI). Apresentou 72% de homologia em nível de nucleótidos, através da região de codificação examinadas e 83% de identidade (90% de positivos) em nível de aminoácidos e possui homologia com o mesmo gene presente na canola (*Brassica napus*). Foi realizada uma análise de Northern deste transcrito onde mostrou que o oleoil-ACP thioesterase amplificado é fortemente expressa em 20 semanas após a antese (WAA) no mesocarpo do fruto de *E. oleifera*. Este resultado é consistente pelos estudos até o momento terem mostrado que em todas as espécies de planta, até agora examinadas, foi encontrado a atividade da enzima oleoil-thioesterase. Este gene que codifica esta enzima em *E. oleifera* foi encontrado por ter considerável similaridade de sequências com outros thioesterases clonados anteriormente em outras espécies. O fato de o fruto amadurecer totalmente a cerca de 28 WAA considerando que esta transcrição é expressa apenas na 20 WAA, no entanto, os autores sugerem que pode haver outros thioesterases ativo nesta espécie em outros estágios de maturidade do fruto.

San e Shah (2005) estudaram a expressão diferencial de genes e caracterização de cDNA em tecido específico em clones de *E. guineensis* e *E. oleifera*, sendo possível demonstrar suas especificidades e separar em três grupos. O grupo um composto por quatro clones (KT3, KT4, KT5 e KT6) que são genes específicos da amêndoa tipo tenera de *E. guineensis* e *E. oleifera*. O segundo grupo é representado pelo clone FST1, que tem o gene específico do mesocarpo em planas do tipo tenera entre *E. guineensis* e *E. oleifera*. E o terceiro grupo representado pelos clones MLT1, MLT2 e MLO1 que são genes expressos no mesocarpo e na folha. O sequenciamento de nucleotídeos e busca de homologias de dados no GenBank revelou que os clones KT3 e KT4

codificam para a mesma proteína de maturação PM3. Enquanto os clones MLT1 e MLT2 codificam para S-ribonuclease ligação proteica e fibrilina, respectivamente. Os outros clones (KT5, KT6, FST1 e MLO1) não apresentaram qualquer homologia significativa com qualquer proteína conhecida.

Ismail et al. (2009) verificaram que a análise funcional do promotor Sesquiterpene Synthase (Sesqpro) em *Elaeis oleifera* revelou expressão tecidual não específica e ativação do mesmo sob condições de estresse, demonstrando assim que a ação da sesqpro é específica e foi influenciada por indução de ácido jasmônico. A fim de se investigar as sequências promotoras do gene da sintase sesquiterpeno que são responsáveis pela indução da atividade de GUS por ácido jasmônico exógeno, todas as construções promotoras foram testadas no que diz respeito à indutibilidade de expressão do GUS. Sesquiterpeno é um composto de defesa secretado em plantas superiores contra fungos ou infecções microbianas e é catalisada pela enzima sintetase de sesquiterpeno. Para estudar os elementos e fatores de transcrição relacionados com a atividades de defesa, 100 μ M de ácido jasmônico foi adicionado ao meio de cultura líquido de tecidos transformados. O ácido jasmônico está entre os indutores exógenos mais aplicados para a indução de mecanismos de defesa de culturas de tecidos vegetais (VAN ETTEN et al., 1994). Ácido jasmônico e seus derivados éster metílico (jasmonato de metilo, MJ) são usados para induzir a produção de metabólitos secundários em plantas (SANZ et al., 2000). Os mesmos foram propostos como compostos principais de sinalização na indução do acúmulo de vários metabólitos secundários (SZABO; THELEN; PETERSEN, 1999).

Os benefícios na saúde são potenciais de carotenóides como anti câncer e agentes antioxidantes e têm sido recentemente demonstrada. A *Elaeis oleifera* em particular, é conhecida por ser a mais rica fonte natural de caroteno. No entanto, as espécies não têm sido exploradas comercialmente devido ao seu rendimento extremamente baixo de óleo. Rasid et al. (2008) descrevem o

isolamento de um clone de cDNA que codifica o gene fitoeno-sintetase (psy) a partir de *E. oleifera* através de amplificação por RT-PCR. O estudo também foi realizado para se obter um padrão de expressão de palma de óleo psy no desenvolvimento de frutos por análise RT-PCR. Os resultados indicaram que o gene é altamente regulado ao longo de desenvolvimento do fruto do óleo de palma. O padrão de expressão psy se mostrou bem correlacionado com acúmulo de luteína em tecidos jovens e α - e β -caroteno nos tecidos mais velhos do mesocarpo. Esta observação demonstrou que o psy em palma de óleo era altamente regulado para o desenvolvimento de tecidos e o acúmulo de carotenos para armazenamento.

Zaki et al. (2010) desenvolveram dez marcadores microssatélites que foram caracterizados a partir de 1500 sequências da biblioteca genômica de *E. oleifera*. Os marcadores foram utilizados para avaliar a diversidade genética de coleções de germoplasma de *E. oleifera* de quatro países da América do Sul (Colômbia, Costa Rica, Panamá e Honduras). A maioria dos marcadores mostraram transferibilidade para o dendezeiro enquanto dois marcadores mostrou transmissibilidade para a família *Arecaceae*. Mostrando que a taxa de transferibilidade entre espécies de palmeiras através de marcadores SSR revela resultados promissores para estudos genéticos comparativos. Os autores concluem que estes marcadores microssatélites para *E. oleifera* deverão tornar-se ferramentas úteis para determinar a estrutura da população e a conservação das populações.

Bhore e Shah (2012) estudaram a transformação genética de *Elaeis oleifera* em embriões zigóticos imaturos para verificar a expressão do gene Tioesterase Proteína Transportadora (PASTA), principal gene envolvido na biossíntese de ácidos graxos plastidial, sendo este regulador do acúmulo do ácido palmítico. Em *Elaeis oleifera* o teor de ácido palmítico no mesocarpo do fruto, pode ser minimizado por silenciamento gênico pós-transcricional do gene

PATE. Os autores chegaram a supor que o silenciamento de genes PATE pós-transcricional em tecido do mesocarpo do fruto em *E. oleifera* pode aumentar os níveis de ácidos palmitoléico (C16:1), esteárico (C18:0) e oléico (C18:1).

Um estudo genômico recente elucidada em parte a origem das espécies de palma de óleo. Singh et al. (2013a) descrevem a sequência do genoma de palma de óleo Africano de 1,8 gigabase (Gb), a fonte predominante de produção de óleo em todo o mundo. Um total de 1.535Gb de dados de sequência e transcriptoma montado a partir de 30 tipos de tecido foram utilizados para prever, pelo menos, 34,802 genes, incluindo genes da biossíntese de óleo e homólogos de WRINKLED1 (WRI1), e outros reguladores da transcrição, que são altamente expressos no núcleo. Os autores relatam ainda no mesmo estudo sobre o *draft* de sequências para *Elaeis oleifera*, que tem o mesmo número de cromossomos ($2n=32$) e produz híbridos interespecíficos férteis com *E. guineensis*, mas parece ter divergido no Novo Mundo. Duplicações segmentais de braços cromossômicos definem a origem paleotetraploide das palmeiras. As sequências de palma de óleo permitem a descoberta de genes para características importantes, bem como alterações epigenéticas e somaclonais que restringem o uso de clones em plantações comerciais, e, portanto, devem ajudar a alcançar a sustentabilidade para os biocombustíveis e óleos comestíveis.

Estudos visando identificar QTLs associados à qualidade e características associadas ao óleo da palma vêm sendo desenvolvidos. O estudo realizado para detecção de QTLs para composição de ácidos graxos foi realizado em híbridos interespecíficos (SINGH et al., 2009). Os autores não conseguiram concluir se as ligações marcador/QTL serão válidas para *E. oleifera*. A provável ausência de rearranjo cromossômico entre a *E. oleifera* e *E. guineensis* facilitaria a identificação, validação e utilização de marcadores de QTLs em ambas as espécies. Os QTLs favoráveis identificados podem ser utilizados para análise de coleção de germoplasma para incorporação em programas de melhoramento

genético existentes. Mais tarde, foram descobertos 19 QTLs associados com a composição de ácido graxo em materiais interespecíficos (MONTROYA et al., 2013). Segundo os autores estes QTLs constituem ferramentas úteis para aplicação em programas de melhoramento, no entanto, a maioria dos QTLs associados a características de produtividade que foram identificados são responsáveis por apenas uma percentagem baixa ou moderada da variância observada.

Neste sentido, o desenvolvimento de técnicas robustas que permitam a genotipagem de milhares de marcadores em milhares de amostras em um simples experimento, como por exemplo, a seleção genômica ampla (GWS) é um grande avanço em pesquisas de melhoramento genético vegetal. Uma vez obtidos estes marcadores, são estimados os seus efeitos com base em dados fenotípicos de uma população de estimação. Com base nos efeitos são obtidos os modelos de predição dos valores genéticos genômicos. Uma vez realizada a validação dos modelos em populações modelo, são aplicados em populações de seleção (CROSS et al., 2015; CROSSA et al., 2011; WONG; BERNARDO, 2008).

Com os avanços recentes de ferramentas biotecnológicas e principalmente devido ao desenvolvimento de tecnologias de genotipagem em larga escala (WENZL et al., 2004; JENKINS; GIBSON, 2002), novos marcadores moleculares como SNPs (*Single Nucleotide Polymorphism*) e DArTs (*Diversity Array Technology*) têm permitido eficiente avaliação genética em nível molecular de um amplo número de espécies. Essas tecnologias devido ao seu baixo custo por *data point* têm permitido que grande número de marcadores sejam usados para várias culturas sem a necessidade de conhecimento prévio sobre o genoma estudado (JEHAN; LAKHANPAUL, 2006; HASTBACKA et al., 1994).

O rendimento de frutos em palma de óleo tipo tenera é maior que em frutos do tipo dura, consistindo assim como a base para a produção de óleo de palma comercial em todo o Sudeste Asiático. O mapeamento e a identificação do gene responsável por essa característica no fruto (gene Shell) foram realizados por Sing et al. (2013b). Foi realizado um mapeamento de homozigose usando o pedigree AVROS e resequenciamento completo do genoma. Onde os genes candidatos aparecem como regiões com baixa diversidade nos indivíduos homozigotos puros. A classificação varietal de palma de óleo apresenta frutos em três formas distintas, dura (com espessura larga de casca – homozigoto tipo Sh+ Sh+), pisífera (sem casca – homozigotos Sh- Sh-) e tenera (espessura fina da casca – heterozigoto tipo Sh+Sh-), um híbrido entre dura e pisífera. Esta descoberta contribuiu para novas estratégias moleculares para genotipagem do gênero *Elaeis*, minimizando o tempo de seleção no campo e evitando a contaminação de plantas do tipo Dura, sendo possível também usar a estratégia de separação de plantas com base nos diferentes tipos de frutos e assim possibilitar plantações varietais separadamente (SINGH et al., 2013b).

Com base na sequência do genoma e dados de transcriptoma de *E. guineensis* foi identificado um gene específico, denominado Shell, que determina a natureza da casca do fruto. O tipo tenera é heterozigoto para o gene Shell, uma combinação que se traduz por um rendimento de óleo por fruto 30% maior que o tipo dura (homozigoto dominante), o óleo de palma tem um ciclo reprodutivo muito longo, são necessários até seis anos para que os produtores determinem o tipo de muda. Quando utilizado marcadores genéticos em estudos prévios para elucidar o conhecimento sobre determinada espécie é possível estimar o processo de seleção e reduzir a superfície cultivada, conciliando assim o aumento crescente de óleos e biocombustíveis e o plantio sustentável da espécie (SING et al., 2013b).

Estudos recentes mostram o mapeamento e identificação do gene *Virescens* (*Vir*) que é responsável pela cor do exocarpo no fruto, sendo responsável também como indicador de maturação, para o gênero *Elaeis*. Na pesquisa foram identificados também cinco alelos mutantes independentes de *VIR* em mais de 400 acessos da África Subsaariana que respondem pela dominância negativa *virescens* no fenótipo. Cada mutação resulta na interrupção prematura do domínio carboxi-terminal de *VIR*. A identificação de *VIR* permitirá seleção de plantas ainda em fases iniciais, um ganho de 3-6 anos antes de os frutos serem produzidos, avançando na introgressão de material de elite (SINGH et al., 2014). Entretanto, a descoberta da base genética do papel de *VIR* no fenótipo *virescens* abre caminho para o desenvolvimento de testes genéticos para coloração de frutos bem antes do plantio e para a introgressão da característica desejável em materiais de reprodução de elite. E em conjunto com a identificação do gene *Shell*, permite aos produtores desenvolver linhagens que sejam homozigóticas paternas (pisífera) para uso em programas de melhoramento do gênero *Elaeis* sp. e/ou para a produção de sementes comerciais (HAMA-ALI et al., 2015; CROS et al., 2015; XIÃO et al., 2014; ARIAS et al., 2014).

2.6 Desafios e perspectivas para pesquisa em *E. oleifera*

Embora alguns estudos tenham sido realizados, ainda existem muitas lacunas que devem ser preenchidas para que seja possível um aproveitamento mais eficiente do germoplasma de *E. oleifera* conforme antecipado por Parveez et al. (2015).

Uma estratégia importante para os programas de melhoramento é uma correta caracterização dos bancos de germoplasma existentes, o que permite o conhecimento da diversidade genética preservada. Entretanto, devido ao fato das

coleções em geral serem constituídas por grande número de plantas que ocupam extensas áreas, sua caracterização e a manutenção é bastante difícil. Em geral os acessos têm sido avaliados, quanto a algumas poucas características morfo-agronômicas facilmente avaliadas a campo e/ou através de marcadores moleculares. Neste último caso, contudo, os marcadores são em geral avaliados sempre em uma pequena amostra dos acessos mantidos em condição *ex situ* (RIOS et al., 2011; 2012; CUNHA; LOPES, 2010; BARCELOS et al., 2002; MORETZSOHN et al., 2002; BARCELOS; CUNHA; NOUY, 2001; GHESQUIÈRE et al., 1987).

Uma alternativa viável para aumentar o aproveitamento das informações de diversidade nos bancos, é a formação de coleção nuclear, constituída por uma amostra de acessos, que pode representar o espectro de variabilidade genética da coleção inteira. Uma vez formada a coleção, que além de ser uma segurança genética do BAG original, estarão disponíveis materiais elites para serem explorados com melhor e maior acessibilidade.

Uma das técnicas utilizadas para caracterizar acessos em um banco de germoplasma é o uso da análise de divergência genética. Essa caracterização possibilita a definição de estratégias de trabalho, por proporcionar o conhecimento da variabilidade genética das populações, e, também, o monitoramento dos bancos de germoplasma, gerando informações úteis para a preservação e o uso dos acessos. Fundamentais também são os estudos de genômica populacional, informações relativas à composição genética de populações são importantes na definição das estratégias de melhoramento (GRIFFITHS et al., 2006).

Recentemente com o desenvolvimento e aperfeiçoamento de técnicas de genotipagem em larga escala, novos sistemas de marcadores moleculares como SNP (*Single Nucleotide Polymorphism*) e DarT (*Diversity Array Technology*) têm permitido uma eficiente avaliação genética em nível molecular para

virtualmente qualquer espécie. Tais tecnologias proporcionaram considerável redução do custo por data point permitindo assim que um grande número de marcadores possa ser avaliado em várias culturas. Uma vez gerado um grande número de marcadores distribuídos por todo o genoma de um indivíduo, alguns destes marcadores, além de estarem muito pertos, estarão em desequilíbrio de ligação (LD) com QTLs. Esse desequilíbrio de ligação faz referência à associação não aleatória entre dois genes ou entre um QTL e um loco marcador (JEHAN; LAKHANPAUL, 2006; HASTBACKA et al., 1994).

O uso de estratégias seletivas baseadas em marcadores moleculares para aumentar ganhos genéticos por unidade de tempo não é um conceito novo, principalmente para caracteres de difícil fenotipagem e natureza qualitativa. Estudos de simulação demonstraram que a seleção genômica ampla (GWS) aplicada em populações de plantas permite taxas de ganhos por unidade de tempo maior que a seleção fenotípica e a seleção assistida por marcadores (MAS) (ARIAS et al., 2014; CAVALCANTI et al., 2012; WONG; BERNARDO, 2008).

3 CONSIDERAÇÕES FINAIS

Sendo *E. oleifera* uma cultura ainda pouco estudada, os desafios são grandes. Contudo, as pesquisas têm focado no avanço científico na busca em ampliar o conhecimento sobre a espécie e assim possibilitar uma maior inserção no mercado brasileiro como suporte às principais fontes de matéria-prima (soja, gordura animal e óleo de algodão) para produção de biocombustíveis no Brasil.

REFERÊNCIAS

- AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E COMBUSTÍVEIS. **Boletim mensal do biodiesel**. Rio de Janeiro: Processamento de Gás Natural e Produção de Biocombustíveis, 2015. 13 p.
- ARIAS, D. et al. Molecular characterization of oil palm *Elaeis guineensis* Jacq. of different origins for their utilization in breeding programmes. **Plant Genetic Resources - Characterization and Utilization**, Netherlands, v.12, n. 3, p. 341-348, dec. 2014.
- ASEMOTA, O.; SHAH, F. H. Detection of mesocarp oleoyl-thioesterase gene of the South American oil palm *Elaeis oleifera* by reverse transcriptase polymerase chain reaction. **African Journal of Biotechnology**, Nairobi, v. 3, n. 11, p. 595-598, nov. 2004.
- BARCELOS, E. **Características genético-ecológicas de populações naturais de caiaué (*Elaeis oleifera* (H.B.K.) Cortés) na Amazônia brasileira**. 1986. 108 p. Dissertação (Mestrado em Ciências Biológicas) - Instituto Nacional de Pesquisas da Amazônia, Fundação Universidade do Amazonas, Manaus, 1986.
- BARCELOS, E. et al. **A cultura do *E. guineensis***. Brasília: Embrapa, 1995. 68 p. (Coleção plantar, 32).
- BARCELOS, E. et al. Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. **Pesquisa Agropecuária Brasileira**, Brasília, v. 37, n. 8, p. 1105-1114, ago. 2002.
- BARCELOS, E.; CUNHA, R. N. V. da.; NOUY, B. Recursos genéticos de dendê (*Elaeis guineensis*) e *Elaeis oleifera* disponíveis na Embrapa e sua utilização. In: FURLAN JÚNIOR, J.; MULLER, A. A. (Ed.). **Agronegócio do dendê: uma alternativa social, econômica e ambiental para o desenvolvimento sustentável da Amazônia**. Belém: Embrapa Amazônia Oriental, 2001. p. 131-143.
- BARCELOS, E.; NUNES C. D. M.; CUNHA, R. N. V. Melhoramento genético e produção de sementes comerciais de dendezeiro. In: VIEGAS, I. J. M.; MULLER, A. A. (Ed.). **A cultura do dendezeiro na Amazônia Brasileira**. Belém: Embrapa Amazônia Oriental, 2000. p. 145-170.

BACK, K.; CHAPPELL, J. Cloning and bacterial expression of a sesquiterpene cyclase from *Hyoscyamus muticus* and its molecular comparison to related terpene cyclase, **The Journal of Biological Chemistry**, v. 270, n. 13, p. 7375-7381, mar. 1995.

BECKER, B. “Recuperação de áreas desflorestadas da Amazônia: será pertinente o cultivo da palma de óleo (Dendê)?”. **Confins**, [s.n.], n. 10, 2010. Disponível em: <<http://confins.revues.org/6609>>. Acesso em: 15 nov. 2014.

BERGAMIN FILHO, A. et al. Análise temporal do amarelecimento fatal do dendezeiro como ferramenta para elucidar sua etiologia. **Fitopatologia Brasileira**, Brasília, v. 23, n. 3, p. 391-396, set. 1998.

BHORE, S. J.; SHAH, F. H. Genetic transformation of the American Oil Palm (*Elaeis oleifera*) immature zygotic embryos with antisense Palmitoyl-Acyl Carrier Protein Thioesterase (PATE) Gene. **World Applied Sciences Journal**, Paquistão, v. 16, n. 3, p. 362-369, Dec. 2012.

CAMILLO, J. **Diversidade genética, conservação *in vitro* de germoplasma e análise do conteúdo de DNA nuclear em palma de óleo {*Elaeis guineensis* Jacq. e *Elaeis oleifera* (Kunth) Cortés}**. 2012. 120 p. Tese (Doutorado em Agronomia e Medicina Veterinária) - Universidade de Brasília, 2012.

CAVALCANTI, J. J. V. et al. Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. **Revista Brasileira de Fruticultura**, Jaboticabal, v. 34, n. 3, p. 840–846, set. 2012.

CHIA, G. S. et al. Germinação *in vitro* de pólen de híbridos interespecíficos entre o caiaué e o dendezeiro. **Ciência Rural**, Santa Maria, v. 39, n. 5, p. 1569-1571, ago. 2009.

CLEMENT, C. R. 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and population decline. **Economic Botany**, Bronx, v. 53, n. 2, p. 188-202, Apr. 1999.

CORLEY, R.; TINKER, P. **The oil palm**. 4. ed. Oxford: Wiley- Blackwell, 2003. 592 p.

CRONQUIST, A. **An integrated system of classification of flowering plants**. Nova York: Columbia University Press, 1981. 1262 p.

CROSSA, J. et al. Genomic selection and prediction in plant breeding. **Journal of Crop Improvement**, New York, v. 25, n. 3, p. 239–261, Apr. 2011.

CROS, D. et al. Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). **Theoretical and Applied Genetics**, Berlin, v. 128, n. 3, p. 397-410, mar. 2015.

CUNHA, R. N. V.; LOPES, R. **BRS Manicoré: híbrido interespecífico entre caiaué e o dendezeiro**. Manaus: Embrapa Amazônia Ocidental, 2010. 4 p. (Comunicado Técnico, 85).

ESCOBAR, R. Preliminary results of the collection and evaluation of the American oil palm (*Elaeis oleifera* H.B.K. Cortes) in Costa Rica, Panama and Colombia.” In: PUSHPARAJAH, E.; CHEW, P. (Ed.). **The oil palm in agriculture in the eighties**. Kuala Lumpur: Incorporated Society of Planters, 1981. p. 79-97.

FARIA, I. D. et al. **Utilização de óleo vegetal refinado como combustível: aspectos legais, técnicos, econômicos, ambientais e tributários**. Brasília: Senado Federal, 2010. 64 p. (Textos para discussão, 73).

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Base de dados referentes a 2013**. Washington: FAO, 2013. Disponível em: <<http://faostat.fao.org>>. Acesso em: 24 jan. 2015.

GHEQUIÈRE, M. et al. Polymorphisme enzymatique che *Elaeis oleifera* H. B. K. (*Elaeis melanococca*) analyse des populations du Bassin amazonien. **Oléagineux**, França, v. 42, n. 8, p. 143-153, Aug. 1987.

GRIFFITHS, R. R. et al. Psilocybin can occasion mystical experiences having substantial and sustained personal meaning and spiritual significance. **Psychopharmacology**, Oxford, v. 187, n. 3, p. 268–283, Aug. 2006.

HAMA-ALI, E. O. et al. Association between basal stem rot disease and simple sequence repeat markers in oil palm, *Elaeis guineensis* Jacq. **Euphytica**, Wageningen, v. 202, n. 2, p. 199-206, mar. 2015.

HARDON, J. J.; RAO, V.; RAJANAIDU, N. A review of oil-palm breeding. In: RUSSELL, G. E. (Ed.). **Progress in plant breeding**. London: Butterworths, 1985. p. 139–163.

HASTBACKA, J. et al. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. **Cell**, Cambridge, v. 78, n. 1, p. 1073–1087, Sep. 1994.

ISMAIL, I. et al. Functional analysis of the *Elaeis oleifera* sesquiterpene synthase promoter reveals non-specific tissue expression and activation under stress condition. **American Journal of Plant Physiology**, New York, v. 4, n. 1, p. 24–37, jan. 2009.

ISSARIYAKUL, T. et al. Biodiesel production from mixtures of canola oil and used cooking oil. **Chemical Engineering Journal**, Lausanne, v. 140, n. 1–3, p. 77–85, July 2008.

JEHAN, T.; LAKHANPAUL, S. Single nucleotide polymorphism (SNP) – methods and applications in plant genetics: a review. **Indian Journal of Biotechnology**, India, v. 5, n. 4, p. 435–459, Oct. 2006.

JENKINS, S.; GIBSON, S. High-throughput SNP genotyping. **Comparative and Functional Genomics**, Chichester, v. 3, n. 1, p. 57–66, Feb. 2002.

MARTINS, L. H. P.; MIRANDA, I. P. A.; NUNES, C. D. Morfologia polínica de populações amazônicas de *Elaeis oleifera*. **Acta Amazônica**, Manaus, v. 33, n. 2, p. 159–166, abr./jun. 2003.

MOHD-DIN, A.; RAJANAIDU, N.; JALANI, B. S. Performance of *Elaeis oleifera* from Paramá, Costa Rica, Colombia and Honduras in Malaysia. **Journal of Oil Palm Research**, Malásia, v. 12, n. 1, p. 71–80, June 2000.

MONTOYA, C. et al. Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from *Elaeis oleifera* (H.B.K.) Cortés and oil palm (*Elaeis guineensis* Jacq.). **Tree Genetics & Genomes**, Amsterdam, v. 9, n. 3, p. 1207–1225, Oct. 2013.

MORETZSOHN, M. C. et al. Genetic diversity of Brazilian oil palm (*Elaeis oleifera*) germoplasm collected in the Amazon Forest. **Euphytica**, Wageningen, v. 124, n. 1, p. 35–45, Mar. 2002.

OOI, S. C. et al. Oil palm genetic resources: native *E. oleifera* populations in Brazil offer promising sources. **Pesquisa Agropecuária Brasileira**, Brasília, v. 16, n. 3, p. 385–395, maio 1981.

- PARVEEZ, G. K. A. et al. Biotechnology of oil palm: strategies towards manipulation of lipid content and composition. **Plant Cell Reports**, Berlin, v. 34, n. 4, p. 533–543, Apr. 2015.
- RAO, V.; CHIN, C. W.; RAJANAIDU, N. Biology and performance of Surinam *Elaeis oleifera* (H.B.K.) Cortes, **Elaeis**, Kajang, Selangor, v. 1, n. 32, p. 109–118, mar. 1989.
- RAJANAIDU, N. The oil palm (*Elaeis guineensis*) collection in Africa. In: SOH, A. C.; RAJANAIDU, N.; NASIR, M. (Ed.). **International workshop on oil palm germoplasm and utilization**. Bangi: Palm Oil Research Institute of Malaysia, 1986a. p. 59-83.
- RAJANAIDU, N. *Elaeis oleifera* collection in Central and South America. In: WORKSHOP PALM OIL RESEARCH INSTITUTE OF MALAYSIA, 1986, Bangi. **Proceedings...** Bangi: Kuala Lumpur, 1986b. p. 84-94.
- RAJANAIDU, N.; AINUL, M. M. Conservation of oil palm and coconut genetic resources. In: NORMAH, M. N.; CHIN, H. F.; BARBARA, M. R. **Conservation of tropical plant species**. New York: Springer Science, 2013. Cap. 10, p. 189-212.
- RAJANAIDU, N.; TAN, B. K.; RAO, V. Analysis of fatty acid composition (FAC) in *Elaeis guineensis*, *Elaeis oleifera*, their hybrids and its implications in breeding. **Porim Bulletin**, Malásia, v. 7, n. 7, p. 9-20, Apr. 1983.
- RAMOS, E. J.; VEIGA, A. S.; FURLAN JUNIOR, J. **Comportamento do dendezeiro ao amarelecimento fatal em área de replantio, sob diferentes formas de manejo**. Belém: Embrapa Amazônia Oriental, 2006. 22 p. (Documentos, 243).
- RASID, O. A. et al. RT-PCR amplification and cloning of partial DNA sequence coding for oil palm (*Elaeis oleifera*) phytoene synthase gene. **Asia Pacific Journal of Molecular Biology and Biotechnology**, Malásia, v. 16, n. 1, p. 17-24, 2008.
- RENARD, J. L.; NOIRET, J. M.; MEUNIER, J. Sources and ranges of resistance to *Fusarium* wilt in the oil palms *Elaeis guineensis* and *Elaeis melanococca*. **Oléagineux**, França, v. 35, n. 8-9, p. 387-392, Sept. 1980.

RIOS, S. A. et al. Caracterização fenotípica e diversidade genética em subamostras de caiaué (*Elaeis oleifera*) de origem Coari. In: CONGRESSO BRASILEIRO DE MELHORAMENTO DE PLANTAS, 6., 2011, Búzios. **Anais...** Búzios: SBMP, 2011. 1 CD-ROM.

RIOS, S. da A. et al. **Recursos genéticos de palma de óleo (*Elaeis guineensis* Jacq.) e *E. oleifera* (*Elaeis oleifera* (H.B.K.) Cortés)**. Manaus: Embrapa Amazônia Ocidental, 2012. 39 p. (Documentos, 96).

RIVAS, M.; BARBIERI, R. L.; MAIA, L. C. Plant breeding and *in situ* utilization of palm trees. **Ciência Rural**, Santa Maria, v. 42, n. 2, p. 261-269, fev. 2012.

SAN, C. T.; SHAH, F. H. Differential gene expression and characterization of tissue-specific cDNA clones in oil palm using mRNA differential display. **Molecular Biology Reports**, Dordrecht, v. 32, n. 4, p. 227-235, dec. 2005.

SANTOS, E. A. dos. **Caracterização de dendezeiros subespontâneos com base na produção de frutos e cachos**. 2010. 61 p. Dissertação (Mestrado em Produção Vegetal) - Universidade Estadual de Santa Cruz, Bahia, 2010.

SANZ, M.K. et al. Enhancement of tessaric acid production in *Tessaria absinthoides* cell suspension cultures. **Plant Cell Reports**, Berlin, v. 19, n. 9, p. 821-824, jul. 2000.

SHAH, F. H.; CHA, T. S. A mesocarp-and species-specific cDNA clone from oil palm encodes for sesquiterpene synthase. **Plant Science**, Limerick, v. 154, n. 2, p. 153–160, May. 2000.

SINGH, R. et al. Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. **BMC Plant Biology**, Londres, v. 9, p. 114, Aug. 2009.

SINGH, R. et al. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. **Nature**, London, v. 500, n. 7462, p. 335–339, July. 2013a.

SINGH, R. et al. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. **Nature**, London, v. 500, n. 7462, p. 340–344, July. 2013b.

- SINGH, R. et al. The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. **Nature Communications**, New York, v. 5, n. 4106, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24978855>>. Acesso em: 25 maio 2014.
- STEELE, C.L. Regulation of oleoresinosis in Grand Fir (*Abies grandis*). **Plant Physiology**, Rockville, v. 116, n. 4, p. 1497–1504, apr. 1998.
- SZABO, E.; THELEN, A.; PETERSEN, M. Fungal elicitor preparations and methyl jasmonate enhance rosmarinic acid accumulation in suspension cultures of *Coleus blumei*. **Plant Cell Reports**, Berlin, v. 18, n. 6, p. 485-489, feb. 1999.
- TURNER, H. N.; YOUNG, S. S. Y. **Quantitative genetics in sheep breeding**. Ithaca: Cornell University, 1969. 332 p.
- VAN ETTEN, H. D. et al. Two classes of plant antibiotics: Phytoalexins versus phytoanticipins. **Plant Cell**, Rockville Pike, v. 6, n. 9, p. 1191-1192. Sep. 1994.
- WENZL, P. et al. Diversity arrays technology (DArT) for whole-genome profiling of barley. **Proceedings of the National Academy of Sciences**, Washington, v. 101, n. 26, p. 9915–9920, Feb. 2004.
- WESSELS-BOER, J. G. **The indigenous palms of Suriname**: volume 5. Leiden: E.J. Brill, 1965. 172 p.
- WONG, C. K.; BERNARDO, R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, Berlin, v. 116, n. 6, p. 815-824, apr. 2008.
- XIAO, Y. et al. Exploiting transcriptome data for the development and characterization of gene-based SSR markers related to cold tolerance in oil palm (*Elaeis guineensis*). **BMC Plant Biology**, London, v. 14, n. 384, dec. 2014.
- ZAKI, N. M. et al. Development and characterization of *Elaeis oleifera* microsatellite markers (Pembangunan dan Pencirian Penanda Mikrosatelit *Elaeis oleifera*). **Sains Malaysiana**, Malásia, v. 39, n. 6, p. 909-912, Jan. 2010.

CAPITULO 2 Diversidade e estrutura genética de populações nativas de *Elaeis oleifera* inferidas por meio de marcadores SNPs

RESUMO

Neste trabalho foram avaliados os padrões de diversidade genética e de estrutura de subpopulações de *E. oleifera* na bacia do rio Amazonas. Quinhentas e cinquenta e duas plantas coletadas em 06 diferentes regiões amostrando 19 subpopulações foram genotipadas usando a plataforma DArTSeq. Para caracterizar os sítios polimórficos, alinhamos as sequências em que foram descobertos SNPs, ao genoma de referência disponível ao público (RG) (divididos em 130 intervalos 5M), utilizando BLASTN. Isso nos permitiu verificar a distribuição de todo o genoma dos marcadores, e atribuir os SNPs para um dos 16 cromossomos. Os dados moleculares foram então usados para estimar as frequências alélicas, a heterozigosidade, as estatísticas F, tamanhos efetivos de população (NE) e da genética entre as subpopulações e entre os indivíduos. A distância genética entre os indivíduos foi estimada de acordo com a proporção de alelos comuns e agrupamento realizado utilizando o método de neighbor-joining. Estruturação populacional foi então avaliada através de uma análise bayesiana de agrupamento (Estrutura) e AMOVA. O número mais provável de clusters na análise de estrutura foi determinado pelo valor Δk . A plataforma de genotipagem forneceu mais de 1.500 SNPs de alta qualidade (*Call Rate* > 0,90; *MAF* > 0,05). Destes 812 foram efetivamente mapeados contra o RG com uma média de 6,25 por SNPs bin, proporcionando relativamente boa cobertura de todos os cromossomos. Com base neste conjunto de todo o genoma de marcadores SNP, encontramos apenas a diversidade genética moderada (0,301) entre as subpopulações. O índice F, que mede o coeficiente de endogamia, foi relativamente baixo para todas as subpopulações (0,073 em média). Por outro lado, a heterozigosidade observada (H_o) foi moderada (0,185 em média), que pode ser um indicativo de que as subpopulações venham a se reproduzir principalmente pelos cruzamentos entre indivíduos não aparentados. O índice F_{ST} foi, no entanto, elevado (0,315 em média), o que demonstra a existência de estruturação, ou a diferenciação genética entre as subpopulações de *E. oleifera*. Análise da estrutura revelou a existência de 02 a 03 maiores aglomerados e a AMOVA indicou que apenas 54% da variância molecular ocorrem dentro das populações. Para fins de reprodução, o teste de diferentes subpopulações em cruzamentos híbridos deve provar ser mais eficaz com base na distribuição de SNP variação dentro e entre as subpopulações.

Palavras-chave: *Elaeis oleifera*. Estrutura Genética. Marcadores SNPs.

ABSTRACT

Here, we evaluated the genetic diversity patterns and structure of subpopulations of *E. oleifera* in the Amazon River basin. Five hundred and fifty two plants collected in 06 different regions and sampling 19 subpopulations were genotyped using the DArTSeq platform. To characterize the polymorphic sites, we aligned the sequences on which SNPs were discovered, to the public available reference genome (RG) (portioned in 130 5M intervals) using BLASTN. This allowed us to check for genome-wide distribution of the markers, and to assign the SNPs to one of the 16 chromosomes. The molecular data were then used to estimate the allelic frequencies, the heterozygosity, the F statistics, population effective sizes (N_e) and the genetic among the subpopulations and among individuals. The genetic distance between the individuals was estimated according to the proportion of common alleles and grouping performed using the neighbor-joining method. Population structuration was then evaluated through a bayesian clustering analysis (Structure) and AMOVA. The most probable number of clusters in the Structure analysis was determined by ΔK value. The genotyping platform provided over 1.500 high-quality SNPs (Call Rate > 0.90; MAF > 0.05). Of these 812 were effectively mapped against the RG with an average of 6.25 SNPs per bin, providing relatively good coverage of all chromosomes. Based on this genome-wide set of SNP markers, we found only moderate genetic diversity (0.301) among the subpopulations. The F index, which measures the inbreeding coefficient, was relatively low for all subpopulations (0.073 on average). Conversely, the observed heterozygosity (H_o) was moderate (0.185 on average) which may be indicative that the subpopulations reproduce mainly by crosses between unrelated individuals. The F_{ST} index was, however, high (0.315 on average), demonstrating the existence of structuration, or genetic differentiation among the *E. oleifera* subpopulations. Structure analysis indeed revealed the existence of 02 to 03 major clusters and the AMOVA indicated that only 54% of the molecular variance occurs within populations. Consequently, for ex situ conservation, the collections should include as many sites as possible. For breeding purposes, testing of different subpopulations in hybrids crosses should prove to be more effective based on the distribution of SNP variation within and among subpopulations.

Keywords: *Elaeis oleifera*. Genetic Estructure. SNPs Markers.

1 INTRODUÇÃO

A Floresta Amazônica abrange cerca de 40% do território brasileiro, e é um dos maiores reservatórios mundiais de biodiversidade (AZEVEDO et al. 2007). Estima-se que mais de 5.000 espécies arbóreas existam na região, sendo muitas exploradas comercialmente (FERRAZ et al., 2004). Outras apesar de não exploradas comercialmente, podem conter valiosos bioativos que podem vir a ser exploradas no futuro, dentro do contexto da bioeconomia. Desse modo, é importante que além de uma extensa lista de atividades relacionadas à catalogação e mapeamento das espécies, busque-se uma melhor compreensão das características demográficas, da dinâmica populacional, e dos padrões de diversidade e estruturação genética de espécie chave. Sabe-se que a estrutura genética das populações de plantas reflete as interações de muitos processos diferentes, tais como a história evolutiva de longo prazo de espécies (tais como mudanças na distribuição, a fragmentação do habitat e isolamento da população), mutação, reprodutiva, geográfica e isolamento ecológico, deriva genética, sistema de cruzamento, fluxo gênico e seleção (AMORIN et al. 2015). Por isso, informações detalhadas sobre os padrões de estruturação e diversidade permitem ser traçados planos de produção e de gestão de conservação sustentáveis para espécies nativas. Tais empreendimentos certamente colocarão o Brasil em novo patamar no que concerne a bioeconomia.

Dentre as inúmeras espécies nativas da região Amazônica, o *Elaeis oleifera* classe *Liliopsida* (Monocotiledônea), ordem *Arecales* (*Palmales*), família *Areceaceae* (*Palmae* ou *Palmaceae*), subfamília *Arecoideae*, tribo *Cocoseae* (*Cocoinaeae*) e subtribo *Elaeidinae*, conhecida popularmente como caiaué, vem ganhando atenção devido a seu potencial de incorporação em programas de melhoramento de palma de óleo, ou dendê, que é a mais importante fonte de óleo vegetal no mundo atualmente. Este óleo se presta a

diversos fins, atendendo demandas de diversos setores, com destaque para o de alimentos e o de biocombustíveis. O caiaué é uma importante fonte de variabilidade genética para os programas de melhoramento genético do dendzeiro. Esta espécie é fonte de resistência a algumas pragas e doenças que limitam o cultivo do dendê; apresenta uma taxa anual de crescimento do tronco bem menor que este o que lhe confere baixo porte da estirpe, e que pode reduzir o custo de exploração no que tange à colheita; e apresenta elevado teor de ácidos graxos insaturados, o que dá maior fluidez ao óleo nas condições naturais; além disso, existem indicações de que seus híbridos com palma de óleo possam apresentar maior tolerância que este ao déficit hídrico, aos solos encharcados e facilidade no manejo (BARCELOS; CUNHA; NOUY, 2001).

No programa de melhoramento genético do dendzeiro desenvolvido pela Embrapa, a exploração do caiaué tornou-se imprescindível devido a este apresentar resistência ao Amarelecimento Fatal (AF). Esta enfermidade letal, de etiologia desconhecida, é hoje o maior limitante da produção de dendê no estado do Pará, onde se encontra mais de 90% de toda a área com esta cultura no Brasil. O híbrido interespecífico entre estas espécies também apresenta resistência ao AF. Apesar de conhecida e estudada há décadas, pouco se sabe sobre essa doença, não existindo método de controle e com uma história de milhares de hectares de dendê já dizimados (FRANQUEVILLE, 2001).

Embora o caiaué seja nativo das Américas Central e do Sul, sendo encontrado disperso em vários países, inclusive no Brasil (RAJANAIDU, 1986), muito pouco se conhece sobre os padrões de estruturação e diversidade em populações naturais. A quase totalidade das informações sobre a diversidade da espécie e sua estruturação está voltada para características de grandes coleções de germoplasma (BARCELOS et al. 2002; MORETHZSOHN et al. 2002). O BAG da Embrapa, alvo de alguns estudos de cunho genético é a segunda maior coleção de *E. oleifera* do mundo, tendo sido formada inicialmente por 226

acessos coletados durante a década de 1980, em parceria com o CIRAD (*Coopération Internationale en Recherche Agronomique Pour le Développement*), ao longo da bacia do Rio Amazonas (RAJANAIDU; AINUL, 2013).

Os marcadores genéticos são ferramenta eficientes para descrever os padrões da variabilidade genética de uma determinada população e, com o seu uso, é possível avaliar como esta variabilidade encontra-se distribuída dentro e entre as populações (TELLES et al., 2003). Contudo, embora marcadores dominantes como aqueles usados para se caracterizar a diversidade da coleção de germoplasma da Embrapa (BARCELOS et al. 2002; MORETHZSOHN et al. 2002) permitam se obter parâmetros informativos, como o número de alelos por loco, a porcentagem de locos polimórficos, estes não permitem a obtenção de informações quanto a heterozigosidade observada, heterozigosidade esperada segundo o equilíbrio de Hardy-Weinberg ou o computo dos coeficientes F de Wright, parâmetros comumente utilizados para caracterização da variabilidade genética em nível populacional (GUSSON; SEBBENN; KAGEYAMA, 2005). Atualmente com o desenvolvimento de novas plataformas de genotipagem, especialmente daquelas baseadas no princípio de genotipagem-por-sequenciamento (DAVEY et al. 2011), pode-se genotipar rapidamente um grande número de amostras para virtualmente milhares de marcadores codominantes como os SNPs, a custos bastante competitivos. Há assim atualmente a possibilidade de se empreender um estudo detalhado de genômica populacional de *E. oleifera* obtendo informações até então inéditas para a espécie, e abrindo caminho para estudos mais elaborados de genômica populacional da espécie como por exemplo o mapeamento de assinaturas de seleção natural ou cálculo de parentesco realizado.

Frente ao exposto, o objetivo deste trabalho foi caracterizar a diversidade e a estrutura genética de 19 populações nativas de *E. oleifera*, por

meio de marcadores SNPs. Espera-se que como mencionado anteriormente que as informações padrões de estruturação e diversidade possam permitir a elaboração de planos de gestão de conservação para a espécie.

2 MATERIAL E MÉTODOS

2.1 Material Biológico

Um total de 552 plantas do BAG de caiué, mantido pela Embrapa Amazônia Ocidental no Campo Experimental do Rio Urubu - CERU, e que representam 19 diferentes populações coletadas dentro de seis grandes regiões (MORETZSOHN et al., 2002), foram utilizadas neste estudo. O número de plantas amostradas por região/população encontra-se na Tabela 1.

2.2 Extração de DNA, purificação e quantificação

O DNA total foi extraído a partir de tecido foliar adulto congelado utilizando o protocolo CTAB, com modificações (DOYLE; DOYLE, 1990). O DNA purificado foi então quantificado em NanoDrop® (consideraram-se amostras de boa qualidade aquelas que apresentaram razão 260/280 nm \geq 1,8) e por comparação da intensidade de fluorescência de amostras tratadas como brometo de etídeo (0,5 μ g/mL) com a fluorescência emitida pela série de diluição do padrão de DNA λ em eletroforese em géis de agarose 0,8%, sob luz ultravioleta (UV). As amostras de DNA foram então tratadas com RNase e posteriormente diluídas a 100 ng/ μ l e mantidas a -4°C até o envio para genotipagem.

2.3 Genotipagem-por-sequenciamento em plataforma DArTSeq

A genotipagem foi realizada pela empresa DArT Pty® (Yarralumla, Austrália) usando a plataforma DArTSeq (Figura 1). O DNA genômico das 552 plantas foi enviado do Brasil para a Austrália em gelo-seco, e, depois de

verificada a quantidade e qualidade, foi tratado com uma combinação de enzimas de restrição (uma de corte frequente, *Bst*NI e uma de corte raro, *Pst*I) para reduzir a complexidade do DNA a ser sequenciado. As amostras de DNA de complexidade reduzida foram então individualmente etiquetadas por meio da ligação de adaptadores *Pst*I contendo *barcodes* especificamente desenvolvidos para cada uma das 552 plantas. Os produtos resultantes foram amplificados via PCR e posteriormente as amostras foram misturadas em um *flowcell* e sequenciadas pela plataforma *Illumina HiSeq 2000* (SE 100bp). O adaptador *Pst*I incluiu um primer de sequenciamento, de modo que todos os *tags* gerados foram sempre lidos a partir do sítio de restrição *Pst*I.

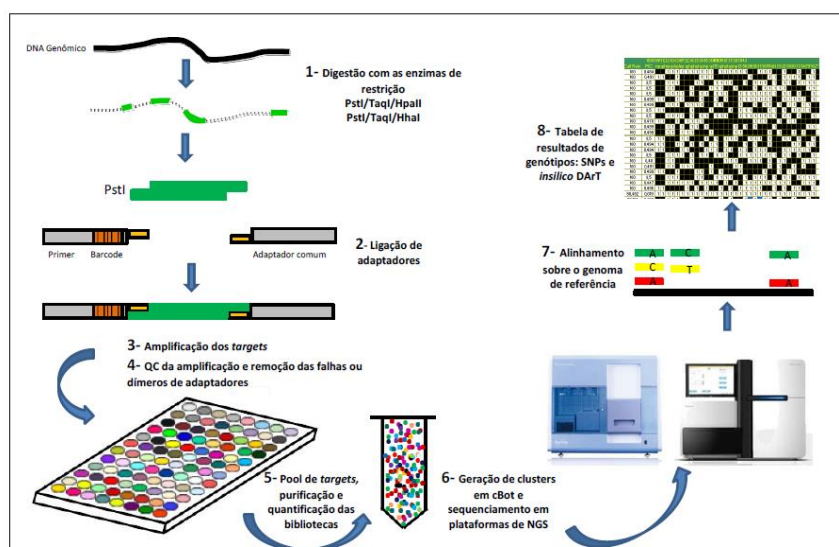


Figura 1 Fluxograma do procedimento DArT-Seq baseado em NGS (*Next Generation Sequencing*)

Fonte: (SANSALONI, 2012).

As sequências resultantes (mais de 2 milhões de *reads* por amostra foram gerados) foram filtradas e alocadas a seus respectivos *datasets* sendo os *barcodes* retirados posteriormente. As sequências foram trimadas em 69 pb (5

pb do sitio de restrição mais 64 bases com *Q score* mínimo de 10). As sequências foram utilizadas para se gerar uma sequência consenso de baixa cobertura para ser utilizada como referência na descoberta de SNPs pelo alinhamento dos *reads* de 69 pb usando o programa *Bowtie* v0.12.

2.4 Mapeamento dos marcadores SNPs

Um conjunto de 5.365 SNPs gerados pelo *pipeline* DArTPty® foi filtrado com base em *Call Rate* (>0,90) e *MAF* (>0,05). O número de marcadores foi reduzido para 1.667, sendo este o número de marcadores utilizado no presente estudo. Estes marcadores foram alinhados contra o genoma de referência de *E. guineensis* (SING et al., 2013) por meio da ferramenta BLASTN (parâmetros padrão). Os resultados foram filtrados, aceitando-se somente sequências em regiões únicas no genoma e de no máximo dois *gaps* por sequência.

2.5 Análise de Variabilidade Genética

A fim de se verificar a variabilidade e a estrutura genética das populações nativas de *E. oleifera*, os dados obtidos foram utilizados para se calcular a matriz de coeficientes de dissimilaridade par-a-par através do coeficiente de Coincidência Simples (SNEATH; SOKAL, 1973), por meio do *software* DARwin v5.5 (PERRIER; JACQUEMOUD-COLLET, 2006). Para computar a dissimilaridade entre duas unidades usando um *dataset* que incluía dados perdidos, o critério de deleção de dados par-a-par, foi adotado ao invés de eliminar variáveis (marcadores) ou unidades que apresentavam dados perdidos da análise. Neste caso, foi estipulado que um mínimo de 70% de dados válidos seria requerido para computar a dissimilaridade entre duas unidades. A partir dos

dados gerados foram então computadas estatísticas descritivas relacionadas à diversidade das populações, como: (1) dissimilaridade média, (2) variância dos dados, (3) distribuição de frequência de dissimilaridade, (4) populações mais diversas e (5) populações mais relacionadas.

Em seguida, a matriz de coeficientes de dissimilaridade de Coincidência Simples foi utilizada para: (6) construir um dendrograma usando o método de análise *Neighbour-joining* não ponderado (SAITOU; NEI, 1987), implementado pelo programa DARwin v 5.0.158 (PERRIER; JACQUEMOUD-COLLET, 2009). Posteriormente, foi utilizado o *software* Fig Tree v1.4.0 (RAMBAUT, 2012) com a finalidade de (7) definir os agrupamentos entre as populações de *E. oleifera* usando escalas distintas de cores no dendrograma e (8) na Análise de Coordenadas Principais (PCoA), e, (9) análise fatorial para obter representações gráficas bidimensionais do relacionamento das 19 subpopulações em termos de dissimilaridade através da PCoA.

A diversidade genética das populações total foi analisada através das estimativas da (10) frequência alélica, (11) heterozigosidade esperada (H_e) e observada (H_o), (12) conteúdo de informação polimórfica (PIC), (13) índice de fixação de Wright e (14) tamanho efetivo populacional com auxílio do programa Power Marker v3.25 (LIU; MUSE, 2005). Foram calculadas as (15) percentagens de locos polimórficos, (16) índice de diversidade de Shannon (NEI, 1972) e (17) matriz de distância genética de Nei (NEI, 1972) através do programa *GenAlex* v6.5 (PEAKALL; SMOUSE, 2012), estimativas estas que podem mostrar como está a diversidade entre populações. Foi estimada (18) a correlação cofenética (CCC) entre os valores de dissimilaridade originais e os valores de dissimilaridades graficamente representados no dendrograma de modo a determinar a precisão com que a árvore representa a diversidade entre populações de *E. oleifera*.

Os dados originais foram também utilizados para se analisar (19) a partição da variação genética total entre regiões Amazônicas amostradas, entre populações e populações dentro de regiões, por meio de uma análise de variância molecular (AMOVA) (EXCOFFIER; SMOUSE; QUATTRO, 1992) por meio do programa Genes (CRUZ, 2013). Os níveis de significância para os componentes de variância foram calculados adotando-se o critério de 1000 permutações.

Análise discriminante (20) foi realizada para averiguar a estrutura genética de populações nativas de *E. oleifera* pelo método de PCoA. A PCoA para populações foi calculada baseando-se na matriz gerada a partir do índice de Coincidência Simples através do programa DARwin.

2.6 Análise de Estruturação genética das populações

Com o objetivo de avaliar a variabilidade e o grau de estruturação da população, foram estimados (21) os parâmetros F_{IT} , F_{IS} e F_{ST} , sendo este último utilizado por meio da avaliação da correlação de todos os alelos de diferentes indivíduos na mesma subpopulação. O parâmetro F_{IT} é a correlação entre alelos dentro de indivíduos em todas as populações, representando o coeficiente de endogamia dos indivíduos em relação ao conjunto de populações, já o parâmetro F_{IS} expressa a correlação dos alelos dentro de indivíduos dentro da população à qual estes indivíduos pertencem, ou seja, representa o grau de endogamia dentro de populações.

O programa Structure v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000) foi utilizado para (22) inferir a estruturação das populações estudadas, através de agrupamento baseado no modelo Bayesiano. A estrutura das populações foi classificada em “K” clusters, de acordo com suas similaridades genéticas. O número de clusters (K) testado variou de 1 a 10, com

10 interações cada. O número de passos do comprimento de *burnin* foi de 10.000 e o de repetições MCMC (*Markov chain Monte Carlo*) de 100.000. Para determinar (23) o número de grupos genéticos (K mais provável) utilizou-se o critério proposto por Evanno, Regnaut e Goudet (2005), através do programa Structure Harvester v0.6.93 (EARL; vonHOLDT, 2012).

3 RESULTADOS

3.1 Parâmetros Genéticos Populacionais - Índice de Diversidade

A partir da análise da 1.667 SNPs e distribuídos ao longo dos 16 pares de cromossomos em 552 plantas foi possível constatar que a diversidade genética em *E. oleifera* é apenas moderada, uma vez que a dissimilaridade média foi de apenas 0,301 com variação de 0,024 a 0,410. A maior dissimilaridade ocorreu entre indivíduos da população de BR 174 e Tonantins. Quando a dissimilaridade é observada a nível populacional as que apresentam maior dissimilaridade entre seus indivíduos são Manicoré (0,358) e Careiro (0,356) e com menor dissimilaridade as populações de Autazes (0,020) e Careiro (0,032) (Tabela 1).

Os resultados da estatística descritiva, referente aos 1.667 marcadores SNPs para as 552 subamostras divididas em 19 populações estão representados na Tabela 2, onde são apresentados os parâmetros: tamanho da amostra (N), número efetivo de alelos por locos (Ae), número de alelos (Na) heterozigidade esperada (He), heterozigidade observada (Ho) e índice de fixação (f).

O número efetivo de alelos, que é o número de alelos igualmente frequentes necessário para se atingir o valor determinado de He variou de 1,140 – Benjamin Constant a 1,440 – Novo Aripuanã e, considerando o desvio padrão apresentado para cada estimativa, todas as populações apresentaram diferenças entre si.

As populações Caldeirão, Manacapuru, Anori, B. Constant e Barcelos apresentam estimativas de coeficiente de endogamia negativas, podendo ser um indicativo de ausência de endogamia nas populações. Para as demais populações o coeficiente de endogamia estimado foi positivo, com exceção apenas da população Iranduba (Tabela 2). A população que apresentou maior coeficiente de endogamia foi a BR-174 (0,256). Contudo, a média do índice de fixação

Tabela 1 Dissimilaridades baseadas no coeficiente de Coincidência Simples entre indivíduos e dentro de populações e regiões, para 552 indivíduos de *E. oleifera*, por meio de 1.667 marcadores SNPs. N = número da amostra. Análise realizada pelo programa Darwin5

Regiões	Populações	N	Média	Máxima	Mínima
	Caldeirão	18	0,174	0,206	0,075
Região de Manaus ¹	Careiro	68	0,201	0,356	0,032
	Manacapuru	3	0,083	0,091	0,072
	Irاندوبا	6	0,162	0,252	0,102
	¹	95	0,212	0,356	0,032
Rio Amazonas ²	Amatari	30	0,206	0,299	0,085
	Autazes	28	0,213	0,336	0,020
	Maués	32	0,222	0,301	0,083
	²	89	0,229	0,336	0,020
Rio Solimões ³	Anori	9	0,186	0,226	0,087
	Benjamin Constant	3	0,108	0,119	0,095
	Coari	54	0,201	0,257	0,088
	Tefé	14	0,211	0,276	0,082
	³	93	0,237	0,321	0,082
Rio Negro ⁴	Acajatuba	29	0,165	0,262	0,086
	Barcelos	2	0	0	0
	⁴	63	0,173	0,286	0,059
Região Caracará ⁵	BR 174	35	0,153	0,303	0,036
	Vila Moderna	18	0,126	0,312	0,047
	⁵	53	0,148	0,314	0,035
Rio Madeira ⁶	Manicoré	140	0,186	0,358	0,060
	Novo Aripuanã	19	0,217	0,291	0,097
	⁶	159	0,198	0,358	0,060
Total de indivíduos:		552	-	-	-
Geral:		-	0,301	0,410	0,024

(Tabela 2) para as 19 populações foi $f = 0,062$, indicando um nível relativamente baixo de endogamia.

Na Tabela 3 estão apresentadas as estimativas para número de alelos privados, conteúdo de informação polimórfica (PIC), maior frequência alélica e o tamanho efetivo populacional (N_e) para as 19 populações de *E. oleifera*. É possível verificar o número de alelos privados por populações, sendo a população V. Moderna a única que não apresentou nenhum alelo privado. As populações que apresentaram maior número de alelos privados foram Coari (56 alelos), Manicoré (37 alelos), Tonantins (33 alelos) e Tefé (26 alelos).

Os valores de PIC foram obtidos para todas as populações, sendo mais altos para as populações de Careiro (0,243), Acajatuba (0,242) e Manicoré (0,241). Tendo apenas as populações Manacapuru, Iranduba, Anori, Benjamin Constant e Barcelos valores de PIC inferiores a 0,200. A literatura relata que valores de PIC inferiores a 0,25 são pouco informativos (BOTSTEIN et al., 1980). Como a estimativa de PIC está fortemente relacionada ao número de amostras por população, e dada ao pequeno número amostral de populações supramencionadas, sendo estas populações as menos representativas amostralmente, os valores de PIC não podem ser diretamente comparados.

Tabela 2 Estatística descritiva dos 1.667 SNPs caracterizados em 552 subamostras de *E. oleifera*: tamanho da amostra (N), número de alelos (Na), número efetivo de alelos por locos (Ae), heterozigosidade esperada (He), heterozigosidade observada (Ho), e índice de fixação (f) para as 19 populações. Valores entre parênteses são estimativas de desvio padrão. Análise realizada pelo programa GenAIEx

Regiões	Populações	N	Na	Ae	Ho	He	f
1	Caldeirão	18	1,691 (0,011)	1,323 (0,009)	0,206 (0,006)	0,188 (0,005)	-0,076 (0,006)
	Careiro	68	1,969 (0,004)	1,374 (0,008)	0,191 (0,004)	0,231 (0,004)	0,160 (0,007)
	Manacapuru	3	1,326 (0,012)	1,218 (0,009)	0,193 (0,008)	0,126 (0,005)	-0,464 (0,009)
	Irاندuba	6	1,558 (0,012)	1,289 (0,008)	0,175 (0,006)	0,175 (0,005)	0,002 (0,011)
2	Amatari	30	1,879 (0,008)	1,385 (0,009)	0,197 (0,004)	0,233 (0,004)	0,103 (0,007)
	Autazes	28	1,915 (0,007)	1,378 (0,008)	0,182 (0,004)	0,236 (0,004)	0,197 (0,008)
	Maués	32	1,896 (0,007)	1,429 (0,009)	0,223 (0,004)	0,256 (0,004)	0,106 (0,007)
3	Anori	9	1,694 (0,011)	1,367 (0,009)	0,233 (0,006)	0,220 (0,005)	-0,063 (0,008)
	B, Constant	3	1,257 (0,013)	1,140 (0,009)	0,134 (0,006)	0,105 (0,004)	-0,257 (0,010)
	Coari	54	1,881 (0,008)	1,392 (0,009)	0,205 (0,004)	0,235 (0,005)	0,098 (0,006)
	Tefé	14	1,732 (0,011)	1,377 (0,009)	0,173 (0,004)	0,225 (0,005)	0,174 (0,009)
	Tonantins	12	1,801 (0,010)	1,377 (0,008)	0,184 (0,004)	0,231 (0,004)	0,148 (0,010)
4	Acajatuba	29	1,837 (0,009)	1,321 (0,008)	0,193 (0,005)	0,197 (0,004)	0,033 (0,006)
	Barcelos	2	1,232 (0,011)	1,164 (0,009)	0,146 (0,007)	0,103 (0,004)	-0,409 (0,012)
	Moura	32	1,765 (0,010)	1,300 (0,008)	0,169 (0,004)	0,183 (0,004)	0,038 (0,006)

“Tabela 2, conclusão”

Regiões	Populações	N	Na	Ae	Ho	He	f
5	BR 174	35	1,853 (0,009)	1,255 (0,007)	0,112 (0,003)	0,166 (0,004)	0,256 (0,009)
	V. Moderna	18	1,810 (0,010)	1,216 (0,006)	0,142 (0,004)	0,148 (0,003)	0,044 (0,007)
6	Manicoré	140	1,990 (0,002)	1,362 (0,008)	0,205 (0,004)	0,222 (0,004)	0,103 (0,005)
	N. Aripuanã	19	1,868 (0,008)	1,440 (0,009)	0,249 (0,005)	0,259 (0,005)	0,035 (0,007)
Média			1,735 (0,003)	1,321 (0,002)	0,185 (0,001)	0,197 (0,001)	0,062 (0,002)

Legenda: Regiões: 1 (Região de Manaus), 2 (Rio Amazonas), 3 (Rio Solimões), 4 (Rio Negro), 5 (Região Caracará) e 6 (Rio Madeira).

É possível observar na Tabela 3 que a frequência alélica máxima (indicador da riqueza alélica) apresenta médias superiores a 0,800 para a maioria das populações (Manacapuru, Iranduba, Maués, Anori, B. Constant. Coari, Tefé, Tonantins, Barcelos, Moura, BR 174 e V. Moderna).

Tabela 3 Estimativas de diversidade genética. Tamanho da amostra (N), número de alelos privados (AP), conteúdo de informação de polimorfismo (PIC), estimativas de frequência alélica máxima (F_{máx}) e tamanho efetivo populacional (N_e) em 19 populações de *E. oleifera*, por meio de 1.667 marcadores SNPs. Análise realizada pelo programa GenAlEx

Regiões	Populações	N	AP	PIC	F _{máx}	N _e
Região de Manaus	Caldeirão	18	3	0,228	0,799	19,48
	Careiro	68	17	0,243	0,790	58,62
	Manacapuru	3	1	0,147	0,857	5,59
	Irاندuba	6	1	0,178	0,843	5,98
Rio Amazonas	Amatari	30	8	0,229	0,796	27,19
	Autazes	28	2	0,233	0,792	23,39
	Maués	32	7	0,223	0,805	28,93
Rio Solimões	Anori	9	4	0,190	0,825	9,60
	Benjamin Constant	3	12	0,133	0,872	4,03
	Coari	54	56	0,218	0,808	49,18
Rio Negro	Tefé	14	26	0,214	0,812	11,92
	Tonantins	12	33	0,224	0,800	10,45
	Acajatuba	29	2	0,242	0,787	28,07
	Barcelos	2	7	0,153	0,844	3,38
Região Caracaráí	Moura	32	4	0,225	0,802	30,82
	BR 174	35	13	0,224	0,805	27,86
Rio Madeira	Vila Moderna	18	-	0,227	0,800	17,24
	Manicoré	140	37	0,241	0,791	126,92
	Novo Aripuanã	19	8	0,228	0,799	18,35

A análise de PCoA para representação de agrupamento entre as populações de *E. oleifera* (Figura 2) mostrou ótima concordância de agrupamento entre regiões geográficas de coletas, é perceptível quando comparado com mapa de localização geográfica de regiões amazônicas de Moretzsohn et al. (2002), onde as regiões mais próximas geograficamente são: região Caracaráí, rio Negro, rio Solimões, rio Madeira, rio Amazonas e região de

Manaus. Observa-se que as populações V. Moderna e BR 174 são as mais distantes e pertencentes a mesma região de coleta (região Caracarái). Concordância semelhante pode-se perceber para as populações do rio Madeira, rio Amazonas e região de Manaus. Sendo as regiões do rio Solimões e rio Negro os que apresentam maior distância de agrupamento de suas populações. Corroborando com a matriz de distância genética de Nei (Tabela 5), onde as maiores distâncias são observadas entre populações pertencentes à região do rio Solimões, particularmente à população Benjamin Constant.

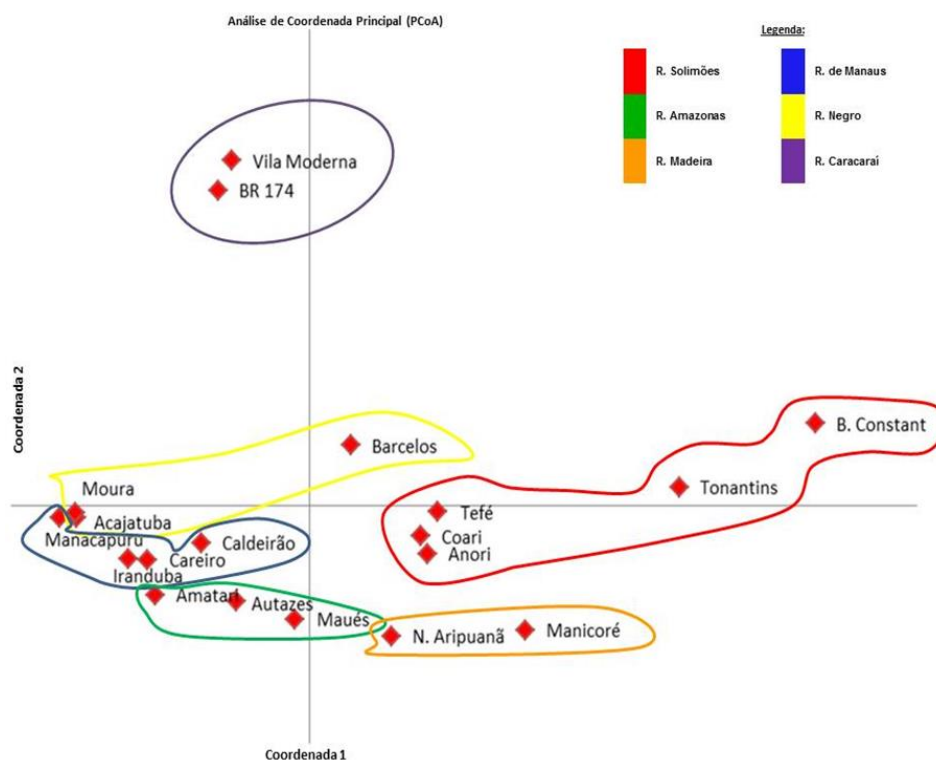


Figura 2 Representação Gráfica da Análise de Coordenadas Principais (PCoA) para as 19 populações nativas de *E. oleifera* em seis regiões amazônicas. Análise realizada pelo programa GenAlEx

É possível observar no dendrograma (Figura 3) a formação de seis grupos, com base no ponto de corte tomando por base a dissimilaridade média. A correlação cofenética foi de 97%, o que mostra elveada correlação entre a matriz de dissimilaridade e a árvore. O grupo 1 foi formado por populações coletadas no rio Madeira e rio Amazonas. Os grupos 2 e 3 foram compostos por subamostras pertencentes a região de coleta rio Solimões. Já o grupo 4 foi formado por populações pertencentes as regiões do rio Negro, rio Amazonas e Manaus. O grupo 5 foi composto por subamostras de Caracaraí e o grupo 6 com 9 amostras do rio Amazonas.

As regiões que melhor agruparam suas populações, com base na análise de agrupamento (Figura 3) foram Caracaraí e rio Solimões, onde é possível observar o agrupamento de todos os indivíduos pertencentes às regiões de origem de coleta.

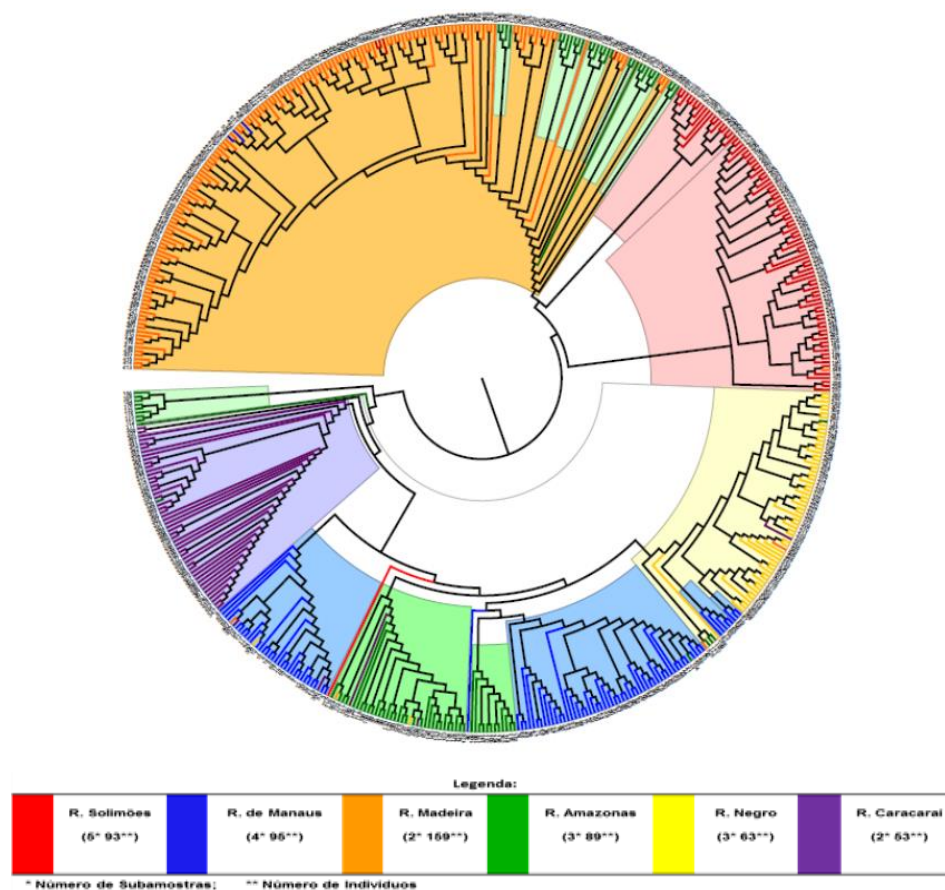


Figura 3 Dendrograma gerado a partir de distâncias genéticas pelo Índice de Coincidência Simples e pelo método não ponderado de *Neighbour-joining*, representando o agrupamento das 552 amostras de *E. oleifera*. Análise realizada pelos programas Darwin5 e FigTree

A região do Rio Madeira é o grupo com maior número de genótipos sendo possível observar o uniforme agrupamento. Essa ocorrência pode ser explicada pela proximidade de regiões ao longo dos rios amazônicos e assim, existir a possibilidade de fluxo gênico entre as populações (Figura 3).

O índice de Shannon-Wiener (Tabela 4) tem sido empregado especificamente em estudos genéticos como medida de diversidade dentro da

população e se assemelha a um índice de riqueza genotípica. Quantifica o grau de certeza com que se prevê a proximidade genética entre indivíduos. Sendo assim quanto menor o valor do índice, menor o grau de incerteza e, portanto, a diversidade da amostra é baixa. Como podem ser observadas na Tabela 4, as populações Manacapuru, B. Constant e Barcelos, são aquelas que apresentam os menores índices. Isso pode, contudo, refletir a baixa quantidade de amostras das populações. A diversidade tende a ser mais alta quanto maior o valor do índice, e este comportamento pode ser observado nas populações de Tonantins (0,624), Maués (0,616), Coari (0,613) e N. Aripuanã (0,611), sendo estas populações as que obtiveram Índice de Shannon acima de 0,600.

Na análise de porcentagem de locos polimórficos para cada população (Tabela 4), observa-se que as populações mais polimórficas são Manicoré (98,98%), Careiro (96,88%) e Autazes (91,54%). Porém, a maioria das populações apresentou acima de 70% de locos polimórficos, indicando que existe alta variabilidade entre as populações. Este resultado contrasta com os índices de dissimilaridade entre plantas que indica uma diversidade apenas moderada. Coerentemente a AMOVA, indica que a maior fonte de variação ocorre entre populações.

A diversidade média avaliada para cada região corrobora com a diversidade encontrada nas populações e assim a região do rio Madeira (34,32%) e de Manaus (18,98%) apresentam maior diversidade em suas populações amostradas, corroborando com resultados de porcentagem de locos polimórficos, confirmando mais uma vez com resultados da AMOVA, onde a segunda maior fonte de variação ocorre dentro de regiões (Tabela 4).

A matriz de distância genética de Nei (Tabela 5) apresenta a menor distância genética (0,02) entre as populações Acajatuba e Moura e BR 174 e Vila Moderna, respectivamente. Este resultado pode indicar que está havendo uma homogeneização entre estas populações que poderá levar à diminuição da

variabilidade genética. Já as maiores distâncias genéticas analisadas ocorrem entre Benjamin Constant e Manacapuru com 0,35, seguida de Vila Moderna e Benjamin Constant com 0,34. A população de Benjamin Constant encontra-se distante geograficamente das populações de Manacapuru e Vila Moderna, havendo assim impedimento para o fluxo gênico. Pelo dendrograma gerado, esse comportamento pode ser observado, onde as populações pertencentes ao Rio Solimões agrupam-se distantes das populações das regiões de Manaus e Caracarái.

Tabela 4 Diversidade genética dentro de populações e regiões de *E. oleifera*. Número de Amostras (NA), frequências alélicas baseadas no Índice de Shannon (SHa), Percentagem de locos polimórficos por população (%LP), diversidade média dentro de populações (P), diversidade média dentro de regiões (R). Análise realizada pelos programas Genes e GenA1EX

Região	População	NA	SHa	% LP	P	R
R. de Manaus	Caldeirão	18	0,467	69,11	3,17	28,461
	Careiro	68	0,579	96,88	18,98	
	Manacapuru	03	0,288	32,69	0,18	
	Irاندوبا	06	0,426	55,85	1,08	
R. Amazonas	Amatari	30	0,575	87,88	8,30	28,967
	Autazes	28	0,584	91,54	8,29	
	Maués	32	0,616	89,62	9,49	
R. Solimões	Anori	09	0,538	69,41	1,90	32,829
	B. Constant	03	0,314	29,21	0,28	
	Coari	54	0,613	88,12	15,03	
	Tefé	14	0,594	73,25	4,16	
	Tonantins	12	0,624	80,14	3,67	
R. Negro	Acajatuba	29	0,486	83,74	5,93	14,411
	Barcelos	02	0,246	24,72	0,13	
	Moura	32	0,459	76,48	6,48	

“Tabela 4, conclusão”

Região	População	NA	SHa	% LP	P	R
R. Caracaráí	BR 174	35	0,467	85,30	8,05	11,272
	V. Moderna	18	0,404	80,98	2,72	
R. Madeira	Manicoré	140	0,554	98,98	34,32	42,167
	N. Aripuanã	19	0,611	86,80	5,24	
Total	-	552	0,762	-	-	-
Média	-	-	-	77,72	-	-

Os resultados da AMOVA considerando suas regiões geográficas (Tabela 6) indicam que, a maior percentagem de variação é devida à variação dentro de populações: 54,20%; seguida pela percentagem de variação entre populações e dentro de regiões: 24,40%.

Tabela 5 Distância Genética entre 19 populações de *E. oleifera* pelo método de Nei 1972, a partir de 1.667 marcadores SNPs. Análise realizada pelo programa GenAlEX

	Populações																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	0,08	-																
3	0,14	0,11	-															
4	0,12	0,03	0,14	-														
5	0,11	0,03	0,11	0,06	-													
6	0,14	0,07	0,14	0,11	0,03	-												
7	0,13	0,08	0,14	0,13	0,04	0,03	-											
8	0,12	0,13	0,19	0,16	0,14	0,14	0,12	-										
9	0,30	0,28	0,35	0,31	0,28	0,28	0,27	0,25	-									
10	0,12	0,12	0,17	0,15	0,12	0,12	0,10	0,05	0,23	-								
11	0,14	0,13	0,18	0,16	0,13	0,13	0,11	0,07	0,23	0,06	-							
12	0,20	0,18	0,23	0,21	0,18	0,17	0,15	0,15	0,06	0,14	0,13	-						
13	0,10	0,07	0,06	0,09	0,06	0,10	0,10	0,15	0,30	0,14	0,15	0,20	-					
14	0,22	0,18	0,23	0,22	0,18	0,18	0,18	0,20	0,31	0,18	0,18	0,22	0,20	-				
15	0,11	0,08	0,05	0,11	0,07	0,10	0,11	0,16	0,31	0,15	0,15	0,20	0,02	0,21	-			
16	0,18	0,15	0,19	0,18	0,16	0,16	0,16	0,19	0,31	0,17	0,16	0,22	0,15	0,22	0,15	-		
17	0,21	0,18	0,22	0,21	0,18	0,18	0,18	0,21	0,34	0,19	0,19	0,24	0,18	0,24	0,18	0,02	-	
18	0,19	0,17	0,25	0,21	0,16	0,14	0,09	0,13	0,27	0,12	0,13	0,13	0,21	0,24	0,22	0,23	0,24	-
19	0,15	0,11	0,18	0,16	0,08	0,05	0,04	0,11	0,26	0,10	0,11	0,14	0,14	0,20	0,15	0,18	0,20	0,05

Legenda: Populações: (1= Caldeirão, 2= Careiro, 3= Manacapuru, 4= Iranduba, 5= Amatari, 6= Autazes, 7= Maués, 8= Anori, 9= Benjamin Constant, 10= Coari, 11= Tefé, 12= Tonantins, 13= Acajatuba, 14= Barcelos, 15= Moura, 16= BR 174, 17= Vila Moderna, 18= Manicoré e 19= Novo Aripuanã).

Tabela 6 Análise de Variância Molecular (AMOVA) para 552 subamostras de *E. oleifera* representando 19 famílias agrupadas em seis regiões geográficas da Floresta Amazônica Brasileira, por meio de 1.667 marcadores SNPs. Análise realizada pelo programa Genes

FV	GL	SQ	QM	CV	PV	P
Regiões	5	77,80	15,55	0,11	24,40	**
Populações/Regiões	13	20,66	1,58	0,10	21,40	**
Dentro	533	137,44	0,25	0,26	54,20	**
Total	551	235,90	-	0,47	100	-

** significativo a 1% de probabilidade pelo teste F. FV: Fonte de Variação, GL: Grau de Liberdade, SQ: Soma dos Quadrados, QM: Quadrados Médios, CV: Coeficiente de Variação, PV.: Porcentagem da Variação, P: Probabilidade.

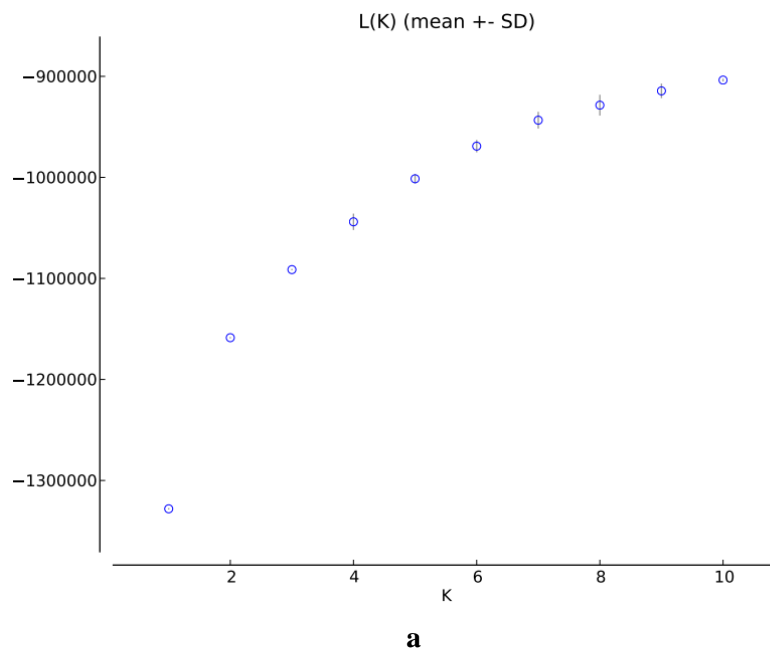
3.2 Parâmetros Genéticos Populacionais - Análise de Estruturação Genética

A análise da estrutura genética das populações tem sido baseada em princípios subjacentes à estatística F de Wright (WRIGHT, 1978). Os valores de F_{IS} , F_{ST} , F_{IT} e fluxo gênico (Nm) estão apresentados na tabela 7, os quais são utilizados para quantificar a extensão de endogamia dentro das subpopulações (F_{IS}), o nível de diferenciação genética entre as subpopulações (F_{ST}), a redução média em heterozigosidade de um indivíduo em relação a toda população (F_{IT}) e quantificar a dispersão ou migração de genes de uma população para outra (Nm).

O índice de fixação dentro das populações (F_{IS}) estimado com os 1.667 locos de SNPs foi de 0,073, conforme a tabela 7. Observa-se que as estimativas de f (Tabela 2) e F_{IS} (Tabela 7) apresentam comportamentos semelhantes para ambos os conjuntos de dados. Uma vez mais se evidencia que as populações de *E. oleifera* apresentam baixo nível de endogamia. Para as estimativas de F_{ST} o valor obtido por meio dos SNPs foi de 0,315. Este valor é considerado alto e,

portanto é um claro indicativo de diferenciação genética. Já para as estimativas de F_{IT} o valor foi de 0,353 e fluxo gênico (Nm) de 0,895.

Afim de se melhor compreender a estrutura da população, procedeu-se com a análise via Structure. O número mais provável de aglomerados (K) na estrutura foi determinado utilizando o método Δk . Segundo o critério proposto o número mais provável de cluster é de dois ou três. O modelo com $K=3$ é tomado como o melhor modelo para explicar a estrutura genética das populações de *E. oleifera* (Figura 4), onde são observados os agrupamentos das populações. Isto sugere forte diferenciação genética inter-populacional



(...continua...)

“Figura 4, Continuação”

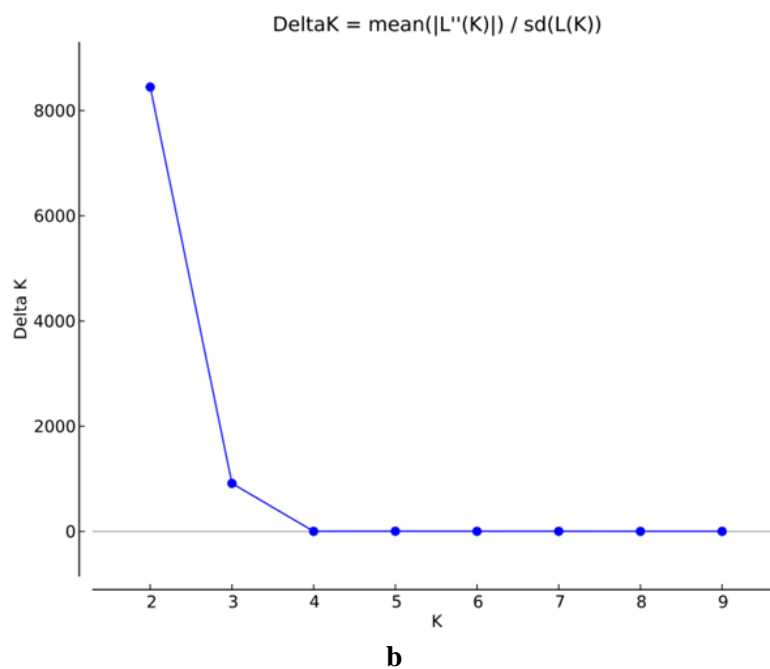


Figura 4 Média dos valores de: a) $\text{LnP}(D)$ de Pritchard et al. (2000) e b) Evanno et al. (2005) para 10 repetições de simulações no programa Structure com $k = 1$ a 10 para 552 indivíduos de *E. oleifera*. Análise realizada pelo programa Structure Harvester

Tabela 7 Estatísticas F - Wright e fluxo gênico (Nm) para 19 populações de *E. oleifera*. Análise realizada pelo programa GenAlEx

	Média	Desvio Padrão
F_{IS}	0,073	0,005
F_{IT}	0,353	0,006
F_{ST}	0,315	0,004
Nm	0,895	0,030

3.3 Cobertura do genoma de *Elaeis oleifera* por marcadores SNPs

Para caracterizar o espaço gênico de *E. oleifera*, foram consideradas apenas *reads* de 69 pb alinhadas contra o Genoma de Referência (GR) de *Elaeis guineensis* para os 16 cromossomos. Foram utilizados 812 marcadores SNPs para o alinhamento no GR. O genoma foi particionado em 130 intervalos de 5 Mpb cada (Figura 7), com média de 6,25 e máximo de 20 marcadores para os 130 intervalos. É possível verificar que os marcadores são bem distribuídos no genoma, possibilitando maior cobertura para explorar a espécie.



Figura 5 Estrutura genética de populações de *E. oleifera* quando avaliadas por $K = 2$ pelo Programa Structure

Legenda: Grupo 1: cor vermelha, Grupo 2: cor verde. 552 indivíduos e 19 populações de *Elaeis oleifera*. (Populações: 1 = Caldeirão, 2 = Careiro, 3 = Manacapuru, 4 = Iranduba, 5 = Amatari, 6 = Autazes, 7 = Maués, 8 = Anori, 9 = Benjamin Constant, 10 = Coari, 11 = Tefé, 12 = Tonantins, 13 = Acajatuba, 14 = Barcelos, 15 = Moura, 16 = BR 174, 17 = Vila Moderna, 18 = Manicoré e 19 = Novo Aripuanã).

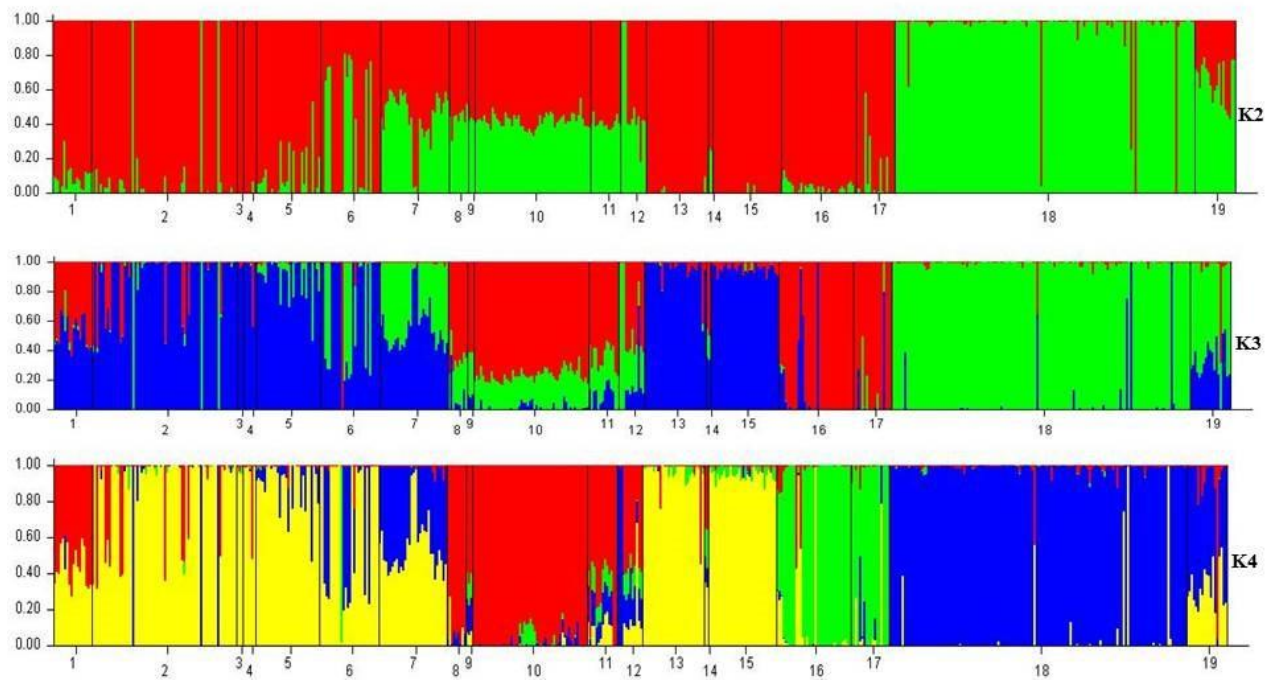


Figura 6 Estrutura genética de populações de *E. oleifera* pertencentes a seis regiões amazônicas, quando avaliadas de 2 a 4 níveis de estruturação genética pelo Programa Structure

Legenda: Populações: (1 = Caldeirão, 2 = Careiro, 3 = Manacapuru, 4 = Irlanduba, 5 = Amajari, 6 = Autazes, 7 = Maués, 8 = Anori, 9 = Benjamin Constant, 10 = Coari, 11 = Tefé, 12 = Tonantins, 13 = Acajatuba, 14 = Barcelos, 15 = Moura, 16 = BR 174, 17 = Vila Moderna, 18 = Manicoré e 19 = Novo Aripuanã).

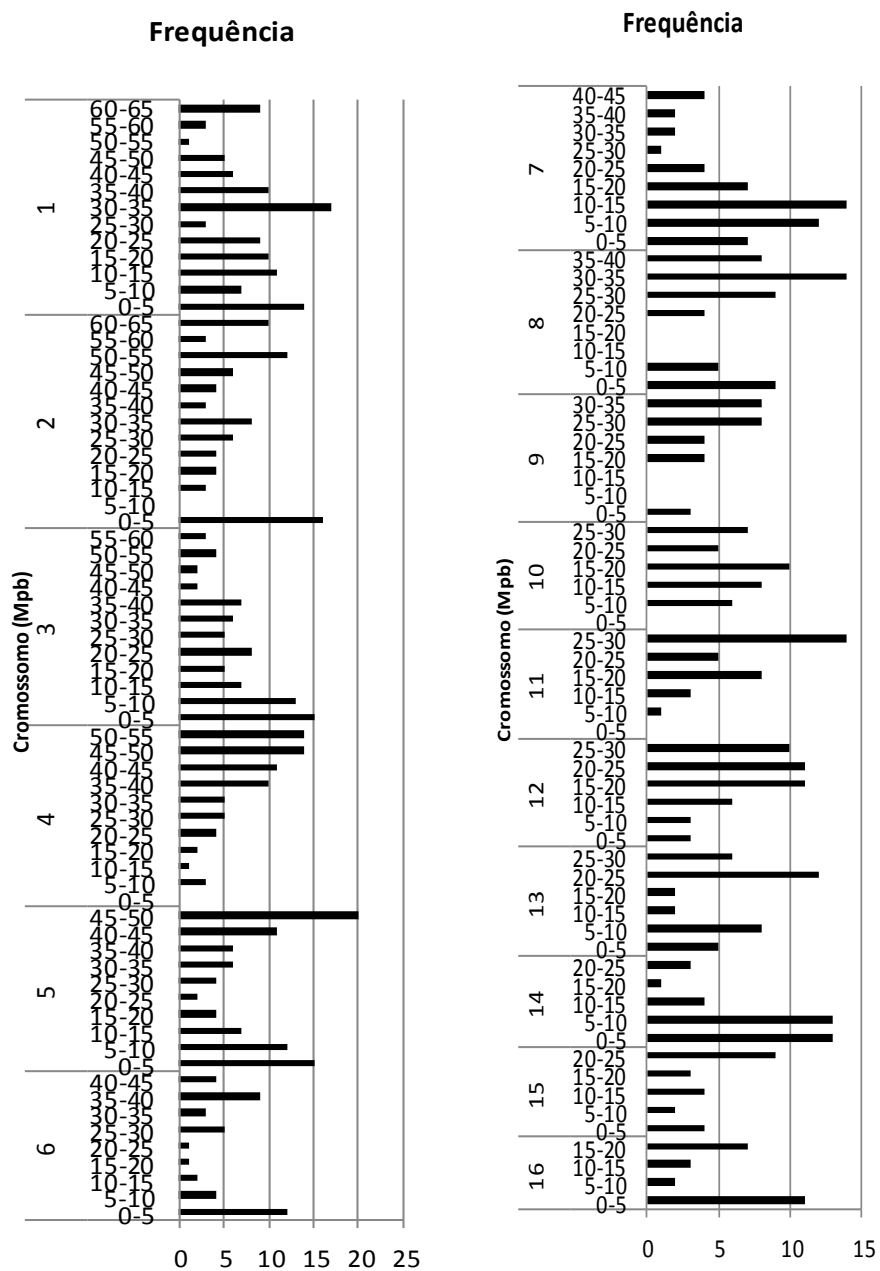


Figura 7 Análise da frequência de marcadores SNPs no genoma de referência *Elaeis guineensis*. Os 16 cromossomos foram divididos em 130 intervalos de 5Mb

4 DISCUSSÃO

4.1 Diversidade genética entre populações

O conhecimento sobre a distribuição da variabilidade genética entre e dentro de populações de *E. oleifera* é essencial para a adoção de estratégias eficientes de conservação desta espécie que é nativa da Amazônia e que apresenta interesse comercial, além de prover subsídios para se amostrar sua diversidade populacional. Além disso, é de interesse ter conhecimento sobre a diversidade genética do caiaué, pois tem sido utilizado em programas de melhoramento genético de *E. guineensis*. O programa de melhoramento genético da palma de óleo da Embrapa tem como foco atual a obtenção de híbridos interespecíficos entre *E. guineensis* e *E. oleifera* (RIOS et al., 2012), considerando as características de interesse do caiaué, a exemplo da tolerância a pragas e doenças, em especial ao Amarelecimento Fatal e *Fusariose*, além do porte reduzido e elevado teor de ácidos graxos insaturados. Marcadores moleculares têm sido muito utilizados em estudos de diversidade e de genômica populacional, e também em aplicações diretas no melhoramento de plantas no que se refere à seleção e ao agrupamento dos materiais mais promissores para formação de populações básicas (FERREIRA; GRATTAPAGLIA, 1998).

Quando se avalia a diversidade de *E. oleifera* a nível populacional, é possível observar que as populações de Manicoré (0,358) e Careiro (0,356) são as mais diversas. Em contrapartida as populações Manacapuru e B. Constant, são as menos diversas entre as populações com 0,091 e 0,119, respectivamente (Tabela 1). Essa informação será importante para direcionar os programas de melhoramento que buscam a obtenção de híbridos interespecíficos.

A endogamia estimada para as populações de *E. oleifera* foi bastante reduzida. Esse comportamento é comum em espécies alógamas com alta

densidade populacional, como ocorre com *E. oleifera* na Amazônia pois a maior densidade de indivíduos pode permitir maior número de recombinações, resultando em altos níveis de diversidade genética. O coeficiente de endogamia com valores de estimativas negativas nas populações de *E. oleifera* indicam ausência de endogamia, mas essa estimativa pode estar sendo superestimada, em razão do tamanho amostral das mesmas, visto que os índices negativos foram apresentados para todas as populações com baixa representatividade amostral (Manacapuru, Anori, B. Constant e Barcelos), com exceção da população Caldeirão, que apresentou índice negativo e sua população é formada por 18 indivíduos, número amostral considerado moderado quando comparado com as demais populações e seus respectivos índices. Quando o tamanho da população é pequeno, mesmo em cruzamentos ao acaso, uma diminuição de heterozigose pode ocorrer (BEAUMONT; HOARE, 2003). Este fato pode enviar os resultados para as populações com pequeno tamanho amostral. Como estas, no entanto são poucas, pode se considerar que em geral o nível de endogamia é baixo em *E. oleifera*.

As populações que apresentaram maiores números de alelos privados foram Coari, Manicoré, Tonantins e Tefé (Tabela 3). As populações Moura, Caldeirão, Acajatuba e Autazes apresentaram números baixos de alelos privados apesar do tamanho amostral. Já para as populações B. Constant e Barcelos, apesar de apresentarem número de amostras muito baixo, foram identificados números moderados de alelos privados. Quanto maior o número de alelos privados encontrado em um genótipo, mais divergente é sua origem, sendo sua identificação de especial interesse para o lançamento de cultivares comercial com base genética mais ampla, portanto, constituem em um dos parâmetros para se avaliar a perda de diversidade genética. Esse comportamento entre as populações de *E. oleifera* é bem diferente de pesquisas com outras palmeiras da região Amazônica como a Pupunha (*Bactris gasipaes* Kunth.) (RODRIGUES,

2007), o Tucumã (*Astrocaryum tucuma*) (BERNARDES, 2008) e o Jauari (*Astrocaryum jauari*) (OLIVEIRA, 2012), apesar de apresentarem a diversidade genética avaliadas por marcadores microssatélites o número de alelos privados é inferior ao encontrado para *E. oleifera*, é necessário também que se leve em consideração o número de populações e *loci* microssatélites utilizados nos estudados. Possivelmente, o pequeno número de indivíduos amostrados nestes estudos, assim como o limitado número de marcadores, tenha limitado a detecção de alelos privados.

As estimativas de frequência alélica para as 19 populações de *E. oleifera* foram consideradas altas, podendo-se inferir que existe uma grande diversidade dentro das populações estudadas independentemente do número de amostras. Segundo Frankel et al. (1996), espécies que apresentam menor variação das frequências alélicas ou maior equidade gênica podem manter as populações menos susceptíveis à fixação e à perda de alelos, quando submetidas a perturbações e a gargalos genéticos. Assim, populações com maior diversidade genética estariam mais protegidas dos efeitos de deriva do que as que têm alelos em baixa frequência (baixa equidade gênica), estando estes susceptíveis à perda.

Como mostra a tabela 3, em geral o tamanho efetivo populacional foi inferior ao número de amostras por população. Moraes et al. (1997) ressalta que as estimativas dos tamanhos efetivos populacionais são indicadores instantâneos da representatividade genética das amostras. Sendo assim, o conhecimento do tamanho efetivo populacional (N_e) é imprescindível quando o objetivo é preservação de germoplasma, coleta de sementes e/ou conservação *in situ*, já que mede a representatividade genética dos indivíduos amostrados na população em relação a uma população panmítica ideal. Portanto, fatores dinâmicos que afetam a distribuição das frequências alélicas devem ser levados em consideração, tais como, flutuações do tamanho populacional entre gerações, variação de fertilidade entre os indivíduos, estrutura de idades e sobreposição de gerações e

tamanho de vizinhança (FRANKEL; BROWN; BURDON, 1995). O tamanho efetivo de uma população (N_e) depende do número de indivíduos que efetivamente participam na reprodução e de suas contribuições relativas para a geração seguinte (CROSSA; VENCOVSKY, 1994).

A análise de dispersão gráfica (PCoA) (Figura 2) e agrupamento (dendrograma) (Figura 3), referente às populações de *E. oleifera*, mostram concordância entre as técnicas ao alocarem as populações seguindo a região de coleta, corroborando os resultados de Moretzsohn et al. (2002). Comportamento semelhante foi observado também com marcadores PAVs, em estudos de diversidade genética de subamostras mantidas na coleção de germoplasma da Embrapa (Capítulo 3).

Apesar do índice de Shannon (Tabela 4) considerar valores próximos de zero como baixa diversidade amostrada CHAGAS et al., (2015) relatam a diversidade em *E. guineensis* com média de 0,285. Essa baixa diversidade pode estar relacionada à pequena quantidade de populações avaliadas. Neste estudo os autores avaliaram apenas uma população, diversidade esta que pode ser comparada com média obtida para as populações de Manacapuru, Benjamin Constant, Barcelos e Vila Moderna devido ao baixo número de amostras. Para Pádua (2011) o valor do índice de Shannon pode variar entre 0 e 1, sendo 1 o ápice de diversidade de uma população; dessa forma, as populações em estudo apresentam moderados níveis de diversidade genética. Os valores estimados para o índice de Shannon evidenciaram moderada a elevada diversidade genética entre as populações de *E. oleifera* analisadas, variando de 0,246 a 0,624.

Tais comparações, juntamente com o fato de que não foram observados indivíduos geneticamente semelhantes, revelam forte evidência de que populações de *E. oleifera* apresentam níveis moderados de variabilidade genética intrapopulacional.

O resultado observado na matriz de distância genética sugere a existência de estruturação genética espacial na população, visto que progênies procedentes de árvores próximas tendem a apresentar menor distância entre si, como ocorre com as populações Acajatuba e Moura (Rio Negro) e populações de BR 174 e Vila Moderna (Região Caracará), do que de árvores distantes, comportamento este que pode ser observado para as populações Benjamin Constant e Manacapuru (Rio Solimões e Região de Manaus) e Vila Moderna e Benjamin Constant (Região Caracará e Rio Solimões). Assim a em programas de melhoramento seleção deve priorizar progênies mais divergentes a fim de procurar explorar a heterose ou vigor híbrido (BORÉM; MIRANDA, 2005).

4.2 Estrutura Populacional de *Elaeis oleifera*

A análise de estrutura genética das populações tem sido baseada em princípios subjacentes à estatística F de Wright (WRIGHT, 1978). O índice de fixação (f) é um dos parâmetros mais importantes em genética de populações, por medir o balanço entre homozigotos e heterozigotos nas populações. A explicação para populações que contêm maior número de locos em homozigose e menor em heterozigose deve está associado ao sistema reprodutivo e, ou deriva genética. As estimativas de índice de fixação revelaram valores baixos na maioria das populações, apresentando 42% das estimativas de f superior a 0,100 (Tabela 2), corroborando com valor médio calculado para todas as populações (Tabela 7), indicando baixo coeficiente de endogamia (HARTL; CLARK, 2010).

A divergência genética acima de $F_{ST}=0,25$ é considerada alta e vista como indicação de presença de alta diferenciação interpopulacional (YEH, 2000). Uma importante consequência em populações estruturadas em subpopulações é a redução na proporção média de genótipos heterozigotos relativos ao esperado na população única de acasalamento ao acaso (HARTL;

CLACK, 2010). A medida de F_{ST} pode variar de zero (onde não são averiguadas nenhuma divergência genética) a um (ocorrendo fixação de alelos alternativos em diferentes subpopulações). Geralmente esses valores observados apresentam-se menores que um, sendo esse valor máximo observado geralmente muito menor que um (HARTL; CLARK, 2010). Segundo os autores, Wright sugere valores médios para interpretação de F_{ST} sendo estes de 0 a 0,05 indicação de pequena diferenciação genética, 0,05 a 0,15 indicação moderada de diferenciação genética, 0,15 a 0,25 indicação de grande diferenciação genética e valores acima de 0,25 indicam uma diferenciação genética muito grande, corroborando com resultados deste estudo onde $F_{ST}=0,307$.

Estimativas de fluxo gênico segundo Wright (1978), com valores menores que 1 mostram isolamento genético. A estimativa obtida neste estudo foi de $Nm=0,895$, média de todas as populações. Esse valor pode ser um indicativo de que esteja ocorrendo isolamento genético causado por deriva, indicando a necessidade de se rever e/ou estabelecer estratégias de conservação da espécie. De acordo com Smouse e Sork (2004), o valor do fluxo gênico calculado a partir da divergência genética reflete o fluxo gênico que ocorreu durante um longo período. A estimativa não indica se está havendo fluxo gênico em determinado evento reprodutivo, mas calcula os níveis de fluxo gênico que devem ter ocorrido para produzir os padrões observados de estrutura genética.

Com base na análise Bayesiana de possíveis clusters entre $k=1$ e $k=10$ para as 552 amostras, a probabilidade de log dos dados pareceu começar a estabilizar em $k=4$, é a partir deste que aumenta o desvio padrão entre as iterações independente da estrutura para cada cluster putativo. O maior valor do parâmetro estimado de Δk indicou que o número ideal de agrupamentos foi $k=2$ (Figura 5). No entanto, como apontado por Eckert et al. (2010a), a utilização de Δk para escolher um valor ótimo de $k=2$ é difícil porque, neste caso, compara a

falta de estrutura ($k=1$) para algumas estruturas ($k=2$). O segundo maior valor de Δk indicou como número ótimo de agrupamentos $k=3$.

Os resultados da figura 3 mostram que o número de grupos variou entre $K=2$ e $K=4$ ($\log(\text{Pr} | K)$), quando foram analisadas as 19 populações. Estes resultados são consistentes com os observados por Eckert et al. (2010a, b), Quesada et al. (2010), Cumbie et al. (2011) e Chhatre et al. (2013). Assim, para uma análise mais detalhada, foi selecionando $K=2$ como o número ótimo de *clusters* para definir corretamente a estrutura das 19 populações. Porém, de acordo com os valores de Δk que indicam que o número ótimo de *clusters* para averiguar a estruturação seria $k=4$ percebe-se que essa estruturação seria fraca e optou-se por $k=2$, além de ter sido indicado por Δk com a maior estimativa de diferenciação de estruturação, corroboram com estimativas de $F_{ST}=0,307$.

É possível perceber que as populações nativas de *Elaeis oleifera* possuem uma diversidade moderada, fato este comprovado pelas análises estatísticas de diversidade genética e que essas populações possuem estrutura genética clara como mostra a estatística F de Wright e estruturação pela análise Bayesiana, assim como também mostram os resultados de tamanho efetivo populacional (N_m). Existe variabilidade genética entre e dentro das populações, que podem ser exploradas para fins de conservação, exploração comercial, seleção para aplicação em melhoramento genético da espécie e fornecer uma visão ampla da estrutura como um todo para facilitar futuras aplicabilidades dos melhoristas.

5 CONCLUSÕES

Existe variabilidade genética moderada entre e dentro de populações de *Elaeis oleifera*, que pode ser explorada para a conservação e exploração comercial da espécie.

Existem evidências de estruturação genética nas populações de *E. oleifera*, sendo dois o número de clusters (pool gênico) mais provável;

As informações geradas sobre a estruturação genética destas populações poderão a partir de agora serem incorporadas em estratégias de melhoramento de *Elaeis*.

A diferenciação genética entre populações dentro de regiões e entre regiões foi elevada, podendo ser explicada por um provável efeito fundador e/ou diferentes dinâmicas evolutivas das populações.

Os resultados deste estudo podem nortear novos estudos na área de genética e genômica populacional de *Elaeis*.

REFERÊNCIAS

- AMORIM, J. A. et al. Diversity and genetic structure of mangaba remnants in states of northeastern Brazil. **Genetics and Molecular Research**, Ribeirão Preto-SP, v. 14, n. 1, p. 823-833, feb. 2015.
- AZEVEDO, V. C. R. et al. Genetic Structure and Mating System of *Manilkara huberi* (Ducke) A. Chev., a Heavily Logged Amazonian Timber Species. **Journal Heredity**, Oxford, v. 98, n. 7, p. 646-654, sep. 2007.
- BARCELOS, E., CUNHA, R. N. V. da., NOUY, B. **Recursos genéticos de dendê (*Elaeis guineensis*, Jacq. e *Elaeis oleifera* (Kunth), Cortés) disponíveis na Embrapa e sua utilização**. In: MULLER, A. A., FURLAN JÚNIOR, J. (Ed.). Agronegócio do dendê: uma alternativa social, econômica e ambiental para o desenvolvimento sustentável da Amazônia. Embrapa Amazônia Oriental: Belém – PA. p. 131-143. 2001.
- BARCELOS, E. et al. Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. **Pesquisa Agropecuária Brasileira**, Brasília – DF, v. 37, n. 8, p. 1105-1114, aug. 2002.
- BEAUMONT, A. R., HOARE, K. Genetic structure in natural populations. In: BEAUMONT, A.R., HOARE, K. Biotechnology and genetics in fisheries and aquaculture. **Blackwell Science**. Oxford, p.47-72. 2003.
- BERNARDES, L. G. Diversidade genética de Tucumanzeiro (*Astrocaryum aculeatum* Mart.) com marcadores microssatélites. Manaus: UFAM, 2012. **Dissertação** (Mestrado em Ciências Florestais e Ambientais) – Universidade Federal do Amazonas. 43 f. 2008.
- BORÉM, A., MIRANDA, G.V. **Melhoramento de Plantas**. Universidade Federal de Viçosa, Viçosa - MG. 4ª edição. 525p. 2005.
- BOTSTEIN, D., et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics**, Estados Unidos, v. 32, p. 314-331. 1980.
- CHAGAS, K. P. T., et al. Seleção de marcadores ISSR e diversidade genética em uma população de *Elaeis guineensis*. **Revista Brasileira de Ciências Agrárias**, Recife - PE, v.10, n.1, p.147-152. 2015.

- CHHATRE, V. E., et al. Genetic structure and association mapping of adaptive and selective traits in the east Texas loblolly pine (*Pinus taeda* L.) breeding populations. **Tree Genetics e Genomes**. Amsterdam. v. 9 p. 1161–1178, oct. 2013.
- CROSSA, J., VENCOVSKY, R. Implications of the variance affective population size on the genet conservation of monoecious species. **Theoretical and Applied Genetics**, Stuttgart – Alemanha. v. 89, p. 936-942. 1994.
- CRUZ, C. D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**. Maringá – PR. v.35, n.3, p.271-276. 2013.
- CUMBIE, W., et al. Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. **Heredity**. País de Gales – Reino Unido. v. 107, p.105–114, aug. 2011.
- DAVEY, J. W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, London, v. 12, p. 499–510, jun. 2011.
- DOYLE, J. J., DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**. Rockville. v.12, n.1, p.13-15. 1990.
- EARL, D. A., VONHOLDT, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. 2012. **Conservation Genetics Resources**, Germânia. vol. 4, n. 2, p. 359-361. Core version: vA.2 July, 2014. Plot version: vA.1 November, 2012. Web version: v0.6.94 July 2014.
- ECKERT, A. J., et al. Patterns of population structure and environmental associations to aridity across the range of loblolly pine [*Pinus taeda* (L.), Pinaceae]. **Genetics**. Austin. v. 185, n. 3, p. 969–982, may. 2010a.
- ECKERT, A. J., et al. Back to nature: ecological genomics of loblolly pine ([*Pinus taeda*], Pinaceae). **Molecular Ecology**. Reino Unido. v. 19, n. 17, p. 3789–3805, sep. 2010b.
- EVANNO, G., REGNAUT, S., GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology**, Reino Unido. v. 14, p. 2611–2620. 2005.

EXCOFFIER, L., SMOUSE, P. E., QUATTRO, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**. Austin. v. 131, p. 479-491, 1992.

FERRAZ, I. D. K. et al. Características básicas para um agrupamento ecológico preliminar de espécies madeireiras da floresta de terra firme da Amazônia Central. **Acta Amazônica**, Manaus, v. 34, n. 4, p. 621–633, fev. 2004.

FERREIRA, M. E., GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa-Cenargen. Brasília-DF. 220p. 1998.

FRANKEL, A., et al. **National seismic-hazard maps; documentation**. United States Geological Survey Open-File Report, 96-532, 110p. 1996.

FRANKEL, O. H., BROWN, A. H. D., BURDON, J. J. **The conservation of plant biodiversity**. Cambridge University Press, Cambridge. 1995.

FRANQUEVILLE, H. **La pudrición del cogollo de la palma aceitera en América Latina: revisión preliminar de hechos y logros alcanzados**. Cirad – Departamento de Cultivos Perennes. França. 37p. 2001.

GUSSON, E., SEBBENN, A. M., KAGEYAMA, P. Y. Diversidade e estrutura genética espacial em duas populações de *Eschweilera ovata*. **Scientia Forestalis**, Piracicaba-SP, v. 67, p. 123-135, abr. 2005.

HARTL, D. L., CLARK, A. G. **Princípios de genética de populações**. 4 edição. Editora Artmed. Porto Alegre-RS. 659 p. 2010.

LIU, K., MUSE, S. V. PowerMarker: Integrated analysis environment for genetic marker data. **Bioinformatics**. Oxford. v. 21, p. 2128-2129. 2005.

MORAES, M. L. T., et al. Avaliação da densidade básica da madeira e de sua relação com os caracteres de crescimento, em uma população base de *Eucalyptus camaldulensis* Dehnh. In: Conferência IUFRO sobre Silvicultura e Melhoramento de Eucaliptos, **Resumos...** Salvador-BA. 1997.

MORETZSOHN, M. C., et al. Genetic diversity of Brazilian oil palm (*Elaeis oleifera*) germoplasm collected in the Amazon Forest. **Euphytica**. Wageningen. v. 124, p. 35-45, 2002.

NEI, M. Genetic distance between populations. **American Naturalist**. Chicago. v. 106, p. 283–292. 1972.

OLIVEIRA, L. dos S. Diversidade genética de populações naturais de Jauari (*Astrocaryum jauari* Mart.). 2012. **Dissertação** (Mestrado em Agronomia Tropical) – Universidade Federal do Amazonas: UFAM, Manaus-AM. 53 f. 2012.

PÁDUA, J. A. R. Genética da paisagem: áreas prioritárias para manejo e a conservação de *Eremanthus erythropappus* (DC.) Macleish no Estado de Minas Gerais. Lavras: **Dissertação de Mestrado** - Engenharia Florestal, área de concentração em Manejo Ambiental. Universidade Federal de Lavras, Lavras-MG. 80 p. 2011.

PEAKALL, R., SMOUSE, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. **Bioinformatics**. Oxford. v. 28, p. 2537–2539. 2012.

PERRIER, X., JACQUEMOUD-COLLET, J. P. DARwin (Dissimilarity Analysis and Representation for Windows) software, version 5.0. 158. Disponível em: <http://darwin.cirad.fr/>. 2009.

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

QUESADA, T., et al. Association mapping of quantitative disease resistance in a natural population of loblolly pine [*Pinus taeda* (L.)]. **Genetics**. Austin. v. 186, n. 2, p. 677–686. 2010.

RAJANAIDU, N. *Elaeis oleifera* collection in Central and South America. In: **WORKSHOP PALM OIL RESEARCH INSTITUTE OF MALAYSIA**. Proceedings... Bangi. s.l., p. 84-94. 1986.

RAJANAIDU, N., AINUL, M. M. Conservation of oil palm and coconut genetic resources. In: **Conservation of Tropical Plant Species**. cap. 10, p. 189-212. 2013.

RAMBAUT, A. Tree Figure Drawing Tool Version 1.4.0. UK: **Institute of Evolutionary Biology**. University of Edinburgh - UK. 2012.

RIOS, S. de. A., et al. Recursos genéticos de palma de óleo (*Elaeis guineensis* Jacq.) e Caiaué (*Elaeis oleifera* (H.B.K.) Cortés). **Documentos 96**, Embrapa Amazônia Ocidental, Manaus – AM. 39p. 2012.

RODRIGUES, D. P. Diversidade genética e Sistema de reprodução em progênie elites de pupunheira inerme (*Bactris gasipaes* Kunth) com marcadores microssatélites: implicações para o melhoramento do palmito. **Tese** (Doutorado em Biotecnologia) – Universidade Federal do Amazonas: UFAM. Manaus-AM, 103 f. 2007.

SAITOU, N., NEI, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**. Oxford. v. 4, p. 406–425. 1987.

SANSALONI, C. P. Desenvolvimento e aplicações de DArT (*Diversity Arrays Technology*) e genotipagem por sequenciamento (*Genotyping-by-Sequencing*) para análise genética em *Eucalyptus*. **Tese de Doutorado** – Universidade de Brasília. Brasília - DF. 145 p. 2012.

SINGH, R., et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds, **Nature**. London, v. 500, jul. 2013.

SMOUSE, P. E., SORK, V. L. Measuring pollen flow in forest trees: an exposition of alternative approaches. **Forest Ecology and Management**. v. 197. p. 21-38. 2004.

SNEATH, P. H., SOKAL, R. R. Numerical taxonomy: The principles and practice of numerical classification. **Freeman**. San Francisco: W.H. 573p. 1973.

TELLES, M. P. C. et al. Caracterização genética de populações naturais de araticunzeiro (*Annona crassiflora* Mart. – Annonaceae) no Estado de Goiás. **Revista Brasileira de Botânica**. São Paulo-SP. v. 26, n. 1, p. 123-129, jan./mar. 2003.

WRIGHT, S. **Evolution and the Genetics of Population, Variability Within and Among Natural Populations**. The University of Chicago Press, Chicago. 1978.

YEH, F. C. Population genetics. *In: Forest conservation genetics: principles and practice*. A. Young, D. Boshier & Boyle (eds.). p. 21-37. Publishing. Collingwood: CSIRO. 2000.

CAPÍTULO 3 Genotipagem de alta densidade via plataforma DArTSeq fornece novas informações sobre a extensão e a estrutura da diversidade genética da coleção de germoplasma brasileira de *Elaeis oleifera*

RESUMO

Este estudo teve por objetivo avaliar a diversidade genética entre subamostras de *Elaeis oleifera* (caiaué) que compõem o Banco Ativo de Germoplasma (BAG) da Embrapa, por meio de marcadores PAVs. A genotipagem dos 552 indivíduos de caiaué utilizados neste estudo foi realizada em plataforma DArTSeq. Após a genotipagem, os marcadores foram então selecionados por meio da aplicação de filtros de qualidade baseados em *Call Rate* ($>0,90$), *Minor Allele Frequency* (MAF $> 0,05$) e *Q-value* (> 2). Ao final, 3.187 marcadores PAVs foram selecionados e utilizados neste estudo. Estes foram utilizados para se calcular a matriz de diversidade par-a-par por meio do Índice de Jaccard. O agrupamento das subamostras foi realizado por meio do método UPGMA, e pela Análise de Coordenadas Principais (PCoA). A partição da variação genética total entre regiões, subamostras dentro de regiões e entre subamostras foi avaliada por meio de análise de variância molecular (AMOVA). A utilização de uma bateria maior de marcadores moleculares forneceu resultados diferentes daqueles até então obtidos. Observou-se a existência de moderada diversidade genética entre subamostras do BAG de *E. oleifera* (0.4079). Os resultados obtidos reforçam a hipótese que a Amazônia Brasileira deve ser um centro de diversificação secundário da espécie, e, dada à história dispersiva da espécie, a maior parte da diversidade concentra-se dentro de subamostras. A técnica de agrupamento UPGMA mostrou tendência de agrupamento de acordo com a região de coleta. A AMOVA evidenciou que a maior diversidade genética de *E. oleifera* ocorreu dentro de subamostras; sendo Manicoré a população mais diversa. Os resultados deste estudo podem embasar futuras ações de conservação e enriquecimento da coleção de germoplasma, além de indicar populações a serem exploradas para o desenvolvimento de novos híbridos interespecíficos.

Palavras-chave: Palma de Óleo. *Elaeis oleifera*. Diversidade Genética. DArTSeq. PAVs.

ABSTRACT

This study aimed to evaluate the genetic diversity among samples of *Elaeis oleifera* (caiaué) that make up the Active Germplasm Bank (BAG) of Embrapa, through PAVs markers. Genotyping of 552 caiaué of subjects used in this study was carried out by DArT. After genotyping markers were then selected by applying quality filters based Call Rate (> 0.90) Minor Allele Frequency (MAF > 0.05) and Q-value (> 2). In the end, 3,187 PAVs markers were selected and used in this study. These markers were used to calculate the dissimilarity matrix based on the Jaccard's index. Grouping of accessions were performed by the UPGMA method, as was also using Principal Coordinates Analysis (PCoA). The partition of the total genetic variation between regions within subsamples of regions and among samples was analysed by molecular analysis of variance (AMOVA). The use of a larger battery molecular markers provided different results from those obtained so far. It was observed the existence of moderate genetic diversity (0.4079) among germoplasm subsamples *E. oleifera*. The results support the hypothesis that the Brazilian Amazon should be a secondary center of diversification of the species, and, given the history of dispersive species, most of the diversity are concentrated in subsamples. UPGMA clustering technique showed a tendency of clustering according to the collection region. The AMOVA evidenced that most genetic diversity of *E. oleifera* occurred within sub-samples; being Manicoré the most diverse population. The results of this study can base future conservation and enrichment actions of the germplasm collection, and indicate populations/subsamples to be explored for the development of new interspecific hybrids.

Keywords: Oil Palm. *Elaeis oleifera*. Genetic Diversity. DArTSeq. PAVs.

1 INTRODUÇÃO

A Palma de Óleo Americana (*Elaeis oleifera*), conhecida popularmente por ‘Caiaué’ é amplamente distribuída nas Américas Central e do Sul, principalmente na região da Amazônia. No Brasil, uma ampla coleção de germoplasma nativo foi estabelecida com base em uma série de expedições organizadas pela Embrapa (Brasil) e pelo CIRAD - *Centre de Coopération Internationale em Recherches Agronomique pour le Développement* (França).

Esta coleção vem sendo estudada desde então com a finalidade de se verificar a extensão e a organização da diversidade genética amostrada, uma vez que o conhecimento do modelo de variação entre os materiais é importante na definição das estratégias a serem utilizadas nos programas de melhoramento da espécie (OOI et al., 1981). Esses estudos foram em sua maioria realizados mediante emprego de marcadores moleculares (BARCELOS, 1998; MORETHZSOHN et al., 2002), pois a caracterização molecular da diversidade genética de coleções de germoplasma proporciona informações rápidas e precisas para a seleção de genitores a serem usados para formar novas populações de melhoramento. Estas novas populações podem, por exemplo, ser formadas pelo intercruzamento de plantas com distâncias genéticas conhecidas, favorecendo a obtenção de novas combinações alélicas ou, no caso do melhoramento interespecífico, a priorização de cruzamentos entre genótipos de regiões onde há a ocorrência de maior variabilidade em detrimento de outras.

Barcelos et al. (1998) e Morethzsohn et al. (2002) indicaram previamente que a variabilidade do caiaué amazônico é mais restrita que a diversidade de dendê, embora Barcelos et al. (2002) tenha obtido evidências que a *E. oleifera* era mais diverso que *E. guineensis*. Em comum, esses artigos, entretanto indicaram que a diversidade existente na coleção de germoplasma de *E. oleifera* pode ser explorada em programas de melhoramento. Entretanto, estes

trabalhos se basearam em uma quantidade limitada de marcadores moleculares (37 RFLPS e 96 RAPDs respectivamente). Considerando-se, entretanto, que *E. oleifera* é uma espécie alógama e nativa, o nível de diversidade existente talvez não tenha sido adequadamente verificado com uma quantidade tão reduzida de marcadores. Isso porque ao se utilizar algumas dezenas de marcadores para amostrar a diversidade, não mais que alguns poucos marcadores estarão amostrando cada grupo de ligação. Considerando-se que a espécie possui 16 pares de cromossomos, e uma distribuição aleatória e uniforme dos marcadores, cada grupo de ligação estaria sendo amostrado por apenas 2.31 e 6 marcadores RFPL e RAPD, respectivamente. Desse modo, é possível que grandes porções do genoma não tenham sido amostradas nestes estudos, o que pode levar a estimativas enviesadas de diversidade e relacionamento genético entre os indivíduos avaliados. -Esse pode ser um dos motivos pelos quais Barcelos et al. (2002) e Morethzsohn et al. (2000) obtiveram resultados contrastantes. Ademais, como o foco dos trabalhos envolvendo *E. oleifera* é o melhoramento genético, a diversidade estimada com base em número reduzido de marcadores pode não se correlacionar com a diversidade fenotípica existente, pelo fato de os marcadores não amostrarem densamente todo o genoma da espécie, especialmente se a extensão média do desequilíbrio de ligação em *E. oleifera* for bastante reduzida como em outras espécies alógamas e não melhoradas.

O desenvolvimento de marcadores para espécies pouco estudadas tem envolvido normalmente etapas de descoberta do polimorfismo, desenvolvimento de ensaio de genotipagem do polimorfismo, e validação dos ensaios em uma população de triagem antes da implantação completa em grandes populações (DAVEY et al. 2011). Esse processo geralmente é oneroso e usualmente resulta na geração de algumas dezenas ou poucas centenas de marcadores. Com o desenvolvimento de novas plataformas de genotipagem, especialmente daquelas baseadas no princípio de genotipagem-por-sequenciamento (DAVEY et al.

2011), pode-se atualmente genotipar rapidamente um grande número de amostras para virtualmente milhares de marcadores a custos bastante competitivos. Há assim atualmente a possibilidade de se estimar de maneira mais precisa o nível de diversidade existente em coleções de germoplasma, mesmo que a espécie em questão não apresente uma grande quantidade de recursos genômicos. Inúmeras coleções de germoplasma ao redor do mundo têm sido extensivamente caracterizadas por meio de baterias robustas de marcadores moleculares (NIMMAKAYALA et al., 2014).

Frente ao exposto, este estudo teve por objetivo avaliar a dissimilaridade entre acessos do Banco Ativo de Germoplasma (BAG) de *E. oleifera* mantido pela Embrapa por meio da genotipagem-por-sequenciamento de 206 amostras.

2 MATERIAL E MÉTODOS

2.1 Material Biológico

Um total de 552 plantas do BAG de caiaué que representam 206 subamostras coletadas em 19 locais distintos (populações) compreendendo seis grandes regiões foi utilizado para este estudo, como citado no capítulo anterior. Destes materiais 159, 28 e 19 subamostras estão representadas por três, dois e um indivíduos, respectivamente. O número de plantas amostradas por região/população/subamostra encontra-se no material suplementar (Tabela S1).

As etapas de extração de DNA, purificação e quantificação estão descritas no capítulo anterior.

2.2 Genotipagem-por-sequenciamento em plataforma DArTSeq

A descrição dos procedimentos utilizados para se gerar os marcadores utilizados neste estudo pode ser encontrada na seção de material e métodos do capítulo 2.

As sequências resultantes (mais de 2 milhões de *reads* por amostra foram gerados) foram filtradas e alocadas a seus respectivos *datasets* sendo os *barcodes* retirados posteriormente. As sequências foram trimadas em 69 pb (5 pb do sítio de restrição mais 64 bases com Q score mínimo de 10). O *pipeline* proprietário da DArT P/L foi utilizado para se realizar o score dos marcadores do tipo PAV (*presence/absence variants*). Casos em que a presença ou ausência de *reads* do fragmento resultante da clivagem pela enzima de restrição não era clara (i.e. menos de 6x) foram registrados como dados perdidos.

2.3 Análise de dados

Apenas marcadores que passaram pelo controle de qualidade, onde o [*Call Rate* ($>0,90$), *Minor Allele Frequency* (MAF $> 0,05$) e *Q-value* (> 2)] foram utilizados nas análises de diversidade genética. Após a filtragem foi possível selecionar 3.187 marcadores PAVs.

Para se avaliar a extensão da diversidade amostrada na coleção de germoplasma de *E. oleifera* os dados obtidos foram utilizados para se calcular a matriz de coeficientes de dissimilaridade par-a-par de Jaccard (JACCARD, 1908) por meio do programa Darwin v5.5 (PERRIER; JACQUEMOUD-COLLET, 2006). Para computar a dissimilaridade entre duas unidades (plantas) usando um *dataset* que incluía dados perdidos o critério de deleção de dados par-a-par foi adotado ao invés de eliminar marcadores ou plantas que apresentavam dados perdidos na análise. Neste caso, foi estipulado que um mínimo de 70% de dados válidos seria requerido para computar a dissimilaridade entre duas unidades.

A partir dos dados gerados foram então computadas estatísticas descritivas relacionadas à diversidade da coleção, como: (1) dissimilaridade média, (2) variância dos dados, (3) distribuição de frequência de dissimilaridade, (4) par de subamostras mais diverso e (5) par de subamostras mais relacionados.

Em seguida a matriz gerada com base nos coeficientes de dissimilaridade de Jaccard foi utilizada para: (6) construir um dendrograma usando o método de análise UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*), (7) realizar uma análise fatorial para obter representações gráficas bidimensionais do relacionamento das 206 subamostras em termos de dissimilaridade e (8) realizar análise de agrupamento baseado no método de otimização de Tocher Modificado (RAO, 1952), onde é identificado o par de indivíduos mais similares a partir da matriz de dissimilaridade, esses formarão

os grupos iniciais e a inclusão de novos indivíduos, é tomada por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo permitido da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo (CRUZ; FERREIRA; PESSONI, 2011). A correlação cofenética foi computada entre os valores de dissimilaridade originais e os valores de dissimilaridades graficamente representados no dendrograma de modo a determinar a precisão com que a árvore representa a diversidade entre as subamostras de *E. oleifera*.

Os dados originais foram também utilizados para se analisar a partição da variação genética total entre regiões Amazônicas amostradas, entre subamostras e subamostras dentro de regiões, a fim de se verificar a estrutura da diversidade genética amostrada na coleção de germoplasma de *E. oleifera*, por meio de uma análise de variância molecular (AMOVA) (EXCOFFIER; SMOUSE; QUATTRO, 1992) utilizando o programa GENES (CRUZ, 2013). Os níveis de significância para os componentes de variância foram calculados adotando-se o critério de 1000 permutações.

3 RESULTADOS

3.1 Diversidade Genética

Observou-se a existência de moderada diversidade genética entre subamostras do BAG de *E. oleifera*. Os resultados da análise realizada com base em 206 subamostras (Tabela S1) por meio de 3.187 marcadores moleculares revelaram uma dissimilaridade média de 0,4079 com variação de 0,0016 a 0,6064 (Figura 1). A menor dissimilaridade ocorreu entre subamostras de Careiro, enquanto a maior dissimilaridade observada foi entre BR 174 e Benjamin Constant. Quando a análise foi realizada com base em 552 indivíduos (Tabela S1), a dissimilaridade média encontrada foi de 0,4113, com variação de 0,0009 a 0,6113 (Figura S1). A menor dissimilaridade observada foi entre Caldeirão e Acajatuba e a maior dissimilaridade observada entre Manicoré e BR 174.

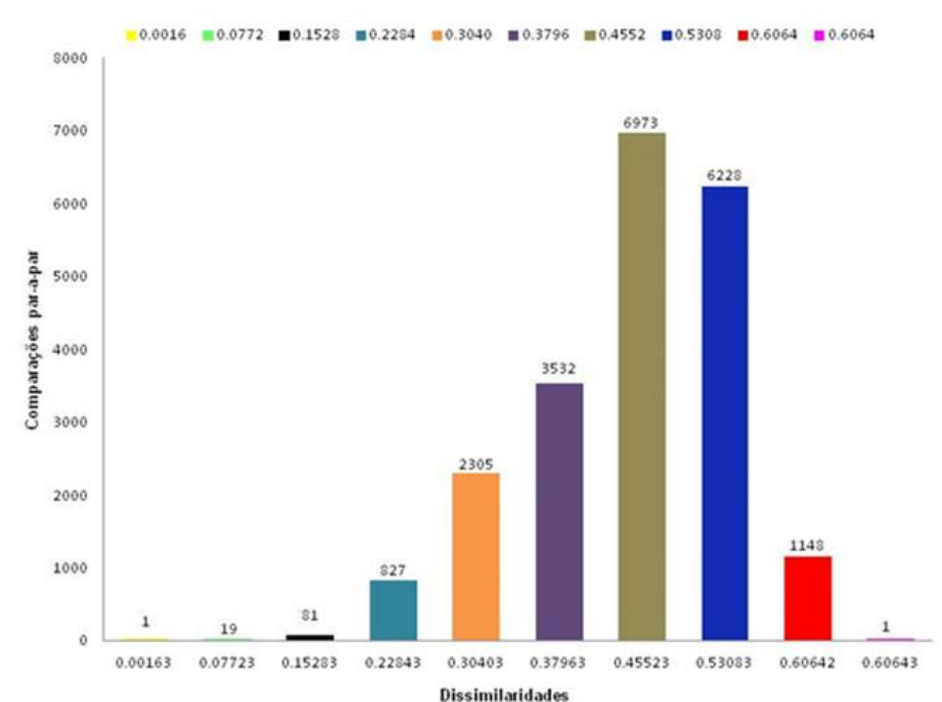


Figura 1 Distribuição de frequência das estimativas de distância genética, utilizando o coeficiente de dissimilaridade de Jaccard, obtidas entre os pares de 206 subamostras de *E. oleifera*. Análise realizada através do Programa Darwin v5.5

Fonte: (PERRIER; JACQUEMOUD-COLLET, 2006).

3.2 Organização/Estrutura da diversidade amostrada no BAG

A Análise das Coordenadas Principais (PCoA) demonstrou que os dois primeiros componentes explicam 39,91% do total da variação encontrada para as 206 subamostras (Figura 2). Valores aproximados foram encontrados na análise com 552 plantas (36,63%) – Figura S2. É possível observar na Figura 3 que a distribuição dos grupos está fortemente relacionada à região de coleta. Comportamento semelhante é observado na Figura S3 para 552 indivíduos

apresentando as mesmas tendências de agrupamento, porém com mais consistência, devido ao número maior de amostras na análise.

A representação gráfica da PCoA (Figura 2) permitiu a caracterização das subamostras baseadas na localização geográfica de coleta. Admitindo a separação de seis regiões geográficas da Floresta Amazônica. Verifica-se nas Figuras 2 e S2 que o agrupamento por dispersão gráfica mostra uma tendência de agrupamento geográfico, porém, a região do R. Madeira, R. Caracaraí e R. Solimões apesar de mostrarem uma maior tendência de agrupamento entre as demais, apresentam subamostras dispersas do agrupamento, assim como demonstrado nas técnicas de agrupamentos hierárquicos e aglomerativo para ambos os grupos de indivíduos.

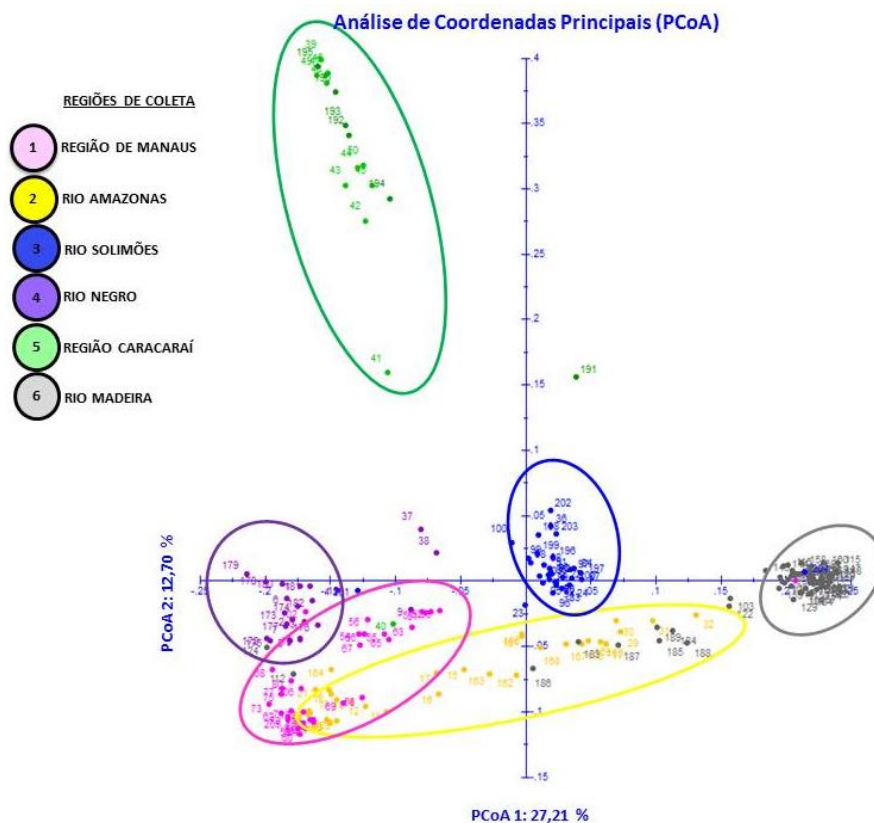


Figura 2 Dispersão Gráfica de 206 subamostras de *E. oleifera* obtida pelo método de Análise das Coordenadas Principais (PCoA), utilizando o coeficiente de dissimilaridade de Jaccard. Análise realizada através do Programa Darwin v5.5

Fonte: (PERRIER; JACQUEMOUD-COLLET, 2006).

Os agrupamentos gerados por meio do método de UPGMA (Figuras 3 e S3) apresentaram elevado coeficiente de correlação cofenética em ambas às situações ($r = 0,93$ e $r = 0,95$ $p < 0,0001$), indicando boa concordância das distâncias genéticas plotadas na árvore com aquelas estimadas por Jaccard. A tendência de agrupamento de acordo com a região de coleta pode ser mais uma vez percebida, com pequenas exceções (raios de cores distintas em meio a agrupamentos). É possível ainda se observar na Figura 3 elevada consistência de

agrupamento entre as subamostras do Rio Madeira, Região Caracará e R. Solimões, corroborando os resultados obtidos por meio da análise de PCoA. É possível verificar também uma dispersão nos agrupamentos pertencentes às regiões do Rio Amazonas, Manaus e Negro. No entanto, quando se analisa o agrupamento por meio de com 552 indivíduos (Figura S3), observa-se que ambas se agrupam com melhor consistência. Nota-se ainda, o deslocamento das subamostras de Barcelos em um grupo distante das demais pertencentes à mesma região.

Com relação à consistência do agrupamento na análise com 206 subamostras (Figura 3) pode-se observar que as mesmas estão dispostas em sete grupos distintos, baseados no corte utilizando a diversidade média da espécie como *threshold*. O Grupo 1 é o maior, composto por 86 indivíduos e concentrando 41,74% dos mesmos, e inclui subamostras pertencentes às regiões de Manaus, Amazonas e Negro. No Grupo 2 estão alocados três indivíduos de Tefé, pertencentes às sub-regiões Tefé, Tefé e Lago Caiambé, respectivamente, participando apenas de 1,45% de toda a amostragem. O Grupo 3 aloca 24 subamostras (11,65%), originárias do R. Solimões. Já o Grupo 4 é o segundo maior grupo, composto por 71 subamostras do R. Madeira juntamente com subamostras pertencentes a Maués e Autazes – R. Amazonas, representando 34,46% da coleção de germoplasma. Um grupo distante que alocou duas subamostras pertencentes à Tonantins e uma a Benjamin Constant formam o Grupo 5, com 1,45% de representatividade em relação às 206 subamostras. A Região Caracará por apresentar uma boa consistência de agrupamento constitui o Grupo 6 (17 subamostras) com 8,25%. E por fim, duas subamostras (0,97%) pertencentes à população de Barcelos alocaram-se no Grupo 7. Esse comportamento reflete e corrobora os resultados observados por meio de outras técnicas de dispersão e aglomerativa.

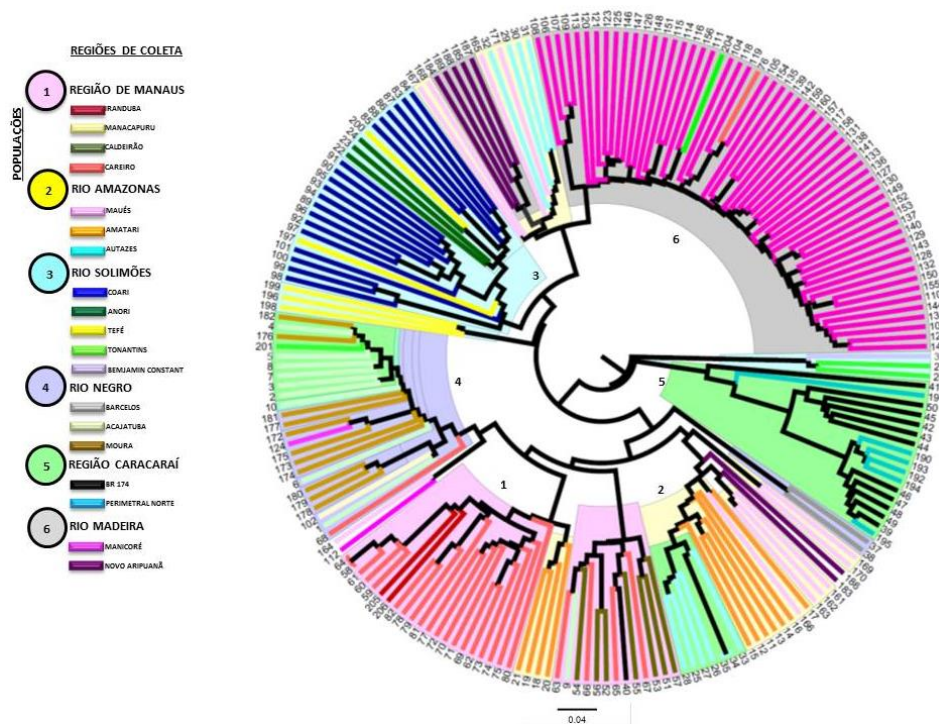


Figura 3 Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para 206 subamostras de *E. oleifera* que compõe o BAG da Embrapa. Análise realizada através do Programa Darwin v5.5

Fonte: (PERRIER; JACQUEMOUD-COLLET, 2006).

A consistência do agrupamento entre as subamostras de BR 174 e Perimetral Norte é notória, ainda que uma única subamostra de BR 174 tenha se agrupado com outras da R. de Manaus. Logo para a região do R. Madeira as subamostras de Manicoré foram alocadas todas no mesmo grupo, e sendo alocadas entre as mesmas uma amostra de Tonantins – R. Solimões e uma de Careiro – R. Manaus, tendo a subamostra N. Aripuanã agrupamento junto à subamostras pertencentes ao R. Amazonas.

Quando empregado o método de agrupamento de Tocher Modificado para as 206 subamostras, observa-se a formação de nove grupos, com as maiores

concentrações de subamostras nos grupos 1 e 2, representando 71,94% das subamostras, onde não são apresentadas subamostras apenas da Região Caracaráí. Nota-se também que uma subamostra de Careiro se manteve em agrupamento unitário (Tabela 1), esses resultados encontram-se em consonância com os apresentados nas análises de PCoA e UPGMA. A região do Rio Madeira alocou subamostras nos dois primeiros grupos, tendo a R. de Manaus com subamostras alocadas nos grupos 2 e 3, já a região do R. Amazonas (grupos 1, 2 e 3) e R. de Caracaráí (3, 4 e 9) participaram de três grupos cada, a região do R. Negro de quatro (1, 2, 3 e 8) e R. Solimões de cinco grupos (grupos 1, 2, 5, 6 e 7). Os grupos 2, 5 e 7 formados na análise UPGMA (Figura 3) estão em consonância com os grupos 6, 7 e 8 pelo agrupamento de Tocher Modificado.

A distância genética média entre as amostras nos grupos 1 e 2, pelo método de Tocher é 0,29 e 0,25, respectivamente. Baseado na técnica de agrupamento pelo qual é identificado o par de indivíduos mais similares para formação do grupo inicial este resultado indica uma tendência ao agrupamento de acordo com a distribuição geográfica das subamostras. As subamostras Amatari, Autazes encontram-se representada nos três primeiros grupos e subamostras de Manicoré, Tefé, BR 174, Tonantins encontram-se distribuídas em pelo menos três grupos. Este resultado corrobora com a variação apresentada na AMOVA (Tabela 2), onde é possível observar que a maior variação apresentada ocorre dentro de subamostras.

Tabela 1 Agrupamento baseado pelo método de Tocher Modificado, utilizando o coeficiente de dissimilaridade de Jaccard como medida de distância genética, para 206 subamostras de *E. oleifera*

Grupos	N	% do BAG	Subamostras
1	94	45,63	MAN; NAP; AMA; AUT; MAU; ANO; COA; TEF; TON e ACA.
2	54	26,31	MAC; MOU; CAL; IRA; CAR; AMA; AUT; MAN; MAU; TON e ACA.
3	20	9,70	AMA; AUT; CAL; CAR; MOU e BR 174.
4	16	7,76	BR 174 e PEN.
5	13	6,31	COA e TEF.
6	03	1,45	BJC e TON.
7	03	1,45	TEF.
8	02	0,97	BAR.
9	01	0,48	BR 174.

Fonte: Análise realizada pelo Programa Genes (CRUZ, 2013).

Nota: Subamostras: **R. Manaus** (Careiro=CAR; Manacapuru=MAC; Iranduba=IRA; Caldeirão=CAL), **R. Amazonas** (Amatari=AMA; Autazes=AUT e Maués=MAU), **Rio Solimões** (Benjamin Constant=BJC; Coari=COA; Anori=ANO; Tefé=TEF e Tonantins=TON), **R. Negro** (Acajatuba=ACA; Barcelos=BAR e Moura=MOU), **R. Caracaraí** (BR 174=BR 174; Perimetral Norte=PEN) e **R. Madeira** (Manicoré=MAN e Novo Aripuanã=NAP), N = número de indivíduos.

3.3 Partição da variância genética

Os resultados da análise de variância molecular (AMOVA) mostram que a variação genética ocorre em maior significância dentro de subamostras seguida de variação dentro de regiões, revelando uma concordância com os resultados obtidos pelas demais estatísticas estudadas. Pode ser verificado, na Tabela 2, que a maior diversidade genética de *E. oleifera* ocorre dentro de subamostras, com

variabilidade de 57,16% e variação de 24,33% dentro de regiões. Os resultados indicam também que há diferenças significativas entre as regiões ($p < 0,001$), bem como entre as subamostras ($p = 0,001$). Resultados semelhantes são observados na AMOVA para 552 indivíduos (Tabela S2).

Tabela 2 Análise de Variância Molecular (AMOVA) para 206 subamostras de *E. oleifera* representadas por 19 famílias distribuídas em seis regiões geográficas da Floresta Amazônica Brasileira

FV	GL	SQ	QM	CV	PV	P
Regiões	5	31207,68	6241,53	127,21	24,33	**
Subamostra/Região	13	9871,45	759,34	96,73	18,50	**
Subamostra	187	55892,19	298,88	298,88	57,16	**
Total	205	96971,33	473,03	522,84	100	

Fonte: Análise realizada pelo Programa Genes (CRUZ, 2013).

** significativo a 1% de probabilidade pelo teste F. FV: Fonte de Variação, GL: Grau de Liberdade, SQ: Soma dos Quadrados, QM: Quadrados Médios, CV: Coeficiente de Variação, PV: Porcentagem da Variação, P: Probabilidade.

3.4 Erros de alocação de subamostras e de indivíduos dentro de subamostras

Por meio dos resultados das análises de agrupamento foi possível se realizar uma avaliação de erros de alocação de subamostras e de indivíduos dentro de subamostras. Isso porque se espera que indivíduos de uma mesma subamostra apresentem maior similaridade genética entre si que com os indivíduos de outras subamostras. O mesmo raciocínio é válido para as subamostras, uma vez que subamostras coletadas em uma mesma população tendem a ser mais relacionadas geneticamente entre si que com subamostras coletadas em outras subpopulações sob a hipótese de existência de estruturação genética conforme evidenciado no capítulo anterior. As Figuras 4 e 5 mostram a relação entre indivíduos de

subamostras que foram amostras por meio de duas ou três plantas. O dendrograma gerado a partir de dados com duas repetições por subamostra mostra uma discordância de 10,71%, uma vez que das 28 subamostras analisadas em duplicadas, três não apresentaram perfil de agrupamento esperado (duas subamostras de Manicoré e uma de Amajari, tendo sido estas subamostras coletadas em regiões de Rio Madeira e Rio Amazonas) (Tabela 3). Já para a análise de agrupamento com três repetições por subamostra o comportamento é semelhante, sendo observada uma discordância de 17,61%, composto de 159 subamostras e apresentando 28 delas não agrupadas como esperado.

Tabela 3 Percentagem de erro de alocação em subamostras de *E. oleifera*

NIS	NS	NTI	SD	%
2	28	56	3	10,71
3	159	477	28	17,61

Fonte: Análise realizada pelo Programa Genes (CRUZ, 2013).

Nota: Análise com subamostras representadas por dois e três indivíduos, respectivamente, usando o Índice de Jaccard. NIS = número de indivíduos por subamostra, NS = número de subamostra, NTI = número total de indivíduos, SD = número de subamostras em discordância, % = percentagem de erro de alocação de subamostras.

Na figura 5, é possível observar o agrupamento entre as subamostras representadas em triplicatas e algumas discordâncias representadas pela cor amarela. É possível observar dois comportamentos sobre os agrupamentos. Primeiro, onde a subamostra apresenta pelo menos dois de seus indivíduos agrupados e o terceiro deslocado destes. Comportamento este observado em oito subamostras de Manicoré, duas em Acajatuba, Caldeirão, Careiro, Autazes, e uma em Moura, Tonantins, Coari, Iranduba, BR 174 e Amajari. Sendo estas subamostras representativas de todas as regiões geográficas de coletas. Segundo, onde subamostras apresentam os três os indivíduos separadamente,

comportamento este observado em duas subamostras de Careiro e uma de Autazes, BR 174, Coari e Vila Moderna, respectivamente. Totalizando uma discordância entre 28 de 159 subamostras.

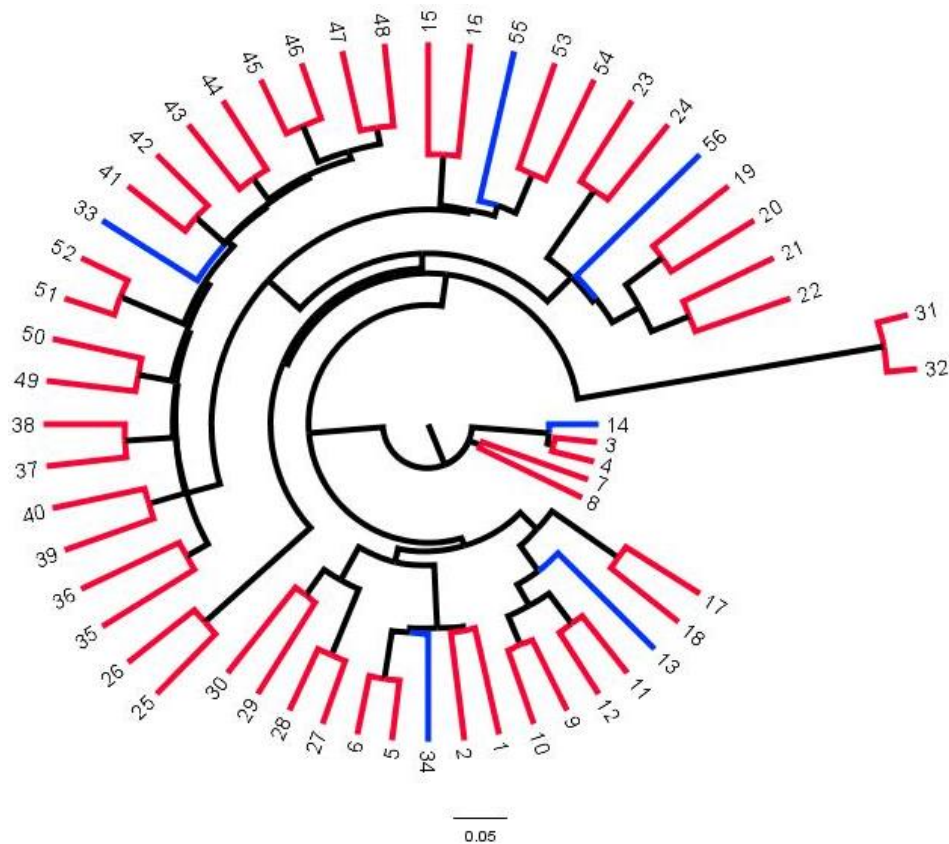


Figura 4 Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para 29 subamostras (representadas por 2 indivíduos cada) de *E. oleifera* que compõe o BAG da Embrapa, evidenciando possíveis erros de alocação de indivíduos em subamostra

Fonte: (PERRIER; JACQUEMOUD-COLLET, 2006).

Nota: Cor vermelha representa subamostras com os dois indivíduos agrupados e cor azul representa subamostras com indivíduos em discordância de agrupamento. Análise realizada através do Programa Darwin v5.5

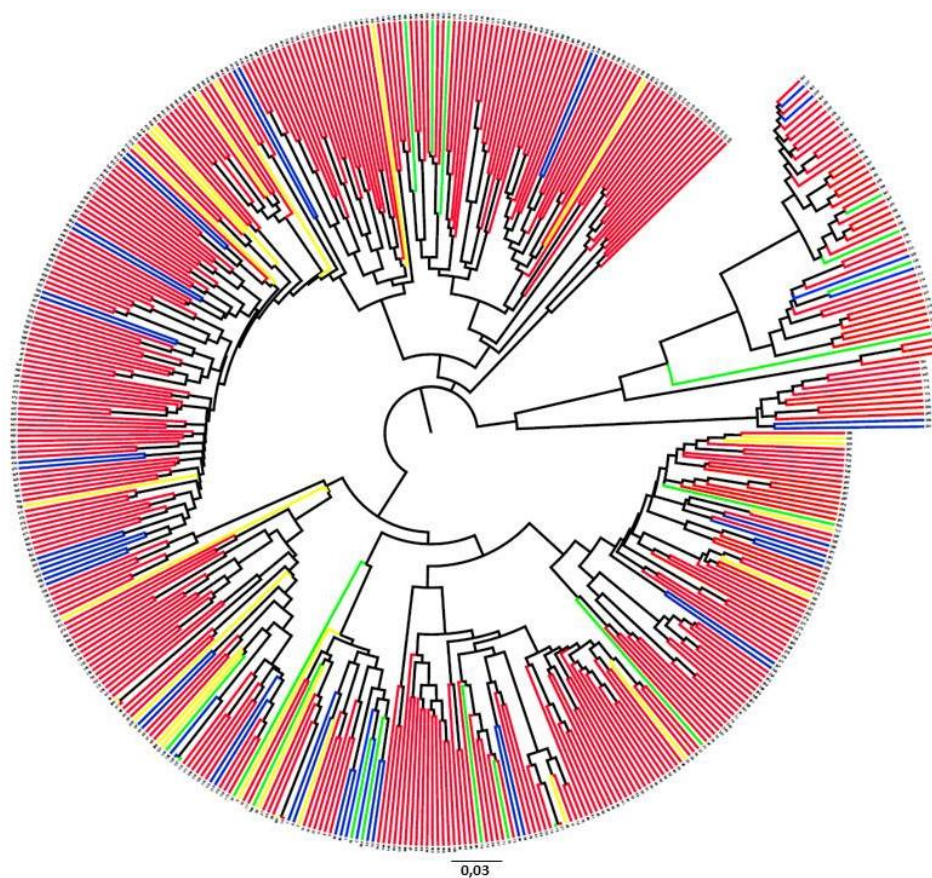


Figura 5 Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para 159 subamostras (representadas por 3 indivíduos cada) de *E. oleifera* que compõe o BAG da Embrapa, evidenciando possíveis erros de alocação de indivíduos em subamostra

Fonte: (PERRIER; JACQUEMOUD-COLLET, 2006).

Nota: Cor vermelha representa subamostras com os três indivíduos agrupados, cor verde representa subamostras que apresentam seus três indivíduos separados, cor azul representa subamostras que apresentam pelo menos dois dos três indivíduos agrupados sendo o terceiro representado pela cor amarelo. Análise realizada através do Programa Darwin v5.5

4 DISCUSSÃO

4.1 A extensão da diversidade genética de *E. oleifera* é menos restrita que se supunha

Observou-se a existência de moderada diversidade genética entre subamostras do Banco de Ativo de Germoplasma de *E. oleifera*. Embora os dados indiquem que esta variabilidade é apenas moderada, é interessante destacar que genótipos coletados na Amazônia Brasileira têm sido apontados como entre os mais diversos da espécie *E. oleifera*. Este fato já havia sido indicado por Barcelos et al. (2002) que ao avaliarem a diversidade genética em populações da espécie, identificaram quatro grandes grupos também relacionados de acordo com as regiões geográficas: 1. Brasil; 2. Peru; 3. América Central/Norte da Colômbia e 4. Suriname/Guiana Francesa, sendo a população Brasileira uma das mais diversas. Resultados semelhantes foram encontrados por Araya, Alvarado e escobar (2009), ao caracterizar a diversidade genética do banco de germoplasma de *E. oleifera* da empresa ASD - Agricultural Services & Development na Costa Rica por meio de oito marcadores microssatélites. Rios et al. (2012), com base no estudo de características fenotípicas, também relatam a existência de variabilidade entre as subamostras de *E. oleifera* do Banco Ativo de Germoplasma da Embrapa.

Embora Barcelos et al. (2002) tenham indicado que *E. oleifera* apresentava diversidade maior que *E. guineensis* (gene diversity de 0,157 e 0,135, respectivamente) Moretzsohn et al. (2002) demonstraram a diversidade genética da espécie americana (dissimilaridade média - 0,325) era mais restrita do que aquela observada em *E. guineensis* (dissimilaridade média - 0,528), mesmo considerando-se as diferenças de amostragem das mesmas. Os resultados apresentados neste estudo demonstram uma dissimilaridade média de 0,407 e,

portanto mais alinhado com os resultados apresentados por Moretzsohn et al. (2002) embora, podendo ser considerada menos restrita do que se supunha. Isso pode estar relacionado ao fato da bacia Amazônica provavelmente não ser o centro de origem de *E. oleifera*, mas sim um centro de diversificação secundária (GHESQUIÈRE et al., 1987; MORETZSOHN et al., 2002). Os resultados de Barcelos et al. (2002) podem ainda ter subestimado a diversidade de *E. guineensis* devido ao limitado número de subamostras analisadas quando comparada com *E. oleifera*. É interessante ressaltar que os pares de subamostras mais e menos divergentes identificados neste estudo não correspondem as mesmas que haviam sido identificadas por Moretzsohn et al. (2002). Esse fato pode ser um indicativo que uma maior quantidade de indivíduos e marcadores utilizados para amostrar a diversidade genética de uma determinada espécie pode prover resultados diferentes e possivelmente mais acurados.

Com base nos resultados aqui apresentados é possível se inferir sobre a existência de maior variabilidade para a população de Manicoré. Esse resultado além de coerente com aquele apresentado por Moretzsohn et al. (2002) que também constaram uma maior diversidade para subamostras de Manicoré e Autazes, ajuda a explicar o sucesso obtido por programas de melhoramento ao utilizar indivíduos da região de Manicoré como genitores de híbridos interespecíficos. Uma vez que estes indivíduos estão entre os mais diversos, a utilização de um grande número destes em cruzamentos com *E. guineensis* permite que melhoritas explorem melhor a diversidade de *E. oleifera* que se utilizassem um grande número de indivíduos provenientes de populações diferentes e menos diversas. Além disso, a prática adota pode ainda contribuir para maximizar a heterose ao cruzar indivíduos bastante distintos de *E. oleifera* com indivíduos igualmente diversos de *E. guineensis*.

4.2 O uso de uma bateria de alta densidade de marcadores moleculares proporcionou uma maior discriminação de subamostras por meio de técnicas dispersivas, aglomerativas e de otimização

Observou-se, de maneira generalizada, uma tendência de agrupamento por origem geográfica de coleta, uma vez que subamostras de uma mesma região tendem ao agrupamento conjunto (Figuras 2 e 3). Esses resultados contrastam com aqueles obtidos por Mohd Din, Rajanaidu e Jalani (2000) e Moretzsohn et al. (2002) e podem estar relacionados a dois fatores: (i) a amplitude da amostragem (abrangendo 206 subamostras pertencentes a 6 regiões distintas e assim uma amostragem completa do BAG) e (ii) a utilização de uma densidade de marcadores nos permite amostrar regiões do genoma antes não exploradas, contribuindo para estimativas diferentes de dissimilaridade.

Quando se analisou a diversidade a partir de 206 subamostras, observou-se que a maior e a menor dissimilaridade foram encontradas em pares de subamostras pertencentes a regiões de coletas diferentes. Este fato pode estar associado ao número reduzido de plantas por subamostra, principalmente Benjamin Constant e Manacapuru que continham apenas três plantas no BAG e Barcelos com apenas duas plantas representando a subamostra. E nesta análise as subamostras estão sendo representadas por uma única amostra. Comportamento este também observado por Martins, Miranda e Nunes (2012), ao analisarem a morfologia polínica em plantas de *E. oleifera*, utilizando grãos de pólen de quatro indivíduos de cada uma das 10 populações estudadas. Os autores relatam que a análise permitiu a distribuição das 10 populações em apenas três grupos e que apesar da técnica ter se mostrado eficiente, quando associada a outras análises como de isoenzimas que permitiu o mesmo grupo de populações apresentarem-se distribuídas em seis grupos (GHESQUIÈRE et al., 1987) e DNA nuclear pela análise de marcadores RFLP (*Restriction Fragment*

Length Polymorphism), a formação de cinco grupos com o mesmo conjunto de dados (BARCELOS et al., 1999), pode representar uma ferramenta importante na discriminação de populações.

Foi possível observar que houve diferença nos resultados em relação à consistência e melhor resolução dos agrupamentos entre as duas análises. Na análise com 206 amostras apresentaram pares mais e menos dissimilares diferentes dos encontrados na análise com 552 amostras. A análise com 206 subamostras apresentou resultados semelhantes com os encontrados por Moretzsohn et al. (2002), onde também foram utilizados números aproximados de indivíduos.

Um ponto interessante a ser destacado é que os resultados deste trabalho, embora corroborem com aqueles obtidos por Moretzsohn et al. (2002), demonstram que a diversidade genética em *E. oleifera* é de fato um pouco maior do que se supunha. Tal fato pode estar relacionado à quantidade de marcadores utilizados nos estudos. Enquanto Moretzsohn et al. (2002) fizeram uso de 96 marcadores RAPDs, o presente estudo fez uso de 3.187 marcadores. Provavelmente, um pequeno número de marcadores pode não ser suficiente para amostrar todo o genoma, considerando-se sua extensão. Assim, a utilização de um maior número de marcadores pode ter favorecido a identificação de diferenças entre subamostras imperceptíveis para uma quantidade limitada de marcadores. Isso porque ao mesmo se utilizar algumas centenas de marcadores (e.g. 200), para amostrar a diversidade, não mais que algumas dezenas de marcadores, a diversidade estimada com base em número reduzido de marcadores pode não se correlacionar com a diversidade fenotípica existente, pelo fato de os marcadores não amostrarem densamente todo o genoma da espécie.

A análise de coordenadas principais (PCoA) (Figuras 2 e S2) representam as distâncias genéticas geradas pelos marcadores PAVs e

confirmaram o agrupamento por origem geográfica o que ratifica os resultados obtidos pelos métodos anteriores. Porém, as duas primeiras coordenadas explicaram apenas 39,91% e 36,63 da variação entre as amostras, respectivamente. Resultados semelhantes foram encontrados por Arias et al. (2013), com 44% de variação e formação de quatro grupos distintos ao avaliarem a dispersão de 43 famílias de *Elaeis guineensis* com base em 30 marcadores microssatélites - SSRs (*Simple Sequence Repeats*). Com base no exposto, acredita-se que a dispersão gráfica conseguiu representar a variabilidade genética existente entre as subamostras de *E. oleifera*.

Além disso, as plantas que compõe o Banco Ativo de Germoplasma são oriundas de poucas sementes coletadas em um pequeno número de matrizes. O efeito fundador pode explicar o padrão de agrupamento, e associada a ele, o fenômeno do fluxo gênico através dos cursos hídricos (provavelmente com forte efeito antrópico). Esses resultados indicam ainda, que provavelmente, apenas uma pequena parcela da diversidade da espécie esteja representada no BAG, ao se coletar subamostras apenas nas regiões dos cursos hídricos de maior destaque na região. A fim de se ampliar a diversidade da coleção e fazendo uso destes dados, uma estratégia seria coletar novas subamostras em regiões distantes dos rios (menos sujeitas assim a efeito antrópico e fundador), priorizando-se a amostragem de um grande número de subamostras e de indivíduos destas subamostras (ESCOBAR, 1982).

4.3 O conhecimento detalhado da organização da diversidade genética da coleção permite se traçar estratégias de conservação, coleta e melhoramento mais eficientes

Atualmente com o desenvolvimento de novas plataformas de genotipagem, especialmente baseadas no princípio de genotipagem-por-

sequenciamento, pode-se genotipar rapidamente um grande número de amostras para virtualmente milhares de marcadores. Há assim a possibilidade de se estimar de maneira precisa o nível de diversidade existente em coleções de germoplasma mesmo que a espécie em questão não apresente uma grande quantidade de recursos genômicos (SANSALONI et al., 2010). Além disso, é certo que o conhecimento sobre a organização da diversidade genética da coleção de germoplasma de *E. oleifera* é de fundamental importância para embasar futuras ações de conservação e enriquecimento da coleção de germoplasma, além de indicar populações a serem exploradas para o desenvolvimento de híbridos interespecíficos entre *E. oleifera* e *E. guineensis*.

O conhecimento da diversidade genética dentro do germoplasma melhorado auxilia na organização da variabilidade desses materiais e potencializa a maximização dos ganhos genéticos em programas de melhoramento. Os dados obtidos neste estudo podem, por exemplo, ser utilizado a fim de delinear novas expedições de coleta de germoplasma para se ampliar a base genética mantida na coleção de germoplasma. Como anteriormente mencionando, a fim de se ampliar a diversidade da coleção uma estratégia seria coletar novas subamostras em regiões distantes dos rios, priorizando-se a amostragem de um grande número de subamostras e de indivíduos destas subamostras. Os dados obtidos neste estudo podem ainda ser utilizados para se delinear coleções nucleares para *E. oleifera*. Esta coleção de tamanho reduzido, mas que contém grande parte da variabilidade da coleção total tende a facilitar o manejo e a caracterização morfológica e agrônômica das subamostras.

Os *fingerprintings* moleculares, de maneira rápida e precisa, podem ser utilizados na definição de relações genéticas entre os materiais, na alocação desses em grupos heteróticos (PARENTONI et al., 2001). Logo, para fins de melhoramento, caso o interesse seja a geração de híbridos interespecíficos de *E. oleifera* com *E. guineensis* e na hipótese de não se possuir recursos para se testar

todas as combinações possíveis, pode-se direcionar os cruzamentos ao se selecionar subamostras a serem testadas com base em sua dissimilaridade com as demais. Assim pode-se priorizar uma determinada população a exemplo da Manicoré que se mostrou bastante diversa, ou indivíduos dentro de algumas dezenas de subamostras bastante diversas. Essa estratégia tende a ser mais eficiente que o teste ao acaso de subamostras, por permitir se explorar a diversidade da coleção de germoplasma de maneira organizada e dirigida.

Adicionalmente, a caracterização molecular de genótipos tem sido amplamente utilizada em processos de proteção de cultivares e no monitoramento da pureza genética de sementes (PADILHA, 2002), representando novas oportunidades tecnológicas devido às possibilidades de aumento de retorno dos investimentos aplicados nos programas de melhoramento. Tal procedimento pode vir a se tornar realidade para palma de óleo, uma vez que os marcadores e os perfis genotípicos gerados neste trabalho podem ser utilizados, por exemplo, para se monitor a eficiência dos cruzamentos protegidos realizados, por exemplo, no processo de geração de híbridos interespecíficos, ou na triagem de lotes de sementes para se garantir pureza genética dos híbridos resultantes dos programas de melhoramento.

5 CONCLUSÕES

As subamostras de *E. oleifera* apresentam variabilidade genética moderada e relacionada a sua região de origem geográfica, porém, essa diversidade é menos restrita do que se supunha.

O uso de um maior número de marcadores moleculares em *E. oleifera*, resultou em taxa de diversidade maior que encontrada anteriormente.

Os resultados obtidos neste estudo reforçam a hipótese que a Amazônia Brasileira deve ser um centro de diversificação secundário e que dada à história dispersiva da espécie a maior parte da diversidade concentra-se dentro de subamostras.

A população de Manicoré é a mais divergente amostrada na coleção de germoplasma da Embrapa.

A maior parte da variabilidade genética é encontrada dentro de subamostras.

A caracterização genética das subamostras permite se traçar planos para futuras ações de conservação e enriquecimento da coleção de germoplasma, além de indicar populações a serem exploradas para o desenvolvimento de híbridos interespecíficos entre *E. oleifera* e *E. guineensis*.

REFERÊNCIAS

- ARAYA, E., ALVARADO, A., ESCOBAR, R. Use of DNA markers for fingerprinting compact clones and determining the genetic relationship between *E. oleifera* germoplasm origins. San José, ASD. **Agricultural Services & Development**. Costa Rica. 21p. 2009.
- ARIAS, D., et al. Morpho-agronomic and molecular characterisation of oil palm *Elaeis guineensis* Jacq. material from Angola. **Tree Genetics e Genomes**. Amsterdam, v.9, p.1283–1294, oct. 2013.
- BARCELOS, E. Étude de la diversité génétique du genre *Elaeis* (*E. oleifera* (Kunth) et *E. guineensis* Jacq.), par marqueurs moléculaires (RFLP et AFLP). **These (Doutorado)** – École Nationale Supérieure Agronomique de Montpellier, France. 137p. 1998.
- BARCELOS, E., et al. Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. **Pesquisa Agropecuária Brasileira**, Brasília-DF, v. 37, n. 8, p. 1105-1114, aug. 2002.
- BARCELOS, E., et al. Molecular Markers Applied to the Biogeography of *Elaeis* (Palmae). In: HENDERSON, A., BORCHSENIUS, R., ed. Evolution, variation, and classification of palms. Bronx: New York Botanical Garden, 1999. p.191-201. Embrapa Amazônia Ocidental: CIRAD. 1999.
- CRUZ, C. D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**, Maringá – PR, v.35, n.3, p.271-276, jul/set. 2013.
- CRUZ, C. D., FERREIRA, F. M., PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco-MG. Suprema, 620p. 2011.
- DAVEY, J. W., et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, London, v. 12, p. 499–510, jun. 2011.
- ESCOBAR, R. Preliminary results of the collection and evaluation of the American oil palm *Elaeis oleifera* (H.B.K, Cortes) in Costa Rica. In: *The oil palm in agriculture in the eighties*, Vol. 1 (Ed. by E. Pushparajah & Chew P. S.), pp. 79–93, **Incorporated Society of Planters**. Kuala Lumpur. 1982.

- EXCOFFIER, L., SMOUSE, P. E., QUATTRO, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**. Austin. v. 131, p. 479-491, jun. 1992.
- GHESQUIÈRE, M., et al. Polymorphisme enzymatique chez *Elaeis oleifera* H. B. K. (*E. melanococca*). **Oléagineux**, Paris, v. 42, n. 4, p.143-150. 1987.
- JACCARD, P. Nouvelles recherches sur la distribution florale. **Bulletin de la Société vaudoise des sciences naturelles**. v. 44. p. 223–270. 1908.
- MARTINS, L. H. P., MIRANDA, I. P. de A., NUNES, C. D. Morfologia polínica de populações amazônicas de *Elaeis oleifera*. **Acta amazonica**. Manaus – AM. v. 33. n. 2. p. 159-166. 2012.
- MOHD DIN, A., RAJANAIDU, N., JALANI, B. S. Performance of *E. oleifera* from Panamá, Costa Rica, Colombia and Honduras in Malasya. **Journal of Oil Palm Research**, Kuala Lumpur, v.12, n. 1, p.71-80, Jun. 2000.
- MORETZSOHN, M. C., et al. Genetic diversity of Brazilian oil palm (*Elaeis oleifera*) germoplasm collected in the Amazon Forest. **Euphytica**, Wageningen, v. 124, p. 35-45, mar. 2002.
- MORETZSOHN, M. C., et al. RAPD linkage mapping of the Shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). **Theoretical and Applied Genetics**, Stuttgart – Alemanha, v. 100, n. 1, p. 63-70, jan. 2000.
- NIMMAKAYALA, P., et al. Linkage disequilibrium and population-structure analysis among *Capsicum annum* L. cultivars for use in association mapping. **Molecular Genetic and Genomics**, Berlin, v. 289, p. 513–521, aug. 2014.
- OOI, S. C., et al. Oil palm genetic resources: native *E. oleifera* populations in Brazil offer promising sources. **Pesquisa Agropecuária Brasileira**, Brasília – DF, v. 16, n. 3, p. 385-395, mai/jun. 1981.
- PADILHA, L. Marcadores moleculares semiautomatizados na caracterização e determinação da diversidade genética entre linhagens de milho tropical. **Tese** (Doutorado em Agronomia/Fitotecnia) – Universidade Federal de Lavras, Lavras - MG. 85 f. 2002.

PARENTONI, S. N. et al. Melhoramento para tolerância ao alumínio como fator de adaptação a solos ácidos. In: NASS, L. L. et al. **Recursos genéticos e melhoramento**. Fundação Mato Grosso. Rondonópolis- MT. p. 783-851. 2001.

PERRIER, X., JACQUEMOUD-COLLET, J. P. DARwin Software: <http://darwin.cirad.fr/darwin>. 2006.

RAO, C. R. **Advanced statistical methods in biometric research**. New York. Hafner Press, 390 p. 1952.

RIOS, S. da. A., et al. Recursos genéticos de palma de óleo (*Elaeis guineensis* Jacq.) e caiaué (*Elaeis oleifera* (H.B.K.) Cortés). **Documentos 96**. Embrapa Amazônia Ocidental. Manaus-AM. 39p. 2012.

SANSALONI, C. P., et al. Methodology A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. **Plant Methods**, London, p. 6-16, jun. 2010.

ANEXOS

ANEXO A - Material Suplementar

Tabela S1. Origem e número das plantas de palma de óleo Americana coletadas no BAG da Embrapa, pertencentes a seis regiões Amazônicas e distribuídas em 19 populações.

Análise com 206 amostras					
Origem/População	Nº/Plantas	Origem/População	Nº/Plantas	Origem/População	Nº/Plantas
Região de Manaus:		Rio Solimões:		Região Caracará:	
Careiro	25	Tefé	05	BR 174 (Caracará)	01
Manacapuru	01	Anori	03	(Km 157)	01
Irاندuba	02	Tonantins	04	(Km 365)	04
Caldeirão	07	B. Constant	01	(Km 490)	05
Rio Amazonas:		Coari	19	(Km 500)	01
Amatari	11	Rio Negro:		Vila Moderna	06
Autazes	11	Acajatuba	10	Rio Madeira:	
Maués	11	Moura	11	Manicoré	58
		Barcelos	02	Novo Aripuanã	07
Análise com 552 amostras					
Região de Manaus:		Rio Solimões:		Região Caracará:	
Careiro	68	Tefé	14	BR 174 (Caracará)	03
Manacapuru	03	Anori	09	(Km 157)	03
Irاندuba	06	Tonantins	12	(Km 365)	12
Caldeirão	18	B. Constant	03	(Km 490)	14
Rio Amazonas:		Coari	54	(Km 500)	03
Amatari	30	Rio Negro:		Vila Moderna	18
Autazes	28	Acajatuba	29	Rio Madeira:	
Maués	32	Moura	32	Manicoré	140
		Barcelos	02	Novo Aripuanã	19

Tabela S2. Análise de Variância Molecular (AMOVA) para 552 indivíduos de *E. oleifera* representadas por 19 populações distribuídas em seis regiões geográficas da Floresta Amazônica Brasileira. Análise realizada pelo Programa Genes (CRUZ, 2013).

FV	GL	SQ	QM	CV	PV	P
Regiões	5	82628,00	16525,60	124,27	23,61	**
Subamostra/Região	13	21855,95	1681,22	105,88	20,11	**
Subamostra	533	157848,08	296,15	296,15	56,26	**
Total	551	262332,05	476,10	526,30	100	

** significativo a 1% de probabilidade pelo teste F. FV: Fonte de Variação, GL: Grau de Liberdade, SQ: Soma dos Quadrados, QM: Quadrados Médios, CV: Coeficiente de Variação, PV: Porcentagem da Variação, P: Probabilidade.

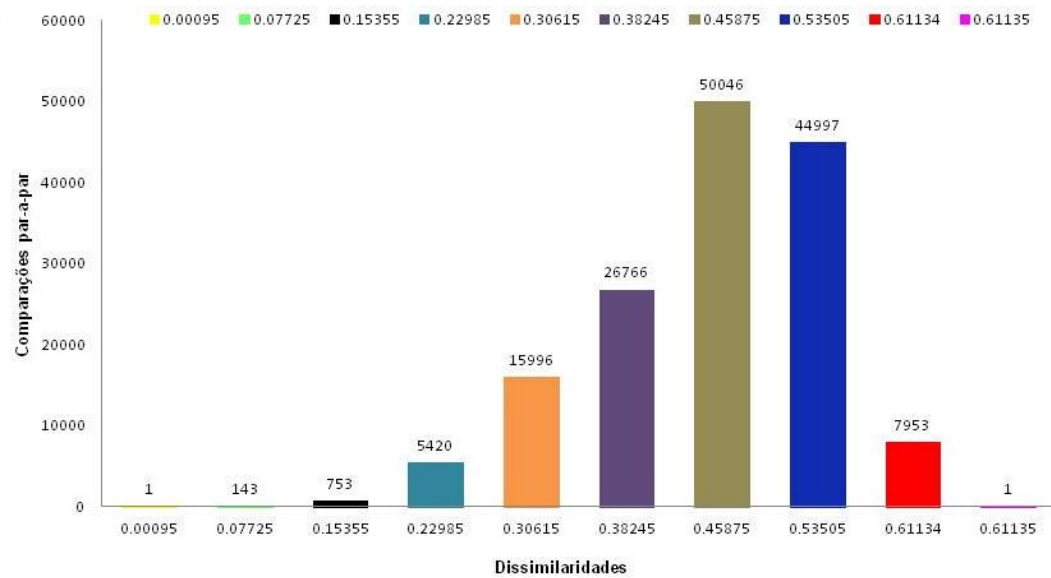


Figura S1. Distribuição de frequência das estimativas de distância genética, utilizando o coeficiente de dissimilaridade de Jaccard, obtidas entre os pares de 552 indivíduos de *E. oleifera*. Análise realizada através do Programa Darwin v5.5 (PERRIER; JACQUEMOUD-COLLET, 2006).

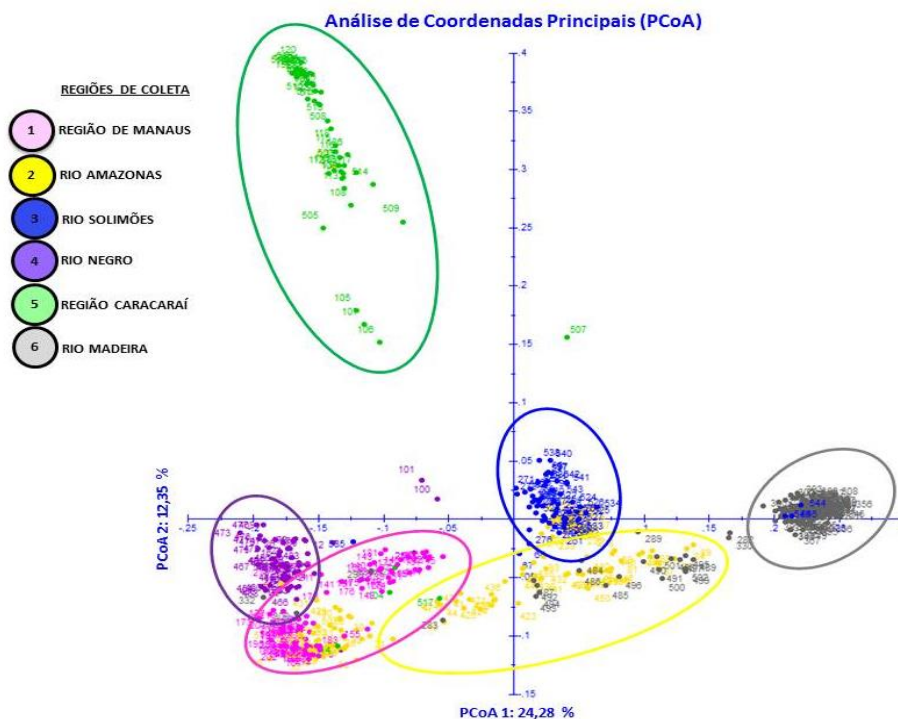


Figura S2. Dispersão Gráfica de 552 indivíduos de *E. oleifera* obtida pelo método de Análise das Coordenadas Principais (PCoA), utilizando o coeficiente de dissimilaridade de Jaccard. Análise realizada através do Programa Darwin v5.5 (PERRIER; JACQUEMOUD-COLLET, 2006).

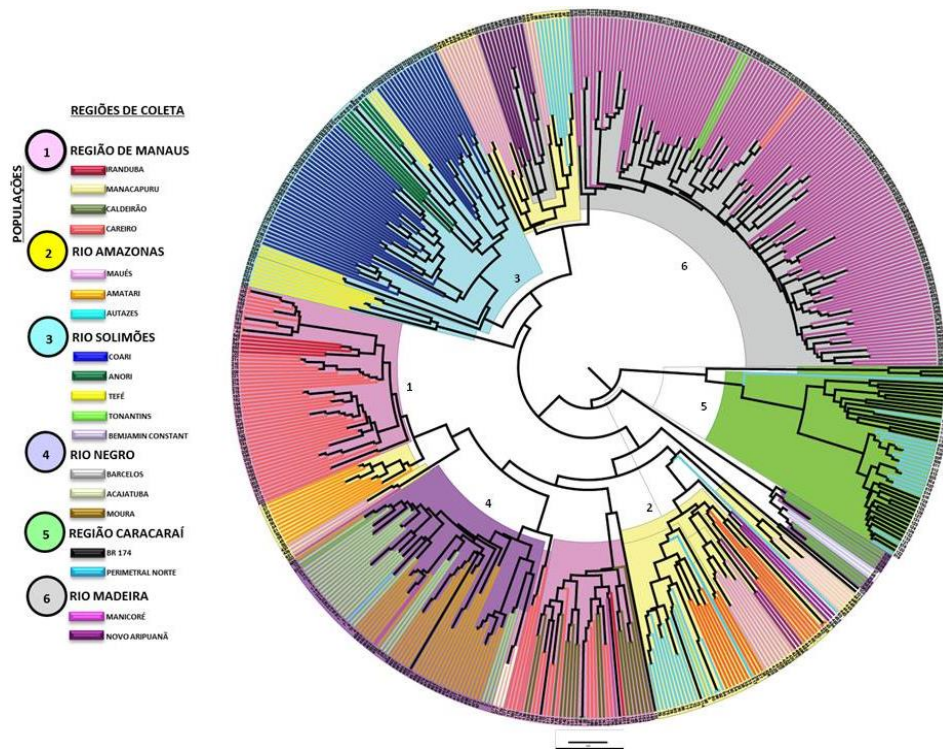


Figura S3. Dendrograma gerado pelo método UPGMA com base no índice de dissimilaridade de Jaccard para os 552 indivíduos de *E. oleifera* que compõe o BAG da Embrapa. Análise realizada através do Programa Darwin v5.5 (PERRIER; JACQUEMOUD-COLLET, 2006).