



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA
EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA
MESTRADO EM BIOTECNOLOGIA APLICADA À AGROPECUÁRIA

JOHNES PINTO SANCHES

**DIVERSIDADE E ESTRUTURA GENÉTICA POPULACIONAL EM ACESSOS DE
MANDIOCA (*Manihot esculenta* Crantz) DA EMBRAPA AMAZONIA ORIENTAL POR
MEIO DE MARCADORES SNPS**

BELÉM/PA

2019

JOHNES PINTO SANCHES

**DIVERSIDADE E ESTRUTURA GENÉTICA POPULACIONAL EM ACESSOS DE
MANDIOCA (*Manihot esculenta* Crantz) DA EMBRAPA AMAZONIA ORIENTAL POR
MEIO DE MARCADORES SNPS**

Dissertação apresentada à Universidade Federal Rural da Amazônia, como parte das exigências do curso de Mestrado em Biotecnologia Aplicada à Agropecuária: Área de concentração: Caracterização, conservação e uso de recursos genéticos, para a obtenção do título de Mestre.
Orientadora: Dra. Elisa Ferreira Moura Cunha.
Coorientadora: Dra. Nelcimar Reis de Sousa.

BELÉM/PA

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Bibliotecas da Universidade Federal Rural da Amazônia
Gerada automaticamente mediante os dados fornecidos pelo(a) autor(a)

- S211d Sanches, Johnes Pinto Sanches
DIVERSIDADE E ESTRUTURA GENÉTICA POPULACIONAL EM ACESSOS DE MANDIOCA
(*Manihot esculenta* Crantz) DA EMBRAPA AMAZONIA ORIENTAL POR MEIO DE MARCADORES
SNPS / Johnes Pinto Sanches Sanches. - 2019.
45 f. : il. color.
- Dissertação (Mestrado) - Programa de PÓS-GRADUAÇÃO em Biotecnologia Aplicada à Agropecuária
(PPGBAA), Campus Universitário de Belém, Universidade Federal Rural Da Amazônia, Belém, 2019.
Orientador: Profa. Dra. Elisa Ferreira Moura Cunha Cunha
Coorientador: Profa. Dra. Nelcimar Reis de Sousa.
1. Recursos Genéticos. 2. Euphorbiaceae. 3. Genômica. I. Cunha, Elisa Ferreira Moura Cunha, *orient.*
II. Título
-

CDD 631.5233

JOHNES PINTO SANCHES

JOHNES PINTO SANCHES

**DIVERSIDADE E ESTRUTURA GENÉTICA POPULACIONAL EM ACESSOS DE
MANDIOCA (*Manihot esculenta* Crantz) DA EMBRAPA AMAZONIA ORIENTAL POR
MEIO DE MARCADORES SNPS**

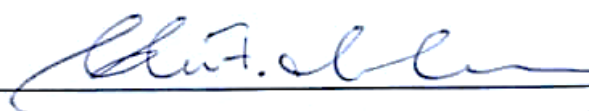
Dissertação apresentada à Universidade Federal Rural da Amazônia, como parte das exigências do curso de Mestrado em Biotecnologia Aplicada à Agropecuária: Área de concentração: Caracterização, conservação e uso de recursos genéticos, para a obtenção do título de Mestre.

Orientadora: Dra. Elisa Ferreira Moura Cunha.

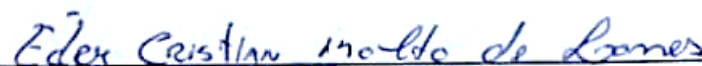
Coorientadora: Dra. Nelcimar Reis de Sousa

Data da Aprovação 20/02/2019

Banca examinadora



Elisa Ferreira Moura Cunha – Orientadora
Embrapa Amazônia Oriental



Éder Cristian Malta de Lanes – 1º Examinador
Instituto Tecnológico Vale – ITV



Rui Alberto Gomes Junior – 2º Examinador
Embrapa Amazônia Oriental



Maria do Socorro Padilha de Oliveira – 3º Examinadora
Embrapa Amazônia Oriental

LISTA DE ILUSTRAÇÕES

- Figura 1 - Parte do Banco Ativo de Germoplasma de mandioca (*Manihot esculenta* Crantz) da Embrapa Amazônia Oriental, localizado em sua sede em Belém, Pará. 16
- Figura 2 - Passos para a construção de bibliotecas de Genotipagem por Sequenciamento (GBS - *Genotyping by Sequencing*)..... 18
- Figura 3 - Mapa ilustrativo dos estados e o ponto de coleta (pontos vermelhos) dos acessos de mandioca coletados para compor o Banco Ativo de Germoplasma da Embrapa Amazônia Oriental em Belém, PA. 19
- Figura 4 - Distribuições e a densidade de 3.300 marcadores SNPs neutrais e independentes encontrados nos acessos conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, na qual foram mapeados fisicamente nos 18 cromossomos do genoma de mandioca (*Manihot esculenta* Crantz), gerados a partir tecnologia de Genotipagem por Sequenciamento. 28
- Figura 5 - Gráficos mostrando o número ideal de agrupamentos genéticos (k) considerando os 3.300 marcadores SNPs neutrais independentes analisando acessos de mandioca conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental gerados a partir da tecnologia de Genotipagem por Sequenciamento. A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará. A escolha k ideal baseia-se em erros de entropia cruzada da média \pm sd (LEA, I e II) e Critério de Informação Bayesiana (BIC) (DAPC, III e IV). 31
- Figura 6 - Agrupamentos genéticos dos acessos de mandioca do Banco Ativo de Germoplasma da Embrapa Amazônia Oriental (BAG) de acordo com a função snmf do pacote LEA, analisados com 3.300 marcadores SNPs neutrais independentes gerados pela tecnologia de Genotipagem por sequenciamento (GBS). A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará. (SU) Açucarada, (SW) mansa e (B) mandioca. 32
- Figura 7 - Agrupamentos genéticos a partir dos 3.300 marcadores SNPs neutrais independentes mapeados nos acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma (BAG) da Embrapa Amazônia Oriental de acordo com a Análise Discriminante dos Componentes Principais (DAPC). A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará. 33
- Figura 8 - Comparação da distribuição estabelecidos pelo pacote LEA e DAPC do pacote adegenet dos acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, considerando 3.300 SNPs neutrais e independentes gerados a partir da técnica de Genotipagem por sequenciamento (GBS). A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará 33
- Figura 9 - Dispersão de 397 acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, considerando 3.300 SNPs neutrais e independentes gerados a partir da técnica de Genotipagem por sequenciamento (GBS) com base na Análise de Componentes Principais - PCA. CP1: componente principal 1, CP2: componente principal 2. 35
- Figura 10 - Dispersão de 246 acessos de mandioca (*Manihot esculenta* Crantz) coletados no estado do Pará conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental,

considerando 3.300 SNPs neutrais e independentes gerados a partir da técnica de Genotipagem por sequenciamento (GBS) com base na Análise de Componentes Principais - PCA. CP1: componente principal 1, CP2: componente principal 2..... 35

LISTA DE TABELAS

- Tabela 1 - Local de coleta dos acessos de mandioca conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental com seus respectivos números de variedades de mandiocas bravas, mansas e açúcaradas que foram genotipados por meio da técnica de Genotipagem por Sequenciamento (GBS). Os acessos sem os dados de passaporte foram designados como ND.....20
- Tabela 2 - Distribuição genômica dos 3.300 marcadores SNPs neutrais e independentes encontrados nos acessos conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, na qual foram mapeados fisicamente nos 18 cromossomos do genoma de mandioca (*Manihot esculenta* Crantz), gerados a partir tecnologia de Genotipagem por Sequenciamento.27
- Tabela 3 - Porcentagem de transição e transversão dos 3.300 marcadores SNPs neutrais e independentes encontrados nos acessos conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, na qual foram mapeados fisicamente nos 18 cromossomos do genoma de mandioca (*Manihot esculenta* Crantz), gerados a partir tecnologia de Genotipagem por Sequenciamento.28
- Tabela 4 - Análise da Variância Molecular (AMOVA) entre e dentro dos grupos estabelecidos pelo pacote LEA utilizando os 3.300 marcadores SNPs neutrais e independentes gerados pela técnica de Genotipagem por Sequenciamento. A) considerando todos os 397 acessos de mandioca. B) 246 acessos coletados no estado do Pará. Df: Graus de liberdade; MS: Média dos Quadrados; %var: % Diferenciação Genética; F_{ST} : Índice de Fixação.37
- Tabela 5 - Medidas de diversidade genética dos grupos considerando todos os acessos e grupos estabelecido pelos acessos coletados no Pará, gerados pelo LEA. Cada agrupamento foi comparado com os grupos estabelecidos pelos tipos de mandioca. N: número de acessos; H_O : heterozigosidade observada; H_E : heterozigosidade esperada; F: coeficiente de endogamia; π : diversidade de nucleotídeos. Todas as estimativas são apresentadas ao longo de intervalos de confiança de 95% (CI). (SU) Açucarada, (SW) mansa e (B) mandioca.....38

RESUMO

A mandioca (*Manihot esculenta* Crantz) é uma cultura originária da região norte do Brasil, amplamente utilizada como fonte de subsistência às famílias dessa região e do mundo. Apresenta alta variabilidade genética, sendo cultivada em todo o Brasil e em outros países. Apesar da existência de vários estudos utilizando marcadores moleculares da mandioca, poucos são baseados na identificação de marcadores de polimorfismo de nucleotídeo único (SNPs). O presente estudo tem como objetivo avaliar os níveis de diversidade e a estrutura genética de acessos de mandioca conservados no Banco Ativo de Germoplasma (BAG) da Amazônia Oriental por meio de marcadores moleculares SNPs. Foram analisados 397 acessos de mandioca pertencentes ao BAG, que foram coletados principalmente em áreas no estado do Pará. As bibliotecas de DNA genômico foram construídas e os fragmentos foram sequenciados usando a plataforma Illumina. Utilizando o pipeline *Tassel-GBS* V5.2.44, as sequências foram identificadas e submetidas à análise de controle de qualidade para remover adaptadores e regiões de baixa qualidade. Inicialmente, foram identificados 670.835 SNPs, distribuídos em 18 cromossomos e 2 *scaffolds*. Posteriormente, foram removidos os *scaffolds*, inserções e deleções, retenção de dados perdidos e SNPs de acordo com a menor frequência alélica do alelo alternativo (MAF <0,05). Em seguida, os dados foram exportados para o ambiente R para filtragem adicional, considerando o desequilíbrio de ligação, o equilíbrio de Hardy-Weinberg e o desvio externo do F_{ST} usando o pacote R2vcftools. Após a filtragem adicional, foram identificados 3.300 SNPs neutros e independentes, com densidade média de 158,90kb por cromossomo. A estrutura genética foi estabelecida por meio de duas abordagens complementares, onde foram encontrados três grupos genéticos ($K = 3$) em ambas as análises e foi possível observar correspondência com a classificação com relação a um tipo de mandioca (brava, mansa ou açúcarada). Os parâmetros de diversidade foram iguais, dentro do intervalo de confiança, para os grupos 2 e 3 de dois agrupamentos genéticos e nos grupos das bravas e mansas, entretanto, os grupos 1 ($H_E = 0,48$) e o grupo das açúcaradas ($H_E = 0,33$) apresentaram menores valores de diversidade genética em comparação aos outros grupos. A Análise de Variância Molecular (AMOVA) revelou que há diferenciação genética considerando todos os acessos ($F_{ST} = 0,13$; $P < 0,001$) e a partir do grupo determinado pelos tipos de mandioca foi de $0,12$; $P < 0,001$. Analisando os acessos coletados no estado do Pará, os valores foram similares tanto para os grupos genéticos ($F_{ST} = 0,15$; $P < 0,001$), quanto para os grupos de acordo com os tipos ($F_{ST} = 0,14$; $P < 0,001$). Portanto, os marcadores SNPs neutrais foram eficazes no estudo da diversidade e estrutura genética da mandioca, onde foi possível observar a diversidade genética e estrutura genética presente no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, fornecendo uma melhor compreensão da variabilidade genética e suporte para o desenvolvimento de estudos genômicos de mandioca.

Palavras-chave: Recursos Genéticos. Euphorbiaceae. Genômica.

ABSTRACT

Cassava (*Manihot esculenta* Crantz) is a crop originating in the northern region of Brazil, widely used as a source of subsistence for families in this region and the world. It has high genetic variability, being cultivated throughout Brazil and in other countries. Despite the existence of several studies using cassava molecular markers, few are based on the identification of single nucleotide polymorphism markers (SNPs). The present study aims to evaluate the levels of diversity and the genetic structure of cassava accessions conserved in the Active Bank of Germplasm (BAG) of the Eastern Amazon using molecular SNPs markers. 397 cassava accessions belonging to the BAG were analyzed, which were collected mainly in areas in the state of Pará. Genomic DNA libraries were built and the fragments were sequenced using the Illumina platform. Using the Tassel-GBS V5.2.44 pipeline, the strings were identified and subjected to quality control analysis to remove low quality adapters and regions. Initially, 670,835 SNPs were identified, distributed in 18 chromosomes and 2 scaffolds. Subsequently, scaffolds, insertions and deletions, lost data retention and SNPs were removed according to the lower allele frequency of the alternative allele ($MAF < 0.05$). Then, the data were exported to the R environment for additional filtering, considering the link imbalance, the Hardy-Weinberg balance and the external deviation of the F_{ST} using the R2vcftools package. After additional filtration, 3.300 neutrals and independents SNPs were identified, with an average density of 158.90kb per chromosome. The genetic structure was established through two complementary approaches, where three genetic groups ($K = 3$) were found in both analyzes and it was possible to observe correspondence with the classification regarding the type of cassava (wild, soft or sweet). The diversity parameters were the same, within the confidence interval, for groups 2 and 3 of two genetic groups and in the bravas and meek groups, however, groups 1 ($H_E = 0.48$) and the sugary group ($H_E = 0.33$) had lower values of genetic diversity compared to the other groups. The Molecular Variance Analysis (AMOVA) revealed that there is genetic differentiation considering all accessions ($F_{ST} = 0.13$; $P < 0.001$) and from the group determined by the types of cassava it was 0.12; $P < 0.001$. Analyzing the accessions collected in the state of Pará, the values were similar both for the genetic groups ($F_{ST} = 0.15$; $P < 0.001$), and for the types collected in Pará ($F_{ST} = 0.14$; $P < 0.001$). Therefore, the neutral SNPs markers were effective in studying the diversity and genetic structure of cassava, where it was possible to observe the genetic diversity and genetic structure present in the Active Germplasm Bank of Embrapa Amazônia Oriental, providing a better understanding of genetic variability and support for the development of genomic studies of cassava.

Keywords: Active Germplasm bank. Euphoiaceae. Genomic.

SUMÁRIO

1. INTRODUÇÃO	11
1. OBJETIVOS	13
1.1. Geral.....	13
1.2. Específicos	13
2. REVISÃO DE LITERATURA	14
2.1. A mandioca	14
3.1. Marcadores moleculares	16
3.2. Genotipagem por sequenciamento	17
4. MATERIAL E MÉTODOS	19
4.1. Material Vegetal	19
4.2. Extração de DNA e Genotipagem por Sequenciamento (GBS).....	21
4.3. Análise de dados de sequenciamento.....	22
4.4. Análise da estrutura genética	24
4.5. Análise da variância molecular (AMOVA) e índices de diversidade genética	24
5. RESULTADOS E DISCUSSÃO	26
5.1. Identificação de SNPs e Processamento dos dados	26
5.2. Análise da estrutura genética	29
5.3. Análise da variância molecular (AMOVA) e índices de diversidade genética	36
REFERÊNCIAS	40

1. INTRODUÇÃO

A mandioca (*Manihot esculenta* Crantz) é uma planta dicotiledônea pertencente à família Euphorbiaceae e gênero *Manihot*, que possui aproximadamente 98 espécies, sendo *Manihot esculenta* a mais cultivada. A espécie é originária da América do Sul e a sua possível domesticação ocorreu na Região Amazônica e é cultivada principalmente por pequenos produtores de países de clima tropical e subtropical (OLSEN, 2004; LEOTÁRD *et al.*, 2009). A mandioca produz uma raiz tuberosa de importância para a alimentação humana, usada como subsistência para muitas famílias (FERREIRA *et al.*, 2008). A nível mundial, a espécie é considerada a principal fonte de carboidratos para mais de 800 milhões de pessoas, representando grande importância para a alimentação humana (LEBOT, 2009; FERREIRA *et al.*, 2008).

A espécie apresenta imensa diversidade de usos, tanto de raízes quanto de parte aérea, com destaque para as raízes tuberosas, ricas em carboidratos (CARVALHO e SCHAAL, 2001; FERREIRA *et al.*, 2008).

Diversos estudos de caracterização de genótipos de mandioca têm se concentrado na análise de descritores morfológicos e no potencial cianogênico das raízes para classificar os genótipos como mansa ou brava. No entanto, o uso de descritores morfológicos não permite uma clara distinção entre os acessos de mandioca, possivelmente devido à capacidade limitada de discriminação (BENESI *et al.* 2010; KAWUKI *et al.* 2011; VIEIRA *et al.*, 2013; AFONSO *et al.*, 2014).

Os marcadores moleculares têm tido cada vez mais destaque como uma ferramenta essencial para identificação e caracterização de genótipos, contribuindo para o melhoramento genético de plantas e subsidiando o melhor planejamento em coleções de germoplasma (HE *et al.*, 2014).

Dentre os marcadores, os de polimorfismo de nucleotídeo simples (SNP – *Single Nucleotide Polymorphism*) destacam-se pelo alto desempenho na geração de informações, e o sequenciamento de nova geração (NGS – *Next Generation Sequencing*) vem permitindo a descoberta de milhares a centenas de milhares desses marcadores, acessando diretamente no DNA genômico em uma etapa única de sequenciamento paralelamente a uma construção de biblioteca genômica denominada de genotipagem por sequenciamento (GBS - *Genotyping by Sequencing*) que surgiu como uma ferramenta útil, relativamente simples e barata para

contribuir para análises genéticas (ELSHIRE *et al.*, 2011). A técnica de GBS, mesmo com a baixa cobertura das sequências de DNA, tem apresentado sucesso em diversos trabalhos com mandioca, a exemplo da análise genômica populacional (RABBI *et al.*, 2015; ALBUQUERQUE, *et al.*, 2018).

Dessa forma, o objetivo deste trabalho foi avaliar os níveis de diversidade e a estrutura genética de acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma (BAG) da Amazônia Oriental por meio de marcadores moleculares SNPs.

1. OBJETIVOS

1.1. Geral

Avaliar os níveis de diversidade e a estrutura genética de acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma (BAG) da Amazônia Oriental por meio de marcadores moleculares SNPs.

1.2. Específicos

- (i) Identificar um subconjunto de SNPs neutrais e independentes para estudo de genômica populacional de acessos de mandioca;
- (ii) Caracterizar a estrutura genética dos acessos de mandioca conservados no banco de germoplasma, considerando os pressupostos genéticos de uma população ideal e os tipos de mandioca (brava, mansa e açúcarada);
- (iii) Avaliar o nível de variabilidade genética dos acessos conservados no BAG de mandioca da Embrapa Amazônia Oriental.

2. REVISÃO DE LITERATURA

2.1. A mandioca

A mandioca (*Manihot esculenta* Crantz, Euphorbiaceae) pertence ao gênero *Manihot*, sendo a única cultivar das 98 espécies do gênero. A espécie apresenta parte aérea arbustiva lenhosa e perene com altura entre 1 a 5m, pode ser cultivada em pequenas áreas, com diferentes características climáticas e tipos de solo, usando baixo nível tecnológico (EMBRAPA, 2011). É classificada como planta tropical e é adaptada a solos de baixa fertilidade, sendo capaz de ser propagada por estacas ou sementes (LEBOT, 2009).

A espécie apresenta grande variabilidade genética refletindo em uma grande diversidade de usos, tanto de raízes quanto de parte aérea (CARVALHO e SCHAAL, 2001; FERREIRA *et al.*, 2008; MÜHLEN *et al.*, 2010) sendo considerada uma das mais importantes fontes calóricas de alimentação nos trópicos, atrás apenas de cana-de-açúcar, milho, arroz, trigo, batata, beterraba e soja, uma importante fonte de carboidratos para mais de 800 milhões de pessoas no mundo, principalmente nos países tropicais subdesenvolvidos (LEBOT, 2009).

Historicamente, a mandioca vem sendo propagada vegetativamente pela interferência humana, contudo manteve a reprodução sexuada ativa, importante para a variabilidade genética e possibilitando aos melhoristas selecionar acessos de maior importância agrônômica (SILVA *et al.* 2001).

A espécie é originária da América do Sul, e o Brasil é considerado o principal centro de diversidade da espécie, principalmente na Região Amazônica, apresentando grande importância para a agricultura familiar regional, devido ao significativo número de famílias do meio rural que vivem da produção e do processamento da farinha, e de outros produtos desta cultura (OLSEN, 2004). Segundo Valle *et al.* (2014), a mandioca é cultivada para o consumo próprio e para abastecer pequenos negócios locais e regionais.

A mandioca é classificada em três tipos, sendo dois os mais conhecidos, de acordo com o potencial cianogênico: mandioca brava, também conhecida como mandioca para indústria apresenta teor de HCN superior a 100mg.kg^{-1} na polpa crua de raiz e a mandioca mansa conhecida também como aipim, macaxeira, mandioca para mesa e mandioca doce, com teor inferior a 100mg.kg^{-1} (VALLE *et al.*, 2004; VIEIRA *et al.*, 2009). Além dessas, a mandioca açucarada possui alto teor de açúcar livres nas suas raízes e baixo teor de amido em suas raízes e corresponde a um terceiro tipo, existente na Amazônia brasileira, sendo utilizada para a

produção de uma bebida conhecida como mandiocabas e com potencial para a produção de etanol (CARVALHO *et al.*, 2004). Independentemente do tipo, todas são da espécie *M. esculenta*, mandioca, segundo Carvalho e Guerra (2002), é uma espécie diploide, com $2n=36$ cromossomos e meiose regular de $n=18$.

A Embrapa Amazônia Oriental mantém um Banco Ativo de Germoplasma (BAG) de mandioca, contendo acessos tradicionalmente cultivadas em fazendas familiares em diferentes locais do Brasil, dessa forma, promover a preservação e conservação deste importante recurso, algumas instituições brasileiras mostrando-se preocupadas com o uso inadequado e ao risco de redução da sua variabilidade genética, podendo ocasionar perda de genes ou erosão genética desses recursos (OLIVEIRA *et al.*, 2018)

O BAG de mandioca da Empresa Brasileira de Pesquisa Agropecuária da Amazônia Oriental conserva, em condições *ex situ*, o material genético (germoplasma) que foi coletado desde 1940, em diferentes locais do Brasil, sendo constituído de 470 acessos, incluindo acessos de mandioca mansa (SW), brava (B) e açúcarado (mandiocabas) (SU) (VIEIRA *et al.*, 2009; MOURA *et al.*, 2013).

Para o melhor aproveitamento do BAG de mandioca, é de fundamental importância a geração de informações sobre sua variabilidade genética, contribuindo para o melhor gerenciamento do banco e contribuindo para o conhecimento do material conservado. Portanto, segundo Oliveira *et al.* (2014) é essencial conhecer a estrutura genética dos acessos. A manutenção da variabilidade em bancos de germoplasma é de importância fundamental, pois ajuda a reduzir a erosão genética. Uma forma de acessar tais informações seria por meio dos marcadores moleculares.

Figura 1 - Parte do Banco Ativo de Germoplasma de mandioca (*Manihot esculenta* Crantz) da Embrapa Amazônia Oriental, localizado em sua sede em Belém, Pará.



Foto: Johnes Sanches, 2019.

3.1. Marcadores moleculares

Marcadores moleculares são ferramentas extremamente úteis para avaliar a divergência genética entre e dentro de populações de espécies tropicais (JAMNADASS *et al.*, 2009; ARIF *et al.*, 2010). Tornando-se uma ferramenta extremamente útil para a ciência, auxiliando na geração de informações moleculares, os SNPs se destacam como marcadores moleculares gerados a partir de variações de um único nucleotídeo em um local particular no genoma, consequências das substituições, deleções e inserções, formas de variação mais comuns em diferentes genomas, uma importante forma natural de variação genética que ocorre na maioria das vezes em regiões não codificadoras do genoma (CAETANO, 2009). Segundo Ly *et al.* (2013), os marcadores moleculares fornecem um número praticamente ilimitado de informações genéticas.

Os marcadores moleculares podem ser classificados de acordo com o tipo de herança alélica em dominantes (RAPD – *Random Amplified Polymorphic DNA* e AFLP – *Amplified Fragment Length Polymorphisms*), na qual possibilitam a identificação da presença ou ausência

de um determinado alelo e marcadores codominantes (RFLP – *Restriction Fragment Length Polymorphism*, SSR – *Simple Sequence Repeat* e SNPs – *Single Nucleotide Polymorphism*), que permitem diferenciar indivíduos homocigotos e heterocigotos (CLEMENT, 2001).

Dentre os marcadores codominantes, os SNPs apresentam certas vantagens como o elevado potencial para análise automatizada de alto rendimento a custo moderado, o baixo custo por dados gerados, ampla distribuição no genoma e a grande reprodutibilidade aumenta as possibilidades de utilização, aliado ao alto desempenho do sequenciamento de nova geração (NGS – *Next Generation Sequencing*), permite a descoberta de milhares a centenas de milhares de marcadores, executadas diretamente no DNA genômico em uma etapa única de sequenciamento paralelamente a uma construção de biblioteca genômica (ARIF *et al.*, 2010 ; ELSHIRE *et al.*, 2011).

Os marcadores moleculares permitiram a geração de diversos estudos, sendo bastante utilizados no estudo de diversidade genética e estrutura populacional de diversas espécies (WRIGHT *et al.*, 2015; ZHANG *et al.*, 2016; VISSER *et al.*, 2016), incluindo diversos estudos de conjuntos de genótipos de mandioca (COSTA *et al.*, 2003; FERREIRA *et al.*, 2008; MÜHLEN *et al.*, 2013; OLIVEIRA *et al.*, 2014; RABBI *et al.*, 2015; MOURA *et al.*, 2016; WOLFE *et al.*, 2017; OLIVEIRA *et al.*, 2018; ALBUQUERQUE *et al.* 2018).

3.2. Genotipagem por sequenciamento

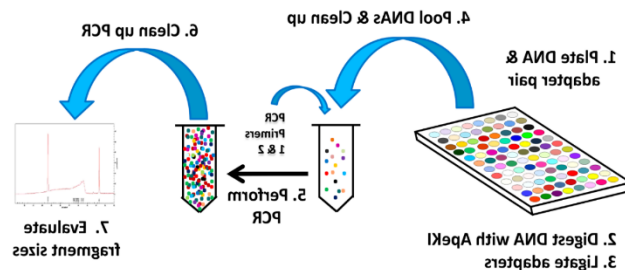
A tecnologia de Genotipagem por Sequenciamento (GBS – *Genotyping by Sequencing*) é uma metodologia de representação reduzida do genoma, proveniente da metodologia RAD-seq – *Restriction-site associated DNA sequencing* (ELSHIRE *et al.*, 2011; MARRANO *et al.*, 2017). A GBS é uma técnica simples, altamente multiplexável, adequada para estudos populacionais, caracterização de germoplasma, dentre outros (DAVEY *et al.*, 2011).

Apesar da alta taxa de dados perdidos devido à baixa cobertura da tecnologia (KIM *et al.*, 2016; MARRANO *et al.*, 2017), a técnica gera rapidamente milhares de pontos de dados de marcadores com custos e espaço de tempo cada vez mais reduzidos, que são distribuídos por todo o genoma (ELSHIRE *et al.*, 2011; DAVEY *et al.* 2011; BEISSINGER *et al.* 2013)

A metodologia envolve a clivagem do DNA genômico com enzimas de restrição, geralmente a *ApeKI*, que é uma endonuclease de restrição do tipo II, a qual reconhece uma sequência degenerada de 5pb (GCWGC, onde W é A ou T) (HAMBLIN *et al.*, 2014,

MARRANO *et al.*, 2017), reduzindo a complexidade do genoma de maneira rápida e fácil. Em seguida são ligados aos adaptadores com *barcode* e comuns, sendo que após a ligação dos adaptadores, as amostras são agrupadas em um único *pool* e é realizada a reação em cadeia da polimerase (PCR – *Polymerase Chain Reaction*). Dessa forma, apenas amostras curtas (<1kb) com a combinação *barcode* e o adaptador comum serão amplificadas em seguida sequenciados utilizando de plataformas de sequenciamento de nova geração (DAVEY *et al.*, 2011; PETERSON *et al.*, 2014) (Figura 2).

Figura 2 - Passos para a construção de bibliotecas de Genotipagem por Sequenciamento (GBS - *Genotyping by Sequencing*).



Fonte: ELSHIRE *et al.* (2011).

A enzima de restrição *ApeKI* geralmente é selecionada por apresentar uma sensibilidade parcial a metilação do DNA, e corta o DNA genômico em baixa frequência, evitando regiões genômicas repetitivas (DAVEY *et al.*, 2011; MARRANO *et al.*, 2017).

Abordagens como o GBS permitem a análise de organismos com genomas grandes e diversos, e, no caso da mandioca, facilita para a descoberta e genotipagem de SNPs, auxiliando na identificação de variedades e no melhor planejamento de manejo da espécie. Dessa forma, os marcadores SNPs surgem como importantes marcadores genéticos na avaliação da diversidade genética de espécies como a mandioca (RABBI *et al.*, 2015; WOLFE *et al.*, 2017), a cevada (ELSHIRE *et al.* 2011). Os SNPs também podem ser empregados no mapeamento de associação de genes candidatos, como para o conteúdo de β -caroteno na mandioca (UDOH *et al.*, 2017).

4. MATERIAL E MÉTODOS

4.1. Material vegetal

Um conjunto de 397 acessos de mandioca foi utilizado, em que 324 acessos foram classificados como mandioca brava, 49 como mansa ou macaxeira e 24 como açúcaradas ou mandiocabas. Essa classificação foi com base na informação fornecida pelos produtores no momento da coleta. Os acessos pertencem ao Banco Ativo de Germoplasma de mandioca da Embrapa Amazônia Oriental, localizado no município de Belém, Pará, Brasil (01° 27' 21" S e 48° 30' 16" W). Os acessos foram coletados em diferentes locais de sete estados, sendo 246 no Pará (61,96%), 18 no Amazonas (4,53%), quatro em Rondônia (1,01%), 33 no Amapá (8,29%), um no Mato Grosso do Sul (0,25%), um no Paraíba (0,25%), dois no Paraná (0,50%) e 92 acessos não possuem dados de local de coleta (23,12%) (Figura 3, Tabela 1).

Figura 3 - Mapa ilustrativo dos estados e o ponto de coleta (pontos vermelhos) dos acessos de mandioca coletados para compor o Banco Ativo de Germoplasma da Embrapa Amazônia Oriental em Belém, PA.

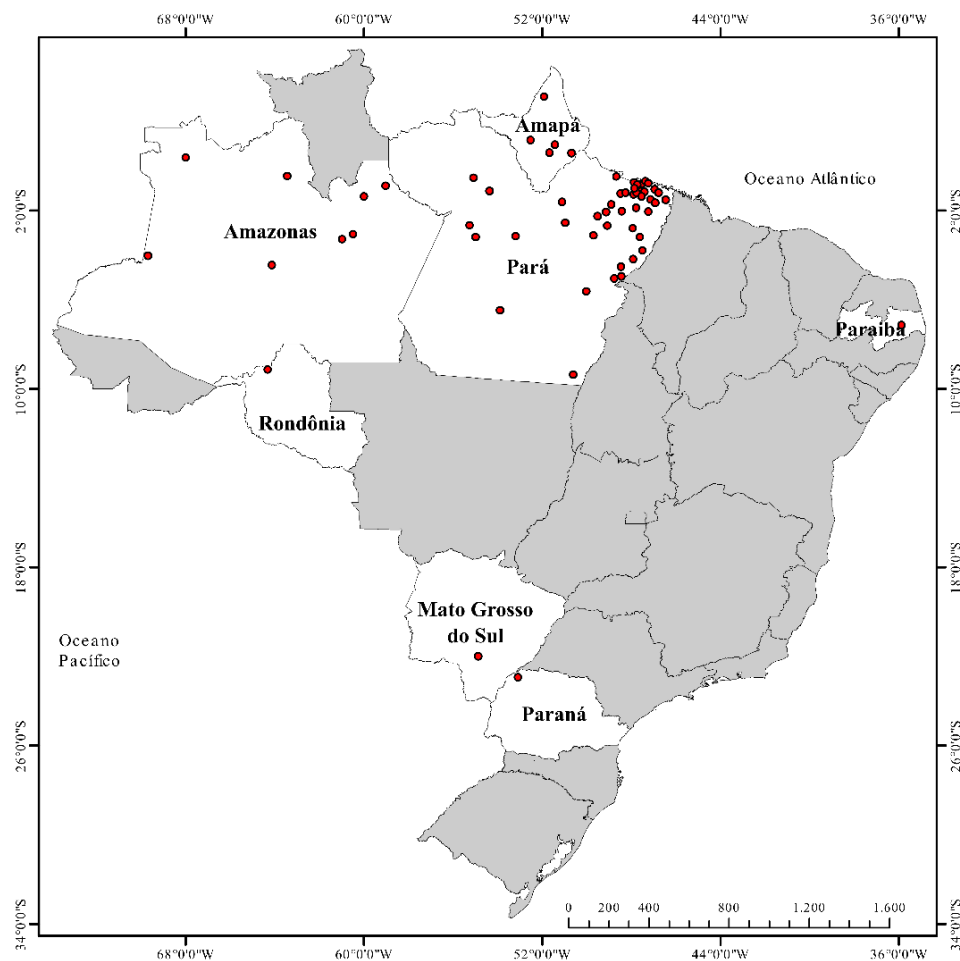


Tabela 1 - Local de coleta dos acessos de mandioca conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental com seus respectivos números de variedades de mandiocas bravas, mansas e açúcaradas que foram genotipados por meio da técnica de Genotipagem por Sequenciamento (GBS). Os acessos sem os dados de passaporte foram designados como ND.

Estado	mandioca			Total	
	Município	Brava (B)	Mansa (SW)		Açucarada (SU)
Amapá					
	Ferreira Gomes	1	2	-	3
	Macapá	1	-	-	1
	Oiapoque	8	-	-	8
	Pedra do Amapari	12	3	-	15
	Porto Grande	6	-	-	6
	Total	28	5	-	33
Amazonas					
	Alto Rio Negro	3	-	-	3
	Carvoeiro	1	-	-	1
	Irاندوبا	1	-	-	1
	Lago Mamai	-	1	-	1
	Manacapuru	-	1	-	1
	Marara	1	-	-	1
	Panacarica	1	-	-	1
	Presidente Figueiredo	1	-	-	1
	Tabatinga	6	1	-	7
	Urucará	1	-	-	1
	Total	15	3	-	18
Pará					
	Abaetetuba	6	-	2	8
	Abel Figueiredo	1	-	-	1
	Acará	11	-	-	11
	Alenquer	2	1	-	3
	Altamira	2	-	-	2
	Baião	2	-	-	2
	Belém	5	2	-	7
	Belterra	4	1	-	5
	Bom Jesus do Tocantins	1	-	-	1
	Bragança	5	1	1	7
	Cametá	3	-	-	3
	Capitão Poço	-	2	-	2
	Castanhal	11	1	1	13
	Curuçá	1	-	-	1
	Dom Eliseu	4	-	-	4
	Igarapé Açu	6	1	4	11
	Igarapé Miri	2	-	-	2
	Ipixuna	1	-	-	1
	Marabá	-	1	-	1
	Maracanã	-	-	1	1
	Marapanim	2	1	2	5
	Medicilândia	1	1	-	2
	Melgaço	26	-	1	27

Moju	15	-	-	15
Monte Alegre	1	-	-	1
Nova Timboteua	3	-	-	3
Ourém	2	-	-	2
Paragominas	18	2	-	20
Portel	2	-	-	2
Rondon do Pará	8	6	-	14
Salinópolis	-	-	1	1
Santa Bárbara	-	-	1	1
Santa Luzia	2	-	-	2
Santa Maria	1	-	-	1
Santana do Araguaia		1	-	1
Santarém	30	5	1	36
São Domingo do Capim	2	-	-	2
São Francisco do Pará	-	1	1	2
São João de Pirabas	-	-	5	5
Soure	1	-	-	1
Terra Alta	3	-	-	3
Tracuateua	8	-	3	11
Ulianópolis	-	2	-	2
Viseu	-	1	-	1
Total	192	30	24	246
Mato Grosso do Sul	1	-	-	1
Itaporã	1	-	-	1
Paraíba	1	-	-	1
Lagoa Seca	1	-	-	1
Paraná	2	-	-	2
Loanda	2	-	-	2
Rondônia	4	-	-	4
Calama	4	-	-	4
ND	81	11	-	92
Total	324	49	24	397

Fonte: O Autor (2019)

4.2. Extração de DNA e Genotipagem por Sequenciamento (GBS)

As amostras de tecido foliar foram processadas no Laboratório de Genética Molecular da Embrapa Amazônia Oriental, Belém, PA. As amostras foram mantidas em freezer -80°C até o momento da extração. O DNA genômico foi extraído conforme o protocolo de Doyle e Doyle (1990), com modificações descritas por Moura *et al.* (2013). O tecido foliar, depois de lavado com hipoclorito 10% e água destilada, foi macerado em cadinho de porcelana previamente resfriado, adicionando nitrogênio líquido para romper as paredes celulares e para liberar constituintes celulares. Foram adicionados 3mL do tampão de extração (CTAB 2%,

NaCl 5M, EDTA 0,5M, PVP, Tris-HCl 1M e água estéril) foram adicionados a maceração. Em seguida, o macerado e o tampão de extração foram homogeneizados em vortex e os tubos foram incubados em banho-maria à 65°C por 1 hora, sendo agitados de 10 em 10 minutos. Em seguida, foi adicionada solução de clorofórmio e álcool isoamílico (24:1; v/v) e invertido levemente por 100 vezes para a separação das fases. Posteriormente, foram homogeneizados e centrifugados a 10.000rpm durante 10 minutos.

Após retirar o sobrenadante, foi adicionado 5mL de álcool 95% gelado e retornado a centrifuga por mais dez minutos a 10.000 rpm. Finalizando este passo, foi realizada a lavagem do precipitado com álcool 70% e centrifugado por 10 minutos a 10.000 rpm. O DNA obtido foi armazenado em tampão TE (10mM Tris-HCl pH 8,0; 1mM EDTA), acrescido de RNase 10ng.µL⁻¹. As amostras foram aplicadas em gel de agarose a 1%, corado com brometo de etídio, utilizando-se três DNAs *lambda* (Invitrogen, São Paulo, Brasil) nas concentrações de 50, 100 e 200ng.µL⁻¹. Os perfis dos géis foram visualizados em transiluminador sob luz ultravioleta e em seguida, fotodocumentados para avaliação da concentração e a integridade da molécula de DNA.

A genotipagem por sequenciamento foi preparada utilizando a enzima de restrição *ApeKI*, gerando os marcadores moleculares do tipo SNP de acordo com metodologia descrita por Elshire *et al.* (2011). A enzima de restrição *ApeKI* foi selecionada porque apresenta uma sensibilidade parcial a metilação do DNA e corta o DNA genômico a baixa frequência, evitando assim as regiões genômicas repetitivas e é extremamente útil na promoção da ligação eficiente do adaptador com o DNA (Peterson *et al.*,2014).

4.3. Análise de dados de sequenciamento

As seqüências dos acessos foram identificadas a partir dos adaptadores *barcode*, com pelo menos três nucleotídeos de diferença, para identificar cada indivíduo e alinhar com a seqüência do genoma de referência da mandioca v6.1, disponível no banco de dados *Phytozome* v12.0 (<http://www.phytozome.net/cassava.php>) (GLAUBITZ *et al.*, 2014). Com isso, foi aplicado o chamado *Discovery Pipeline* para obtenção dos SNPs, usando o *pipeline* TASSEL-GBS v5.2.44 (BRADBURY *et al.* 2007), seguindo os algoritmos de bioinformática descritos por Elshire *et al.* (2011) e Glaubitz *et al.* (2014). O TASSEL *GBSSeqToTagDBPlugin* foi

utilizado com tamanho de 64 nucleotídeos e com pontuação mínima de qualidade de 20 ($mxTagL = 64$, $mntagL = 20$, $mnQS = 20$, $mxTagNum = 100000000$).

As *tags* GBS foram exportadas do banco de dados no formato FASTQ através do TASSEL *TagExportToFastqPlugin* e foram alinhadas com a sequência do genoma de referência da mandioca v6.1, disponível no banco de dados *Phytozome* v12.0 (<http://www.phytozome.net/cassava.php>). Os alinhamentos foram importados no formato SAM (alinhamento de sequência/mapa) usando o TASSEL *SAMToGBSdbPlugin* através dos parâmetros ($aProp = 0,0$ e $aLen = 0$). Em seguida, os SNPs foram chamados a partir dos alinhamentos importados, usando a função TASSEL *DiscoverySNP CallerPluginV2*, onde foi considerado a cobertura de loco mínima entre taxa de 0,1 e frequência de alelo menor (MAF) mínima de 0,001 e máximo de 64 *tags* permitidas para alinhar por site de corte ($mnLCov = 0,1$ e $mnMAF = 0,001$). Para a análise de qualidade e profundidade, foi usado o TASSEL *SNPQualityProfilerPlugin*.

O TASSEL *ProductionSNP CallerPluginV2* foi utilizado para exportar dados SNP no formato HAPMAP com o comprimento dos *kmer* máximo de 64 *tags* ($kmerLength = 64$). Posteriormente, foi realizada a retirada das inserções e deleções e retenção de locos com um máximo de 20% de dados perdidos, em seguida os dados foram filtrados considerando menor frequência alélica do alelo alternativo (MAF - *Minor Allele Frequency*) igual ou superior a 0,05, e por fim, houve a retirada dos *scaffolds*. Dessa forma, foram eliminados marcadores com MAF muito baixa, pois de acordo com Glaubitz *et al.* (2014), esses marcadores podem representar erros de sequenciamento.

As informações sobre os marcadores SNPs filtrados e as variantes genômicas encontradas foram então codificadas e salvas em formato VCF - *Variant Calling Format*, e em seguida foram exportados para o ambiente do software R (3.6.0) (R Foundation for Statistical Computing). Foi utilizado o pacote *r2vcftools* (<https://github.com/nspope/r2vcftools>) – *VCFtools* para realizar a filtragem para manter apenas sites bialélicos com no máximo 20% de dados ausentes ($max.alleles = 2$ e $min.alleles = 2$) (DANECEK *et al.*, 2011) –, posteriormente, foi usado filtragem dos loci em desequilíbrio de ligação (LD, $r^2 < 0,3$) e fortes desvios do Equilíbrio de Hardy-Weinberg (HWE, $p < 0,0001$).

Em seguida, foram removidos os locos potenciais sob seleção, detectados através de escaneamento do genoma com teste de *F_{ST} outliers*, aplicado após ajustes taxa de falso

positivo (FRANCIS *et al.*, 2016). Posteriormente, a partir dos bancos de dados neutrais e independentes, foi realizada uma seleção dos acessos coletados no estado do Pará, em virtude de ser considerado centro de diversidade. Dessa forma, foram gerados dois bancos de dados, um considerando todos os acessos e outro analisando acessos coletados no estado do Pará. Em seguida, foram comparados com a classificação de acordo com o tipo de mandioca (brava (B), mansa (SW) e açucarada (SU)).

4.4. Análise da estrutura genética

A genética populacional foi avaliada por meio de duas metodologias complementares: a função *snmf*, do pacote LEA (v2.0) (FRICHOT *et al.*, 2014; FRICHOT; FRANÇOIS, 2015), que segue os pressupostos de uma população ideal e o DAPC, do pacote *adegenet* (v2.0.1) (JOMBART & DEVILLARD 2011), que não segue os pressupostos genéticos. O número de populações ancestrais (k) testadas variou entre 1 e 10, com 10 repetições para cada valor k , e o melhor k foi estabelecido com base nos erros de entropia cruzada e validação cruzada (FRICHOT *et al.*, 2014). Para gerar o DAPC, os dados foram submetidos inicialmente a análise de componentes principais (PCA - *Principal Components Analysis*) (JOLLIFFE, 1986) a partir da função *glPca*, que permite a identificação de agrupamentos genéticos, e, posteriormente, o número de grupos genéticos foi avaliado usando a função *find.clusters* para identificar o valor ótimo de k a partir do Critério de Informação Bayesiana (BIC - *Bayesian Information Criterion*). O valor K com o menor BIC foi selecionado como o número ideal de *clusters*.

4.5. Análise da variância molecular (AMOVA) e índices de diversidade genética

A análise da variância molecular (AMOVA) entre os grupos estabelecidos pelo pacote LEA (v2.0) e pelo tipo de mandioca foi calculada usando 1.000 permutações, pelo modelo proposto por Excoffier *et al.* (1992), através da função *poppr.amova* do pacote *poppr* v. 2.7.1 (KAMVAR *et al.*, 2015). As estatísticas de diferenciação genética foram geradas usando o pacote *Hierfstat* (GOUDET, 2005). O grau de diferenciação genética (F_{ST}) foi estabelecido por meio dos pacotes *StAMPP* (PEMBLETON *et al.* 2013) e *Hierfstat* (GOUDET, 2005), utilizando 1.000 *bootstraps*, a fim de buscar maior confiabilidade no resultado.

As estimativas de diversidade genética para cada loco neutral e independente foram realizadas conforme YANG *et al.*, 2010: heterozigosidade observada (H_O), heterozigosidade esperada (H_E), coeficientes de endogamia (F), média de diversidade de nucleotídeos por local (π). Todas estas análises foram calculadas usando o pacote r2vcftools.

5. RESULTADOS E DISCUSSÃO

5.1. Identificação de SNPs e Processamento dos dados

Neste trabalho foram sequenciados e genotipados, usando a técnica de GBS, 397 acessos de mandioca, na qual foram mapeados 670.835 SNPs com base no genoma de referência de mandioca versão 6.1, disponível no banco de dados *Phytozome* v12.0 (<http://www.phytozome.net/cassava.php>), distribuídos em 18 cromossomos e dois *scaffolds* que não foram ancorados em cromossomos (GOODSTEIN *et al.* 2012; PROCHNIK *et al.* 2012).

Após a filtragem com base em desequilíbrio de ligação (LD, $r^2 < 0,3$), fortes desvios do Equilíbrio de Hardy-Weinberg (HWE, $p < 0,0001$) e remoção dos locos potenciais sob seleção, detectados através de escaneamento do genoma com teste de *F_{ST} outliers* (Figura suplementar 1), foram selecionados 3.300 SNPs, representando 0,49% do número original, os quais estão distribuídos nos 18 cromossomos, com média de 183,33 SNPs por cromossomo. A maioria dos SNPs foi mapeado fisicamente no cromossomo 7 (7,33%, 242 SNPs), com a maior densidade de marcadores (1 a cada 111,94kb), e o menor valor de densidade foi observado no cromossomo 3 (1 a cada 200,27kb) (Figura 4 e Tabela 2).

A distribuição dos SNPs identificados no genoma de mandioca foi relativamente uniforme, considerando uma densidade média de marcadores de 1 SNP a cada 158,90kb. Albuquerque *et al.*, (2018) relataram uma maior densidade em relação ao presente trabalho (26,85 SNPs/kb), mas porque utilizaram número maior de SNPs, por terem realizado filtragens menos rígidas.

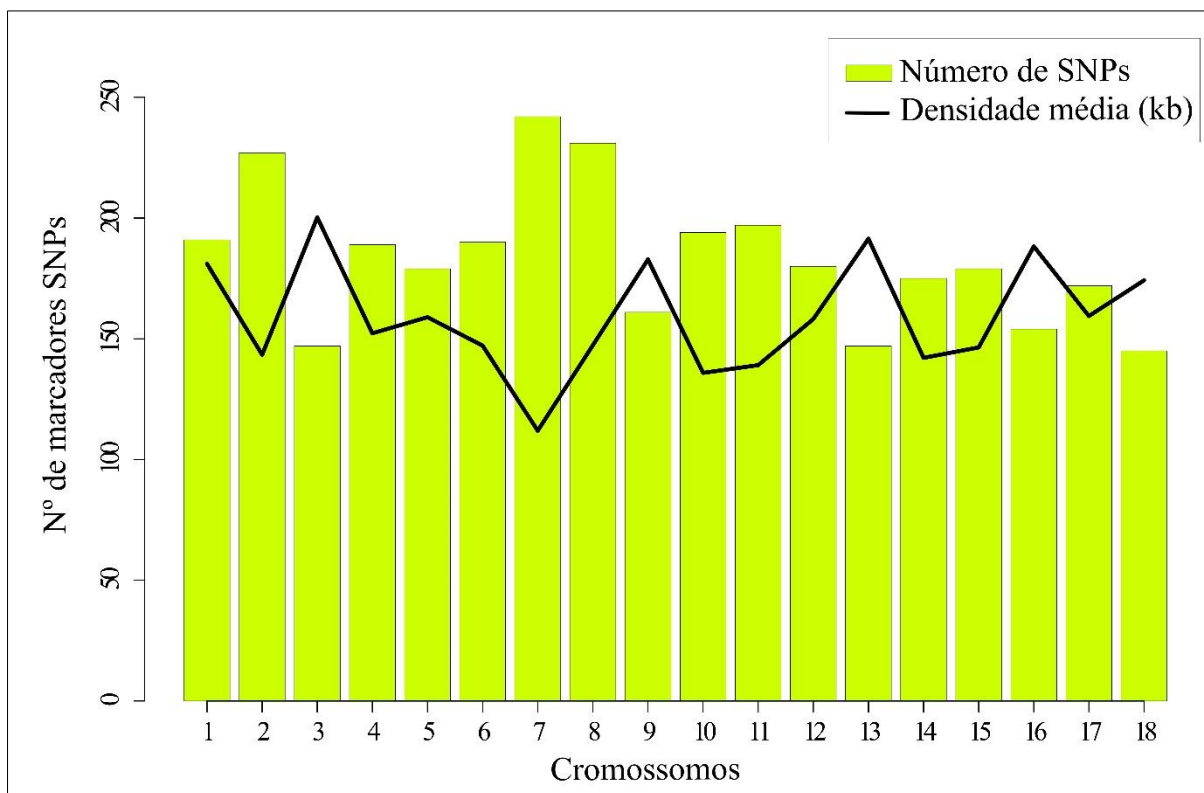
A proporção entre transições e transversões foi de 1,58. Os SNPs do tipo transição (61,24%, 2.021 SNPs) foram mais frequentes do que as transversões (38,76%, 1.279 SNPs) (Tabela 3). As transições A/G (33,33%, 1.100 SNPs) ocorreram em maior frequência, enquanto as transversões G/C (6,39%, 211 SNPs) ocorreram com menor frequência. Albuquerque *et al.* (2018) também encontraram porcentagem próxima à do presente trabalho (transições = 55,73% e transversões 44,27%), confirmando que as mutações do tipo transição são mais frequentes.

Tabela 2 - Distribuição genômica dos 3.300 marcadores SNPs neutrais e independentes encontrados nos acessos conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, na qual foram mapeados fisicamente nos 18 cromossomos do genoma de mandioca (*Manihot esculenta* Crantz), gerados a partir tecnologia de Genotipagem por Sequenciamento.

Cromossomos	Tamanho do Cromossomo (Mb)	Número de SNPs	Porcentagem de SNP	Densidade média (kb)
1	34,58	191	5,79	181,05
2	32,55	227	6,88	143,39
3	29,44	147	4,45	200,27
4	28,78	189	5,73	152,28
5	28,45	179	5,42	158,94
6	27,96	190	5,76	147,16
7	27,09	242	7,33	111,94
8	34,03	231	7,00	147,32
9	29,44	161	4,88	182,86
10	26,36	194	5,88	135,88
11	27,40	197	5,97	139,09
12	28,47	180	5,45	158,17
13	28,14	147	4,45	191,43
14	24,87	175	5,30	142,11
15	26,21	179	5,42	146,42
16	28,99	154	4,67	188,25
17	27,42	172	5,21	159,42
18	25,26	145	4,39	174,21
Média		183,33		158,90

Fonte: O Autor (2019).

Figura 4 - Distribuições e a densidade de 3.300 marcadores SNPs neutrais e independentes encontrados nos acessos conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, na qual foram mapeados fisicamente nos 18 cromossomos do genoma de mandioca (*Manihot esculenta* Crantz), gerados a partir tecnologia de Genotipagem por Sequenciamento.



Fonte: O Autor (2019).

Tabela 3 - Porcentagem de transição e transversão dos 3.300 marcadores SNPs neutrais e independentes encontrados nos acessos conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, na qual foram mapeados fisicamente nos 18 cromossomos do genoma de mandioca (*Manihot esculenta* Crantz), gerados a partir tecnologia de Genotipagem por Sequenciamento.

	Transição			Transversão		
	A/G	C/T	A/T	A/C	G/T	G/C
Número de SNPS	1100	921	393	370	305	211
Frequência	33,33%	27,91%	11,91%	11,21%	9,24%	6,39%
Total (percentagem)	2021 (61,24%)			1279 (38,76%)		

Fonte: O Autor (2019).

5.2. Análise da estrutura genética

A análise da estrutura genética é extremamente importante para entender a diversidade genética e facilita os estudos subsequentes da mandioca (ALBUQUERQUE *et al.* 2018; ALVES-PEREIRA *et al.* 2018). Neste trabalho, duas abordagens (função *snmf* do programa LEA e *find.cluster* do pacote *adegenet* para gerar a DAPC) foram utilizadas para análise de estruturação genética e indicaram a presença de três grupos genéticos (K=3) em ambas abordagens. O *snmf* estabeleceu o K com base nos erros de entropia cruzada e validação cruzada. A DAPC estabeleceu o número de K por meio do Critério de Informação Bayesiana (BIC), onde o melhor K é aquele que apresenta o menor valor de BIC (Figura 5).

As duas abordagens de clusterização genética (LEA e DAPC) apresentaram resultados consistentes quanto à detecção de uma estrutura genética entre os acessos de mandioca utilizando marcadores SNPs, onde identificaram que os 397 poderiam ser agrupados em três grupos, e os resultados do LEA coincidiam com os resultados do DAPC. Alves-Pereira *et al.* (2018), ao analisarem mandioca silvestres, bravas e mansas, a partir de 14 microssatélites nucleares, também encontraram três grupos genéticos, a partir do BIC.

Por meio do LEA, os 397 acessos foram agrupados em: grupo 1 (23 acessos), grupo 2 (107 acessos) e grupo 3 (267 acessos). Quando se utilizaram somente os 246 acessos coletados no estado do Pará, o LEA distribuiu os acessos em grupo 1 (21 acessos), grupo 2 (53 acessos) e grupo 3 (172 acessos). Fazendo um paralelo com a classificação de acordo com os tipos, as açúcaradas ficaram no grupo 1 e os grupos 2 e 3, em ambas as análises, apresentaram mistura de bravas e mansas (Figura 6). Esse comportamento é evidenciado também na Análise Discriminativa do Componente Principal (DAPC) (Figura 7). Dessa forma, é possível observar o isolamento de um conjunto de acessos dos grupos 1, que correspondem aos acessos de mandioca açúcarada, na qual apresentam distinções em relação as mansas e bravas, como o acúmulo de açúcar em suas raízes e um baixo nível de amido, apresentando funções diferentes na indústria, com a produção de etanol (CARVALHO *et al.*, 2004; SOUZA *et al.*, 2013).

A DAPC apresentou o mesmo padrão estabelecido pelo LEA, dessa forma, mesmo utilizando diferentes abordagens, o número de grupos genéticos é similar e com grande correspondência dos acessos em cada grupo, e, desse modo, os grupos estabelecidos pelo LEA foram utilizados para as análises posteriores (Figura 8).

Na análise de componentes principais, considerando os 397 acessos, os dois primeiros componentes principais explicaram 7,51% da variação genética total, valor próximo foi obtido para os dois primeiros componentes principais quando se usou somente os acessos do estado do Pará (8,83%). Avaliando todos os acessos, o primeiro e o segundo componente principal apresentaram valores baixos (CP1 = 4,47% e CP2 = 3,03%) (Figura 9). Com os acessos coletados no estado do Pará, obteve-se CP1 = 4,46% e CP2 = 4,37% (Figura 10). Albuquerque *et al.* (2018), também encontraram valores baixos nos dois primeiros componentes principais (CP1 = 6,4% e CP2 = 2,9%).

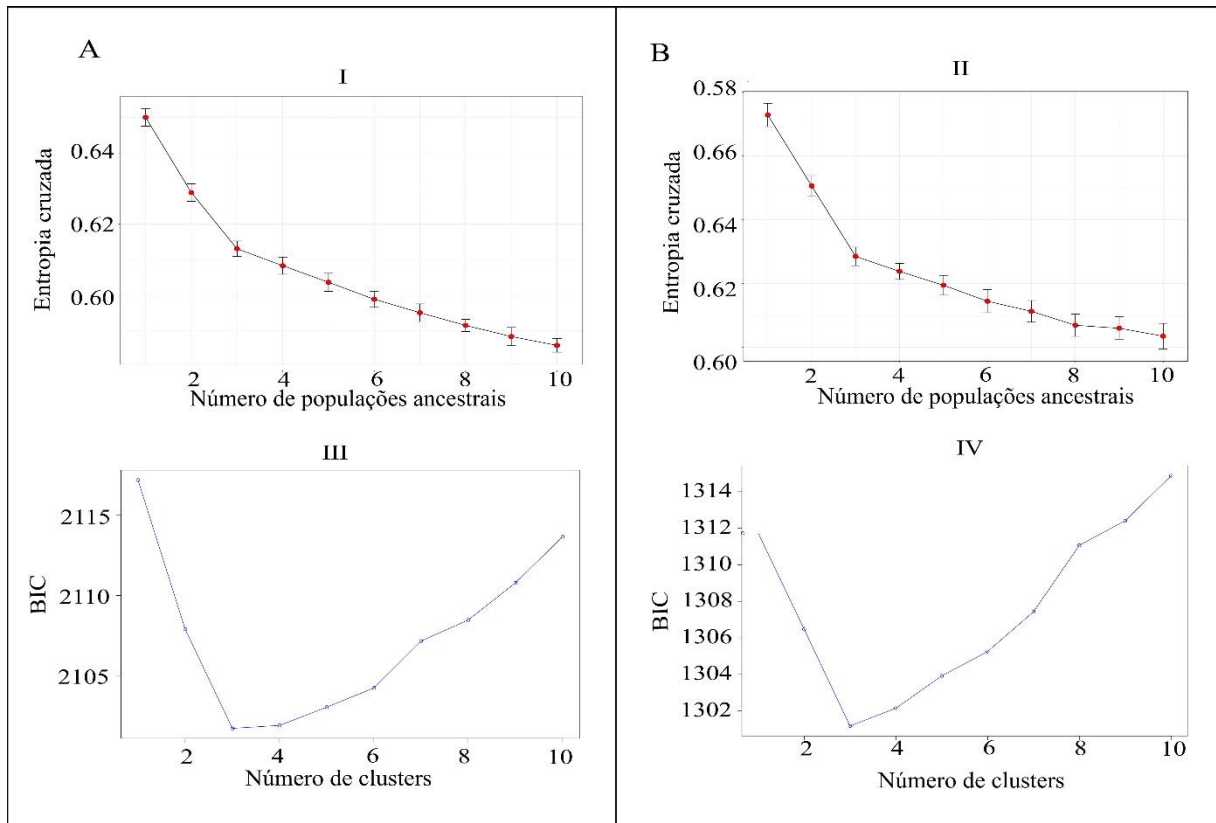
Segundo Oliveira *et al.* (2014), a análise de DAPC apresenta vantagens em comparação com outras abordagens clássicas, como as abordagens bayesianas, que são implementadas nos softwares structure (PRITCHARD *et al.*, 2000), admixture (ALEXANDER *et al.*, 2009) e o pacote LEA (FRICHOT *et al.*, 2014; FRICHOT; FRANÇOIS, 2015) por várias razões, dentre elas o fato de os caracteres burlarem os pressupostos genéticos. Todavia, para análise de genética de populações, de acordo com Allendorf *et al.* (2010), devemos considerar os pressupostos genéticos, sendo representados pela variabilidade genética neutral. Por conseguinte, este trabalho é o primeiro a relatar o uso de SNPs neutrais e independentes para uso de germoplasma de mandioca.

Neste trabalho foram encontrados três agrupamentos genéticos considerados as duas abordagens (LEA e DAPC). Oliveira *et al.* (2014), encontraram um valor de k superior (K = 30) ao encontrado neste trabalho ao avaliarem 1.280 acessos do Banco de Germoplasma de mandioca da Embrapa Mandioca e Fruticultura, que possui acessos coletados principalmente na região Nordeste do Brasil e com base em 402 marcadores SNPs. Albuquerque *et al.* (2018), também encontraram k superior a este trabalho (K = 22) a partir do BIC estudando 1.580 acessos de mandioca com 20.610 SNPs, entretanto nesses dois trabalhos, os autores não utilizaram SNPs neutrais e independentes, o que pode explicar o número alto de agrupamentos genéticos.

Dessa forma, a presença de estruturação genética estabelecida pelos marcadores neutrais e independentes entre os acessos de mandioca atende às expectativas, devido apresentar correspondência com os tipos estabelecidos pelo conhecimento tradicional, pois certas características específicas intencionalmente selecionadas pelos agricultores também podem levar à estruturação genética da espécie. Portanto, a diferenciação genética foi mais evidenciada para o grupo de mandioca açucarada (Figura 6), enquanto houve uma mistura entre mandiocas

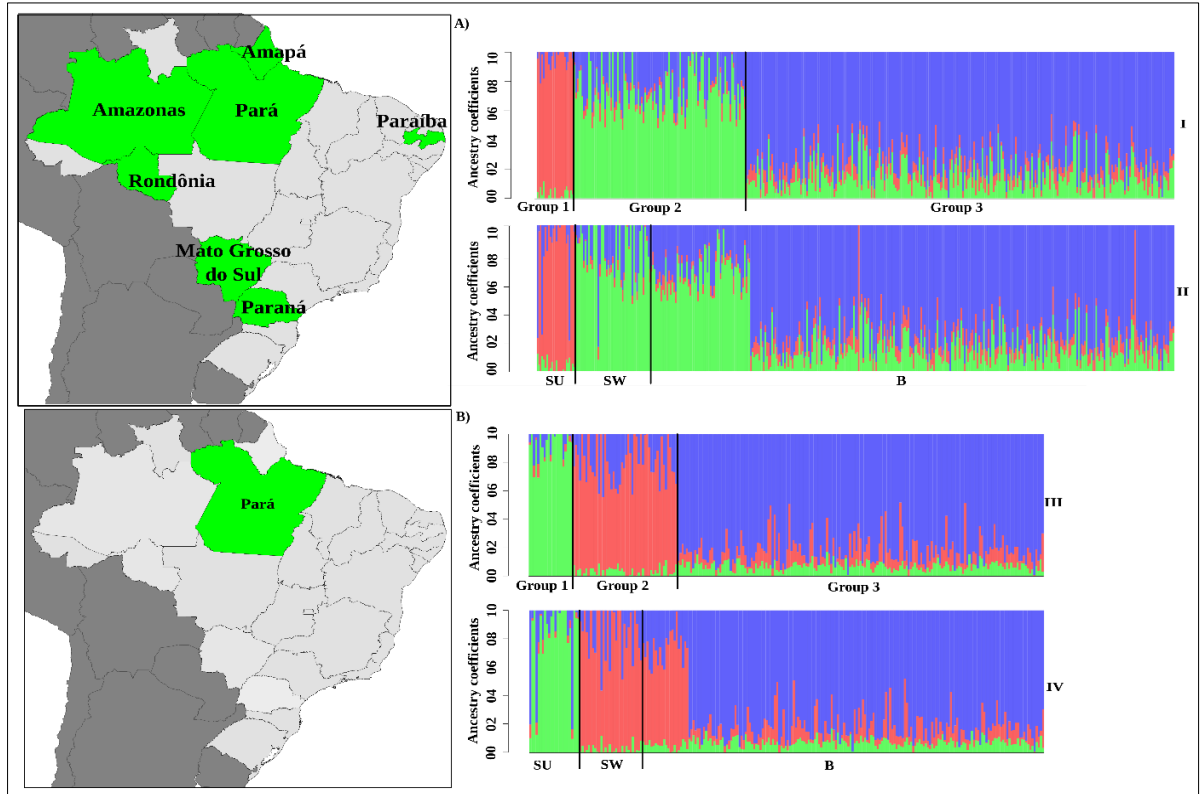
mansas e bravas nos grupos gerados pela metodologia LEA, o que é justificado devido ao intercâmbio genético correspondendo troca de germoplasma entre os agricultores, principalmente utilizando esses dois tipos que são mais comercializáveis.

Figura 5 - Gráficos mostrando o número ideal de agrupamentos genéticos (k) considerando os 3.300 marcadores SNPs neutrais independentes analisando acessos de mandioca conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental gerados a partir da tecnologia de Genotipagem por Sequenciamento. A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará. A escolha k ideal baseia-se em erros de entropia cruzada da média \pm sd (LEA, I e II) e Critério de Informação Bayesiana (BIC) (DAPC, III e IV).



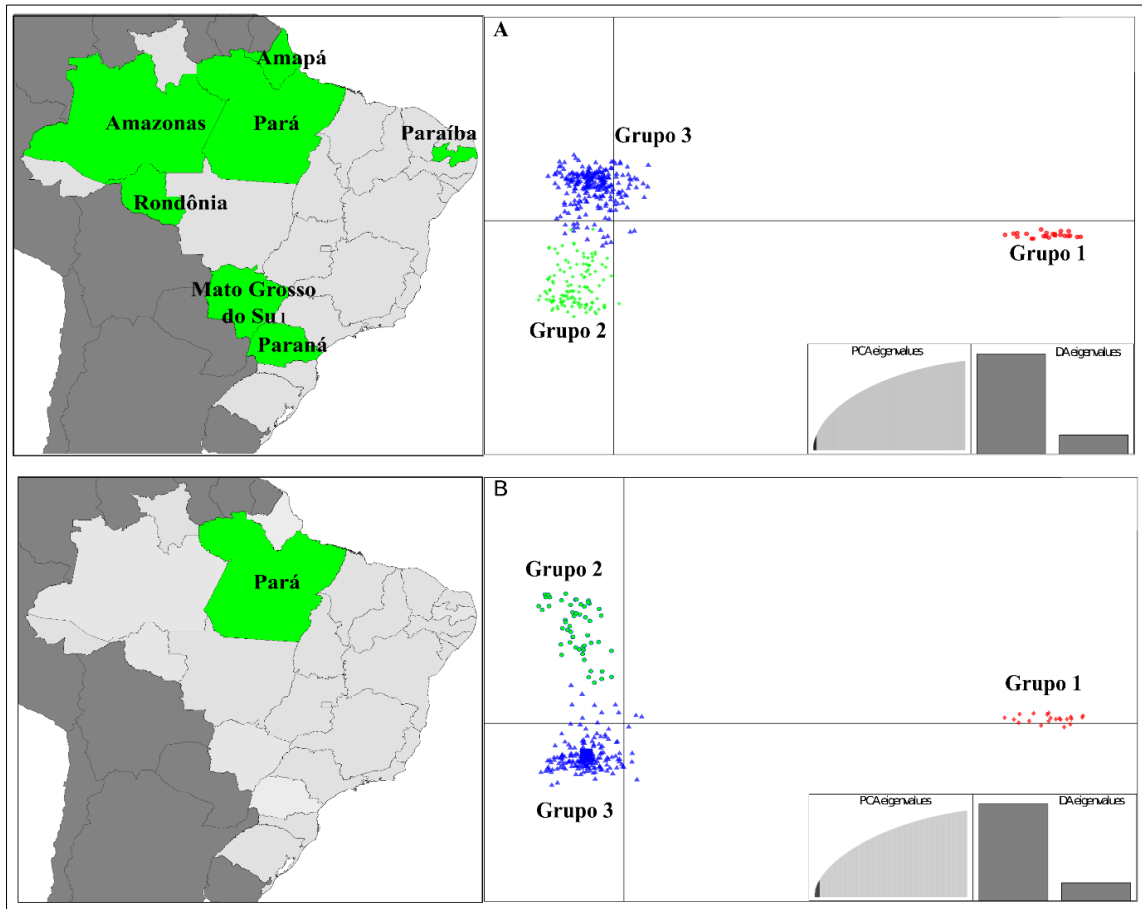
Fonte: O Autor (2019).

Figura 6 - Agrupamentos genéticos dos acessos de mandioca do Banco Ativo de Germoplasma da Embrapa Amazônia Oriental (BAG) de acordo com a função *snmf* do pacote LEA, analisados com 3.300 marcadores SNPs neutrais independentes gerados pela tecnologia de Genotipagem por sequenciamento (GBS). A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará. (SU) Açucarada, (SW) mansa e (B) mandioca.



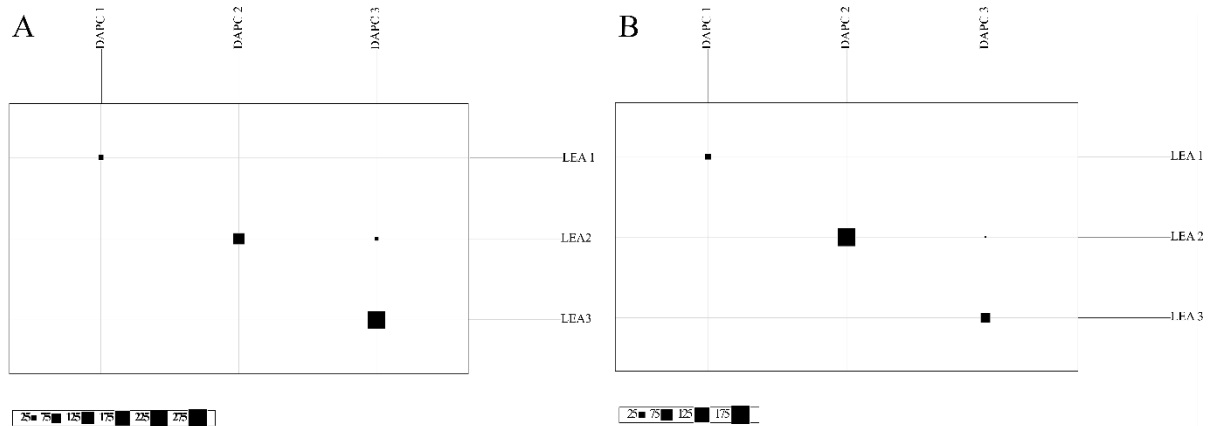
Fonte: O Autor (2019).

Figura 7 - Agrupamentos genéticos a partir dos 3.300 marcadores SNPs neutrais independentes mapeados nos acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma (BAG) da Embrapa Amazônia Oriental de acordo com a Análise Discriminante dos Componentes Principais (DAPC). A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará.



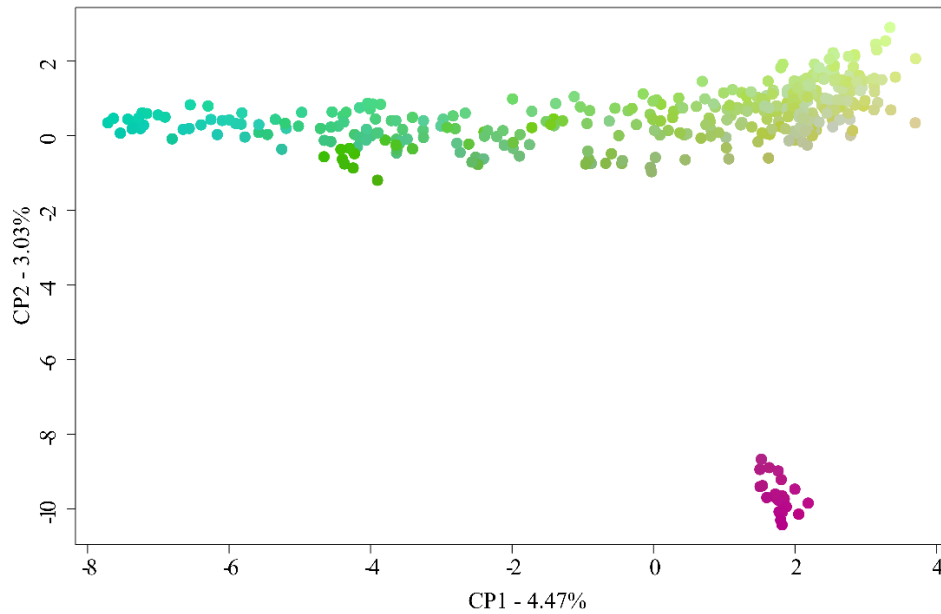
Fonte: O Autor (2019)

Figura 8 - Comparação da distribuição estabelecidos pelo pacote LEA e DAPC do pacote *adegenet* dos acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, considerando 3.300 SNPs neutrais e independentes gerados a partir da técnica de Genotipagem por sequenciamento (GBS). A) considerando todos os 397 acessos. B) 247 acessos coletados no estado do Pará



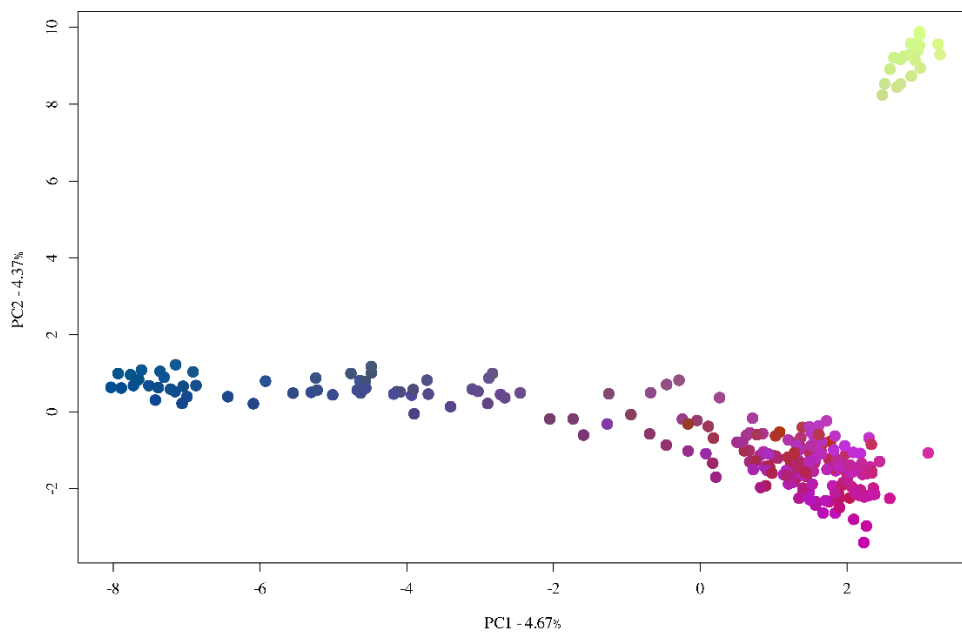
Fonte: O Autor (2019)

Figura 9 - Dispersão de 397 acessos de mandioca (*Manihot esculenta* Crantz) conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, considerando 3.300 SNPs neutrais e independentes gerados a partir da técnica de Genotipagem por sequenciamento (GBS) com base na Análise de Componentes Principais - PCA. CP1: componente principal 1, CP2: componente principal 2.



Fonte: O Autor (2019)

Figura 10 - Dispersão de 246 acessos de mandioca (*Manihot esculenta* Crantz) coletados no estado do Pará conservados no Banco Ativo de Germoplasma da Embrapa Amazônia Oriental, considerando 3.300 SNPs neutrais e independentes gerados a partir da técnica de Genotipagem por sequenciamento (GBS) com base na Análise de Componentes Principais - PCA. CP1: componente principal 1, CP2: componente principal 2.



Fonte: O Autor (2019)

5.3. Análise da variância molecular (AMOVA) e índices de diversidade genética

Considerando todos os acessos do BAG de mandioca, foi aplicada submetidos a Análise de Variância Molecular (AMOVA) para verificar o grau de diferenciação genética entre os grupos, e a diferenciação genética (F_{ST}) foi de 0,13 entre os grupos genéticos gerados pelo pacote LEA e foi de 0,12 para os grupos determinados pelos tipos de mandioca (Tabela 4A). Quando se considerou somente os acessos coletados no estado do Pará, os valores foram similares tanto para a análise considerando os grupos genéticos gerados pelo pacote LEA ($F_{ST} = 0,15$; $P < 0,001$), quanto para os grupos considerando os tipos de mandioca ($F_{ST} = 0,14$; $P < 0,001$) (Tabela 4B). Segundo Wright (1978), os valores de F_{ST} variando de 0 a 0,05 indicam valores baixos, 0,05–0,15 moderados, 0,15–0,25 e altos acima de 0,25. Dessa forma, a AMOVA revelou que a diferenciação genética entre os grupos foi moderada.

Em relação a variação genética, a maior parte da variação se concentrou dentro dos grupos (variando de 84,17% a 87,24%) (Tabela 4). Albuquerque *et al.* (2018) fizeram a AMOVA entres os tipos estabelecidos pelo DAPC e obtiveram 85,15% de variação genética dentro dos grupos, valores semelhantes aos encontrados neste trabalho. Alves-Pereira *et al.* (2011), obtiveram 86,8% de variação genética dentro de tipos de variedades de mandioca (brava e mansa), utilizando marcadores microsatélites em mandioca.

Os parâmetros de diversidade apresentaram medidas similares de heterozigosidade observada (H_O) em relação à esperada (H_E), com exceção do grupo 1 e do grupo de mandioca açucarada (SU) que apresentaram valores de $H_O > H_E$ (Tabela 5), refletindo diretamente no coeficiente de endogamia. A diversidade de nucleotídeos encontrou valores semelhantes em todas as análises (Tabela 5). As medidas de diversidade genética foram iguais, dentro do intervalo de confiança, para os grupos 2 e 3 dos dois agrupamentos genéticos e nos grupos das bravas e mansas, sugerindo extensa troca de germoplasma de mandioca no Brasil e ampla diversidade molecular devido esses dois grupos apresentam maior número de acessos do tipo brava, sendo o principal tipo comercializado para a produção de farinha, por outro lado, os grupos 1 ($H_E = 0,48$) e as açucaradas (SU) ($H_E = 0,33$) apresentaram maiores valores de diversidade genética (Figura 5), apresentando uma fonte de exploração pelos produtores, que de acordo com Carvalho *et al.* (2004), a açucarada apresenta grande potencial para a produção de etanol e pode ser utilizada para a produção de bebidas.

Tabela 4 - Análise da Variância Molecular (AMOVA) entre e dentro dos grupos estabelecidos pelo pacote LEA utilizando os 3.300 marcadores SNPs neutrais e independentes gerados pela técnica de Genotipagem por Sequenciamento. A) considerando todos os 397 acessos de mandioca. B) 246 acessos coletados no estado do Pará. Df: Graus de liberdade; MS: Média dos Quadrados; %var: % Diferenciação Genética; F_{ST} : Índice de Fixação.

A								
TODOS OS ACESSOS								
Fonte de variação	Grupos determinados pelo método LEA				Tipos de mandioca			
	GL	MS	%var	F_{ST}	GL	MS	%var	F_{ST}
Entre populações	2	61.16,89	12,76	0,13	2	4.088,21	12,06	0,12
Dentro de populações	394	416,4	87,24		394	426,7	87,93	
Total	396	445,19	100		396	445,2	100	

B								
ACESSOS COLETADOS NO PARÁ								
Fonte de variação	Grupos determinados pelo método LEA				Tipos de mandioca			
	Df	MS	%var	F_{ST}	Df	MS	%var	F_{ST}
Entre populações	2	4.704,85	15,83	0,15	2	7.207,21	14,56	0,14
Dentro de populações	243	406,15	84,17		243	100.896,86	85,44	
Total	245	441,24	100		245	108.104,08	100	

Fonte: O Autor (2019)

Os valores de heterozigosidade observada e esperada, no geral, foram semelhantes aos obtidos em outros trabalhos com acessos de mandioca. Por exemplo, Oliveira *et al.* (2014) estudando 1.280 acessos de mandioca do Banco de Germoplasma da Embrapa Mandioca e Fruticultura, com base em 402 marcadores SNPs, encontraram excesso de heterozigotos, com média de 32% de heterozigosidade

Alves-Pereira *et al.* (2018), encontraram nas mandiocas do tipo brava elevada heterozigosidade observada e valor negativo para o coeficiente de endogamia ($H_O = 0,69$, $F = -0,16$), porém com déficit de heterozigotos no grupo das mandiocas mansas ($H_O = 0,573$, $F = 0,065$). Albuquerque *et al.* (2018) encontraram $H_O = 0,23$ e $H_E = 0,29$, o que caracteriza excesso de homozigotos, onde também foi encontrado nos grupos 3 e nos grupos classificados como mandiocas bravas deste trabalho (Tabela 5), o que corresponde a características de uma planta com reprodução alógama e propagação assexuada.

Tabela 5 - Medidas de diversidade genética dos grupos considerando todos os acessos e grupos estabelecido pelos acessos coletados no Pará, gerados pelo LEA. Cada agrupamento foi comparado com os grupos estabelecidos pelos tipos de mandioca. N: número de acessos; H_O : heterozigosidade observada; H_E : heterozigosidade esperada; F : coeficiente de endogamia; π : diversidade de nucleotídeos. Todas as estimativas são apresentadas ao longo de intervalos de confiança de 95% (CI). (SU) Açucarada, (SW) mansa e (B) mandioca.

	Grupos genéticos	N	H_O (CI)	H_E (CI)	F (CI)	π (CI)
Grupos formados pelo método de LEA	Grupo 1	23	0,48 (0,45/0,51)	0,36 (0,36/0,36)	-0,33 (-0,42/-0,24)	-
	Grupo 2	107	0,24 (0,23/0,25)	0,24 (0,24/0,24)	-0,01 (-0,04/0,03)	0,24 (0,23/0,24)
	Grupo 3	267	0,23 (0,23/0,24)	0,25 (0,25/0,25)	0,08 (0,06/0,10)	0,25 (0,25/0,26)
Tipos de mandioca	SU	24	0,33 (0,31/0,35)	0,27 (0,27/0,27)	-0,22 (-0,31/-0,14)	0,21 (0,20/0,21)
	SW	49	0,26 (0,25/0,27)	0,25 (0,25/0,25)	-0,04 (-0,09/0,01)	0,23 (0,22/0,24)
	B	324	0,23 (0,23/0,24)	0,25 (0,25/0,25)	0,08 (0,06/0,10)	0,25 (0,25/0,26)
Grupos formados pelo método de LEA	Grupo 1	21	0,48 (0,44/0,52)	0,36 (0,36/0,36)	-0,33 (-0,43/-0,22)	-
	Grupo 2	53	0,26 (0,25/0,27)	0,24 (0,24/0,24)	-0,06 (-0,10/-0,02)	0,23 (0,23/0,24)
	Grupo 3	172	0,23 (0,22/0,24)	0,25 (0,25/0,25)	0,08 (0,05/0,11)	0,25 (0,25/0,26)
Tipos de mandioca	SU	24	0,33 (0,31/0,35)	0,27 (0,27/0,27)	-0,22 (-0,31/-0,14)	0,21 (0,20/0,21)
	SW	30	0,28 (0,27/0,30)	0,26 (0,26/0,26)	-0,09 (-0,15/-0,04)	0,23 (0,22/0,24)
	B	192	0,23 (0,23/0,24)	0,25 (0,25/0,25)	0,08 (0,05/0,10)	0,25 (0,25/0,26)

Fonte: O Autor (2019)

7. CONCLUSÕES

Os marcadores SNPs neutrais mostraram-se eficazes em agrupar os acessos do Banco Ativo de Germoplasma de mandioca da Embrapa Amazônia Oriental.

O Banco Ativo de Germoplasma Embrapa Amazônia Oriental apresenta alta diversidade e estruturação genética, em que o grupo das mandiocas açúcaradas apresenta maior diferenciação.

REFERÊNCIAS

- ALLENDORF, Fred W.; HOHENLOHE, Paul A.; LUIKART, Gordon. Genomics and the future of conservation genetics. **Nature reviews genetics**, v. 11, n. 10, p. 697-709, 2010.
- ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, Kenneth. Fast model-based estimation of ancestry in unrelated individuals. **Genome research**, v. 19, n. 9, p. 1655-1664, 2009.
- ALVES-PEREIRA, A., PERONI, N., ABREU, A. G., GRIBEL, R., CLEMENT, C. R. Genetic structure of traditional varieties of bitter manioc in three soils in Central Amazonia. **Genetica**, v. 139, n. 10, p. 1259-1271, 2011.
- ALVES-PEREIRA, Alessandro et al. Patterns of nuclear and chloroplast genetic diversity and structure of manioc along major Brazilian Amazonian rivers. **Annals of botany**, v. 121, n. 4, p. 625-639, 2018.
- ANDREWS, Kimberly R. et al. Harnessing the power of RADseq for ecological and evolutionary genomics. **Nature Reviews Genetics**, v. 17, n. 2, p. 81, 2016.
- DE ALBUQUERQUE, Hilçana Ylka Gonçalves et al. Genetic diversity of *Manihot esculenta* Crantz germplasm based on single-nucleotide polymorphism markers. **Annals of applied biology**, v. 173, n. 3, p. 271-284, 2018.
- BOARI, A. de J. et al. First report of *Phytophthora* sp. causing storage root rot and foliage blight of cassava in Brazil. **Plant Disease**, v. 102, n. 5, p. 1042-1042, 2018.
- BRADBURY, Peter J. et al. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633-2635, 2007.
- DE CARVALHO, Reginaldo; GUERRA, Marcelo. Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. **Hereditas**, v. 136, n. 2, p. 159-168, 2002.
- CARVALHO, Luiz Joaquim Castelo Branco; SCHAAL, Barbara Anna. Assessing genetic diversity in the cassava (*Manihot esculenta* Crantz) germplasm collection in Brazil using PCR-based markers. **Euphytica**, v. 120, n. 1, p. 133-142, 2001.
- DANECEK, Petr et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156-2158, 2011.
- DAVEY, John W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, v. 12, n. 7, p. 499-510, 2011.

- DE OLIVEIRA, Eder Jorge et al. Potential of SNP markers for the characterization of Brazilian cassava germplasm. **Theoretical and Applied Genetics**, v. 127, n. 6, p. 1423-1440, 2014.
- DOYLE, Jeff J.; DOYLE, Jane L. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, n. 13, p. 39-40, 1990.
- ELSHIRE, Robert J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS one**, v. 6, n. 5, p. e19379, 2011.
- EXCOFFIER, Laurent; SMOUSE, Peter E.; QUATTRO, Joseph M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**, v. 131, n. 2, p. 479-491, 1992.
- FERREIRA, Cláudia Fortes et al. Molecular characterization of cassava (*Manihot esculenta* Crantz) with yellow-orange roots for beta-carotene improvement. **Crop Breeding and Applied Biotechnology**, v. 8, n. 1, 2008.
- FRICHOT, Eric; FRANÇOIS, Olivier. LEA: An R package for landscape and ecological association studies. **Methods in Ecology and Evolution**, v. 6, n. 8, p. 925-929, 2015.
- FRICHOT, Eric et al. Fast and efficient estimation of individual ancestry coefficients. **Genetics**, v. 196, n. 4, p. 973-983, 2014.
- GLAUBITZ, Jeffrey C. et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **PLoS one**, v. 9, n. 2, p. e90346, 2014.
- GOODSTEIN, David M. et al. Phytozome: a comparative platform for green plant genomics. **Nucleic acids research**, v. 40, n. D1, p. D1178-D1186, 2012.
- GOUDET, Jérôme. Hierfstat, a package for R to compute and test hierarchical F-statistics. **Molecular Ecology Notes**, v. 5, n. 1, p. 184-186, 2005.
- JOLLIFFE, Ian T. Principal components in regression analysis. In: **Principal component analysis**. Springer, New York, NY, 1986. p. 129-155.
- JOMBART, Thibaut; DEVILLARD, Sébastien; BALLOUX, François. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC genetics**, v. 11, n. 1, p. 1-15, 2010.

KAMVAR, Zhian N.; BROOKS, Jonah C.; GRÜNWALD, Niklaus J. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. **Frontiers in genetics**, v. 6, p. 208, 2015.

LEBOT, Vincent et al. NIR determination of major constituents in tropical root and tuber crop flours. **Journal of Agricultural and Food Chemistry**, v. 57, n. 22, p. 10539-10547, 2009.

LÉOTARD, Guillaume et al. Phylogeography and the origin of cassava: new insights from the northern rim of the Amazonian basin. **Molecular Phylogenetics and Evolution**, v. 53, n. 1, p. 329-334, 2009.

LIMA, Aline Medeiros et al. Expression profiles of defense genes in cassava storage roots upon exposure to *Phytophthora* sp., causal agent of soft root rot disease. **Physiological and Molecular Plant Pathology**, v. 104, p. 23-30, 2018.

LUIKART, Gordon et al. The power and promise of population genomics: from genotyping to genome typing. **Nature reviews genetics**, v. 4, n. 12, p. 981-994, 2003.

MARRANO, Annarita et al. SNP-discovery by RAD-sequencing in a germplasm collection of wild and cultivated grapevines (*V. vinifera* L.). **PloS one**, v. 12, n. 1, p. e0170655, 2017.

MOURA, Elisa Ferreira et al. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. **Acta Amazonica**, v. 43, n. 4, p. 461-467, 2013.

MOURA, Elisa Ferreira et al. Molecular characterization of accessions of a rare genetic resource: sugary cassava (*Manihot esculenta* Crantz) from Brazilian Amazon. **Genetic Resources and Crop Evolution**, v. 63, n. 4, p. 583-593, 2016.

MÜHLEN, G. S. et al. Distribuição da variabilidade genética e diferenciação de variedades mansas e bravas de mandioca (*Manihot esculenta* Crantz, Euphorbiaceae) do Brasil baseada em marcadores moleculares SSR. In: **2nd International Meeting on Amazonian Archaeology, Brazilian Archaeological Society and Amazonas State University, Manaus, Amazonas, Brasil**. 2010.

MÜHLEN, G.S.;Pereira, A.A.; Clement, C.R.; Valle, T. L. Genetic diversity and differentiation of Brazilian bitter and sweet manioc varieties (*Manihot esculenta* Crantz, Euphorbiaceae) based on SSR molecular markers. *Tipiti: Journal of the Society for the Anthropology of Lowland South America*, v.11, p.66-73, 2013

NASSAR, Nagib. Cassava, *Manihot esculenta* Crantz and wild relatives: Their relationships and evolution. **Genetic Resources and Crop Evolution**, v. 48, n. 5, p. 429-436, 2001.

OLSEN, Kenneth M. SNPs, SSRs and inferences on cassava's origin. **Plant Molecular Biology**, v. 56, n. 4, p. 517-526, 2004.

OYESIGYE, Elias et al. Single nucleotide polymorphism (SNP) diversity of cassava genotypes in relation to cassava brown streak disease in Mozambique. **Plant Genetic Resources**, v. 16, n. 6, p. 533-543, 2018.

POLAND, Jesse A.; RIFE, Trevor W. Genotyping-by-sequencing for plant breeding and genetics. **The Plant Genome**, v. 5, n. 3, 2012.

PEMBLETON, Luke W.; COGAN, Noel OI; FORSTER, John W. St AMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. **Molecular ecology resources**, v. 13, n. 5, p. 946-952, 2013.

PETERSON, Gregory W. et al. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. **Diversity**, v. 6, n. 4, p. 665-680, 2014.

PRITCHARD, Jonathan K.; STEPHENS, Matthew; DONNELLY, Peter. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, n. 2, p. 945-959, 2000.

PROCHNIK, Simon et al. The cassava genome: current progress, future directions. **Tropical plant biology**, v. 5, n. 1, p. 88-94, 2012.

RABBI, Ismail Y. et al. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). **BMC genetics**, v. 16, n. 1, p. 1-11, 2015.

ROULLIER, Caroline et al. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. **Proceedings of the National Academy of Sciences**, v. 110, n. 6, p. 2205-2210, 2013.

SILVA, Rainério Meireles da et al. Biologia reprodutiva de etnovariedades de mandioca. **Scientia Agricola**, v. 58, n. 1, p. 101-107, 2001.

SOUZA, Hugo Antonio Lima de et al. Physicochemical properties of three sugary cassava landraces. **Ciência Rural**, v. 43, n. 5, p. 792-796, 2013.

VASCONCELOS, Arthur A. et al. Determination of Cu, Fe, Mn e Zn Using FAAS in Cassava (*Manihot esculenta* Crantz) Roots from Eastern Amazon. **REVISTA VIRTUAL DE QUIMICA**, v. 9, n. 6, p. 2316-2331, 2017.

VALLE, Teresa Losada; LORENZI, José Osmar. Variedades melhoradas de mandioca como instrumento de inovação, segurança alimentar, competitividade e sustentabilidade: contribuições do Instituto Agronômico de Campinas (IAC). **Cadernos de Ciência & Tecnologia**, v. 31, n. 1, p. 15-34, 2014.

VIEIRA, Eduardo Alano et al. Caracterização fenotípica e molecular de acessos de mandioca de indústria com potencial de adaptação às condições do Cerrado do Brasil Central. **Semina: Ciências Agrárias**, v. 34, n. 2, p. 567-581, 2013.

YANG, Jian et al. Common SNPs explain a large proportion of the heritability for human height. **Nature genetics**, v. 42, n. 7, p. 565-569, 2010.

WOLFE, M. D., DEL CARPIO, D. P., ALABI, O., EZENWAKA, L. C., IKEOGU, U. N., KAYONDO, I. S., JANNINK, J. L. Prospects for genomic selection in cassava breeding. *The Plant Genome*, v. 10, n. 3, 2017.

WRIGHT, S. Evolution and the genetics of populations: variability within and among natural populations. 4: University of Chicago press. **Chicago, IL, USA**, 1978.

Figura suplementar 1 – Distribuição das estimativas de F_{ST} para a detecção dos 39 locos outliers (vermelho), após as filtrações por desequilíbrio de ligação e desvio de equilíbrio de Hardy-Weinberg.

