

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

Genômica populacional da cajazeira (*Spondias mombin* L.)

Allison Vieira da Silva

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

**Piracicaba
2021**

Allison Vieira da Silva
Engenheiro Agrônomo

Genômica populacional da cajazeira (*Spondias mombin* L.)
versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

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

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Silva, Allison Vieira da

Genômica populacional da cajazeira (*Spondias mombin* L.) / Allison Vieira da Silva. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

52 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Árvore tropical 2. Cajá 3. Amazônia 4. Diversidade genética I. Título

DEDICATÓRIA

*A todos os pesquisadores que acreditam no potencial
e contribuem para a conservação da espécie.*

*A todas as famílias, que durante a safra,
têm a espécie como fonte de renda.*

AGRADECIMENTOS

Agradeço primeiramente à minha família, em especial à minha avó, Severina Bezerra dos Santos, que em sua juventude amargou a vida de retirante e com a experiência que acumulou, sempre me incentivou nos estudos e me apoiou, mesmo sofrendo, em todas as decisões que tomei em prol do meu crescimento.

À professora Marta, ao professor Sérgio, ao professor Humberto, à professora Auxiliadora, à professora Marineide, à professora Marisa e à professora Daniele que durante o ensino fundamental e médio na Escola Professor Alfredo Freire, reconheceram em mim o potencial (do qual eu duvidava) de um aluno do ensino superior e me mostraram que era possível chegar à faculdade.

À todos os amigos que fiz na Escola Professor Alfredo Freyre que sempre apoiaram e incentivaram o meu progresso e que mesmo distantes sempre me enviaram mensagens de incentivo e conforto.

À Universidade Federal Rural de Pernambuco (UFRPE), que me acolheu e fez de mim um engenheiro agrônomo. A todos os amigos que fiz durante o curso de Agronomia e que levarei para a vida, Claudio Henrique, Talmo Henrique, Danielson Ramos, Lucas Lima e Patryk.

Agradeço aos professores do curso de Agronomia da UFRPE que foram decisivos para o meu desenvolvimento como pessoa e profissional. A Angélica Virginia Valois Montarroyos, a minha primeira orientadora que, com toda a paciência que possui, me guiou durante os primeiros passos como pesquisador.

Ao professor Edson Ferreira da Silva que me adotou academicamente, me orientou por quatro anos e me mostrou que eu era capaz de conquistar mais do que eu desejava.

Agradeço ao professor Wellington Ferreira do Nascimento, que me acolheu em sua casa durante o processo de seleção da pós-graduação e contribuiu diretamente para a conquista da minha aprovação.

À bióloga Ieda Ferreira de Oliveira que me iniciou na rotina laboratorial e me forneceu todo o conhecimento de laboratório necessário para a minha carreira acadêmica.

A todos os amigos que ganhei em Piracicaba, em especial aos meus amigos de departamento, Igor, Cristiane, Vitor e Flaviane, que contribuíram imensamente para o desenvolvimento da minha dissertação e me deram suporte emocional durante todo o processo.

A todos que tornaram possível a coleta a campo do material vegetal, o Prof. Dr. Edson Ferreira da Silva, o Prof. Dr. Wellington Ferreira do Nascimento, o Prof. Dr. Santiago Linorio Ferreyra Ramos, ao pesquisador Dario Dantas Amaral, ao doutorando Igor Araújo Santos de Carvalho, ao graduando em Agronomia Adolfo Melo e todos os guias colaboradores que tivemos durante as coletas.

A todos que contribuíram nas atividades laboratoriais, o doutorando Igor Araújo Santos de Carvalho, a Profa. Dra. Doriane Picanço Rodrigues, a Dra. Flaviane Malaquias Costa, a Dra. Carolina Grandó e ao Dr. Carlos Eduardo de Araújo Batista.

A todos que contribuíram para realização das análises dos dados, o Dr. Alessandro Alves Pereira, a Dra. Flaviane Malaquias Costa, a Dra. Maria Imaculada Zucchi, o doutorando Igor Araújo Santos de Carvalho, o doutorando Víctor Hugo de Mello Pessoa e a Dra. Cristiane Hayumi Taniguti.

Agradeço em especial às duas pessoas que mais contribuíram (em todas as fases) para o desenvolvimento da dissertação e que sem elas a conclusão do trabalho não seria possível. À minha orientadora Elizabeth Ann Veasey, que esteve sempre presente em todas as etapas do projeto e do curso, que sempre esteve e está disposta a auxiliar no meu desenvolvimento e no desenvolvimento da dissertação e que sempre me deu segurança quanto a sua disponibilidade para contribuição. À minha colega de curso e equipe, a mestranda Caroline Bertocco Garcia, que junto comigo, passou por todas as etapas de desenvolvimento do trabalho e contribuiu de forma ativa em todas elas como profissional e amiga.

Agradeço à Esalq/USP por toda a infraestrutura oferecida no campus. Ao Departamento de Genética, pela riqueza inestimável a mim repassada em forma de conhecimento por todos os excelentes professores integrantes do corpo docente. À FAPESP pelo financiamento com o qual foi possível realizar o projeto e ao CNPq pela bolsa de mestrado que tornou possível a minha estadia na cidade de Piracicaba.

*“Viver é um desafio
Desafiar é viver
Por isso eu vou vivendo
Sempre buscando aprender
Para não ser devorado
Pela falta de saber.
Se posso dou um sorriso
Se não posso, um lamento
Mas não fico esperando
Sonhando sou avarento
E busco sonhar meus sonhos
Até no sopro do vento.
Nas gotas fracas da chuva
Que a terra vai borrifando
E faz levantar o cheiro
De chuva que vou cheirando
Eu sonho dias melhores
E levo a vida cantando.”
(Nildo Cordel)*

SUMÁRIO

RESUMO.....	8
ABSTRACT	9
1. INTRODUÇÃO.....	11
2. OBJETIVOS	13
2.1. Objetivo Geral.....	13
2.2. Objetivos específicos	13
3. REVISÃO DE LITERATURA.....	15
3.1. Distribuição geográfica e aspectos gerais	15
3.2. Bancos de germoplasma e estudos de diversidade genética da cajazeira	17
3.3. Genotipagem por sequenciamento e marcadores SNPs.....	18
4. MATERIAIS E MÉTODOS.....	21
4.1. Amostragem	21
4.2. Extração e quantificação de DNA	22
4.3. Montagem da biblioteca de GBS.....	23
4.4. Obtenção dos marcadores SNPs	24
4.5. Análises dos marcadores SNPs	25
5. RESULTADOS.....	27
6. DISCUSSÃO.....	37
7. CONCLUSÕES	43
REFERÊNCIAS	45

RESUMO

Genômica populacional da cajazeira (*Spondias mombin* L.)

A cajazeira (*Spondias mombin*) é uma espécie arbórea, frutífera, alógama, nativa da região que compreende o Sul do México ao Leste do Brasil. A espécie tem grande potencial para exploração econômica no Brasil e já é bem consolidada comercialmente nas regiões Norte e Nordeste do país. A maior parte dos frutos comercializados é proveniente do extrativismo, uma vez que a espécie é classificada como semi-domesticada e não possui áreas expressivas de cultivo comercial. O trabalho foi realizado com o objetivo de estudar a estrutura e a diversidade genética da espécie *S. mombin* nas Regiões Norte e Nordeste do país por meio de marcadores SNPs (*Single Nucleotide Polymorphism*). No estudo foram avaliadas nove populações naturais da espécie, cinco populações provenientes da Região Nordeste e quatro populações da Região Norte. Na Região Nordeste foram avaliadas as populações de Paudalho-PE (PD), São Lourenço da Mata-PE (SM), Mata de São João-BA (MS), Areia-PB (AR) e Chapadinha-MA (CP). Na Região Norte foram avaliadas quatro populações do Estado do Amazonas, são elas: Iranduba (IR), Silves (SV), Presidente Figueiredo (PF) e Novo Airão (NV). As análises de estruturação genética demonstraram que existe divergência genética significativa entre as Regiões Norte e Nordeste. Na análise de DAPC (Análise Discriminante de Componentes Principais) foram formados três grupos: o grupo G1 composto pelas populações do Nordeste, exceto a população de Chapadinha-MA (CP) que formou um grupo separado (G3), e o grupo G2 composto pelas quatro populações coletadas no Estado do Amazonas. O teste de Mantel realizado com base na distância geográfica entre as populações e a divergência genética (F_{st}) identificou alta correlação positiva ($r = 0,78$; $p < 0,001$), o que indica que a divergência genética apresentada pelas populações coletadas é altamente influenciada pela distância geográfica. Foi possível identificar que os valores de divergência genética foram menores entre populações de uma mesma região e maiores quando populações da Região Nordeste foram comparadas com populações da Região Norte, o que indica que a ocorrência de fluxo gênico é maior dentro das regiões e menor entre as Regiões Norte e Nordeste do Brasil. Nas análises de diversidade genética foi constatado que as populações da Região Norte apresentam (em média) maior número de alelos em relação às populações da Região Nordeste, com número médio de alelos de 2722 para a Região Norte e 2509 para a Região Nordeste. A Região Norte também é detentora da maior diversidade genética (H_e) com valor médio de 0,1860 contra o valor de 0,1059 apresentado pela Região Nordeste. Embora possua maior diversidade genética, a Região Norte apresentou coeficiente de endogamia (f) positivo em três das quatro populações, variando de 0,0855 para a população de Novo Airão-AM a 0,2421 para a população de Presidente Figueiredo-AM. Na Região Nordeste apenas a população de Areia-PB apresentou valor positivo de f (0,1581), todas as outras populações apresentaram valores negativos sendo o menor valor obtido (-0,8146) na população de Chapadinha-MA (CP). Os resultados obtidos contribuem para o entendimento da distribuição da variação genética apresentada pela espécie, para a avaliação do estado de conservação das populações estudadas e fornece ferramentas que podem ser utilizadas para a conservação e no desenvolvimento de estratégias para ampliação da variabilidade genética da cajazeira.

Palavras-chave: Frutífera tropical, Cajá, Amazônia, Diversidade genética, Estrutura genética

ABSTRACT

Population genomics of cajazeira (*Spondias mombin* L.)

The cajazeira (*Spondias mombin*), an allogamous species, is a fruit tree native to the region that comprises the South of Mexico to the East of Brazil. The species has great potential for economic exploitation in Brazil and is already well consolidated commercially in the North and Northeast regions of the country. Most of the commercialized fruits comes from extraction since the species is classified as semi-domesticated and does not have significant commercial cultivation. The work was carried out with the aim of studying the structure and genetic diversity of *S. mombin* species in the North and Northeast regions of the country using SNP (Single Nucleotide Polymorphism) markers. Nine natural populations of the species were evaluated, five populations from the Northeast region and four populations from the North region. In the Northeast region, the populations of Paudalho-PE (PD), São Lourenço da Mata-PE (SM), Mata de São João-BA (MS), Areia-PB (AR) and Chapadinha-MA (CP) were evaluated. In the North Region, four populations from the State of Amazonas were evaluated, namely: Iranduba (IR), Silves (SV), Presidente Figueiredo (PF) and Novo Airão (NV). The analyzes of genetic structure showed a significant genetic divergence between the North and Northeast regions. In the analysis of DAPC (Discriminant Analysis of Principal Components) three groups were formed: the G1 group composed of the populations of the Northeast, except the population of Chapadinha-MA (CP) which formed a separate group (G3), and the G2 group composed of the four populations collected in the state of Amazonas. The Mantel test carried out based on the geographical distances between the populations and the genetic divergence (F_{st}) identified a high positive correlation ($r = 0.78$; $p < 0.001$), which indicates that the genetic divergence presented by the populations is highly influenced by the geographical distance. The values of genetic divergence were lower among populations from the same region and higher when populations from the Northeast region were compared with those from the North region, which indicates that the occurrence of gene flow is greater within regions than among the North and Northeast regions of Brazil. In the genetic diversity analysis, populations from the North region showed (on average) a greater number of alleles in relation to populations from the Northeast region, with an average value of 2722 for the North region and 2509 for the Northeast region. The North region also showed the greatest genetic diversity (H_e), with an average value of 0,1860, against the value of 0,1059 presented by the Northeast region. Although it has greater genetic diversity, the North region showed a positive inbreeding coefficient (f) in three of the four populations, with the lowest value (0,0855) for the population of Novo Airão-AM and highest value (0.2421) for the population of Presidente Figueiredo-AM. In the Northeast region only the population of Areia-PB presented a positive value of f (0.1581). All the other populations showed negative values, with the lowest value (-0.8146) obtained for the population of Chapadinha-MA (CP). The results obtained contribute to the understanding of the distribution of the genetic variation presented by the species, to assess the conservation status of the studied populations and provide tools that can be used for their conservation and in the development of strategies to increase the genetic variability of the cajazeira.

Keywords: Tropical fruit tree, Yellow mombin, Amazon, Genetic diversity, Genetic structure

1. INTRODUÇÃO

Na Amazônia são encontradas diversas espécies frutíferas com potencial econômico. Algumas dessas espécies possuem valor local, ou seja, são apreciadas e comercializadas localmente e outras já são consumidas em maior volume e têm importância nacional, como o cupuaçu (*Theobroma grandiflorum*) e o açaí (*Euterpe oleracea*), por exemplo. A cajazeira (*Spondias mombin*) é a espécie com maior potencial para exploração agroindustrial dentro do gênero *Spondias* (Silva et al., 2004). A espécie é nativa da região que compreende o sul do México ao Sudeste do Brasil (Mitchell e Daly, 2015). No Brasil a espécie ocorre na Amazônia em suas terras firmes ou várzeas, assim como em zonas mais úmidas da Mata Atlântica. A espécie se encontra em estado semi-domesticado e acredita-se que era cultivada pelos indígenas desde antes da chegada dos primeiros europeus à Amazônia (Clement, 1999).

Dentre as partes da planta, os frutos da cajazeira possuem maior valor agregado, podendo ser consumidos in natura ou processados em diversos produtos. Os frutos são drupas de coloração amarelada brilhante quando maduros, com mesocarpo suculento (Lorenzi, 1992). A polpa do cajá possui elevados níveis de potássio, magnésio, fósforo e cobre comparada a outras frutas consumidas no Brasil, a cajazeira possui compostos fenólicos, antioxidantes e carotenoides em maior concentração, o que confere à fruta elevado valor nutricional auxiliando na prevenção de diversas doenças (Tiburski et al., 2011). No entanto, devido à falta de conhecimento técnico que ainda existe a respeito da domesticação do cajá, a exploração dessa frutífera ainda é realizada, em grande parte, por meio do extrativismo. Dessa forma, estudos a respeito da caracterização desse recurso genético são de extrema contribuição nos avanços de pesquisas envolvendo a espécie (Silva et al., 2017a).

A Amazônia brasileira tem como principais causas diretas de desmatamento a pecuária, a agricultura de larga escala e a agricultura de corte e queima (Rivero e Almeida, 2009). O modelo de exploração predatória dos recursos florestais, baseado na substituição de áreas de floresta por extensões de pasto e área cultivada não leva em consideração as particularidades dos diversos espaços ecológicos da região (Vieira et al., 2008). Como resultado da prática desse modelo insustentável nas florestas tropicais, há perda da diversidade genética e alterações desfavoráveis na frequência de alelos com valor adaptativo ou comercial. Esse processo é conhecido como erosão genética (O'Neill et al., 2001). A preservação *ex situ* e *in situ* de germoplasma de espécies de *Spondias* que apresentam valor comercial ainda é bem restrita e considerada vulnerável, de maneira que a preservação não acompanha o acelerado processo de erosão genética causado por ação antrópica, podendo levar a perda de indivíduos agronomicamente valiosos (Silva et al., 2004).

Para avaliar o atual estado de conservação da espécie e traçar estratégias para sua conservação, é imprescindível a realização de estudos para quantificar e caracterizar a diversidade genética existente em populações naturais. Em trabalhos realizados para estudos da diversidade da

espécie foram utilizados marcadores morfológicos, bioquímicos (isoenzimas) e moleculares do tipo RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism), e ISSR (Inter Simple Sequence Repeat) (Silva et al., 2009a; Lima et al., 2011; Magalhães et al., 2013; Silva et al., 2014; Silva et al., 2017a,b). Nenhum dos estudos realizados anteriormente possuem uma amostragem tão extensa geograficamente quanto a realizada no presente estudo. Além do mais, nenhum dos marcadores antes utilizados possuem uma amostragem do genoma tão ampla quanto a amostragem realizada por marcadores do tipo SNPs (*Single Nucleotide Polymorphism*).

O presente estudo tem como objetivo fornecer informações a respeito da diversidade e estruturação genética apresentada pela espécie na Região Norte e Nordeste do Brasil através de uma amostragem ampla e um conjunto denso de marcadores SNPs, sendo o primeiro estudo com tal magnitude desenvolvido com a espécie. As informações geradas irão auxiliar na identificação de populações em vulnerabilidade genética nas duas regiões e permitir o desenvolvimento de estratégias para conservação *in situ* e *ex situ* da variabilidade genética da espécie a fim de contribuir para a sua conservação e no desenvolvimento de futuros programas de melhoramento genético.

2. OBJETIVOS

2.1. Objetivo Geral

Caracterizar a diversidade e estrutura genética de populações de cajazeira com uma amostragem ampla a partir de marcadores SNPs obtidos com a técnica de *genotyping by sequencing* (GBS).

2.2. Objetivos específicos

Estimar a diversidade genética entre e dentro das populações de cajazeira nas Regiões Norte e Nordeste do Brasil a partir de marcadores SNPs;

Avaliar o padrão espacial de diversidade genética e fluxo gênico das populações de cajazeira nas Regiões Norte e Nordeste do Brasil;

Comparar os níveis de diversidade genética e endogamia encontrados nas populações de cajazeira da Região Norte e Nordeste do Brasil;

Fornecer o entendimento da distribuição geográfica da diversidade genética de populações de cajazeira a fim de fornecer informações a respeito do manejo e conservação *in situ* e *ex situ* da variabilidade genética nas Regiões Norte e Nordeste do Brasil.

3. REVISÃO DE LITERATURA

3.1. Distribuição geográfica e aspectos gerais

Spondias L. é um gênero de espécies frutíferas pertencente à família das Anacardiaceae, e possui 18 espécies nativas da América, Ásia e Oceania (Mitchell e Daly, 2015). Entre estas espécies, a cajazeira (*Spondias mombin* L.) destaca-se pelo valor agroindustrial atrelado ao seu fruto que pode ser consumido in natura ou em forma de produtos processados. O centro de origem da espécie *S. mombin* é incerto, sendo considerada por Leon e Shaw (1990), citado por Sacramento e Souza (2009), uma espécie nativa da América tropical. Mitchell e Daly (2015) consideram a cajazeira nativa da região que compreende o sul do México ao Sudeste do Brasil. León et al. (2012) consideram a espécie como nativa da Mesoamérica. Clement (1999) relata que a espécie já era cultivada por indígenas na Amazônia antes mesmo da chegada dos descobridores das Américas em 1492. Na Região Nordeste foi descrita pela primeira vez em 1587 no Tratado Descritivo do Brasil pelo português Gabriel Soares de Souza (Carvalho e Alves, 2008, citados por Sacramento e Souza, 2009). A cajazeira tem a Mata Atlântica, a Amazônia Ocidental do Estado do Acre e as regiões de fronteira do Brasil com o Peru e a Bolívia como centro de diversidade no Brasil (Mitchell e Daly, 1995, citados por Fonseca et al., 2017). No Brasil a espécie ocorre na Amazônia em suas terras firmes ou várzeas, assim como em zonas mais úmidas da Mata Atlântica (Figura 1) (Sacramento e Souza, 2009).

A cajazeira é uma espécie de porte arbóreo, ereta, podendo atingir mais de 20 m de altura, com tronco revestido por casca muito grossa, acinzentada, rugosa, saliente, apresentando fendas (Lorenzi, 1992). As folhas da cajazeira são compostas, alternas, imparipinadas, com 5 a 11 pares de folíolos (Figura 2) (Sacramento e Souza, 2009).

A espécie *S. mombin* ($2n = 2x = 30$) apresenta reprodução assexuada (vegetativa) e sexuada, é alógama e depende de polinização cruzada (realizada principalmente por abelhas) para maior sucesso reprodutivo, apresentando dicogamia protândrica e alto valor para a razão pólen/óvulo (Souza et al., 2002; Ramos, 2009; Zórtea, 2019). No Brasil, segundo Ramos (2009), as plantas são hermafroditas, com a presença de flores estaminadas sendo a espécie classificada como andromonoica. As panículas possuem cachos compostos por várias flores (centenas) pedunculadas, com cinco sépalas, cinco pétalas, dez estames, gineceu com ovário formado por cinco carpelos que coincidem com o número de lóculos (Lozano, 1986). Com relação ao fruto da cajazeira, é uma drupa elipsoidal, glabra ou levemente pilosa de tamanho variado, com coloração amarelada brilhante quando maduros (Figura 2) (Silva e Silva, 1974).

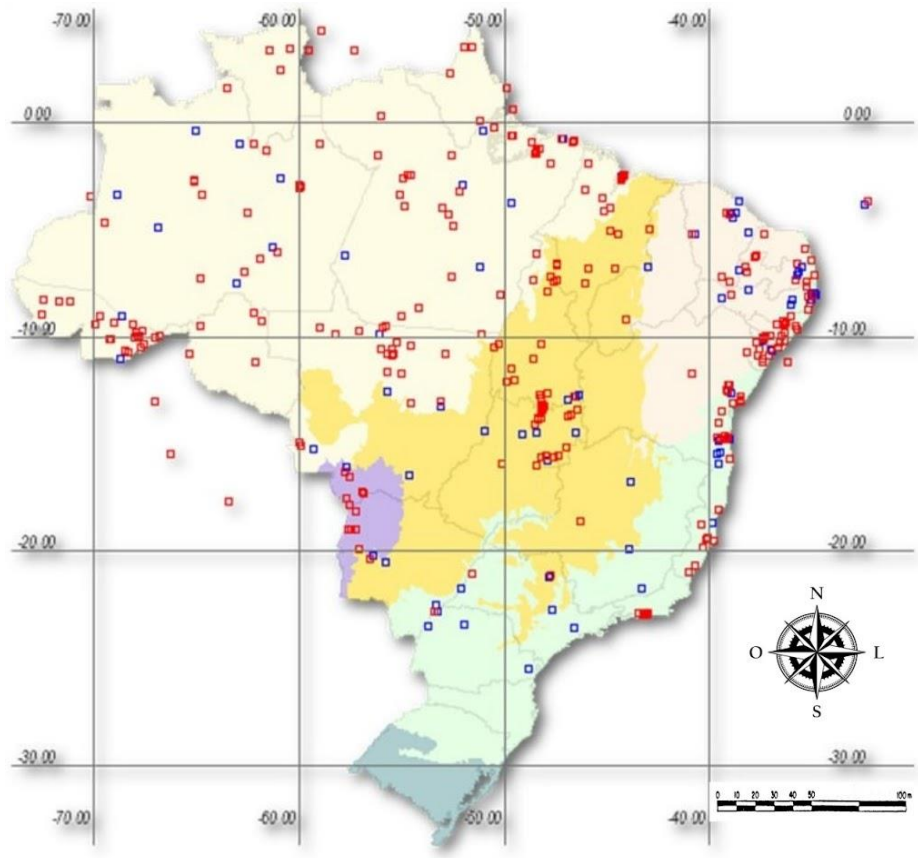


Figura 1. Mapa de distribuição da espécie *Spondias mombin*. Fonte: SpeciesLink (<http://smlink.cria.org.br/>)



Figura 2. Aspectos gerais dos aspectos gerais da árvore, fruto, folha e flores de cajazeira (*Spondias mombin*).

3.2. Bancos de germoplasma e estudos de diversidade genética da cajazeira

O Brasil é o país com o maior número de acessos de *S. mombin* mantidos em bancos de germoplasma, todos via *ex situ* (Santana, 2010). A EMEPA (Empresa Estadual de Pesquisa Agropecuária da Paraíba) possui 21 acessos da espécie coletados nos Estados da Paraíba e Pernambuco (Bosco, 2000, citado por Santana, 2010). O Instituto Agronômico de Pernambuco possui 33 acessos coletados na Zona da Mata e na região metropolitana do Estado (Bezerra et al., 2011). De acordo com as informações disponíveis no banco de dados Alelo (<http://alelobag.cenargen.embrapa.br/AleloConsultas>), a Embrapa (Empresa Brasileira de Pesquisa Agropecuária) possui três bancos de germoplasma ativo *ex situ* da espécie: a Embrapa Meio-Norte com 38 acessos (sem informação quanto ao local de coleta), a Embrapa Amazônia Oriental com 35 acessos coletados no Estado do Pará (municípios de Belém, Igarapé-Açu e Tome-Açu) e a Embrapa Agroindústria Tropical, que possui sete acessos coletados no Estado do Ceará e Piauí.

Silva et al. (2009a) estudaram a diversidade e a estrutura genética de populações de cajazeiras da Zona da Mata de Pernambuco por meio de isoenzimas e identificaram grande diversidade genética nas populações (H_e entre 0,530 e 0,574) e valores baixos de endogamia (f) que variaram entre -0,08 e -0,37. Silva et al. (2009b) avaliaram a diversidade genética de cajazeiras do banco de germoplasma do IPA (Instituto Agronômico de Pernambuco) com marcadores ISSR e detectaram grande variabilidade genética entre os acessos, que apresentaram 51,69% de média para similaridade genética. Lima et al. (2011), com base na técnica de RAPD, quantificaram a variação genética entre acessos de cajá pertencente à coleção de germoplasma da Embrapa Meio-Norte e identificaram 68,8% de similaridade genética média entre os acessos avaliados. Mais recentemente, Silva et al. (2017a) avaliaram a diversidade genética entre e dentro de populações naturais de cajazeira ocorrentes no Estado do Mato Grosso utilizando ISSR e observaram maior variação genética dentro de populações (77 %). Silva et al. (2017b) estudaram a diversidade genética de *S. mombin* em três populações do norte do Mato Grosso com o uso de marcadores morfológicos dos frutos e das sementes e obtiveram resultados de dissimilaridade genética que variaram de 0,045 a 0,753.

Apesar dos diversos estudos apontados acima com diferentes abordagens metodológicas, nenhum estudo com marcadores de nova geração de sequenciamento de DNA para obtenção de SNPs, a fim de estimar a estrutura genômica de populações de cajazeira, foi realizado até o momento. Vale ressaltar que o espectro espacial de amostragem no presente estudo é maior do que se verifica nos trabalhos já realizados para a cajazeira, o que permitirá uma maior compreensão dos níveis de diversidade nos diferentes ambientes onde a espécie ocorre, assim como a estrutura genética e filogeográfica de suas populações.

3.3. Genotipagem por sequenciamento e marcadores SNPs

Marcadores moleculares são sequências de nucleotídeos capazes de identificar polimorfismo em sequências nucleotídicas de diferentes indivíduos, seja em regiões expressas e não expressas do genoma (Nadeem et al., 2018). Diversas técnicas são utilizadas para obtenção dos diversos tipos de marcadores moleculares existentes, que são utilizados para o estudo da variabilidade genética ao nível de sequências de DNA (Singh et al., 2014; Nadeem et al., 2018).

Até pouco tempo atrás, estudos com populações naturais de espécies não-modelos, faziam uso de muitos indivíduos amostrados dentro da população (>20) e geralmente utilizavam um pequeno número de marcadores genéticos (<20) (Willing et al., 2012). Os marcadores baseados em polimorfismo de nucleotídeo único (SNP), além de serem os marcadores mais abundantes encontrados no genoma, são versáteis pois além da avaliação da variação neutra, possibilitam o estudo e identificação de regiões do genoma que estão sofrendo seleção natural na população, ou seja, regiões diretamente associadas a adaptação (Cortinovis et al., 2020). Marcadores do tipo SNPs são obtidos a partir do polimorfismo em uma única base ao longo de uma sequência de DNA. Esse tipo de polimorfismo ocorre em média a cada 1.000 nucleotídeos, o que viabiliza a obtenção de dezenas de milhares de marcadores desse tipo ao longo do genoma das espécies. A utilização de muitos marcadores (>500) aumenta significativamente a capacidade de detecção de diferenciação genética, mesmo que o tamanho de amostragem seja relativamente pequeno (Willing et al., 2012). Dessa forma, fica evidente o potencial de aplicação dos marcadores do tipo SNPs em estudos de genômica populacional.

Com a chegada das tecnologias de sequenciamento de nova geração (NGS, *Next Generation Sequencing*), pesquisadores são agora capazes de compreender importantes questões ecológicas e evolutivas de espécies não-modelos em uma escala genômica e precisão inimagináveis há apenas alguns anos atrás (Ekblom e Galindo, 2010). As tecnologias de NGS se destacam por gerar grande quantidade de dados de sequências com um custo satisfatório, o que torna cada vez mais democrático o uso do sequenciamento para os grupos de estudos moleculares (Gilad et al., 2009). A medida em que os custos de sequenciamento vão se tornando mais baixos e as melhorias nos protocolos de laboratório e computacionais vão surgindo, os estudos genômicos vão se tornando cada vez mais acessíveis, como o estudo de polimorfismos de nucleotídeo único (SNPs) para espécies vegetais não-modelo em escala genômica, por exemplo (Gilad et al., 2009; Tsai et al., 2017). Alguns trabalhos utilizando os marcadores do tipo SNPs já foram realizados com espécies amazônicas. Cornejo et al. (2018) estudaram a genômica populacional do cacau (*Theobroma cacao*) por meio de marcadores SNPs para entendimento do processo de domesticação da espécie. Souza et al. (2018) realizaram o estudo da estrutura populacional com seringueira (*Hevea brasiliensis*). Brousseau et al. (2018) estudaram a

genômica e a adaptação microgeográfica da espécie amazônica *Eperua falcata* e, mais recentemente, Alves-Pereira et al. (2019) analisaram a diversidade genômica da palmeira juçara (*Euterpe edulis*), do urucum (*Bixa orellana*) e da mandioca (*Manihot esculenta*).

A técnica de genotipagem por sequenciamento (genotyping-by-sequencing - GBS) é baseada na redução da complexidade do DNA genômico por enzimas de restrição e no uso de adaptadores de DNA do tipo barcode para produção de bibliotecas multiplexadas de amostras que são submetidas posteriormente ao sequenciamento de nova geração (NGS) (Poland e Rife, 2012). Com essa combinação a técnica vem demonstrando a sua capacidade na produção de milhares de marcadores moleculares do tipo SNPs em diversas espécies (Elshire et al., 2011; Poland et al., 2012). Com uma abordagem mais simplificada, possibilidade de multiplexagem de diversas amostras e redução do custo de sequenciamento por amostra, a técnica desenvolvida por Elshire et al. (2011) é amplamente aplicada em estudos populacionais, de caracterização de germoplasmas, melhoramento e em estudos de mapeamento genético. Por não exigir conhecimento prévio a respeito do genoma da espécie estudada, a técnica de GBS é ideal em estudos com espécies menos estudadas, com poucas informações genômicas disponíveis já que a genotipagem é realizada simultaneamente durante a identificação de marcadores SNPs com base nos fragmentos sequenciados em *pipelines* como TASSEL-GBS, UNEAK e UStacks (Kim et al., 2016). Ao longo dos últimos anos as tecnologias de NGS tiveram um aumento crescente em seu uso, seja para projetos de sequenciamento de genomas inteiros, seja para identificação de SNPs, para estudo de diversidade dentro de espécies e para construção de mapas de haplótipos (Metzker, 2010).

4. MATERIAIS E MÉTODOS

4.1. Amostragem

As coletadas foram realizadas em áreas afastadas do centro urbano das cidades às margens de estradas. Foi possível visualizar que, embora desabitadas, as áreas de coleta apresentavam sinais claros de ação antrópica, com pouca presença de árvores de grande porte.

Foram utilizadas folhas jovens de indivíduos adultos coletados aleatoriamente em cada população localizados em nove municípios nos Estados do Amazonas, Maranhão, Pernambuco, Paraíba e Bahia (Tabela 1; Figura 3), envolvendo os biomas da Amazônia (AM), Cerrado (MA), Mata Atlântica (PE e BA) e Caatinga (PB), num total de nove populações.



Figura 3: Mapa com a localização das coletas dos indivíduos de cajazeira (*Spondias mombin*), nos Estados do Amazonas (AM), Maranhão (MA), Pernambuco (PE), Paraíba (PB) e Bahia (BA).

4.2. Extração e quantificação de DNA

O DNA foi extraído das folhas por meio do protocolo descrito por Inglis et al. (2018) com algumas modificações. As amostras de folhas foram maceradas em nitrogênio líquido, de forma manual. Após maceradas, as amostras passaram por quatro pré-lavagens com o tampão contendo sorbitol (100 mM Tris-HCl pH8.0, 0.35 M sorbitol, 5 mM EDTA pH 8.0, 1% (massa/volume) Polyvinylpyrrolidona (PVP). Logo após a adição do sorbitol foi realizada a adição de 1 % (volume do tampão) de 2-mercaptoetanol. As amostras foram centrifugadas em temperatura ambiente a 7.000 x g durante cinco minutos e tiveram o sobrenadante descartado ao fim de cada lavagem. Após as lavagens com sorbitol, foram adicionados 700 µl de tampão de lise CTAB (100 mM Tris-HCl pH 8.0, 3 M NaCl, 2% CTAB, 20 mM EDTA e 1% (do volume do tampão) de PVP). As amostras tiveram a adição de 1 % (do volume do tampão) de 2-mercaptoetanol e seguiram ao banho-maria pré-aquecido a 65 °C, onde permaneceram por 30 minutos sendo levemente homogeneizadas a cada cinco minutos. Depois de serem retiradas do banho-maria, as amostras foram resfriadas por cinco minutos à temperatura ambiente e foram então adicionados 800 µl de cloroformio:álcool isoamílico (24:1 v/v; CIA). As amostras foram homogeneizadas suavemente e centrifugadas a 13.000 x g durante 10 minutos em temperatura ambiente. Ao fim da centrifugação, a fase superior aquosa das amostras foi transferida para um novo tubo contendo 10% do volume, do sobrenadante pipetado, de acetato de sódio 3 M pH 5.2 e 66% do volume de Isopropanol a -20 °C. Após a adição do isopropanol as amostras seguiram para o freezer a -20 °C durante o período de uma hora. O pellet foi obtido por centrifugação a 13.000 x g durante 10 minutos à temperatura ambiente e o sobrenadante foi descartado com cuidado para não perturbar o pellet formado. O pellet obtido foi submetido a uma lavagem com o volume de 1 ml de álcool 70% sob centrifugação a 13.000 x g durante 10 minutos. Ao fim da centrifugação as amostras tiveram o excesso de álcool removido e permaneceram na bancada durante uma hora para secarem. O DNA foi ressuspenso em 25 µl de tampão TE e tratado com 0.1 mg ml⁻¹ de RNase A e mantido em freezer a -20 °C até o momento de uso.

Tabela 1: Relação e origem das populações de cajazeira (*Spondias mombin*) coletadas em seis Estados, nove municípios, em quatro biomas, incluindo número de indivíduos/população e coordenadas geográficas.

Populações	Município/Estado	Nº indivíduos	Bioma	Latitude	Longitude
NV	Novo Airão/AM	5	Amazônia	2°37'20,8"S	60°57'01,7"W
IR	Iranduba/AM	2	Amazônia	3°17'03,4"S	60°11'16,2"W
PF	Presidente Figueiredo/AM	5	Amazônia	2°03'15,5"S	60°01'41,8"W
SV	Silves/AM	3	Amazônia	04°14'41,0"S	42°17'40,0"W
CP	Chapadinha/MA	3	Cerrado	3°44'43,0"S	43°22'20,3"W
PD	Paudalho/PE	6	Mata Atlântica	7°53'48,0"S	35°10'47,0"W
SM	São Lourenço da Mata/PE	6	Mata Atlântica	8°00'10,0"S	35°01'04,0"W
AR	Areia/PB	6	Caatinga	6°59'19,6"S	35°44'11,6"W
MS	Mata de São João/BA	5	Mata Atlântica	12°29'19,6"S	37°59'42,8"W
Total		41			

A quantificação das amostras foi estimada pela intensidade de fluorescência emitida pelo corante GelRed sob UV em géis de agarose a 1%. Essa intensidade foi comparada à de padrões com pesos moleculares e concentrações específicas e conhecidas de DNA do fago lambda (5, 20, 40, 80 e 160 ng). A eletroforese foi realizada em cuba contendo tampão TBE 1x a uma voltagem de 2 V/cm.

4.3. Montagem da biblioteca de GBS

O primeiro passo para a montagem da biblioteca de GBS é a obtenção de amostras de DNA normalizadas, ou seja, as amostras a serem multiplexadas devem possuir concentração de DNA homogênea no grupo de amostras (Poland et al., 2012). Para a montagem da biblioteca da cajazeira, a normalização foi realizada com o objetivo de homogeneizar a concentração das amostras a 20 ng de DNA por microlitro de amostra. Antes de dar início ao processo de normalização, as amostras foram previamente analisadas por meio da técnica de eletroforese em gel de agarose (com concentração de 1%) para quantificação da concentração inicial de DNA. Após a quantificação as amostras receberam

a adição do solvente no qual ocorreu a ressuspensão do pellet de DNA, neste caso o TE, até que possuam concentração próxima de 20 ng por microlitro quando comparada com um marcador de peso molecular com a concentração de 20 ng por microlitro. Depois da primeira normalização realizada por eletroforese em gel de agarose, as amostras foram quantificadas mais uma vez com fluorômetro de alta sensibilidade em equipamento de Invitrogen, o Qubit 4, e as amostras que apresentaram desvios na concentração foram então novamente normalizadas.

Depois de normalizadas as amostras precisam ser digeridas. Nesta etapa o DNA tem a sua complexidade reduzida a partir da sua fragmentação por meio de enzimas de restrição, no processo chamado de digestão. Existem diversas combinações de endonucleases que podem ser utilizadas durante a digestão, porém deve-se sempre utilizar uma enzima com corte específico (raro), que cortará a dupla fita de DNA em regiões específicas deixando uma ponta coesiva para ligação de adaptadores *barcode*, e uma enzima de corte comum (frequente) que irá efetivamente fragmentar a molécula de DNA. Após testes de digestão foram selecionadas as duas enzimas a serem utilizadas na cajazeira, a *Pst*I (enzima de corte específico) e a *Mse*I (enzima de corte comum).

As amostras, agora digeridas, tiveram os adaptadores do tipo *barcode* ligados aos fragmentos gerados na digestão. A ligação foi realizada com o auxílio da enzima NEB T4 DNA ligase #M0202 em tampão NEB Buffer4 na presença de ATP. Para assegurar o sucesso na reação de ligação, a ligase T4 foi utilizada em maior concentração. Após a reação de ligação, cada amostra se ligou a um *barcode* específico. Os *barcodes* são sequências de até seis nucleotídeos que diferem entre si em, pelo menos, uma base e que servem como identificadores, permitindo a multiplexagem de diversas amostras, uma vez que cada amostra possui um *barcode* específico e pode ser identificada posteriormente. Ao fim da multiplexagem, o multiplex (conjunto das amostras identificadas) foi submetido à amplificação por PCR (reação em cadeia da polimerase). O enriquecimento dos fragmentos foi realizado a partir de *primers* complementares aos adaptadores que foram utilizados na preparação da biblioteca de GBS das 41 amostras de cajazeira. A biblioteca de GBS Foi realizada no Laboratório de Diversidade Genética e Melhoramento da Esalq/USP. O sequenciamento dos produtos de PCR foi realizado na plataforma HiSeq2500 Illumina, junto à empresa EcoMol, no Centro de Genômica Esalq/USP.

4.4. Obtenção dos marcadores SNPs

Os marcadores SNPs foram obtidos com o auxílio do programa Stacks v. 1.42 (Catchen et al., 2011). As amostras foram inicialmente submetidas a etapa de `process_radtags`, onde as sequências passam por um controle de qualidade onde sequências que poderiam vir a comprometer a veracidade das análises são eliminadas. Nesta primeira etapa também é realizado o `demultiplex`, onde as sequências

de cada indivíduo são devidamente separadas de acordo com os diferentes *barcodes*. Após a primeira etapa (filtragem e demultiplex), as sequências retidas foram analisadas pelo componente *ustacks* com os parâmetros *-m 3* (define o número de *reads* necessário para formar um *stack*), *-M 2* (especifica o número de *mismatches* permitido entre *stacks* a serem agrupados em um locus) e *-N 2* (máxima distância permitida entre *stacks* primários e *reads* secundárias a serem inseridas nos *stacks*). A próxima etapa foi a criação de um catálogo com todos os possíveis locus identificados e a presença destes locus em cada um dos indivíduos, esta etapa foi realizada pelo *cstacks* com o parâmetro *-n 2* (o parâmetro determina o número de *mismatches* entre o loci das amostras para a formação do catálogo). Após a formação do catálogo, dois componentes foram utilizados para o cruzamento de informações entre os locus obtidos para cada indivíduo e os locus do catálogo (*sstacks*) e remoção dos locus com menor probabilidade (*rxstacks*, *--lnl_lim -10*). O componente *populations* foi utilizado para a filtragem final dos marcadores SNPs. Foram retidos apenas um SNP por tag, com profundidade $\geq 3X$; MAF (frequência alélica mínima) $\geq 0,01$; SNP presente em pelo menos 60% das amostras dentro de populações; SNP presente em pelo menos cinco das populações.

4.5. Análises dos marcadores SNPs

Os marcadores SNPs foram analisados com o software R (R Core Team, 2020) a partir de pacotes especializados. As análises de estruturação genética foram realizadas por meio de DAPC (Análise Discriminante de Componentes Principais) (Jombart et al., 2010). A DAPC foi realizada com o auxílio do pacote *adegenet* (Jombart, 2008) no software R. A análise tem como objetivo identificar grupos com base na estruturação genética, minimizando a variação dentro dos grupos e maximizando a variação entre os grupos. O número de grupos regionais (locais de coleta das populações) não é fornecido na análise e o número ótimo de grupos, que explicam os dados, é calculado pelo método *K-means* onde diversos valores de *k* (grupos) e vários agrupamentos são testados por meio de *Bayesian Information Criterion* (BIC). Os resultados gerados são fornecidos em uma curva de valores BIC em função de *k*, onde o valor ideal de grupos é tido com o valor mais baixo da curva a partir do qual a curva retorna à sua crescente. As análises de agrupamento foram realizadas a partir de árvores filogenéticas construídas pelo método de *Neighbor-Joining*, sem enraizamento, geradas com base na distância genética de Nei (1972) com o auxílio do pacote *ape* (Paradis e Schliep, 2019).

As estimativas dos parâmetros básicos de diversidade genética como número total de alelos (*A*), heterozigosidade observada (H_o), heterozigosidade esperada (H_E) e coeficiente de endogamia (*f*), e AMOVA (Análise de Variância Molecular) foram geradas com o auxílio do pacote *hierfstat* (Goudet e Jombart, 2020) e o pacote *poppr* (Kamvar et al., 2014; Kamvar et al., 2015). Tantos

os parâmetros básicos quanto a AMOVA foram realizados em duas situações, levando em consideração as populações regionais coletadas e os grupos formados a partir da análise de DAPC. Com o auxílio do pacote hierfstat foi gerada uma matriz de F_{st} par a par, sendo a visualização gráfica da matriz gerada com o auxílio do pacote corrplot (Wei e Simko, 2017). O teste de Mantel foi realizado visando avaliar a correlação entre a divergência genética entre populações baseada no valor F_{st} par a par e a distância geográfica par a par das populações no pacote ade4 (Chessel et al., 2004; Dray e Dufour, 2007a; Dray et al., 2007b; Bougeard e Dray, 2018; Thioulouse et al., 2018). A matriz de distância geográfica, gerada a partir das coordenadas geográficas, foi construída com o auxílio do pacote geodist (Padgham e Sumner, 2020).

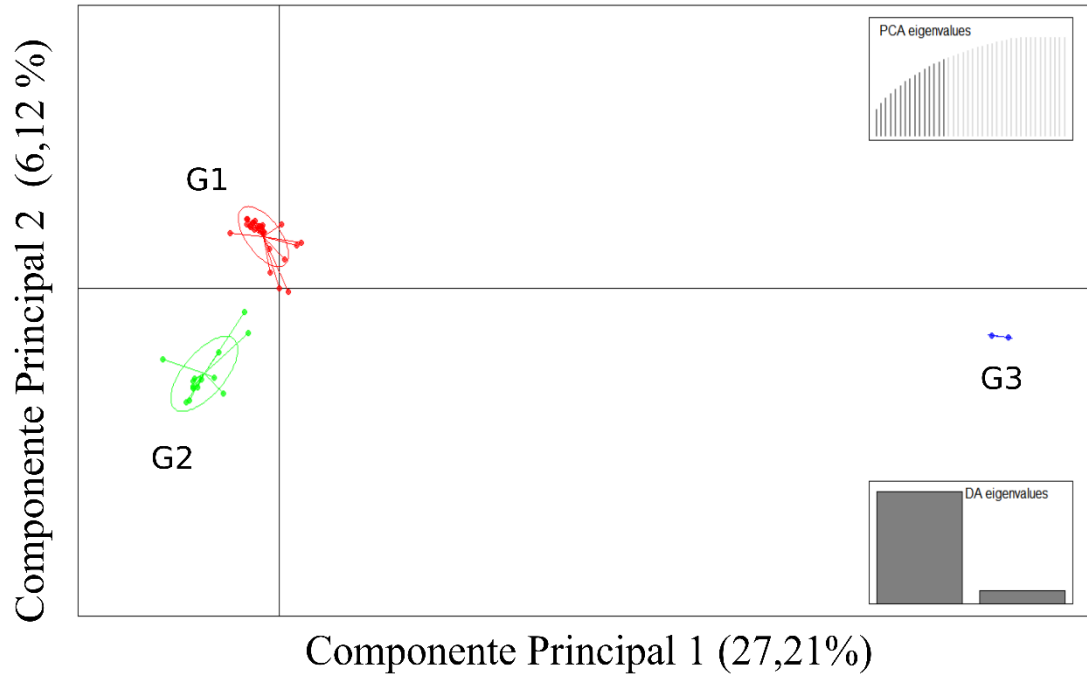
5. RESULTADOS

Os estudos de estrutura e diversidade genética foram realizados com um total de 2003 locos correspondentes a marcadores SNPs para 41 indivíduos de cajazeira. A análise de DAPC foi realizada a partir de um total de 15 componentes principais, os quais explicam 77% da variação total, e duas funções discriminantes. O número ótimo de grupos obtidos a partir da análise de *K-means* corresponde a três grupos (Figuras 4a e 4b). O grupo G1, em vermelho, é composto por 25 indivíduos que compreende os indivíduos das populações de Paudalho-PE (PD), São Lourenço da Mata-PE (SM), Areia-PB (AR), Mata de São João-BA (MS), um acesso da população de Chapadinha-MA (CP) e um acesso da população de Presidente Figueiredo-AM (PF). O grupo G2, em verde, é composto por 14 indivíduos, todos pertencentes às populações coletadas no Estado do Amazonas, Iranduba (IR), Silves (SV), Presidente Figueiredo (PF) e Novo Airão (NV). O grupo G3, em azul, foi composto por apenas dois dos três indivíduos provenientes da população de Chapadinha-MA (CP). Quando realizada a análise de DAPC com $K = 2$, a população de Chapadinha-MA (CP) agrupou com as demais populações do nordeste no grupo G1 (Figura 5).

Foi possível visualizar que os grupos G1 e G2 são mais próximos geneticamente em relação ao grupo G3, que embora seja formado por indivíduos de uma população da Região Nordeste, se apresentou bem isolado das demais populações da sua região de origem. Com a análise de DAPC com apenas uma função discriminante foi possível visualizar o resultado do agrupamento em apenas uma dimensão (Figura 4b). Os grupos G1 e G2 apresentam uma área de sobreposição entre si, com destaque para a maior amplitude de distribuição do grupo G2 em relação aos demais o que demonstra maior variação genética dentro deste grupo em relação aos demais grupos.

Os dados gerados na análise de DAPC foram também visualizados em forma de barras para melhor entendimento das relações entre os grupos a partir de cada indivíduo, onde a probabilidade de agrupamento dos indivíduos pode ser visualizada (Figura 6). Na formação dos grupos ficou explícita a separação dos indivíduos por macrorregiões, embora tenha ocorrido a presença de um acesso de uma população do Estado do Amazonas (PF) dentro do grupo G1, e um dos três indivíduos da população de Chapadinha, no Maranhão, tenha agrupado também com as populações deste grupo. O grupo G1 ainda é majoritariamente formado por indivíduos da região Nordeste (Pernambuco, Paraíba e Bahia) (Figura 7). No grupo G2 temos apenas os indivíduos provenientes da Região Norte, representada pelas populações coletadas no Estado do Amazonas. O grupo G3 se apresenta isolado dos demais, embora um dos indivíduos da população maranhense tenha sido alocado no grupo G1 (Nordeste).

a



b

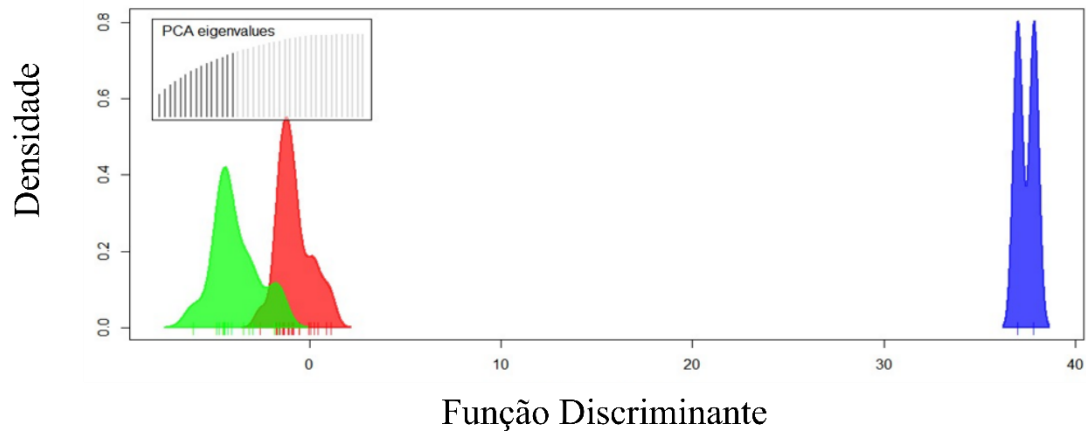


Figura 4: Análise Discriminante de Componentes Principais (DAPC) realizada com os 41 indivíduos de cajazeira (*Spondias mombin*) e 2003 marcadores SNPs. a) Gráfico de dispersão com base em duas funções discriminantes com os três grupos identificados pelo método de K-means. b) DAPC a partir da primeira função discriminante para melhor entendimento da densidade e sobreposição entre os grupos. O grupo G1 é representado na cor vermelha, o grupo G2 é representado na cor verde e o grupo G3 é representado na cor azul.

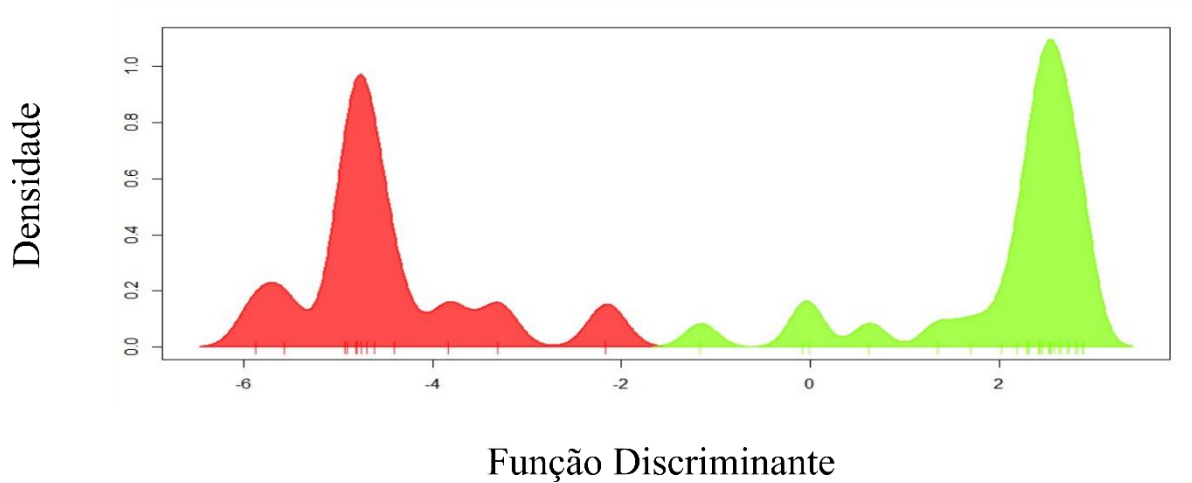


Figura 5: Análise de DAPC gerada com $K = 2$ e 15 componentes principais. O grupo vermelho corresponde ao grupo G1 formado pelas populações de cajazeira (*Spondias mombin*) de Paudalho-PE (PD), São Lourenço da Mata-PE (SM), Mata de São João-BA (SM), Areia-PB (AR) e Chapadinha-MA (CP). O grupo verde corresponde ao grupo G2 composta pelas populações de Iranduba (IR), Novo Airão (NV), Presidente Figueiredo (PF) e Silves (SV), todas provenientes do Estado do Amazonas.

Na análise de probabilidade de agrupamento (*membership probability*), realizada no pacote *ade4* (Jombart, 2008), com base dos dados gerados na análise DAPC e com k ótimo igual a três, novamente observou-se três grupos, sendo o grupo 1 (G1 - vermelho) representado pelas populações do Nordeste, o grupo 2 (G2 - verde) pelas populações do Norte, e o grupo 3 (G3 – azul) pela população do Maranhão (Figura 6). Foi possível identificar que a mistura entre os grupos é baixa, apenas um dos indivíduos da população de Silves (G2) apresenta mistura com o grupo G1. No gráfico a população maranhense aparece entre o grupo com populações da Região Nordeste e o grupo com as populações amazônicas, com um dos indivíduos com 100% de probabilidade de agrupamento no grupo G1, junto com os indivíduos da população de Presidente Figueiredo.

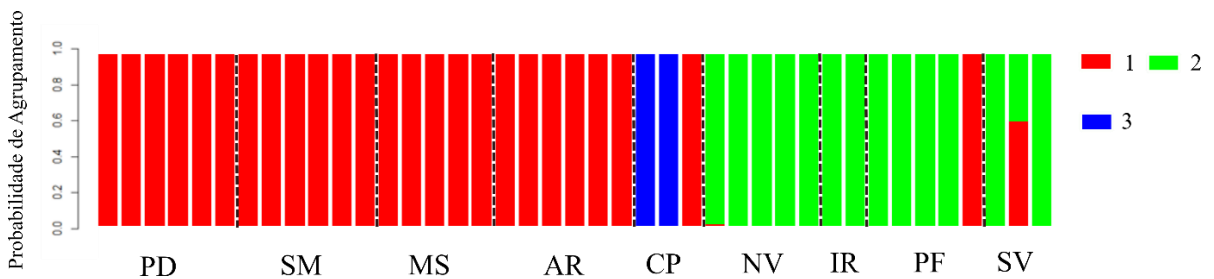


Figura 6: Análise de probabilidade de agrupamento de acordo com os resultados gerados na DAPC. Cada barra delimitada representa um indivíduo, as populações de cajazeira (*Spondias mombin*) estão representadas pelo intervalo formado entre as linhas pontilhadas, sendo PD (Paudalho-PE), SM (São Lourenço da Mata-PE), MS (Mata de São João-Bahia), AR (Areia-Paraíba), CP (Chapadinha-Maranhão), NV (Novo Airão-Amazonas), IR (Iranduba-Amazonas), PF (Presidente Figueiredo-Amazonas), SV (Silves-Amazonas).

A análise de agrupamento foi realizada a partir de uma árvore de distância genética construída pelo método de *Neighbor-Joining* (Figura 7). A árvore foi construída sem enraizamento definido, dessa forma, a análise não foi utilizada para inferências evolutivas e sim para entendimento de como os grupos definidos com base na análise de DAPC se relacionam entre si. Com a análise foi possível identificar que há um padrão de agrupamento por Estados de coleta. Foram formados cinco grupos distintos. Os grupos I, II e III, que equivalem ao grupo G1 da análise de DAPC, classificou as populações da Bahia (BA) (I), Pernambuco (PE) (II) e Paraíba (PB) (III), formando três grupos bem definidos, sem mistura de indivíduos de outros Estados. O grupo IV, que equivale ao grupo G3 da análise de DAPC, foi formado por dois dos três indivíduos da população maranhense, e foi posicionado entre o grupo das populações do Nordeste (G1) e o grupo V, formado pelas populações do Amazonas, equivalente ao G2.

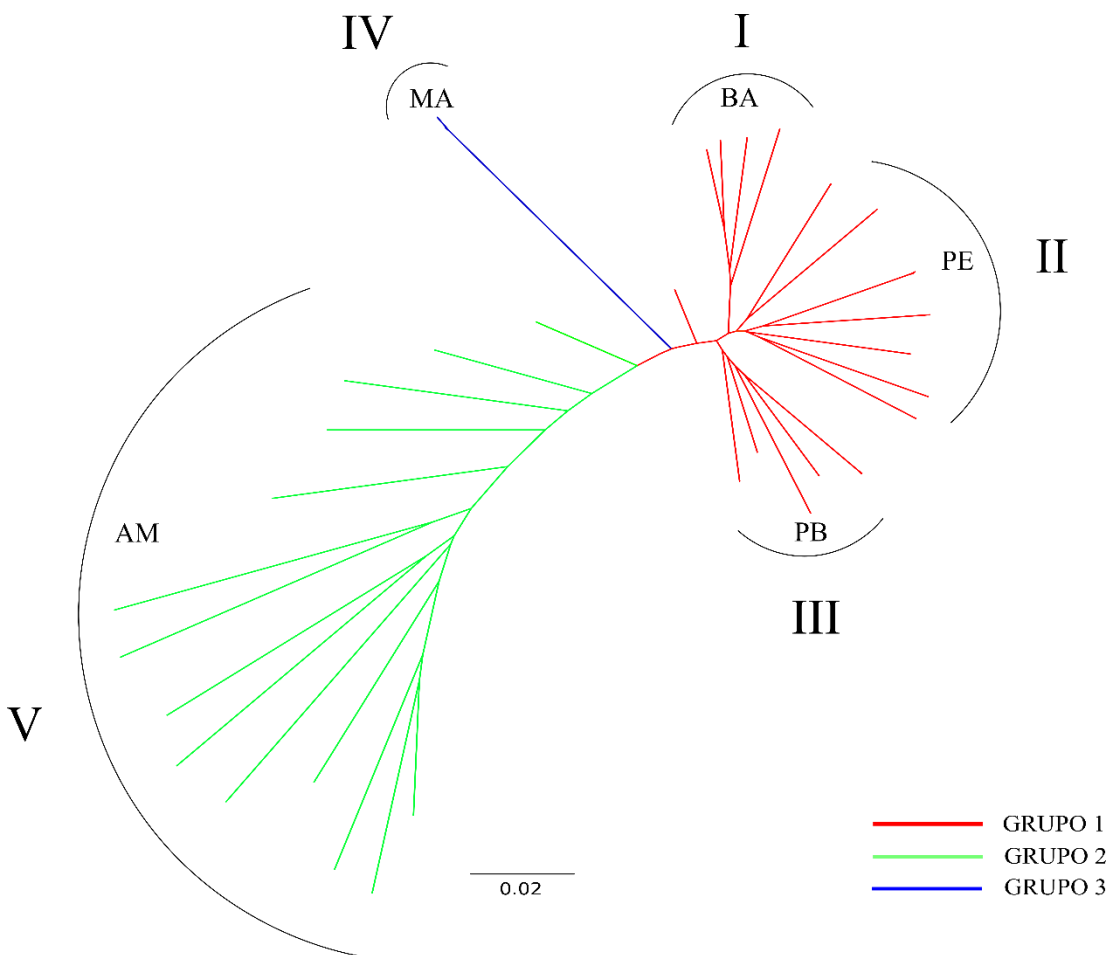


Figura 7: Análise de agrupamento a partir da árvore filogenética construída pelo método de *neighbor-joining* com base nos grupos identificados pela análise DAPC para populações de cajazeira (*Spondias mombin*). Os algarismos romanos (I, II, III, IV e V) representam os grupos obtidos nesta análise, enquanto que as cores representam os grupos identificados pela análise do DAPC.

No dendrograma, também construído pelo método *Neighbor-Joining* (Figura 8) e distância de Nei (1972), é possível observar o comportamento de cada um dos indivíduos em relação aos grupos identificados pela análise DAPC.

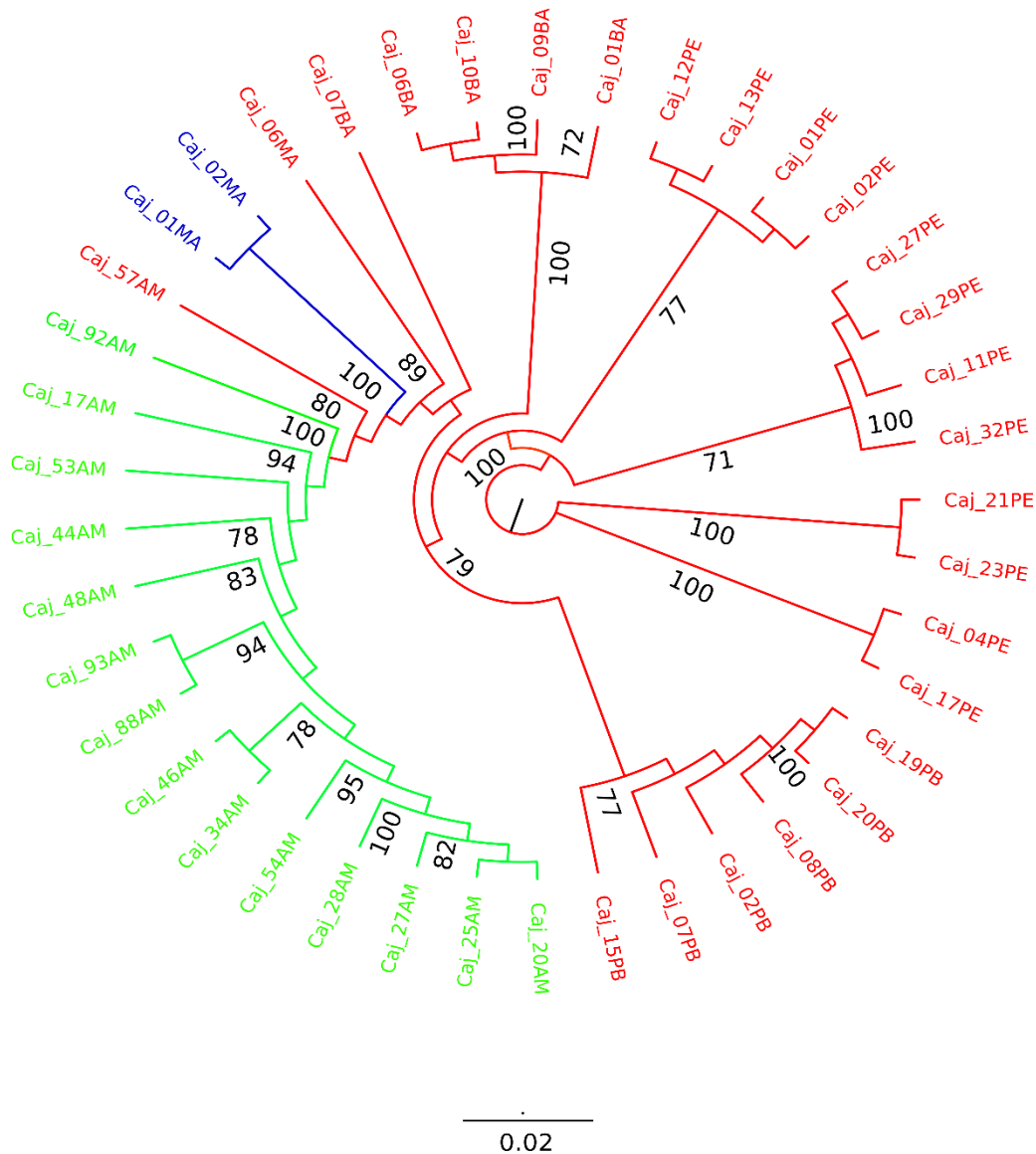


Figura 8: Dendrograma obtido pelo método de *Neighbor-Joining* com representação na mesma coloração dos grupos identificados pela análise DAPC (G1 em vermelho, G2 em verde e G3 em azul) com identificação e posicionamento de cada um dos indivíduos de cajazeira (*Spondias mombin*) analisados.

Na análise de divergência genética (F_{st}) par a par entre os grupos, foi possível identificar elevada diferenciação genética entre todos os grupos, com valores superiores a 0,25 em todos os cenários segundo a escala para F_{st} apresentada por Hartl e Clark (1998). Embora os valores tenham apresentado baixa variação, os maiores valores de F_{st} foram identificados entre os grupos G1 e G3, e G2 e G3, e o menor valor de F_{st} foi identificado entre os grupos G1 e G2 (Tabela 2). A análise de variância

molecular (AMOVA) forneceu resultados semelhantes para os grupos de DAPC e para as populações coletadas, ambos significativos a 5% de probabilidade (Tabela 3). Em ambos os cenários a maior porcentagem de variação aconteceu dentro das populações e dos grupos, 61,4% e 59,3%, respectivamente.

Tabela 2: Resultado da análise de F_{st} par a par entre os grupos identificados pela análise DAPC para a espécie *S. mombin*. O grupo G1 é formado pelas populações de Paudalho-PE (PD), São Lourenço da Mata-PE (SM), Mata de São João-BA (SM), Areia-PB (AR). O grupo verde corresponde ao grupo G2 composta pelas populações de Iranduba (IR), Novo Airão (NV), Presidente Figueiredo (PF) e Silves (SV), todas provenientes do Estado do Amazonas. O grupo G3 é formado por dois indivíduos da população de Chapadinha-MA (CP).

F_{st}	G1	G2	G3
G1	0		
G2	0,33	0	
G3	0,37	0,36	0

Tabela 3: Resultados gerados pela AMOVA para identificação das fontes de variabilidade genética entre e dentro dos grupos identificados na DAPC e das populações regionais coletadas de *S. mombin*.

Fonte de variação	Grau de liberdade	Soma dos quadrados	Coefficiente de variação	Porcentagem da variação	Estatística F
Entre as populações	8	12636,86	259,25	38,56%	0,3856*
Dentro das populações	32	13215,64	412,99	61,44%	
Entre os grupos DAPC	2	2481,53	104,30	40,69%	0,4070*
Dentro dos grupos DAPC	38	5775,89	151,99	59,31%	

*Significativo com p -valor < 0,00005.

A matriz de F_{st} par a par, gerada com todas as populações coletadas, possibilitou a identificação da maior distância genética quando foram comparadas as populações da região Nordeste com as populações da região Norte (Tabela 4; Figura 9). Quando as populações do Nordeste foram comparadas entre si é possível observar que as populações de São Lourenço da Mata (SM) e Paudalho (PD), ambas provenientes do Estado de Pernambuco, apresentaram o menor valor para o F_{st} (0,06). O maior valor de F_{st} dentro da região Nordeste foi obtido quando as populações de Areia (AR), na Paraíba, e Mata de São João (MS), na Bahia, foram comparadas, gerando o F_{st} de 0,21 com divergência genética considerada alta (Hartl e Clark, 1998). Dentro das populações da região Norte foi observado

que o menor valor de F_{st} foi obtido quando se comparou a população de Silves (SV) e a população de Presidente Figueiredo (PF), com o valor de 0,05, que indica baixa divergência genética. A maior divergência genética entre as populações amazonenses foi obtida quando as populações de Iranduba (IR) e Novo Airão (NV) foram comparadas entre si, com valor obtido de 0,22 o que indica alta divergência genética entre as populações.

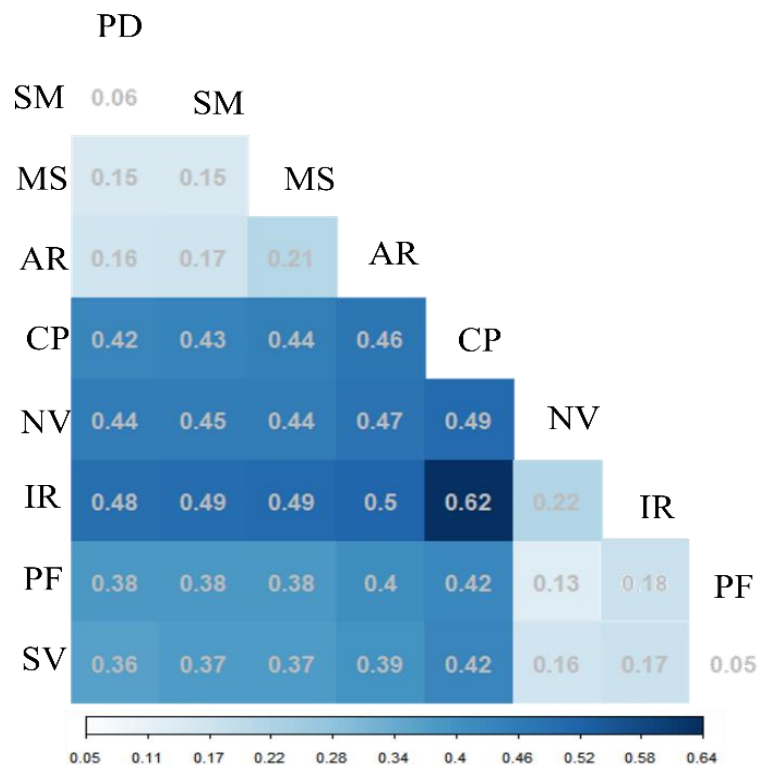


Figura 9: Resultados obtidos a partir da análise par a par de F_{st} entre as populações coletadas da cajazeira (*Spondias mombin*). Todos os valores são significativos para valor de $p < 0,05$. As populações estão representadas pelas siglas PD (Paudalho-PE), SM (São Lourenço da Mata-PE), MS (Mata de São João-Bahia), AR (Areia-Paraíba), CP (Chapadinha-Maranhão), NV (Novo Airão-Amazonas), IR (Iranduba-Amazonas), PF (Presidente Figueiredo-Amazonas), SV (Silves-Amazonas).

A matriz de distância geográfica construída a partir das coordenadas geográficas de coleta das populações apresenta valores de distância entre as populações que variam entre 21,4 km a 2934,4 km (Tabela 4). O teste de Mantel identificou correlação positiva e significativa ($r = 0,78$; p -valor $< 0,001$) entre a distância geográfica apresentada entre as populações e os valores de divergência genética (F_{st}), mostrando a ocorrência de isolamento por distância para as populações de cajazeira.

Tabela 4: Na diagonal superior a matriz par a par de distância geográfica calculada entre as populações coletadas e na diagonal inferior os valores de F_{st} calculados entre as populações. Os valores apresentados são de distância na unidade de quilômetros. As populações são representadas pelas siglas PD (Paudalho-PE), SM (São Lourenço da Mata-PE), MS (Mata de São João-Bahia), AR (Areia-Paraíba), CP (Chapadinha-Maranhão), NV (Novo Airão-Amazonas), IR (Iranduba-Amazonas), PF (Presidente Figueiredo-Amazonas), SV (Silves-Amazonas).

Populações	AR	CP	IR	MS	NV	PD	PF	SM	SV
AR		919.04	2741.17	656.84	2837.75	117.73	2749.53	137.37	2553.47
CP	0.46		1869.07	1133.84	1957.92	1016.66	1861.16	1037.87	1672.22
IR	0.51	0.62		2647.72	112.19	2816.34	137.16	2835.82	205.18
MS	0.21	0.44	0.50		2755.27	594.25	2689.56	593.78	2487.38
NV	0.47	0.49	0.23	0.44		2914.74	120.44	2934.39	285.75
PD	0.16	0.43	0.49	0.15	0.45		2829.20	21.36	2631.60
PF	0.41	0.44	0.19	0.39	0.16	0.38		2849.09	202.19
SM	0.17	0.43	0.49	0.15	0.45	0.06	0.39		2651.36
SV	0.39	0.43	0.21	0.37	0.17	0.36	0.08	0.37	

Para a obtenção dos parâmetros básicos de diversidade foram considerados dois cenários. O primeiro levou em consideração os grupos identificados na análise DAPC. O segundo cenário levou em consideração o local de coleta na definição das populações (Tabela 5). O grupo G2, que apresentou o maior número de alelos (3547) e a maior diversidade genética (0,2031), foi 37% mais diversificado geneticamente que o grupo G1 e 70% mais diversificado que o grupo G3, e apresentou 649 alelos a mais em relação ao grupo G1 e 1307 alelos em relação ao grupo G3. O resultado apresentado pelo grupo G2 está em concordância com a inferência realizada na análise de densidade da DAPC, onde o grupo G2 apresentou distribuição mais ampla em relação a função discriminante, seguido pelo grupo G1 e G2. No entanto, o grupo G2 apresentou o maior valor para o coeficiente de endogamia (f) (0,2280), quase o dobro em relação ao grupo G1. O grupo G3 apresentou o menor valor entre os grupos com coeficiente de endogamia de -0,8529, onde o valor negativo evidencia que a quantidade de locos em heterozigose observados nos indivíduos avaliados é a superior à esperada estatisticamente.

Tabela 5: Resultados da análise de diversidade genética realizada entre os grupos identificados na análise DAPC e as populações avaliadas de cajazeira (*Spondias monbin*). Foram analisados os números de indivíduos (N), o número total de alelos (A), a heterozigosidade observada (H_o), a heterozigosidade esperada (H_e) e o coeficiente de endogamia (f). Populações do Nordeste (NE): PD (Paudalho-PE), SM (São Lourenço da Mata-PE), MS (Mata de São João-Bahia), AR (Areia-Paraíba), CP (Chapadinha-Maranhão); e do Norte (N): NV (Novo Airão-Amazonas), IR (Iranduba-Amazonas), PF (Presidente Figueiredo-Amazonas), SV (Silves-Amazonas).

Grupos	N	A	H_o	H_e	f
G1	25	2898	0,1128	0,1279	0,1187
G2	14	3547	0,1568	0,2031	0,2280
G3	2	2240	0,1101	0,0594	-0,8529
Média	-	2895	0,1266	0,1301	-0,5062
Populações	N	A	H_o	H_e	f
PD	6	2624	0,1205	0,1184	-0,0177
SM	6	2624	0,1254	0,1178	-0,0649
AR	6	2492	0,0833	0,0990	0,1581
MS	5	2562	0,1213	0,1170	-0,0371
CP	3	2244	0,1100	0,0606	-0,8146
Média (NE)	-	2509	0,1121	0,1059	-0,0587
IR	2	2447	0,1790	0,1094	-0,6362
SV	3	2778	0,1873	0,2092	0,1045
NV	5	2806	0,1509	0,1650	0,0855
PF	5	2856	0,1505	0,1986	0,2421
Média (N)	-	2722	0,1669	0,1860	0,1025

Na análise de diversidade genética considerando o local de coleta das populações, foi possível identificar que o resultado foi próximo do obtido com os grupos da DAPC. As populações da região Norte, como Silves, Novo Airão e Presidente Figueiredo, do Amazonas, apresentaram, em média, valores superiores de diversidade genética ($A = 2722$; $H_o = 0,1669$; $H_e = 0,1860$) quando comparadas com as populações da região Nordeste ($A = 2509$; $H_o = 0,1121$; $H_e = 0,1056$) (Tabela 5). A população de Iranduba, embora seja uma população amazonense, apresentou valores divergentes com as demais populações do Estado para o número de alelos, diversidade genética e coeficiente de endogamia, apresentando heterozigosidade observada superior à esperada. As populações da Região Nordeste apresentaram valores próximos para os parâmetros de diversidade exceto pelo alto coeficiente de endogamia apresentado pela população de Areia e o baixo valor do coeficiente de endogamia apresentado pela população de Chapadinha. A população de Chapadinha apresentou o menor número de alelos e a menor diversidade genética entre todas as populações avaliadas no estudo. No entanto, apresentou o menor coeficiente de endogamia.

6. DISCUSSÃO

As árvores filogenéticas permitiram identificar que houve uma tendência de agrupamento das populações por Estado e que dentro dos Estados as populações não apresentam agrupamento por município de coleta. Essa questão fica evidente dentro do grupo formado pelas populações amazonenses e pelas populações pernambucanas. A matriz de F_{st} par a par construída com os grupos identificados na análise de DAPC mostra que os três grupos apresentam diferenciação genética bastante elevada ($F_{st} > 0,25$), segundo Hartl e Clark (1998), em todas as combinações entre os grupos, o que evidencia a segregação identificada nas análises de agrupamento. As análises de agrupamento deixam clara a divisão dos grupos entre a Região Norte do Brasil e da Região Nordeste.

A divergência genética, identificada na matriz par a par de F_{st} entre as nove populações estudadas, permitiu a identificação de menor divergência genética entre as populações provenientes de uma mesma região, tanto entre as populações provenientes do Nordeste quanto entre as populações do Norte do país. A menor divergência genética entre as populações geograficamente mais próximas indica a ocorrência de fluxo gênico entre as populações, uma vez que os valores de F_{st} apresentados entre as populações de uma mesma macrorregião indicam divergência genética de grau moderado (entre 0,05 e 0,15) a alto (entre 0,15 e 0,25). O contrário ocorre quando uma população proveniente da região Nordeste é comparada com uma população proveniente da região Norte do país. Em todos os cenários em que uma população da região Nordeste é comparada com uma população da região Norte, os valores obtidos indicam divergência genética muito alta ($F_{st} > 0,25$), com F_{st} variando de 0,36 a 0,50.

A população de Chapadinha, embora situada na Região Nordeste do Brasil, apresenta valores de F_{st} semelhantes quando contrastada com outras populações da região Nordeste e com populações da região Norte, apresentando o maior valor de divergência genética quando comparada com a população de Iranduba (0,62). Uma possível explicação para os valores apresentados pela população é a distância geográfica que a população apresenta em relação às demais populações, com distância média de 1027 km em relação às demais populações da região Nordeste e 1840 km das populações da região Norte. De acordo com o teste de Mantel, a distância geográfica está significativamente correlacionada ($r = 0,78$) com a divergência genética entre as populações. Dessa forma, a população maranhense, por apresentar valores médios de distância genética de mais de 1000 km entre as populações do Norte e do Nordeste, apresenta o comportamento de isolamento por distância, estando altamente diferenciada em relação às demais populações e formando um terceiro grupo (G3) nas análises de agrupamento realizadas. Esta é a única população amostrada no bioma Cerrado, o que pode também ter levado a adaptações e diferenciações pelos processos evolutivos de seleção natural ou deriva genética.

Na AMOVA foi constatado que a maior fonte de variação genética é encontrada entre os indivíduos dentro das populações e dentro dos grupos de DAPC, 61,4% e 59,3%, respectivamente. A variação genética intrapopulacional mais alta que a variação genética interpopulacional é esperada em espécies perenes alógamas, uma vez que a maior parte da variação está contida nos indivíduos e estas espécies apresentam mecanismos que beneficiam o fluxo gênico entre populações (Hamrick et al., 1992; Raspé e Acquemart, 1998; Paschoa et al., 2017, Francisconi et al., 2021). A cajazeira apresenta reprodução assexuada (vegetativa) e sexuada. É uma espécie alógama e depende de polinização cruzada para maior sucesso reprodutivo, apresentando dicogamia protândrica e alto valor para a razão pólen/óvulo (Souza, 2002; Ramos, 2009). A espécie tem os frutos como estrutura de valor econômico agregado. Em estudos realizados com populações do Nordeste do Brasil, foi identificado que a comercialização dos frutos foi o principal mecanismo atuante no fluxo gênico entre populações distantes (Silva, 2009a). A dispersão dos frutos em longas distâncias e a possibilidade de propagação vegetativa são fatores que favorecem a maior concentração de variação genética dentro de populações de espécies arbóreas (Hamrick et al., 1992).

Embora a variação intrapopulacional tenha sido a maior fonte de variação detectada na AMOVA, a porcentagem da variação de fonte interpopulacional foi também bastante expressiva, com os valores 38,5% e 40,7% entre populações e entre grupos DAPC, respectivamente. Os valores observados na estatística F indicam divergência genética significativa entre os grupos DAPC e as populações coletadas. Os resultados obtidos na AMOVA são semelhantes aos identificados por Magalhães et al. (2013) que analisaram populações de cajazeira do Amazonas a partir de marcadores AFLP e identificaram 47,2% da variação entre populações e 52,8% dentro das populações estudadas. Em um estudo realizado com a espécie *Mangifera indica*, Surapaneni et al. (2013) observaram que 62,25% da variação genética tinha origem na variação entre os indivíduos e Santos et al. (2019), em um estudo realizado com *Anacardium occidentale*, observaram que 78% da variação tinha origem nos indivíduos.

Os resultados obtidos para diversidade genética (H_E) no presente estudo demonstraram que sete das nove populações avaliadas apresentaram valores de H_E superiores aos encontrados por Hamrick et al. (1992) em um estudo realizado com seis espécies de angiospermas arbóreas tropicais, identificando o valor de 0,106 para a média de diversidade genética apresentada pelas espécies. Duas das populações de cajazeira (Chapadinha e Areia) apresentaram valores baixos de diversidade genética. A população de Chapadinha, no entanto, apresentou valor para a heterozigose observada (H_O) superior à esperada (H_E), o que indica a recente introdução de alelos na população que fez com que a quantidade observada de locos em heterozigose fosse superior à esperada. Areia é uma cidade paraibana localizada em região abrangida pelo bioma de Caatinga, a única população proveniente deste bioma no estudo realizado. Além de ter apresentado valor de diversidade genética abaixo da

média esperada, a população apresenta o maior coeficiente de endogamia (f) entre as populações do Nordeste, 0,1581, a única com coeficiente positivo, ou seja, a população apresenta endogamia. Esta população possui grande potencial para fornecimento de genes relacionados ao estresse hídrico. Embora a cidade de Areia, na Paraíba, seja considerada um brejo de altitude, onde há condições edafoclimáticas semelhantes às encontradas no bioma de Mata Atlântica (Machado et al., 2012), espera-se que essa população ainda sofra grande influência do bioma Caatinga. Sendo assim, consideramos de extrema importância a manutenção da diversidade nesta população, aumentando as chances de obtenção do maior número possível de alelos para genes de adaptação às condições climáticas adversas do bioma.

No presente estudo foi possível observar que as populações da região Norte apresentaram valores maiores que os encontrados por Hamrick et al. (1992) para diversidade genética (H_E) em todas as populações coletadas, com valor mínimo de 0,109 e máximo de 0,209. Os valores apontam a região Norte como detentora dos maiores valores para diversidade genética identificados no estudo. No entanto, com exceção da população de Iranduba, todas as outras populações amazonenses demonstraram valores positivos para o coeficiente de endogamia. As políticas de preservação da Floresta Amazônica vêm sendo enfraquecidas a cada ano com o aumento de 9,5% na taxa de desmatamento entre os anos de 2019 e 2020 (INPE, 2020), o que aumenta o estado de alerta sobre as populações da região, que embora possuam maior diversidade genética apresentam maior coeficiente de endogamia e risco de erosão genética.

Em um estudo realizado por Silva et al. (2009a) com quatro populações de Pernambuco a partir de 22 alelos de nove locos de isoenzimas, os valores para diversidade genética variaram de 0,530 a 0,574. Silva et al. (2017b) realizaram um estudo com três populações do norte do Estado do Mato Grosso a partir de 14 *primers* ISSR e identificaram valores de diversidade genética de 0,20 a 0,24. Lins Neto et al. (2013), em um estudo realizado com *Spondias tuberosa* Arruda, a partir de 11 marcadores ISSR, observaram valores de diversidade genética, que variaram de 0,27 a 0,34. Este é o primeiro estudo realizado de diversidade genética com uma espécie do gênero *Spondias* com a utilização de marcadores SNPs. Os valores para diversidade genética observados no presente estudo são inferiores a aqueles identificados anteriormente para a espécie em estudo. O fato pode estar relacionado ao grande número locos avaliados (2003 marcadores SNPs), o que torna os resultados mais precisos. Fischer et al. (2017) em um estudo comparativo entre marcadores SNPs e microssatélites identificaram, em uma mesma amostra, valores de médios de diversidade genética (H_E) de 0,378 quando utilizaram marcadores microssatélites e 0,148 na utilização de SNPs. Já Lemopoulos et al. (2018), também em um estudo comparativo, identificaram o valor médio de diversidade genética de 0,597 com microssatélites e 0,11 quando utilizaram SNPs.

De acordo com Fischer et al. (2017), marcadores do tipo SNPs, derivados da tecnologia de NGS, se mostraram superiores a marcadores do tipo microssatélite (SSR) na realização de estudos genômicos para análises de diversidade genética entre e dentro de populações, sendo que alguns milhares de SNPs são suficientes para realização de análises robustas e não enviesadas de diversidade genética. Lemopoulos et al. (2018), em estudo comparativo entre SNPs e microssatélites na quantificação de diversidade genética e parentesco, concluíram que marcadores do tipo SNPs (em grandes quantidades) são mais informativos e fornecem maior resolução em estudos realizados com populações pequenas e isoladas. Gärke et al. (2012) realizaram um estudo comparativo entre SNPs e microssatélites na identificação de estruturação genética entre populações e verificaram que 70 marcadores SNPs tiveram o mesmo desempenho que 29 marcadores SSR em análises discriminantes baseadas em PCA. Além disso, os autores concluíram que um conjunto com mais de 250 marcadores SNPs apresentou maior desempenho do que um conjunto de 29 marcadores SSR em análises de agrupamento e na identificação de mistura genética entre populações.

O Brasil é o país com o maior número de acessos de *S. mombin* mantidos em bancos de germoplasma, todos mantidos *ex situ* (Santana, 2010). A EMEPA (Empresa Estadual de Pesquisa Agropecuária da Paraíba) possui 21 acessos da espécie coletados nos Estados da Paraíba e Pernambuco (Bosco, 2000, citado por Santana, 2010). O Instituto Agrônomo de Pernambuco possui 33 acessos coletados na Zona da Mata e na região metropolitana do Estado (Bezerra et al., 2011). De acordo com as informações disponíveis no banco de dados Alelo (<http://alelobag.cenargen.embrapa.br/AleloConsultas>), a Embrapa (Empresa Brasileira de Pesquisa Agropecuária) possui três bancos germoplasma ativo *ex situ* da espécie: a Embrapa Meio-Norte com 38 acessos (sem informação quanto ao local de coleta), a Embrapa Amazônia Oriental com 35 acessos coletados no Estado do Pará (municípios de Belém, Igarapé-Açu e Tome-Açú) e a Embrapa Agroindústria Tropical, que possui sete acessos coletados no Estado do Ceará e Piauí. É possível notar que nenhum dos bancos de germoplasma existentes possuem acessos do Estado do Amazonas, que apresentou os maiores valores para a diversidade genética no presente estudo. É urgente a necessidade da criação de um banco de germoplasma para conservação da diversidade genética apresentada pelas populações do Estado do Amazonas, sendo mais indicado a criação de um banco de germoplasma *in situ*. A preservação *in situ* propicia a preservação da fauna e da flora na área a ser preservada, permite que a espécie coevolua em habitat sob as mesmas pressões de seleção as quais são submetidos os seus polinizadores e suas pragas (Scariot e Sevilha, 2007; Ramos et al., 2015), além de dispensar os custos envolvidos com coleta, os custos para implementação de um banco de germoplasma *ex situ* e os custos com manejo. A espécie estudada apresenta estruturação genética em relação a distância geográfica, tendo alta correlação entre distância geográfica e diferenciação genética. As populações amazonenses apresentaram elevada diferenciação genética em relação às populações coletadas na Região Nordeste e vice-versa,

o que indica a presença de alelos nestas populações que são ausentes em outras geograficamente mais distantes. É indicado que os bancos de germoplasma ativos da espécie já implantados introduzam em seu acervo acessos provenientes de outras macrorregiões do país com finalidade de propiciar o aumento da diversidade genética da coleção e maximizar as chances de preservação da maior diversidade genética possível da espécie.

7. CONCLUSÕES

A espécie *S. mombin* apresenta padrão de estruturação genética por isolamento por distância geográfica, com fluxo gênico limitado entre as regiões Norte e Nordeste, porém presente dentro das populações de uma mesma região.

A cajazeira apresenta maior diversidade genética nas populações da região Norte em comparação com a região com Nordeste. Por outro lado, as populações da região Norte apresentaram valores maiores para o coeficiente de endogamia (f) quando comparadas a populações da região Nordeste.

Do ponto de vista da conservação, a espécie possui potencial para a conservação *in situ* com populações que apresentam níveis razoáveis de diversidade genética. Curadores de bancos de germoplasma *ex situ* da espécie devem considerar a inserção no acervo de acessos provenientes de diferentes regiões do país, principalmente do estado do Amazonas, que apresentaram maiores índices de diversidade genética, aumentando a eficiência na conservação da diversidade genética da espécie.

O estudo realizado a partir de uma ampla amostragem da distribuição geográfica da espécie no país e uma ampla cobertura da diversidade genômica contribuiu de maneira significativa para o entendimento do atual estado de conservação de populações naturais da espécie e possibilitará que melhoristas tracem estratégias para a sua conservação e para o melhoramento genético.

REFERÊNCIAS

ALELO CONSULTAS. Disponível em: <http://alelobag.cenargen.embrapa.br/AleloConsultas>. Acessado em 25 dezembro 2020.

ALVES-PEREIRA, A.; NOVELLO, M.; DEQUIGIOVANNI, G.; PINHEIRO, J. B.; BRANCALION, P. H. S.; VEASEY, E. A.; CLEMENT, C. R.; DE SOUZA, A. P.; ZUCCHI, M. I. genomic diversity of three Brazilian native food crops based on double-digest restriction site-associated DNA sequencing. **Tropical Plant Biology**, v. 12, p. 268-281, 2019.

BEZERRA, J. E. F.; LEDERMAN, I. E.; SILVA JUNIOR, J. F.; LIRA JUNIOR, J. S.; MOURA, R. J. M.; SILVA, M. F. F. Banco ativo de germoplasma de fruteiras nativas e exóticas do Instituto Agrônomo de Pernambuco. In: WORKSHOP DE CURADORES DE GERMOPLASMA DO BRASIL, 2011, Campinas, SP. Anais... Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2011.

BOUGEARD, S.; DRAY, S. Supervised multiblock analysis in R with the ade4 package. **Journal of Statistical Software**, v. 86, n. 1, p. 1–17, 2018.

BROUSSEAU, L.; FINE, P. V. A.; DREYER, E.; VENDRAMIN, G. G.; SCOTTI, I. Genomics of microgeographic adaptation in the hyperdominant Amazonian tree *Eperua falcata* Aubl. (Fabaceae). **bioRxiv**, p. 312843, 2019.

CATCHEN, J. M.; AMORES, A.; HOHENLOHE, P.; CRESKO, W.; POSTLETHWAIT, J. H. Stacks: Building and genotyping loci de novo from short-read sequences. **G3: Genes, Genomes, Genetics**, v. 1, n. 3, p. 171–182, 2011.

CHELSEL, D.; DUFOUR, A.; THIOULOUSE, J. The ade4 Package – I: One-Table Methods. **R News**, v. 4, n. 1, p. 5–10, 2004. <https://cran.r-project.org/doc/Rnews/>.

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

CORNEJO, O. E.; YEE, M.-C.; DOMINGUEZ, V.; ANDREWS, M.; SOCKELL, A.; STRANDBERG, E.; LIVINGSTONE, D.; STACK, C.; ROMERO, A.; UMAHARAN, P.; ROYAERT, S.; TAWARI, N. R.; NG, P.; GUTIERREZ, O.; PHILLIPS, W.; MOCKAITIS, K.; BUSTAMANTE, C. D.; MOTAMAYOR, J. C. Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. **Communications Biology**, v. 1, n. 1, p. 1–12, 2018.

CORTINOVIS, G.; FRASCARELLI, G.; DI VITTORI, V.; PAPA, R. Current state and perspectives in population genomics of the common bean. **Plants**, v. 9, n. 3, 2020.

DRAY S.; DUFOUR A. The ade4 Package: Implementing the duality diagram for ecologists. **Journal of Statistical Software**, v. 22, n. 4, p. 1–20, 2007a.

DRAY S.; DUFOUR A.; CHESSEL D. The ade4 Package – II: Two-Table and K-Table Methods. **R News**, v. 7, n. 2, p. 47–52, 2007b. <https://cran.r-project.org/doc/Rnews/>.

EKBLOM, R.; GALINDO, J. Applications of next generation sequencing in molecular ecology of non-model organisms. **Heredity**, v. 107, n. 1, p. 1–15, 2011.

ELSHIRE, R. J.; GLAUBITZ, Q. S.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, v. 6, n. 5, p. 1–10, 2011.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 2 ed. Brasília: Embrapa-Cenargen, 1998. 220 p.

FISCHER, M. C.; RELLSTAB, C.; LEUZINGER, M.; ROUMET, M.; GUGERLI, F.; SHIMIZU, K. K.; HOLDEREGGER, R.; WIDMER, A. Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. **BMC Genomics**, v. 18, n. 1, p. 1–15, 2017.

FONSECA, N.; MACHADO, C. F.; SILVA JUNIOR, J. F.; CARVALHO, R. S.; RITZINGER, R.; ALVES, R. M.; MAIA, M. C. C. Umbu: cajá e espécies afins: *Spondias spp.* Buenos Aires: IICA/PROCISUR, 2017, 29 p.

FRANCISCONI, A. F.; PALHARES, R.; CHARLES, A.; DEQUIGIOVANNI, G.; CARVALHO, I. A. S.; VEASEY, E. A. Genetic structure and diversity identify incipient [*Caryocar villosum* (Aubl.) pers.] domestication of Piquiá along the lower Tapajós River, Brazilian Amazonia. **Genetic Resources and Crop Evolution**, 2021.

GÄRKE, C.; YTOURNEL, F.; BED'HOM, B.; GUT, I.; LATHROP, M.; WEIGEND, S.; SIMIANER, H. Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. **Animal Genetics**, v. 43, n. 4, p. 419–428, 2012.

GILAD, Y.; PRITCHARD, J. K.; THORNTON, K. Characterizing natural variation using next-generation sequencing technologies. **Trends Genetics**, v. 25, p. 463–471, 2009.

GOUDET, J.; JOMBART, T. hierfstat: Estimation and Tests of Hierarchical F-Statistics. R package version 0.5-7, 2020. <https://CRAN.R-project.org/package=hierfstat>

HAMRICK, J. L.; GODT, M. J. W.; SHERMAN-BROYLES, S. L. Factors influencing levels of genetic diversity in woody plant species. **New Forests**, v. 6, p. 95-124, 1992.

HARTL, D.; CLARK, A. **Principles of Population Genetics**. Oxford University Press, 1998. 519 p.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z. X.; LIU, H. K.; LI, Z. Genotyping-by-sequencing (GBS), An ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Frontiers in Plant Science**, v. 5, n. SEP, p. 1–8, 2014.

INGLIS, P. W.; PAPPAS, M.C.R.; RESENDE, L.V.; GRATTAPAGLIA, D. Fast and inexpensive protocols for consistent extraction of high-quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. **PLoS ONE**, v. 13, n. 10, p. 1–14, 2018.

INPE - INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Projeto PRODES: Monitoramento da Floresta Amazônica Brasileira por Satélite, 2019. <http://www.obt.inpe.br/prodes/>.

JOMBART, T. Adegenet: an R package for the multivariate analysis of genetic markers. **Bioinformatics**, v. 24, p. 1403-1405, 2008.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. **BMC Genetics**, v. 11, n. 1, p. 94, 2010.

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

KAMVAR, Z. N.; TABIMA, J. F.; GRÜNWALD, N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. **PeerJ**, v. 2, n. e281, 2014.

KIM, C.; GUO, H.; KONG, W.; CHANDNANI, R.; SHUANG, L. S.; PATERSON, A. H. Application of genotyping by sequencing technology to a variety of crop breeding programs. **Plant Science**, v. 242, p. 14–22, 2016.

LEMOPOULOS, A.; PROKKOLA, J. M.; UUSI-HEIKKILÄ, S.; VASEMÄGI, A.; HUUSKO, A.; HYVÄRINEN, P.; KOLJONEN, M. L.; KOSKINIEMI, J.; VAINIKKA, A. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. **Ecology and Evolution**, v. 9, n. 4, p. 2106–2120, 2019.

LEÓN A. C.; DUQUE A. P.; HAROS, B. R. **Jacotes, jobos, abales o ciruelas mexicanas**. Texcoco: Universidad Autónoma Chapingo, 2012. 128 p.

LIMA, A.T.B; SOUZA, V.A.B; GOMES, R.L.F; LIMA, P.S.C. Molecular characterization of cajá, *Spondias mombin* (Anacardiaceae), by RAPD markers. **Genetics and Molecular Research**, v. 10, p. 2893-2904, 2011.

LINS NETO, E. M. F.; OLIVEIRA, I. F.; BRITTO, F. B.; ALBUQUERQUE, U. P. Traditional knowledge, genetic and morphological diversity in populations of *Spondias tuberosa* Arruda (Anacardiaceae). **Genetic Resources and Crop Evolution**, v. 60, n. 4, p. 1389–1406, 2013.

LORENZI, H. **Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil**. Nova Odessa: Plantarum, 1992. 384 p.

LOZANO, N.B. Desarrollo y anatomía del fruto del jobo (*Spondias mombin* L.). **Caldasia**, v. 14, n. 68, p. 465- 490, 1986.

MACHADO, W. DE J.; PRATA, A. P. DO N. P.; DE MELLO, A. A. Floristic composition in areas of Caatinga and Brejo de altitude in Sergipe state, Brazil. **Check List**, v. 8, n. 6, p. 1089–1101, 2012.

MAGALHÃES, M. A.; LOPES, M. T. G.; COELHO, L. M.; COSTA NETO, P. Q.; BENTES, J. L. S. Genetic diversity within and among populations of Cajazeira. **Revista de Ciências Agrárias - Amazon Journal of Agricultural and Environmental Sciences**, v. 56, p. 61–67, 2013.

METZKER, M. L. Sequencing technologies – the next generation. **Nature Reviews Genetics**, v. 11, p. 31–46, 2010.

MITCHELL, J. D.; DALY, D. C. A revision of *Spondias* L. (Anacardiaceae) in the neotropics. **PhytoKeys**, v. 55, n. 1, p. 1–92, 2015.

NADEEM, M. A.; NAWAZ, M. A.; SHAHID, M. Q.; DOĞAN, Y.; COMERTPAY, G.; YILDIZ, M.; HATIPOĞLU, R.; AHMAD, F.; ALSALEH, A.; LABHANE, N.; ÖZKAN, H.; CHUNG, G.; BALOCH, F. S. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biotechnology and Biotechnological Equipment**, v. 32, n. 2, p. 261–285, 2018.

NEI, M. Genetic distances between populations. **American Naturalist**, v. 106, p. 283-292, 1972.

O'NEILL, G. A.; DAWSON, I.; SOTELO-MONTES, C.; GUARINO, L.; GUARIGUATA, M.; CURRENT, D.; WEBER, J. C. Strategies for genetic conservation of trees in the Peruvian Amazon. **Biodiversity and Conservation**, v. 10, n. 6, p. 837–850, 2001.

PADGHAM, M.; SUMNER, M. D. geodist: Fast, Dependency-Free Geodesic Distance Calculations. R package version 0.0.6, 2020. <https://CRAN.R-project.org/package=geodist>

PARADIS, E.; SCHLIEP, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. **Bioinformatics**, v. 35, p. 526-528, 2019.

PASCHOA, R. P.; CHRIST, J. A.; VALENTE, C. S.; FERREIRA, M. F. DA S.; DE MIRANDA, F. D.; GARBIN, M. L.; CARRIJO, T. T. Genetic diversity of populations of the dioecious *Myrsine coriacea* (Primulaceae) in the Atlantic Forest. **Acta Botanica Brasilica**, v. 32, n. 3, p. 376–385, 2018.

POLAND, J. A.; BROWN, P. J.; SORRELLS, M. E.; JANNINK, J. A. et al. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. **PLoS ONE**, v. 7, n. 2, 2012.

POLAND, J. A.; RIFE, T. W. Genotyping-by-sequencing for plant breeding and genetics. **The Plant Genome Journal**, v. 5, n. 3, p. 92, 2012.

RAMOS, M. C. **Ecologia da polinização de taperebá (*Spondias mombin* L., Anacardiaceae) em área de floresta secundária no município de Santo Antônio do Tauá, Pará, Brasil**. Dissertação (Mestrado em Zoologia) – Universidade Federal do Pará, p. 64, 2009.

RAMOS, S. R. R.; SILVA JÚNIOR, J. F.; BUSTAMANTE, P. G.; MOTA D. M. M.; QUEIRÓZ, M. A. Preservação in situ, conservação ex situ e uso dos recursos genéticos por agricultores e comunidades tradicionais no Brasil. In: VEIGA, R. F. A.; QUEIRÓZ, M. A. **Recursos Fitogenéticos**, Viçosa: Editora UFV, 2015. p. 269-274.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018. Disponível em: <http://www.R-project.org/>. Acesso em 31/08/2019.

RASPÉ, O.; ACQUEMART, A. L. J. Allozyme diversity and genetic structure of European populations of *Sorbus aucuparia* L. (Rosaceae: Maloideae). **Heredity**, v. 81, n. 5, p. 537–545, 1998.

RIVERO, S.; ALMEIDA, O. Pecuária e desmatamento: uma análise das principais causas diretas do desmatamento da Amazônia Nova Economia. **Nova Economia**, v. 19, n. 1, p. 41–66, 2009.

SACRAMENTO, C. K.; SOUZA, F. X. **Cajá** In: Fruticultura tropical: espécie regionais e exóticas. Brasília, DF: Embrapa Informação Tecnológica, 2009. 506 p.

SANTANA, F. F. **Caracterização de genótipos de cajazeiras**. Tese (Doutorado em Produção Vegetal) – Faculdade de Ciências Agrárias e Veterinárias – Universidade Estadual Paulista, p. 97, 2010.

SANTOS, J. O.; MAYO, S. J.; BITTENCOURT, C. B.; DE ANDRADE, I. M. Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí, Brazil. **Plant Systematics and Evolution**, v. 305, n. 10, p. 913–924, 2019.

SCARIOT, A. O.; SEVILHA, A. C. Conservação in situ de recursos genéticos vegetais. In: NASS, L. L. **Recursos genéticos vegetais**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 474-509.

SILVA, A.Q.; SILVA, A. Observações morfológicas e fisiológicas sobre *Spondias tuberosa* Arr. Câm. In: CONGRESSO NACIONAL DE BOTÂNICA, 25., 1974, Mossoró, Anais... Recife: Sociedade Botânica do Brasil, 1974. p. 5-15.

SILVA, B. M.; ROSSI, A. A. B.; DARDENGO, J. F. E.; TIAGO P. V.; SILVEIRA, G. F.; SOUZA, S. A. M. S. Genetic divergences between *Spondias mombin* (Anacardiaceae) genotypes found through morphological traits. **Revista de Biología Tropical**, v. 65, n. 4, p. 1337, 2017a.

SILVA, B. M.; ROSSI, A. A. B.; TIAGO, A. V.; DARDENGO, J. F. E.; SOUZA, S. A. Genetic diversity of cajazeira (*Spondias mombin* L.) in three geographic regions. **Genetics and Molecular Research**, v. 16, n. 1, 2017b.

SILVA, C. A.; COSTA, P. R.; DETONI, J. L.; ALEXANDRE, R. S.; CRUZ, C. D.; SCHMILDT, O.; SCHMILDT, E. R. Divergência genética entre acessos de cajazinho (*Spondias mombin* L.) no norte do Espírito Santo. **Revista Ceres**, v. 61, n. 3, p. 362–369, 2014.

SILVA, E. F. DA; MARTINS, L. S. S.; OLIVEIRA, V. R. DE. Diversity and genetic structure in cajá tree (*Spondias mombin* L.) populations in northeastern Brazil. **Revista Brasileira de Fruticultura**, v. 31, n. 1, p. 171–181, 2009a.

SILVA, C. J. D. **Caracterização genética de cajazeira (*Spondias mombin* L.) (Anacardiaceae) por meio de marcadores moleculares**. Dissertação (Mestrado em Melhoramento Genético de Plantas), Universidade Federal Rural de Pernambuco, 55p, 2009b.

SILVA, J. F.; BEZERRA, J. E. F.; LEDERMAN, I. E.; ALVES, M. A.; MELO NETO, M. L. et al. Collecting, ex situ conservation and characterization of “cajá-umbu” (*Spondias mombin* x *Spondias tuberosa*) germplasm in Pernambuco State, Brazil. **Genetic Resources and Crop Evolution**, v. 51, n. 4, p. 343–349, 2004.

SINGH, U.; DEB, R.; ALYETHODI, R. R.; ALEX, R.; KUMAR, S.; CHAKRABORTY, S.; DHAMA, K.; SHARMA, A. Molecular markers and their applications in cattle genetic research: A review. **Biomarkers and Genomic Medicine**, v. 6, n. 2, p. 49–58, 2014.

SOUZA, F. X.; FRANCA, M. J. P. Avaliação da antese em cajazeira. In: CONGRESSO NACIONAL DE BOTÂNICA, 50., 1999, Blumenau. Programa e Resumos... Blumenau: Sociedade Botânica do Brasil, 1999. p. 207.

SOUZA, F. X. DE; INNECCO, R.; ROSSETTI, G. Influência de porta-enxerto e de método de enxertia no pagamento de enxertos de cajazeira. **Agrotropica**, v. 14, n. 3, p. 85-90, 2002.

SOUZA, L. M.; DOS SANTOS, L. H. B.; ROSA, J. R. B. F.; DA SILVA, C. C.; MANTELLO, C. C.; CONSON, A. R. O.; SCALOPPI, E. J.; FIALHO, J. DE F.; DE MORAES, M. L. T.; GONÇALVES, P. DE S.; MARGARIDO, G. R. A.; GARCIA, A. A. F.; LE GUEN, V.; DE SOUZA, A. P. Linkage disequilibrium and population structure in wild and cultivated populations of rubber tree (*Hevea brasiliensis*). **Frontiers in Plant Science**, v. 9, n. July, p. 1–10, 2018.

SURAPANENI, M.; VEMIREDDY, L. R.; BEGUM, H.; PURUSHOTHAM REDDY, B.; NEETASRI, C.; NAGARAJU, J.; ANWAR, S. Y.; SIDDIQ, E. A. Population structure and genetic analysis of different utility types of mango (*Mangifera indica* L.) germplasm of Andhra Pradesh state of India using microsatellite markers. **Plant Systematics and Evolution**, v. 299, n. 7, p. 1215–1229, 2013.

THIOULOUSE J.; DRAY S.; DUFOUR A.; SIBERCHICOT A.; JOMBART T.; PAVOINE S. Multivariate Analysis of Ecological Data with ade4. **Springer**, 2018.

TIBURSKI, J. H.; ROSENTHAL, A.; DELIZA, R.; GODOY, R. L. O.; PACHECO S. Nutritional properties of yellow mombin (*Spondias mombin* L.) pulp. **Food Research International**, v. 44, n. 7, p. 2326–2331, 2011.

TSAI, W. C.; DIEVART, A.; HSU, C. C.; HSIAO, Y. Y.; CHIOU, S. Y. HUANG, H.; CHEN, H. H. Post genomics era for orchid research. **Botanical Studies**, v. 58, n. 1, 2017.

VIEIRA, I. C. G.; TOLEDO, P. M.; SILVA, J. M. C.; HIGUCHI, H. Deforestation and threats to the biodiversity of Amazonia. **Brazilian Journal of Biology**, v. 68, n. 4 SUPPL., p. 949–956, 2008.

WEI, T.; SIMKO, V. R package "corrplot": Visualization of a Correlation Matrix (Version 0.84), 2017. Available from <https://github.com/taiyun/corrplot>

WILLING, E.-M.; DREYER, C.; VAN OOSTERHOUT, C. Estimates of genetic differentiation measured by F_{st} do not necessarily require large sample sizes when using many SNPs markers. **PLoS ONE**, v. 7, n. 8, e42649, 2012.

ZORTÉA, K. É. M.; ROSSI, A. A. B.; BISPO, R. B.; ROCHA, V. D.; HOOGERHEIDE, E. S. S. Meiotic behavior and pollen viability of *Spondias mombin* L.: Native fruit species of the amazon. **Floresta e Ambiente**, v. 26, n. 3, p. 1–8, 2019.