

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

LAÍS SOUZA SANTOS

MARCADORES MICROSSATÉLITES PARA OS GENOMAS NUCLEAR E
ORGANELAR DE *Amburana cearensis* (Allemão) A.C. Smith

GOIÂNIA, GO

2021



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese

2. Nome completo do autor

LAÍS SOUZA SANTOS

3. Título do trabalho

MARCADORES MICROSSATÉLITES PARA OS GENOMAS NUCLEAR E ORGANELAR DE
Amburana cearensis (Allemão) A.C. Smith

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Thannya Nascimento Soares, Professora do Magistério Superior**, em 19/10/2021, às 12:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **LAÍS SOUZA SANTOS, Discente**, em 21/10/2021, às 21:42, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2418050** e o código CRC **54B23A0F**.

LAÍS SOUZA SANTOS

MARCADORES MICROSSATÉLITES PARA OS GENOMAS NUCLEAR E
ORGANELAR DE *Amburana cearensis* (Allemão) A.C. Smith

Dissertação apresentada ao Programa de Pós- Graduação em Genética e Biologia Molecular, da Universidade Federal de Goiás, como requisito à obtenção do título de mestre em Genética e Biologia Molecular.

Orientadora: Prof^a Dr.^a Thannya Nascimento Soares

Coorientador: Dr. Vidal de Freitas Mansano

Área de concentração: Genética e Biologia Molecular

Goiânia – GO
2021

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

SANTOS, LAIS SOUZA
MARCADORES MICROSSATÉLITES PARA OS GENOMAS
NUCLEAR E ORGANELAR DE *Amburana cearensis* (Allemão) A.C.
Smith [manuscrito] / LAIS SOUZA SANTOS. - 2021.
73 f.

Orientador: Prof. THANNYA NASCIMENTO SOARES; co orientador VIDAL MANSANO DE FREITAS.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e Biologia Molecular, Goiânia, 2021.

Bibliografia. Apêndice.

Inclui tabelas, lista de figuras, lista de tabelas.

1. *Amburana cearensis*. 2. Microsatellite markers. 3. NGS. I. SOARES, THANNYA NASCIMENTO, orient. II. Título.

CDU 575



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 96 da sessão de Defesa de Dissertação de **Lais Souza Santos**, que confere o título de Mestre(a) em **Genética e Biologia Molecular**, na área de concentração em **Genética e Biologia Molecular**.

Ao/s trinta dias do mês de agosto de dois mil e vinte e um, a partir da(s) 14h00, por videoconferência, seguindo portaria CAPES no. 36 de 16 de março de 2020 e recomendação da UFG, realizou-se a sessão pública de Defesa de Dissertação intitulada “**MARCADORES MICROSSATÉLITES PARA OS GENOMAS NUCLEAR E ORGANELAR DE *Amburana cearensis* (Allemão) A.C. Smith**”. Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) Doutor(a) **Thannya Nascimento Soares (ICB/UFG)** com a participação dos demais membros da Banca Examinadora: Professor(a) Doutor(a) **Rosana Pereira Vianello (EMBRAPA)**, membro titular externo; Professor(a) Doutor(a) **Mariana Pires de Campos Telles (ICB/UFG)**, membro titular interno. Durante a arguição os membros da banca não sugeriu alteração de título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido(a) o(a) candidato(a) **aprovado(a)** pelos seus membros. Proclamados os resultados pelo(a) Professor(a) Doutor(a) **Thannya Nascimento Soares**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao(s) **trinta dias do mês de agosto de dois mil e vinte e um**.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Thannya Nascimento Soares, Professora do Magistério Superior**, em 30/08/2021, às 16:56, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mariana Pires De Campos Telles, Professor do Magistério Superior**, em 30/08/2021, às 16:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rosana Pereira Vianello, Usuário Externo**, em 30/08/2021, às 19:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2313529** e o código CRC **469CAE7B**.

Referência: Processo nº 23070.043286/2021-22

SEI nº 2313529

Agradecimentos

Gostaria de agradecer a todos que direta ou indiretamente me acompanharam neste processo e permitiram a realização de mais um sonho, sempre lembrarei com muito carinho de cada suporte e de todas as mãos que foram estendidas para que o meu sonho se tornasse concreto.

Todos os agradecimentos do mundo ao maior amor da minha vida, minha mãe, Maria da Conceição. A pessoa mais incrível que conheço. Obrigada por me permitir sonhar, por segurar minha mão nos momentos mais difíceis, por sempre acreditar em mim, pela segurança, pelo colo, pela paciência, pelos conselhos, por ser você! Sem você nada seria possível, obrigada por sonhar comigo. Te amo!

À minha irmã, por sempre estar presente. Que bom ter tido seu apoio, suas palavras de conforto, seu chamego e sua admiração.

Aos meus amigos por toda torcida, conselhos e dengos, por comemorarem comigo cada pequena conquista, por entenderem minha ausência e por estarem sempre presentes. Um agradecimento especial aos que sempre estiveram ao meu lado: Mayara, Lara, Ítalo, Denner e Vinicius, vocês foram essenciais para essa conquista.

À minha orientadora Prof.^a Dr.^a Thannya Nascimento, obrigada pelo privilégio de ter sido orientada por você! Por todo amparo, por entender minhas necessidades, pelo cuidado, pelos conselhos, pela calma, pelos puxões de orelha e pelas conversas descontraídas. Costumo dizer que você foi a melhor parte do meu mestrado, obrigada por me permitir crescer! Ao meu coorientador, Dr. Vidal Mansano, pela parceria, disponibilidade e por confiar no meu trabalho!

Um agradecimento à Dr.^a Cíntia Targueta e ao Dr. Rhewter Nunes por todo conhecimento compartilhado e pela ajuda e paciência em algumas etapas do meu trabalho. Um agradecimento especial à Dr.^a Adriana Antunes, que foi essencial para que esse trabalho fosse concretizado. Obrigada por sempre ser tão solícita, sem você não seria possível.

Agradeço aos professores do Programa de Genética e Biologia Molecular por todos os ensinamentos. Em especial à Prof.^a Dr.^a Mariana Telles e a Prof.^a Dr.^a Daniela Melo, duas profissionais que me acompanham desde a graduação, que admiro e tenho muito carinho. Obrigada por todas as experiências, nosso convívio contribuiu muito para a minha formação.

Agradeço aos amigos e colegas do Laboratório de Genética e Biodiversidade – LGBio por todo convívio, aprendizado e por contribuírem com o meu desenvolvimento profissional. Um agradecimento especial ao Victor Felipe por trazer leveza ao meu mestrado e sempre deixar

as coisas mais fáceis.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa. Ao Programa de Apoio a Núcleos Emergentes (PRONEM), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), aos Institutos Nacionais de Ciência e Tecnologia (INCT) e à Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) pelo financiamento deste trabalho.

Sumário

Agradecimentos	7
Lista de tabelas	10
Lista de figuras	11
Resumo.....	12
Abstract	14
1. Introdução.....	15
2. Objetivos	19
2.1 <i>Objetivo geral</i>	19
2.2 <i>Objetivos específicos</i>	19
3. Revisão bibliográfica	20
3.1 <i>O gênero Amburana</i>	20
3.2 <i>Marcadores Microsatélites</i>	25
3.3 <i>Desenvolvimento de marcadores microsatélites utilizando tecnologia de sequenciamento de nova geração Illumina</i>	31
4. Material e métodos.....	34
4.1 <i>Extração de DNA e sequenciamento</i>	34
4.2 <i>Identificação das regiões microsatélites e desenho de primers nucleares</i>	34
4.3 <i>Identificação das regiões microsatélites e desenho de primers cloroplastidiais e mitocondriais</i>	36
5. Resultados	37
5.2 <i>Regiões microsatélites e primers nucleares</i>	39
5.3 <i>Regiões microsatélites e primers cloroplastidiais e mitocondriais</i>	47
6. Discussão.....	51
7. Conclusão	54
8. Apêndice	55
9. Referências	64

Lista de tabelas

Tabela 1. Classificação de microssatélites de acordo com o tamanho, o motivo de repetição e sua localização no genoma. Fonte: KALIA et al., 2011; VIEIRA et al., 2016.....	27
Tabela 2. Características das montagens realizadas para <i>Amburana cearensis</i> com informações do genoma nuclear, cloroplastidial e mitocondrial da espécie	38
Tabela 3: Resultados observados na identificação de regiões microssatélites nas sequências genômicas nucleares montadas para a espécie <i>Amburana cearensis</i>	40
Tabela 4: Conjuntos de <i>primers</i> nucleares organizados para amplificação em <i>PCR multiplex</i> . o conjunto 1 foi projetado para amplificação simultânea com 5 pares de <i>primers</i> , o conjunto 2, projetado para amplificação simultânea com 6 pares de <i>primers</i> , o conjunto 3, projetado para amplificação simultânea com pares 7 de <i>primers</i> , o conjunto 4, projetado para amplificação simultânea com 6 pares de <i>primers</i> e o conjunto 5, projetado para amplificação simultânea com 6 pares de <i>primers</i>	42
Tabela 5: Lista dos marcadores microssatélites desenvolvidos para o genoma cloroplastidial de <i>Amburanacearensis</i>	49
Tabela 6: Seleção de marcadores microssatélites para o genoma mitocondrial de <i>A. cearensis</i>	50
Tabela 7: Apêndice com cem marcadores microssatélites nucleares extras para a espécie <i>A. cearensis</i>	55

Lista de figuras

- Figura 1: Representação do clado ADA, com três subclados distintos: Angylocalyceae, Dipterygeae, e Amburaneae. O gênero *Amburana* está presente no último subclado. Fonte: LEITE et al., 2015.....21
- Figura 2: Distribuição geográfica de *Amburana cearensis*. A espécie está localizada em grande parte do território brasileiro e também em países vizinhos, como Paraguai, Argentina e Bolívia. A representação está sendo indicada por um círculo branco. Fonte: SELEME et al., 2015.....22
- Figura 3: A espécie *Amburana cearensis*.....23
- Figura 4: Visualização do impacto da exclusão de sequências de baixa qualidade através do *FASTQC*. As bases filtradas apresentaram um valor de *phred* superior a 20, indicando boa qualidade.....37
- Figura 5: Distribuição dos microssatélites nucleares de *Amburana cearensis* por tamanho de motivo de repetição.....40
- Figura 6: Distribuição dos microssatélites por tamanho de motivo de repetição no genoma cloroplastidial..... 47
- Figura 7: Distribuição dos microssatélites por tamanho de motivo de repetição no genoma mitocondrial.....48

SANTOS. L. S. MARCADORES MICROSSATÉLITES PARA OS GENOMAS NUCLEAR E ORGANELAR DE *Amburana cearensis* (Allemão) A.C. Smith. 2021.73f. Dissertação (Mestrado em Genética e Biologia Molecular) – Instituto de Ciências Biológicas- ICB, Universidade Federal de Goiás, Goiânia, 2021.

Resumo

Amburana cearensis é uma espécie arbórea, conhecida popularmente como cerejeira, amplamente distribuída no Brasil e na Argentina, Bolívia, Paraguai e Peru. A espécie está distribuída em uma faixa de intensa ação antrópica e classificada como quase ameaçada de extinção de acordo com a União Internacional para a Conservação da Natureza (IUCN). A demanda econômica de *A. cearensis* como fonte madeireira e fonte medicinal impulsionou a sua exploração predatória, representando uma ameaça à sobrevivência da espécie. Nesse cenário, a genética de conservação é fundamental para a geração de conhecimentos sobre diversidade genética da espécie e definição de estratégias adequadas para a sua conservação e manejo. O objetivo geral deste trabalho foi desenhar pares de primers com alto potencial para o desenvolvimento de marcadores microssatélites robustos e informativos para os genomas nuclear, cloroplastidial e mitocondrial de *A. cearensis*. Além disso, visando a rapidez e a diminuição dos custos, os pares de primers de microssatélites nucleares foram desenhados para a amplificação em *PCR multiplex*. Para atingir os objetivos propostos, o genoma de *A. cearensis* foi parcialmente sequenciado, com o uso da plataforma *Illumina MiSeq*. A montagem da biblioteca genômica foi realizada a partir do DNA de um indivíduo da espécie e as sequências genômicas obtidas foram analisadas utilizando os *softwares* *FastQC*, *Trimmomatic* e *DipSPAdes*. A busca de regiões microssatélites foi realizada com o uso do programa *QDD*. Foram desenhados trinta pares de primers para regiões microssatélites nucleares, organizados em cinco conjuntos para a amplificação em *PCR multiplex*. Além disso, foram desenhados três pares de primers microssatélites cloroplastidiais e sete mitocondriais. Os pares de primers desenhados no presente estudo apresentam potencial de geração de novos marcadores microssatélites com alta capacidade de discriminação individual e transferibilidade para as outras espécies do gênero *Amburana*. Estes marcadores poderão ser aplicados de forma rápida e eficiente para os diversos tipos de estudos de diversidade genética de *A. cearensis* e do gênero

Amburana.

Palavras chave: *Amburana cearensis*, Marcadores Microsatélites, NGS.

SANTOS. L.S. **MICROSATELLITE MARKERS FOR THE NUCLEAR AND ORGANELLAR GENOMES OF *Amburana cearensis* (Allemão) AC Smith.** 2021.73f. Dissertation (Master in Genetics and Molecular Biology) – Instituto de Ciências Biológicas-ICB, Universidade Federal de Goiás, Goiânia, 2021.

Abstract

Amburana cearensis is a tree species, popularly known in Brazil as cerejeira, widely distributed in Brazil and also in Argentina, Bolivia, Paraguay and Peru. The species is distributed in a range of intense anthropic action and classified as endangered according to the International Union for the Conservation of Nature (IUCN). The economic demand for *A. cearensis* as a wood source and medicinal source has boosted its predatory exploitation, representing a threat to the survival of the species. In this scenario, conservation genetics is essential for generating knowledge about the genetic diversity of the species and defining appropriate strategies for its conservation and management. The general objective of this work was to design primer pairs with high potential for the development of robust and informative microsatellite markers for the nuclear, chloroplast and mitochondrial genomes of *A. cearensis*. Moreover, for quickness and cost savings, nuclear microsatellite primer pairs were designed for amplification in multiplex PCR. To achieve the proposed objectives, the genome of *A. cearensis* was partially sequenced, using the Illumina MiSeq platform. The assembly of the genomic library was performed from the DNA of an individual of the species and the genomic sequences obtained were analyzed using FastQC, Trimmomatic and DipSPAdes software. The search for microsatellite regions was performed using the QDD program. Thirty pairs of primers for nuclear microsatellite regions were designed, organized into five sets for amplification in multiplex PCR. The primer pairs designed in this study have the potential to generate new microsatellite markers with high individual discrimination capacity and transferability to other species of the genus *Amburana*. These markers can be applied quickly and efficiently for different types of genetic diversity studies of *A. cearensis* and *Amburana* genus.

Keywords: *Amburana cearensis*, Microsatellite markers, NGS.

1. Introdução

O Brasil possui a flora mais rica do mundo com mais de 56.000 espécies de plantas, o que corresponde a aproximadamente 19% da flora mundial (FLORA DO BRASIL, 2020). O fortalecimento de setores produtivos, aliado à exploração predatória, exerce alto impacto na biodiversidade, gerando como consequência a vulnerabilidade de populações ou até a extinção de espécies de plantas (FACHIM e GUARIM, 1995; GIULIETTI et al., 2009). A espécie *Amburana cearensis* (Allemão) A.C. Smith. (Leguminosae) pode ser citada como um exemplo nesta situação, uma vez que foi classificada como quase ameaçada de extinção de acordo com a União Internacional para a Conservação da Natureza (IUCN) (CANUTO et al., 2010; SELEME et al., 2015; ALMEIDA et al., 2017; IUCN, 2019; ALVIM et al., 2020; CNC FLORA, 2021). *A. cearensis* é uma árvore conhecida popularmente no Brasil como cerejeira e pertence ao gênero *Amburana*, um gênero sul-americano de plantas lenhosas amplamente distribuído no Brasil e também na Argentina, Bolívia, Paraguai e Peru (LEITE, 2005; SELEME et al., 2015). O gênero *Amburana* compõe o clado ADA, um grupo basal da subfamília Papilionoideae, pertencente à família Leguminosae e à ordem Fabales (CARVALHO, 2007; QUEIROZ, 2009; LPWG et al., 2017).

A demanda econômica de *A. cearensis* como fonte madeireira e fonte medicinal é classificada como atividade predatória, representando uma ameaça à sobrevivência do gênero. Sua madeira possui características vantajosas para a construção civil, fabricação de barris e indústrias nacionais de móveis de luxo, além disso a espécie está distribuída em uma faixa de intensa ação antrópica. Nesse cenário, a genética de conservação é fundamental para a geração de conhecimentos sobre diversidade genética da espécie e definição de estratégias adequadas para conservação e manejo (CARVALHO, 2007; ALVIM et al., 2020).

O sequenciamento de Nova Geração (NGS) gera recursos genômicos que podem ser aplicados para os mais diversos estudos de diversidade genética das espécies. As metodologias de NGS surgiram para atender uma demanda de rapidez e também para superar as limitações do Sequenciamento de Primeira Geração. Seus principais avanços correspondem à capacidade de gerar informações sobre milhões de pares de bases em uma única corrida de forma rápida, mesmo para espécies não modelo (CARVALHO e SILVA, 2010; RAZA e AHMAD, 2016).

Devido à rapidez e facilidade de gerar rascunhos de genoma, as tecnologias de NGS abriram o caminho para a descoberta de marcadores moleculares em grande escala (SINGH et al., 2015; TAHERI et al., 2018). Um marcador molecular pode ser definido como sequências de DNA em regiões específicas do genoma, herdadas de forma mendeliana que codificam ou não quaisquer características do genoma, mas não são afetadas pelo ambiente. (GRATTAPAGLIA e FERREIRA, 1998; ZOLET et al., 2017; DAR et al., 2019). Os marcadores microssatélites são um dos marcadores genéticos mais informativos para análise genética de populações da atualidade, uma vez que possuem atributos desejáveis, que os tornam vantajosos em relação a outros tipos: são encontrados em todos os genomas de organismos procarióticos e eucarióticos; geram segmentos polimórficos pequenos o suficiente para serem detectados por meio da reação de *PCR* (*Polymerase Chain Reaction*); são neutros em relação a efeitos fenotípicos; apresentam codominância; possuem uma alta taxa de mutação por locus por geração (entre 10^{-7} e 10^{-3}); apresentam ampla cobertura do genoma, abrangendo genomas organelares (cloroplasto e mitocôndria) e são passíveis a genotipagem de alto rendimento (ELLEGRAN, 2000; PARIDA et al., 2009; TAHERI et al., 2018; YASODHA et al., 2018; MOKHTAR e ATIA, 2019).

Os trabalhos de SELEME (2014) e BARRAL (2019) possuem grande relevância nos estudos genéticos de *A. cearensis*, pois são pioneiros no desenvolvimento de marcadores microssatélites nucleares (nuSSR). O primeiro realizou uma revisão taxonômica do gênero e descreveu uma nova espécie, a *Amburana eritrosperma*, a terceira espécie do gênero *Amburana*. Além disso, demonstrou as relações filogenéticas entre os clados *Amburana* e *Dipterygeae* e desenvolveu os primeiros marcadores microssatélites para *A. cearensis*. O segundo, gerou uma série de conhecimentos ecológicos, genéticos e de dinâmica populacional para a espécie, além de também desenvolver marcadores microssatélites. Ambos utilizaram a metodologia de construção de bibliotecas genômicas enriquecidas com microssatélites para o desenvolvimento de marcadores. O primeiro identificou 63 regiões microssatélites e desenvolveu 33 pares de *primers*, sendo 4 destes polimórficos. Enquanto que o segundo, identificou 19 regiões microssatélites e desenvolveu 19 pares de *primers*, sendo 18 destes polimórficos, porém com baixa variabilidade genética.

Contudo, visando a otimização de tempo e maior eficiência na genotipagem em larga escala, o desenho de marcadores microssatélites para amplificação em *PCR multiplex* é extremamente vantajoso. A multiplexação envolve a amplificação simultânea de vários loci em uma única reação de PCR e análise subsequente de vários marcadores em um único poço de gel. Isso pode aumentar o conteúdo informativo em uma única reação de PCR e diminuir o custo do uso em larga escala dos marcadores microssatélites. Para o sucesso da multiplexação é necessário um desenho feito sob medida dos pares de *primers* que serão amplificados simultaneamente, utilizando critérios rigorosos para seleção das melhores regiões (LI et al., 2018; GUERRERO-CÓZAR et al., 2020).

Os marcadores microssatélites para regiões organelares são interessantes para complementar a análise genética de *A. cearensis*, uma vez que podem ser aplicados em estudos com objetivos diferentes daqueles que utilizam os marcadores nucleares. Até o momento, não há relatos na literatura de desenhos de marcadores microssatélites específicos para os genomas cloroplastidial e mitocondrial desta espécie. Os genomas organelares diferem do genoma nuclear em vários aspectos, dentre os quais podem-se citar que, em se tratando do cloroplasto, possui genoma de tamanho pequeno, herança uniparental, baixa taxa de substituição de nucleotídeos, natureza haplóide, além de estrutura genômica altamente conservada. A taxa de substituição conservada no DNA do cloroplasto (cpDNA), torna-o importante para estudos evolutivos, ecológicos e filogeográficos (TOMAR et al., 2014; PERVAIZ et al., 2015; DANIELL et al., 2016; XUE et al., 2019). Os marcadores microssatélites de cloroplasto (cpSSR) possuem vantagens como a alta transferibilidade entre espécies e ampla utilização para estimar diversidade genética de plantas (TAKAHASHI et al., 2016; SINGH et al., 2017). Já o genoma mitocondrial de plantas (mtDNA) é grande e variável em tamanho (200 a 2.500 kb), contém muitos íntrons e elementos repetitivos e passa por processos de duplicação e rearranjos do genoma. Sua taxa de substituição de nucleotídeos é mais baixa do que em DNA nuclear, seu uso é popular em estudos evolutivos e filogenéticos (ROSA e PAIVA, 2009; GALTIER, 2011; TOMAR et al., 2014).

Considerando o status de conservação e vulnerabilidade de *A. cearensis* e a vasta possibilidade de aplicação dos marcadores microssatélites em diferentes tipos estudos genéticos

em plantas, é de suma importância obter marcadores polimórficos, informativos, de baixo custo e rápida aplicação para esta espécie.

2. Objetivos

2.1 Objetivo geral

Desenhar marcadores microssatélites para os genomas nuclear, cloroplastidial e mitocondrial de *A. cearensis*. E visando a rapidez e diminuição de custos do uso dos marcadores nucleares, estruturá-los para amplificação em *PCR multiplex*.

2.2 Objetivos específicos

- i. Selecionar regiões microssatélites presentes em sequências genômicas de *Amburana cearensis*, promissoras para o desenvolvimento de marcadores microssatélites polimórficos.
- ii. Desenhar pares de *primers* para amplificação *multiplex* de regiões microssatélites no genoma nuclear de *Amburana cearensis* em *PCR multiplex*.
- iii. Desenhar pares de *primers* para amplificação individual de regiões microssatélites no genoma cloroplastidial e mitocondrial de *Amburana cearensis*.

3. Revisão bibliográfica

3.1 O gênero *Amburana*

Amburana Schwacke & Taubert (1894) é um gênero sul-americano de plantas lenhosas, amplamente distribuído no Brasil e também na Argentina, Bolívia, Paraguai e Peru (SELEME et al., 2015). O gênero é composto por 3 três espécies: *Amburana acreana*, *Amburana cearensis* e *Amburana eritrosperma* (QUEIROZ, 2009; SELEME, 2014; SELEME et al., 2015) e está inserido no clado ADA, um grupo basal da subfamília Papilionoideae, pertencente à família Fabaceae (Leguminosae: Papilionoideae) e à ordem Fabales (CARVALHO, 2007; QUEIROZ, 2009; LPWG et al., 2017).

Amburana é um gênero de ramificação antiga de Leguminosae (Papilionoideae) e faz parte do clado ADA (Figura 1), que inclui três subclados distintos: Angylocalyceae, Dipterygeae, e Amburaneae (este último é composto pelos gêneros *Amburana*, *Cordyla*, *Dussia*, *Myrocarpus*, *Myrospermum*, *Myroxylon* e *Mildbraediodendron*). Este grupo se caracteriza por apresentar flores com apenas uma das pétalas desenvolvidas, sendo diferentes das típicas Papilionoideae com cinco pétalas diferenciadas em vexilo (adaxial), duas alas (laterais) e duas pétalas da quilha (abaxiais) (LEITE et al., 2015).

As três espécies descritas ocorrem em ambientes variados, como savanas e florestas tropicais (SELEME, 2014; SELEME et al., 2015). *A. cearensis* tem a maior distribuição geográfica entre as três espécies do gênero, estando presente nas regiões Nordeste, Sudeste e Centro-Oeste do Brasil, estendendo-se ao Paraguai, Argentina e Bolívia (Figura 2). Possui porte intermediário no gênero, atingindo cerca de 5 a 20 metros de altura, a floração e frutificação ocorrem entre os meses de fevereiro a dezembro. A sua distinção ocorre principalmente pelo caule castanho (*versus* o caule de cor ferrugem de *A. acreana*), pela forma oval dos folíolos (*versus* forma oval-lanceolada em *A. acreana*) e pelo número variável de folíolos por folha, entre 7-15 (*versus* 7-25 em *A. acreana*), e um fruto com endocarpo membranoso ao redor da semente preta (*versus* endocarpo parcialmente desenvolvido em *A. erythrosperma*, que possui semente vermelha). (SELEME et al., 2015; PEREIRA et al., 2017; LATALIZA et al., 2019).

A. cearensis (Figura 3) é conhecida popularmente pelos nomes de cerejeira, cumaru-

do-ceará, cumaré, cumaru-das-caatingas, imburana de cheiro, umburana e cumaru-de-cheiro. Possui caráter ornamental, usada com sucesso no paisagismo principalmente por possuir tronco e ramos lisos de cor avermelhado (LORENZI, 1998; ARAÚJO, 2018). A madeira possui aspecto agradável, excelente para construção civil, fabricação de barris e muito procurada pelas indústrias nacionais de móveis de luxo (CARVALHO, 2007; ALVIM et al., 2020).

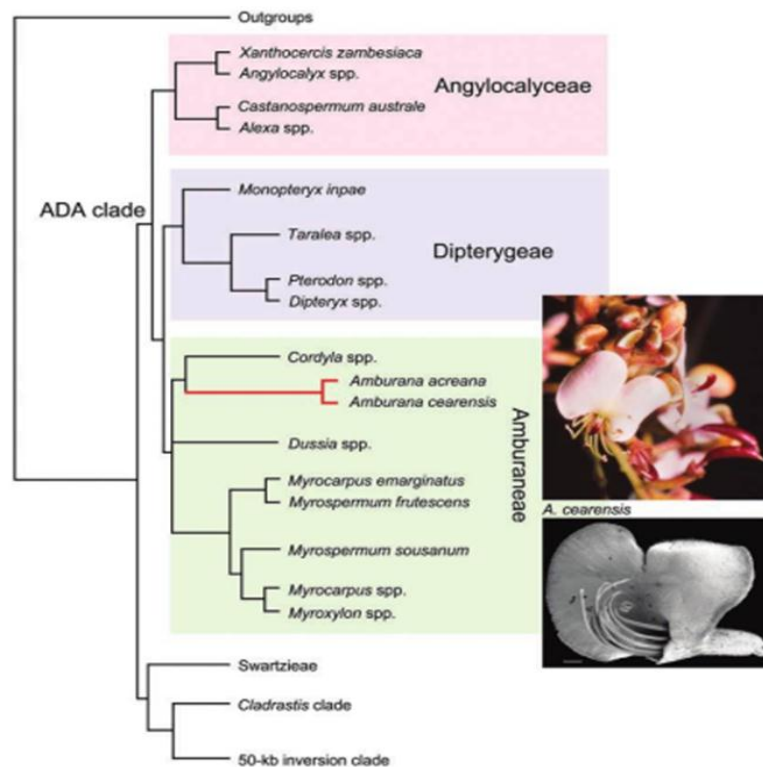


Figura 1: Representação do clado ADA, com três subclados distintos: Angylocalyceae, Dipterygeae, e Amburaneae. O gênero *Amburana* está presente no último subclado. Fonte: LEITE et al., 2015.



Figura 2: Distribuição geográfica de *Amburana cearensis*. A espécie está localizada em grande parte do território brasileiro e também em países vizinhos, como Paraguai, Argentina e Bolívia. A representação está sendo indicada por um círculo branco. Fonte: SELEME et al., 2015.



Figura 3: A espécie *Amburana cearensis*. Fonte: <https://bityli.com/1K7gnw>.

Diferentes partes da planta, como folhas, cascas, caules e sementes, são usados na medicina popular com aplicações medicinais. Há relatos na literatura de aplicações em doenças respiratórias (gripe, resfriado, bronquite e asma), atividade anti-inflamatória, antioxidante e antibacteriana (PEREIRA et al., 2017; BARRAL, 2018; LATALIZA et al., 2019). Industrialmente, foi produzido um fitoterápico pela Universidade Federal do Ceará (UFC), o xarope de cumaru, produzido pelo Programa Farmácias Vivas na faculdade de Farmácia-Escola/UFC e por algumas empresas privadas (CANUTO, 2006; NASCIMENTO, 2018).

As áreas de ocorrência de *A. cearensis* são sabidamente expostas ao desmatamento, principalmente a Mata Atlântica e Cerrado. Além disso, a demanda econômica da espécie como fonte madeireira e fonte medicinal é considerada atividade predatória, representando uma ameaça à sobrevivência do gênero. A espécie foi classificada como quase ameaçada de extinção

de acordo com a União Internacional para a Conservação da Natureza (CANUTO et al., 2010; SELEME et al., 2015; ALMEIDA et al., 2017; IUCN, 2019; ALVIM et al., 2020; CNC FLORA, 2021).

Na literatura há um número considerável de estudos quanto às propriedades químicas e farmacológicas da espécie: PEREIRA et al. (2017) investigou o perfil químico, a citotoxicidade e efeitos neuroprotetores em cultura de células neurais a partir de extratos de sementes de *A. cearensis*. Foi observado a ausência de toxicidade e potencial neuroprotetor contra os danos neuronais induzidos pelo glutamato. OLIVEIRA et al. (2020) investigou a atividade inibitória da serina protease, juntamente com a atividade antibacteriana e modificadora de antibióticos do extrato bruto e sementes de *A. cearensis*. As amostras atuaram como modificadoras de resistência, apresentando sinergismo quando associadas à gentamicina, norfloxacin e penicilina. SILVEIRA (2020) desenvolveu um nanosistema de cumarina, um constituinte natural da espécie, para avaliação da toxicidade e atividade leishmanicida (*Leishmania braziliensis*) *in vitro* e *in vivo*. Como resultado, o estudo identificou características tecnológicas de interesse e estabilidade, não associada à toxicidade e com atividade leishmanicida *in vitro* e *in vivo*.

Os estudos atuais baseados em marcadores moleculares buscam preencher as lacunas de conhecimento, principalmente quanto aos aspectos genéticos da espécie. SANTOS et al. (2007) estudou a dispersão da variabilidade genética no semi-árido brasileiro para as espécies *A. cearensis*, *Myracrodruon urundeuva* e *Schinopsis brasiliensis* com base no marcador de DNA tipo RAPD (*Random Amplified Polymorphic DNA*), indicando que variabilidade genética destas espécies não está uniformemente distribuída por todo o semi-árido brasileiro. SOUZA et al. (2021) verificou a diversidade genética entre 68 genótipos de *A. cearensis* situados em diferentes localidades no estado de Pernambuco, os padrões genéticos foram avaliados por marcadores moleculares ISSR (*Inter Simple Sequence Repeats*). Como resultado, foi descrito que a variabilidade genética destas espécies não está uniformemente distribuída por todo o semi-árido brasileiro.

Os trabalhos de SELEME (2014) e BARRAL (2019) possuem grande relevância nos estudos genéticos de *A. cearensis*, pois são pioneiros no desenvolvimento de marcadores microsatélites. O primeiro realizou uma revisão taxonômica do gênero e descreveu uma nova

espécie, a *Amburana eritrosperma*, a terceira espécie do gênero *Amburana*. Além disso, demonstrou as relações filogenéticas entre os clados *Amburana* e Dipterygeae e desenvolveu os primeiros marcadores microssatélites para *A. cearensis*. O segundo, gerou uma série de conhecimentos ecológicos, genéticos e de dinâmica populacional para a espécie, além de também desenvolver marcadores microssatélites. Ambos utilizaram a metodologia de construção de bibliotecas genômicas enriquecidas com microssatélites para o desenvolvimento de marcadores. O primeiro identificou 63 regiões microssatélites e desenvolveu 33 pares de *primers*, sendo 4 destes polimórficos; Enquanto que o segundo, identificou 19 regiões microssatélites e desenvolveu 19 pares de *primers*, sendo 18 destes polimórficos.

3.2 Marcadores Microssatélites

Um marcador molecular pode ser definido como sequências de DNA em regiões específicas do genoma, herdadas de forma mendeliana, podendo codificar ou não quaisquer características do genoma. Outro conceito também utilizado, compreende marcadores moleculares como um segmento do DNA que apresenta variabilidade, quando comparado à mesma região em outros indivíduos ou organismos (GRATTAPAGLIA e FERREIRA, 1998; ZOLET et al., 2017; MIRANDA et al., 2018; DAR et al., 2019). A investigação utilizando marcadores moleculares ocorre através do polimorfismo presente entre as sequências de nucleotídeos de diferentes indivíduos. Os polimorfismos de DNA surgem como resultado de uma variação no genoma, oriundas de mutações. Estas variações podem se originar de inserções, deleções, duplicações e translocações (BHANDARI et al., 2017; NADEEM et al., 2018).

O estudo de polimorfismos de DNA utilizando marcadores moleculares é um dos avanços mais significativos no campo da genética molecular, pois gera compreensão da organização e evolução do genoma de plantas. Marcadores moleculares em plantas são amplamente empregados em estudos de diversidade genética, inferência filogenética, estrutura genética de populações naturais, conservação e uso de recursos genéticos, melhoramento vegetal, entre outros (ADHIKARI et al., 2017; ZOLET et al., 2017; MIRANDA et al., 2018).

A detecção da variação genética em plantas oferece uma oportunidade para

compreender a base molecular de vários fenômenos biológicos. A escolha do marcador molecular a ser utilizado depende dos objetivos a serem alcançados na pesquisa, da disponibilidade de recursos e da estrutura de laboratório. Alguns marcadores são mais indicados para explorar determinadas aplicações, com base em uma análise de vantagens ou limitações. Alguns requisitos são desejáveis para um bom marcador molecular (KUMAR et al., 2009; ADHIKARI et al., 2017, MIRANDA et al., 2018):

- i. Presença de polimorfismo
- ii. Ocorrência frequente no genoma com distribuição uniforme
- iii. Neutralidade em relação às condições ambientais
- iv. Fácil acesso e disponibilidade
- v. Alta reprodutibilidade
- vi. Herança co-dominante
- vii. Multialelismo

Devido às suas características, os marcadores do tipo microssatélites (*SSRs - Simple Sequence Repeats*) são um dos marcadores genéticos mais informativos para análise genética de populações da atualidade. Caracterizam-se por motivos de repetição em *tandem* de uma a seis bases e por um alto grau de polimorfismo. As sequências únicas que flanqueiam os motivos de microssatélites são usadas para projetar *primers* para amplificação específica de lócus (GRATTAPAGLIA e FERREIRA, 1998; ZANE e BARGELLONI, 2002; PARIDA et al., 2009; MOKHTAR e ATIA, 2019). Os microssatélites são classificados de acordo com o tamanho, motivo de repetição e com sua localização no genoma (Tabela 1) (KALIA et al, 2011; VIEIRA et al., 2016).

Tabela 1: Classificação de microssatélites de acordo com o tamanho, o motivo de repetição e sua localização no genoma. Fonte: KALIA et al.(2011); VIEIRA et al. (2016).

CLASSIFICAÇÃO DOS SSRs		EXEMPLOS
TAMANHO	Observado o número de nucleotídeos por unidade de repetição, os SSRs são classificados em mono-, di-, tri-, tetra-, penta- ou hexanucleotídeos.	Mononucleotídeos (A)n: ...AAAAAAA... Dinucleotídeos (TA)n: ...TATATATA... Trinucleotídeos (TAG)n: ...TAGTAGTAGTAG... Tetranucleotídeos(TAGA)n:...TAGATAGATAGATAG... Pentanucleotídeos(TTTAA)n:...TTTAATTTAATTTAA... Hexanucleotídeos(CAATTT)n:CAATTTCAATTTCAATT
MOTIVO DE REPETIÇÃO	(1) perfeito, se apresentarem repetições de um único motivo; (2) imperfeito, se um par de bases não pertencente ao motivo ocorrer entre repetições; ou (3) composto, quando formado por vários motivos adjacentes e repetitivos.	Perfeito: (TA)n Imperfeito: (AAT)n CAG (AAT)n Composto: (TCA)n (ACG)n
LOCALIZAÇÃO NO GENOMA	Podem ser do tipo: <i>SSR</i> nuclear, <i>SSR</i> mitocondrial ou <i>SSR</i> cloroplástico.	Núcleo: nuSSR Mitocôndria: mtSSR Cloroplasto: cpSSR

n: número de vezes em que o motivo se repete.

Tradicionalmente, o DNA nuclear é usado na avaliação da diversidade e parentesco entre diferentes espécies, em estudos evolutivos a nível de população e genotipagem, na verificação de distâncias genéticas, teste de paternidade, entre outros (GARRIGAN e HAMMER, 2006; ROSA e PAIVA, 2009). Na maioria das espécies de angiospermas, o cloroplasto e a mitocôndria possuem herança materna, essa característica traz relevância para o uso de marcadores microssatélites em regiões organelares, pois são interessantes para complementar a análise genética em estudos com marcadores microssatélites nucleares.

Genomas de plantas contêm grande número de SSRs espalhados por muitos locais do cromossomo, possuindo características desejáveis como a reprodutibilidade, especificidade, hipervariabilidade e natureza codominante (JONES et al., 2009; SUNG et al., 2010; KAPIL et al., 2018). Cloroplastos e mitocôndrias possuem sistemas genéticos e genomas próprios. Os genomas organelares diferem do genoma nuclear em vários aspectos, dentre os quais podem-se mencionar que, em se tratando do cloroplasto, possui genoma de tamanho pequeno, herança uniparental, baixa taxa de substituição de nucleotídeos, natureza haplóide, além de estrutura genômica altamente conservada. A taxa de substituição conservada no DNA do cloroplasto (cpDNA), torna-o importante para estudos evolutivos, ecológicos e filogeográficos (TOMAR et al., 2014; PERVAIZ et al., 2015; DANIELL et al., 2016; XUE et al., 2019). Os marcadores microssatélites de cloroplasto (cpSSR) possuem vantagens como a alta transferibilidade entre espécies e ampla utilização para estimar diversidade genética de plantas (TAKAHASHI et al., 2016; SINGH et al., 2017).

Já o genoma mitocondrial de plantas (mtDNA) é grande e variável em tamanho (200 a 2.500 kb), contém muitos íntrons e elementos repetitivos e passa por processos de duplicação e rearranjos do genoma. Sua taxa de substituição de nucleotídeos é mais baixa do que em DNA nuclear, seu uso é popular em estudos evolutivos e filogenéticos (ROSA e PAIVA, 2009; GALTIER, 2011; TOMAR et al., 2014). As forças evolutivas que moldam o genoma mitocondrial são diferentes das forças que afetam o genoma nuclear, o que pode ser explicado pela transmissão uniparental e pela natureza haplóide das mitocôndrias, que limita a capacidade de eliminação de mutações deletérias (HAVIRD et al., 2017).

O mecanismo evolutivo dos microssatélites é um processo complexo. Possivelmente a

ocorrência de mutações surgem por deslizamento da DNA Polimerase III (*slippage*) nas regiões repetitivas durante o processo de replicação, alterando assim, o número de repetições e, conseqüentemente, o comprimento da sequência de repetição (EISEN, 1999; IDRESS, 2014). Além disso, podem surgir por erros nos processos de recombinação (*crossing over* desigual), falhas no reparo de DNA e retrotransposição (VIEIRA et al., 2016; NADEEM et al., 2017). A presença de polimorfismo ocorre principalmente devido à variação do comprimento alélico, causada por uma diferença no número de unidades de repetição entre alelos (ELLEGREN, 2000). Polimorfismos repetidos geralmente resultam da adição ou deleção de unidades de repetição, assim, diferentes indivíduos exibem variações nas unidades de repetições (VIEIRA et al., 2016).

A identificação e uso de sequências microssatélites geraram um impacto significativo na genética nos últimos dez anos. Tais marcadores possuem atributos desejáveis, que os tornam vantajosos em relação a outros tipos: são encontrados em todos os genomas de organismos procarióticos e eucarióticos, geram segmentos polimórficos pequenos o suficiente para serem detectados por meio da reação de *PCR* (*Polymerase Chain Reaction*). Além disso, são neutros em relação a efeitos fenotípicos, possuem uma alta taxa de mutação por locus por geração entre 10^{-7} e 10^{-3} , podem ser usados em qualquer estágio de desenvolvimento da planta, apresentam reprodutibilidade e abundância relativa, apresentam ampla cobertura do genoma, abrangendo genomas organelares (cloroplasto e mitocôndria) e são passíveis a genotipagem de alto rendimento (ELLEGRAN, 2000; PARIDA et al., 2009; TAHERI et al., 2018).

Essas vantagens forneceram um novo cenário para a detecção de microssatélites, permitindo ampliação e evolução de pesquisas. Marcadores microssatélites fornecem a base para aplicações bem sucedidas em diversos campos como: conservação genética (SOARES et al., 2013; KACZMARCZYK, 2019), genética de populações (SANTOS et al., 2015), análise de diversidade genética (SOARES et al., 2015; HILMARSSON et al., 2017; GUIMARÃES, 2018; SANTOS et al., 2020), estimativa do grau de parentesco entre genótipos (OLIVEIRA et al., 2006; VIEIRA et al., 2016) e mapeamento do genoma (ANTONIO et al., 2019).

Os estudos realizados com marcadores microssatélites podem ser otimizados com metodologia de *PCR multiplex*. Essa metodologia permite que sequências alvos possam ser amplificadas incluindo dois ou mais pares de *primers* na mesma reação (MARKOULATOS et

al., 2002). A multiplexação envolve a amplificação simultânea de vários loci em uma única reação de PCR e análise subsequente de vários marcadores em um único poço de gel. Isso pode aumentar o conteúdo informativo em uma única reação de PCR e diminuir o custo do uso em larga escala dos marcadores microssatélites. Para o sucesso da multiplexação é necessário um desenho feito sob medida dos pares de *primers* que serão amplificados simultaneamente, utilizando critérios rigorosos para seleção das melhores regiões a fim de evitar sobreposição de faixas alélicas nos marcadores com a mesma cor de fluoróforo. (LI et al., 2018; GUERRERO-CÓZAR et al., 2020).

3.3 Desenvolvimento de marcadores microssatélites utilizando tecnologia de sequenciamento de nova geração Illumina

Os marcadores microssatélites são um dos marcadores genéticos mais informativos da atualidade. Anteriormente, o seu desenvolvimento contava com o emprego de metodologias tradicionais, reconhecidas por serem experimentalmente longas, complexas e dispendiosas. Atualmente contamos com as Tecnologias de Sequenciamento de Nova Geração (NGS), que facilitam a identificação eficiente de um grande número de microssatélites (ZALAPA, 2012; TELLES et al., 2017; TAHERI et al., 2018).

O sequenciamento de Sanger (também conhecido por sequenciamento didesoxi), é um método de sequenciamento de DNA de Primeira Geração. Foi criado em 1975 e consagrado como "padrão-ouro". Esse método dominou a indústria por décadas e proporcionou um grande avanço científico como o Sequenciamento do Genoma Humano, que durou 13 anos e gerou um custo de aproximadamente 3 milhões de dólares. As metodologias de Sequenciamento de Nova Geração (também chamadas de Segunda Geração), surgiram para atender uma demanda de rapidez e também para superar as limitações do Sequenciamento de Primeira Geração. O NGS possui abordagem singular para o sequenciamento de DNA, gerando revolução na pesquisa genômica por meio do sequenciamento paralelo massivo que possui capacidade de gerar informações sobre milhões de pares de bases em uma única corrida, mesmo para espécies não modelo, além de ser um método rápido para obter informações genômicas (CARVALHO e SILVA, 2010; RAZA e AHMAD, 2016). O sequenciamento NGS potencializou o caminho para

a descoberta de marcadores genéticos em grande escala (SINGH et al., 2015; TAHERI et al., 2018). A facilidade de sequenciar um genoma completo nesse novo cenário permitiu a realização de estudos comparativos e evolutivos em alta proporção, situação inimaginável há poucos anos (SILVA et al., 2013; SHEN et al., 2019; TINIO et al., 2019; KUMAR et al., 2020).

Outro exemplo da dimensão do NGS é que se o sequenciamento completo do genoma humano fosse realizado hoje, teria duração de menos de um dia utilizando o NGS *Massive Parallel Sequencing* (METZKER, 2009). Vários projetos baseados em NGS foram desenvolvidos nas últimas décadas (WANG et al., 2010; ZHANG et al., 2017; MA et al., 2019; LIU et al., 2020) gerando uma enorme quantidade de sequências disponíveis em bancos de dados públicos e úteis para o subsídio de novos marcadores. Porém, embora o NGS seja mais barato e rápido quando comparado às metodologias tradicionais, possui duas grandes desvantagens: baixa acessibilidade a pequenos laboratórios devido alto custo financeiro e necessidade de infraestrutura e mão de obra qualificada para coleta e interpretação adequada dos conjuntos de dados (VIEIRA et al., 2016; RAZA e AHMAD, 2016).

Atualmente existe uma variedade de plataformas NGS disponíveis no mercado: a tecnologia 454 Roche, *Illumina*, plataforma *SOLID*, *Pacific Biosciences* e *Oxford Nanopore*. A plataforma mais utilizada para o desenvolvimento de marcadores *SSR* é a plataforma *Illumina*. Essa plataforma necessita do preparo de uma biblioteca genômica que realiza um corte enzimático para quebras aleatórias no DNA, resultando em pequenos fragmentos de DNA. São adicionados adaptadores às extremidades de cada fragmento, que são fixados por complementaridade de bases em uma superfície de vidro (*flow cell*) permitindo assim, que genomas complexos sejam amplificados com iniciadores de PCR comuns. A amplificação pode produzir de 100 a 200 milhões de *clusters*, essa etapa é conhecida por clusterização (MARDIS, 2008; METZKER, 2009; MARDIS, 2013).

O método de sequenciamento da plataforma *Illumina* é baseado na tecnologia de terminação reversível. Neste método, as moléculas de DNA e os *primers* são anexados a uma lâmina e amplificados para formação de *clusters* de DNA. Para determinação de sequência, quatro nucleotídeos fluorescentes são adicionados simultaneamente a *flow cell*, cada nucleotídeo tem sua cor correspondente. Uma DNA polimerase incorpora um nucleotídeo fluorescente a fita que está sendo sintetizada por complementaridade de bases e os nucleotídeos

não incorporados são lavados antes da adição de um novo *dNTP* (SINGH et al., 2015). Uma câmera fotografa imagens dos nucleotídeos marcados com fluorescência e, em seguida, o bloqueador terminal 3' OH, é quimicamente removido do DNA, permitindo que o próximo ciclo comece, apenas um nucleotídeo é incorporado por ciclo. O sequenciamento *Illumina* permite até 3 bilhões de leituras por execução com um comprimento de leitura de 50–300 bp (FILICHKIN et al., 2010; SINGH et al., 2015). A qualidade das sequências geradas é avaliada pelo valor de *Phred*, uma medida aplicada em sequenciamentos automatizados de DNA que atribui valores para cada base gerada. Tais valores refletem o nível de confiança da leitura, assim, quanto maior o valor de *Phred*, maior confiabilidade na base correspondente (MARDIS, 2008; OCHOA et al., 2013).

As etapas iniciais para o desenvolvimento de marcadores microssatélites utilizando a tecnologia *Illumina* correspondem, primeiramente, ao preparo de uma biblioteca genômica seguido do sequenciamento de um segmento do genoma estudado. O desenvolvimento desses marcadores envolve a detecção de regiões microssatélites favoráveis, seguida do desenho de *primers*, da validação através da amplificação por PCR e da detecção do polimorfismo (NADEEM et al., 2017).

O desenho do marcador é um componente chave no desenvolvimento bem-sucedido de marcadores microssatélites (ZALAPA et al., 2012). Um *primer* é uma pequena parte do DNA ou RNA a partir do qual inicia-se a síntese do DNA, sua eficiência desempenha um papel vital na sensibilidade e eficiência da PCR (HE et al., 1994; NADEEM et al., 2017). A especificidade do *primer* é um dos fatores mais importantes para uma PCR bem-sucedida, *primers* bem desenhados devem anelar apenas com as sequências de DNA alvo (KALENDAR et al., 2017). Um ponto importante a ser considerado é a escolha ou exclusão de regiões para o desenho de marcadores. Regiões mono-, tri- e hexanucleotídeos geralmente são excluídas deste processo, os mononucleotídicos apresentam maior susceptibilidade a erros de deslizamento durante a etapa de PCR, já os tri- e hexanucleotídeos possuem presença frequente em regiões gênicas, o que acarreta marcadores microssatélites com baixo polimorfismo (XU et al., 2013).

É preciso cautela na projeção de marcadores, bem como seleção de regiões de qualidade e de lócus de melhor desempenho. Em resumo, esforço, despesa e experiência são necessários para o desenvolvimento adequado de marcadores microssatélites, que são úteis em

diferentes estudos no gênero de interesse (ZALAPA et al., 2012).

4. Material e métodos

4.1 Extração de DNA e sequenciamento

Amostras de folhas frescas de um indivíduo da espécie de *A. cearensis* foram coletadas na cidade de Porto Murinho – MS, com o apoio da equipe do Instituto de Pesquisa do Jardim Botânico do Rio de Janeiro, parceiros desta pesquisa. As folhas foram colocadas em sacos plásticos contendo sílica gel, para o transporte até o laboratório de Genética & Biodiversidade, em Goiânia – GO, onde foram conduzidas todas as etapas do presente estudo.

O DNA foi extraído a partir do protocolo CTAB 2% descrito em Doyle e Doyle (1987) e quantificado por comparação visual com uso de marcadores de peso molecular conhecido (10ng/ μ L, 20ng/ μ L, 50ng/ μ L, 100ng/ μ L e 200ng/ μ L) em gel de agarose 1% e tampão para eletroforese TBE (*Tris Borato - EDTA*), na concentração de 1x. O gel foi corado com brometo de etídeo (10mg/mL) e para fins de registro, fotografado utilizando o fotodocumentador *LPIX* (*Loccus Biotecnologia*®).

A montagem da biblioteca genômica foi realizada a partir do DNA de um indivíduo de *A. cearensis*. Para tal, foi utilizado o protocolo do *kit Nextera* (*Illumina*) e a quantificação de DNA foi realizada pelo equipamento *Qubit*® *fluorometer 2.0*, seguido pela fragmentação enzimática do DNA pela enzima *transposase* e de uma reação em *PCR* (*Polymerase Chain Reaction*) para inserção de adaptadores complementares às extremidades dos fragmentos gerados (*P7i2+P5i13*).

A validação desta biblioteca genômica ocorreu por meio do equipamento *Bioanalyzer* (*High Sensitivity DNA Analysis Kit*) e por *PCR* quantitativa (*Kappa Library Quantification Kit Illumina*). O sequenciamento de DNA foi realizado na plataforma *Illumina MiSeq*, utilizando o *kit MiSeq Reagent v3* de 600 ciclos (2x300bp, *reads paired-end- Illumina*).

4.2 Identificação das regiões microssatélites e desenho de primers nucleares

O controle de qualidade dos *reads* oriundos do sequenciamento foi avaliado pelo *software FastQC* (ANDREWS, 2010). Em seguida utilizou-se o *software Trimmomatic* (BOLGER et al., 2014), com objetivo de exclusão de *reads* que apresentavam baixa qualidade e também sequências de adaptadores. Foi realizada a montagem *de novo* com o *software DipSPAdes* (SAFONOVA et al., 2015). A busca de regiões microssatélites foi realizada com o uso do programa *QDD* versão 3.1.2 instalado na máquina virtual *VirtualBox 5.0.14* (MEGLÉCZ et al., 2009). Inicialmente, foi realizada a identificação de regiões microssatélites, seguido da detecção de similaridade entre as sequências extraídas com o *Blast*, e seleção apenas de sequências únicas (*singletons*).

Os parâmetros utilizados para selecionar as sequências para o desenho de *primers* foram propostos com o objetivo de obter as melhores regiões microssatélites para aumentar as chances de polimorfismos. O número mínimo de repetições foi de dez para regiões com dinucleotídeos, seis para tetranucleotídeos e cinco para pentanucleotídeos. Regiões microssatélites do tipo mono-, tri- e hexanucleotídicas não foram buscadas. O tamanho do produto de *PCR* foi estipulado entre 150 pb - 500 pb e o teor de GC de entre 30-60%. Com a intenção de afunilar ainda mais a busca pelas melhores regiões microssatélites, foram aplicados os seguintes filtros: exclusão das regiões cujos motivos de repetição eram do tipo AT, preferência por regiões com os maiores números de repetição em *tandem*, escolha de um *primer* por *contig*, exclusão de regiões cujo motivo de repetição apresentasse muitas bases iguais (ex: AAAAC, AAAG, AAAAT), exclusão das regiões com microssatélites do tipo composto e exclusão de regiões microssatélites associadas a elementos transponíveis.

A próxima etapa foi o desenho de *primers* para amplificação em *PCR Multiplex*. Para tal, utilizou-se o *software fastPCR*, uma ferramenta amplamente utilizada em desenho de *primers* para fins diversos como *PCR multiplex*, *Real Time PCR* e *Single Primer PCR*. O desenho dos pares de *primers* para *PCR multiplex* baseou-se no preenchimento dos seguintes parâmetros: distância da região alvo de no mínimo 50 pb, tamanho máximo do fragmento de *PCR* de até 500 pb, diferença de temperatura de anelamento dentro dos conjuntos *multiplex* de no máximo 2° C e presença de *amplicon* em testes de *PCR in silico*.

4.3 Identificação das regiões microssatélites e desenho de primers cloroplastidiais e mitocondriais

As montagens dos genomas cloroplastidial e mitocondrial foram realizadas a partir do sequenciamento *Illumina* de *A. cearensis*, por meio da separação das sequências nucleares e organelares. As análises do genoma cloroplastidial foram realizadas através do *Fast Plast*, um *pipeline* que realiza todas as etapas de montagem do genoma cloroplastidial em uma única linha de comando. Já a montagem do genoma mitocôndria, foi realizada através do alinhador de sequências *Bowtie*. Em ambas montagens, foi realizada a comparação com sequências de referência de cloroplasto e mitocôndria depositadas no *RefSeq* do *GenBank*.

A busca de regiões microssatélites foi realizada por meio do programa *QDD* versão 3.1.2, instalado na máquina virtual *VirtualBox* 5.0.14 (MEGLÉCZ et al., 2009). Como os genomas cloroplastidiais e mitocondriais possuem menos regiões microssatélites em relação ao genoma nuclear, os critérios de tamanho de motivo e número mínimo de repetições foram flexibilizados para possibilitar a detecção de tais regiões. Assim, o número mínimo de repetições para os motivos buscados foi de quatro repetições para dinucleotídeos e trinucleotídeos e três repetições para tetranucleotídeos, pentanucleotídeos e hexanucleotídeos. Regiões microssatélites do tipo mononucleotídeo não foram buscadas. O tamanho do produto de *PCR* foi estipulado entre 150 pb - 500 pb.

O desenho dos pares de *primers* foi realizado visando a amplificação individual dos loci, ou seja, diferente do que foi realizado para o genoma nuclear. Para tal, foram aplicados os seguintes filtros: regiões com o maior número de repetições em *tandem*, escolha de um *primer* por *contig*, exclusão de regiões cujo motivo de repetição apresentasse muitas bases iguais (ex: AAAAC, AAAG, AAAAT), exclusão de regiões microssatélites do tipo composto e exclusão de regiões microssatélites associadas a elementos transponíveis.

5. Resultados

5.1 Dados do sequenciamento NGS e rascunho dos genomas nuclear, cloroplastidial e mitocondrial

Foram sequenciados 7.5 Gb do genoma de *A. cearensis*, distribuídos em 13.373.658 *reads paired-end*. Após o controle de qualidade, o volume final das sequências foi de 5.3 Gb, distribuídos em 12.782.550 *reads paired-end*, que foram usados para a montagem do genoma de *A. cearensis*. Foi possível observar o impacto da exclusão de sequências de baixa qualidade através do *FASTQC*. As bases filtradas apresentaram um valor de *phread* superior a 20, indicando boa qualidade (Figura 3). As sequências apresentaram distribuição homogênea e ausência de ruídos nas extremidades, esse é um indicativo de que houve retirada de sequências de baixa qualidade e de adaptadores.



Figura 4: Visualização do impacto da exclusão de sequências de baixa qualidade através do *FASTQC*. As bases filtradas apresentaram um valor de *phread* superior a 20, indicando boa qualidade.

Neste estudo, a montagem da sequência genômica nuclear de *A. cearensis* foi realizada pelo *software Dipspades* correspondendo a 285 Mb. De acordo com as características da montagem (Tabela 2), o número total de *scaffolds* foi de 183.015, com comprimento variando entre 500 pb e 122.163 pb. O tamanho médio de *scaffolds* foi de 1560 pb e o valor N50 do comprimento de *scaffolds* foi de 1977. Dos *scaffolds* montados, 99,9% apresentaram tamanho maior que 500 pb e 59,2% apresentaram tamanho maior que 1K.

Tabela 2: Características das montagens dos genomas nuclear, cloroplastidial e mitocondrial da espécie *Amburana cearensis*.

Parâmetros	Núcleo	Cloroplasto	Mitocôndria
Número total de <i>scaffolds</i>	183.015	10	223
Tamanho total dos <i>scaffolds</i> (pb)	285.546.213	136.685	511.590
Maior tamanho de <i>scaffolds</i> (pb)	122.163	46.640	13.592
Menor tamanho de <i>scaffolds</i> (pb)	500	1177	267
Média do tamanho de <i>scaffolds</i> (pb)	1560	10521	1456
Valor N50 do comprimento de <i>scaffolds</i>	1977	16.092	3369
<i>Scaffolds</i> > 500 pb	99,9%	100%	86,5%
<i>Scaffolds</i> > 1K pb	59,2%	100%	63,2%
<i>Scaffolds</i> > 10 K pb	0,1 %	60%	3,1%
<i>Scaffolds</i> A%	33,62	33,19	28,04
<i>Scaffolds</i> T%	34,20	32,77	28,34
<i>Scaffolds</i> C%	16,05	17,05	21,97
<i>Scaffolds</i> G%	16,12	16,98	21,65
<i>Scaffolds</i> N%	0%	0	0

O tamanho do genoma cloroplastidial corresponde a aproximadamente 136 kb. De acordo com as informações da montagem (Tabela 2), foi gerado um *assembly* de apenas 10 *scaffolds*, ou seja, com baixa fragmentação genômica. O tamanho total de *scaffolds* foi de 136.685 pb, variando entre 1.177 pb e 46.640 pb e o valor N50 do comprimento de *scaffolds* foi de 16.092. Todos os *scaffolds* apresentaram tamanho acima de 1k pb e 60% dos *scaffolds* apresentaram tamanho acima de 10k pb. A presença das bases A e T tiveram proporções semelhantes, respectivamente 33,19% e 32,77%, assim como proporções semelhantes entre as bases C e G, respectivamente 17,05% e 16,98%. Todas as bases foram devidamente identificadas, resultando na ausência de detecção de N.

O tamanho do genoma mitocondrial corresponde a aproximadamente 551 kb. De acordo com as informações da montagem (Tabela 2), foi gerado um *assembly* com 223

scaffolds. O tamanho total de *scaffolds* foi de 511.590 pb, variando entre 267 pb e 13.592 pb e o valor N50 do comprimento de *scaffolds* foi de 3.369. Um total de 63,2 % dos *scaffolds* apresentaram tamanho acima de 1k pb e 3,1% com tamanho acima de 10k. Todas as bases foram identificadas na montagem resultando no valor nulo para *scaffolds* N. A porcentagem das bases A e T foi semelhante, respectivamente 28,04% e 28,34%. A porcentagem para C correspondeu a 21,97% e para G 21,65%.

5.2 Regiões microssatélites e primers nucleares

De acordo com os filtros, as regiões mais abundantes do genoma nuclear de *A. cearensis* correspondem às bases adenina (A) com 33.62% e timina (T) com 34.20%. O conteúdo de guanina (G) e citosina (C) correspondeu juntos a 32.18%, sendo a maior porcentagem para G de 16.12% e a menor para C, com 16.05%.

Um total de 18.533 sequências genômicas nucleares montadas para *A. cearensis* apresentaram regiões microssatélites. 16.449 sequências apresentaram uma região microssatélite, 1.858 sequências apresentaram duas, 195 sequências apresentaram três, 28 sequências apresentaram quatro e duas sequências apresentaram cinco. O conjunto informado nas sequências genômicas apresenta 20.878 microssatélites.

Tendo em conta os critérios adotados, a distribuição dos microssatélites por tamanho de motivo de repetição foi de 17.733 regiões com tamanho dinucleotídeo, 2.632 com tamanho tetranucleotídeo e 736 com tamanho pentanucleotídeo (veja as proporções na Figura 4). A variação entre o número de repetições foi de 5- 29 para dinucleotídeo, 5-28 para tetranucleotídeo e 5-38 para pentanucleotídeo (Tabela 3).

Tabela 3: Resultados observados na identificação de regiões microssatélites nas sequências genômicas nucleares montadas para a espécie *Amburana cearensis*.

Tamanhos de repetição	Distribuição	Porcentagem	Número de repetições
Dinucleotídeo	17.733	84,93 %	5 a 29
Tetranucleotídeo	2.632	12,60 %	5 a 28
Pentanucleotídeo	513	2,47 %	5 a 38

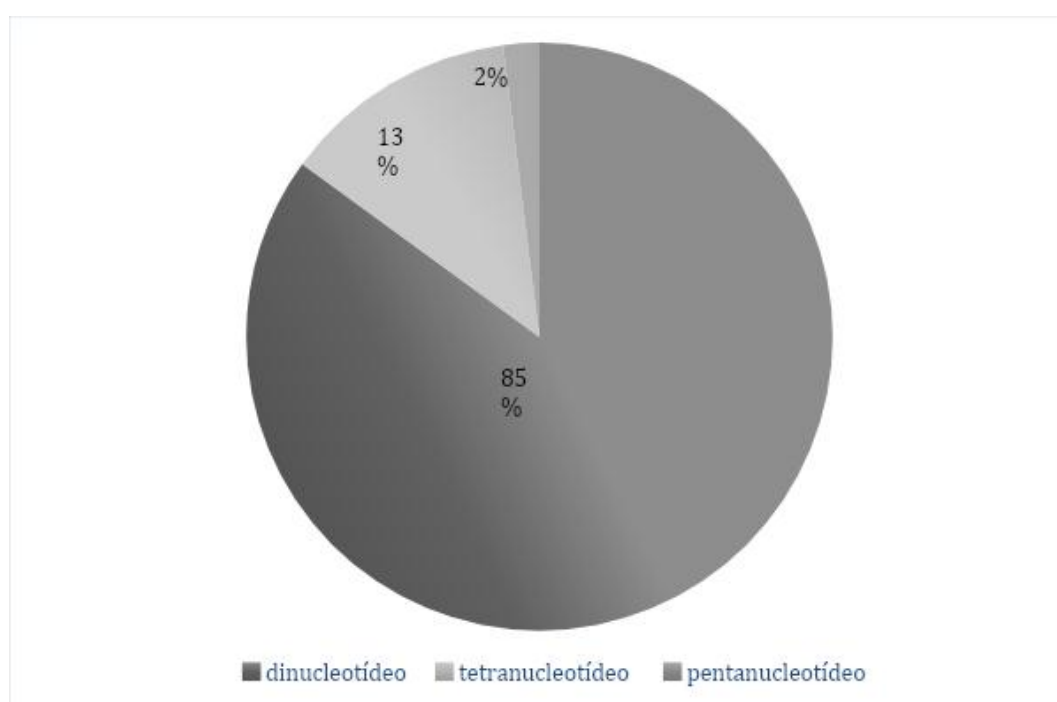


Figura 5: Distribuição dos microssatélites nucleares de *Amburana cearensis* por tamanho de motivo de repetição.

Considerando o total de regiões microssatélites identificadas e todos os tamanhos de motivos, os mais frequentes foram: para dinucleotídeos: AT (50 %) e AG (10%); para tetranucleotídeos: AAAT (12 %), TTTA (7 %), AAAG (6 %) e ATAC (6 %), e para pentanucleotídeos: AAAAG (12 %), AAAAT (10 %), ATTTT (5 %), CTTTT (5 %). Desse

modo, o tamanho de motivo mais frequente foi o dinucleotídeo e o motivo de repetição mais frequente foi AT.

Foram desenhados 28.978 pares de *primers* para 9.374 regiões microssatélites pelo *software QDD*. A distribuição dos microssatélites por tamanho de motivo de repetição foi de 24.223 sequências com tamanho dinucleotídeo, 3.814 com tamanho tetranucleotídeo e 931 com tamanho pentanucleotídeo.

Considerando as regiões microssatélites nucleares identificadas e todos os tamanhos de motivos, os mais frequentes foram, para dinucleotídeos: AT (45%), AG (36%) e AC (17%); para tetranucleotídeos: AAAT (40%), AAAG (24%), ACAT (12%), AAAC (4 %), e para pentanucleotídeos: AAAAG (41%), AAAAT (31%), AAAAC (9%), AAGAG (5%). Desse modo, o tamanho de motivo mais frequente foi o dinucleotídeo e o motivo de repetição mais frequente foi AT.

A partir dos critérios estabelecidos, trinta pares de *primers* foram selecionados para o redesenho para a amplificação em *PCR multiplex*. Desse total, dezessete são dinucleotídeos, dez tetranucleotídeos e 3 pentanucleotídeos. Os *primers* foram organizados em cinco conjuntos para a amplificação (Tabela 4), sendo que o conjunto 1 foi projetado para amplificação simultânea de cinco pares de *primers*, o conjunto 2 para seis pares de *primers*, o conjunto 3 para sete pares de *primers*, o conjunto 4 para seis pares de *primers* e o conjunto 5 para seis pares de *primers*. A temperatura de anelamento variou em 2° C graus, estando na faixa de 60° C - 62° C. Os *primers* desenhados são compostos de regiões di-, tetra- e pentanucleotídicas. Cada *primer* recebeu o prefixo “ACE”, sigla correspondente às iniciais de *A. cearensis*, seguido do número do *primer* e seu indicativo *forward* ou *reverse*. Um apêndice com cem marcadores microssatélites extras foi disponibilizado ao final deste trabalho (pág. 55).

Tabela 4: Conjuntos de *primers* nucleares organizados para amplificação em *PCR multiplex*. o conjunto 1 foi projetado para amplificação simultânea com 5 pares de *primers*, o conjunto 2, projetado para amplificação simultânea com 6 pares de *primers*, o conjunto 3, projetado para amplificação simultânea com pares 7 de *primers*, o conjunto 4, projetado para amplificação simultânea com 6 pares de *primers* e o conjunto 5, projetado para amplificação simultânea com 6 pares de *primers*.

Marcador ID	Sequência do <i>primer</i>	Motivo de repetição	Temperatura de anelamento °C	Tamanho do fragmento(pb)/Tm(°C)
CONJUNTO 1	ACE_01-F e R F: CATTGATCTTTCGTCTTTCACAGC R: TTGACGGCGGATTAACAGTGT	(AGCT) ₅	61 °C	454/78
	ACE_02-Fe R F: GTCATGAAACCATGGAAAGCTC R: AACTCCCTTGAGGCAATGTATCA	(AG) ₁₈	60 °C	308/80
	ACE_03-Fe R F:GAAGAGATAGCAAATTCGGTAGC R: CCAGAGGACCTATCATTAAGCTGT	(AC) ₁₇	61 °C	379/76
	ACE_04-F e R F:CAGAGATCATTTTTCTCCGCAAGAA R: CTGGTCGTAACGAAGAGCCTT	(AG) ₁₇	61°C	212/81
	ACE_07-F e R F: GCAATCTTTGACAAAATCGGAGAC R: CTCGAAAATCAAGCTTCGACTGT	(AGCC) ₅	60°C	131/80

continua

Marcador ID	Sequência do <i>primer</i>	Motivo de repetição	Temperatura de anelamento °C	Tamanho do fragmento(pb)/Tm(°C)	
CONJUNTO 2	ACE_05-F e R	F: TTTTAAGCACCTTCTACCATGACAA R: AGGGATTGTGTCATCTGCAGC	(AG) ₁₇	60°C	220/78
	ACE_06-F e R	F:AAGATCAGAATGGTGACCTTCATTT R: ATCACGCAATCGTATTCATCACT	(AC) ₁₆	60°C	294/79
	ACE_08-F e R	F: TGCATTGACATTCAGCACATATCTA R: GTTTGAACAGTGCTGCTTCTAAGA	(AG) ₁₆	60°C	357/78
	ACE_10-F e R	F:AAAGAAAAGCCTTTCAATTGCACAT R: AGTCCTATTGATCAGCGCTCTT	(AG) ₁₄	60°C	141/78
	ACE_11-F e R	F: TTAGGCGTCCGTGCTTGTAAT R:AATAATTGAATGCTGCGTTTACACC	(AC) ₁₃	61°C	425/74
	ACE_12-F e R	F: TGGTAGCAAAAAGGCCTGATCA R:AAGCAAGGTGTATCTTTTCAGAAGT	(ACTC) ₅	61°C	497/82

continua

Marcador ID	Sequência do <i>primer</i>	Motivo de repetição	Temperatura de anelamento °C	Tamanho do fragmento (pb)/Tm (°C)
ACE_09-F e R	F:ATCACGGCAATAATACTTAGGT R: TGCCCCAATGTTTCTTGTTTACAT	(AG) ₁₄	60°C	225/76
ACE_13-F e R	F:TCCTTAATTTCTTCGAAGGTACCCA R: TGCCCCAATGTTTCTTGTTTACAT	(ACGC) ₅	61°C	289/79
ACE_14-F e R	F:TTACCAATAGAACAGACTGAAGTCC R: TCTCCTTGCTGATGGGATAAACAT	(ACAG) ₅	60°C	345/77
CONJUNTO 3 ACE_15-F e R	F: TCCATGCATTTTCAGTGACTTGC R: TGACGAGCAATTGGAACAATTCTC	(AGCC) ₅	61°C	106/81
ACE_16-F e R	F: CACTGGCTGCATCTTCGAGTA R: ACCACTACACAAAACGGATCTGAA	(AAGAG) ₅	62°C	397/81
ACE_18-F e R	F: CTTACCTCTTCCTGTTGCACTC R:CAAGAGAAATAGATTTCGAGCGTTGA	(ACTC) ₇	60°C	156/79
ACE_25-F e R	F: ATGTTAGGTTCCACGGTAGAAGA R: TGGTTGCTTCATATTGTCAAAAACC	(AC) ₁₃	61°C	486/77

continua

Marcador ID	Sequência do primer	Motivo de repetição	Temperatura de anelamento °C	Tamanho do fragmento(pb)/Tm(°C)	
CONJUNTO 4	ACE_17-F e R	F: TTTGTTCCCTCATCGCTCCCTT R:CGATGACGGTGAATGAGTAGCTAA	(AAGG) ₅	61°C	223/82
	ACE_19-F e R	F:TTGAGCACAATTTACAAAGGGGTTA R: ACCACTATATGCACGTTTTATTCCCT	(ACGC) ₆	60°C	331/78
	ACE_20-F e R	F: CTGCCTCTAAACAATTGCGAAGT R: GTTGAATTTGTTCTGTCCAAGGGA	(AC) ₁₃	61°C	273/76
	ACE_21-F e R	F: GCAGGCTTGCCTAGTTTTTCAT R: AAAATCCC GCGAGCGTGTTAT	(ACGC) ₇	61°C	390/84
	ACE_22-F e R	F: GATTCTGCCAGTAAAAGCGTTCA R: GCACAGCTTGATGGTCATAGAGA	(AC) ₁₄	61°C	131/77
	ACE_28-F e R	F: ATCAGGTGCTGGTAGTTCACC R:AACCTTAATGGGCTGTTAAACA ACT	(ACACC) ₅	61°C	453/85

continua

	Marcador ID	Sequência do <i>primer</i>	Motivo de repetição	Temperatura de anelamento °C	Tamanho do fragmento(pb)/Tm(°C)
CONJUNTO 5	ACE_23-F e R	F:AATGTTAGGTTCCACGGTAGAAGA R: TCCTATGAACCAATATGCGAGCA	(AC) ₁₃	61°C	290/76
	ACE_24-F e R	F:TGTTTAAATTGAAAGCTTACGCCTT R: CACATTCAGTCTGGAACACTTTCC	(AG) ₂₂	60°C	397/78
	ACE_26-F e R	F: TAATTTTATCTGCAATTCCTCCGCT R: CGTAAAAGGTGCTTAAATGCCCA	(AG) ₁₀	60°C	227/75
	ACE_27-F e R	F:TCATCGTCCATAAACATCATGGAGT R: GCCTGTAACCATATATTGCTTTTCC	(AG) ₂₆	60°C	346/79
	ACE_29-F e R	F: GGAAATAAGCACTGCCACAGAA R:GTGCTGGACTTTTTCAAGTTGTAGT	(AG) ₂₂	60°C	170/74
	ACE_30-F e R	F: CCTCCTGGGTCCTAACAAAGT R: TGTGGCTCACTCTTCAGTGGA	(ACACC) 5	60°C	116/75

5.3 Regiões microssatélites e primers cloroplastidiais e mitocondriais

Visando a seleção de marcadores microssatélites promissores e maior chance de polimorfismo, foram estipulados critérios restritivos para seleção de regiões microssatélites. No entanto, os genomas organelares apresentam menores proporções de regiões microssatélites, gerando a necessidade de flexibilização dos filtros utilizados em relação ao tamanho de repetições e aos números mínimos de repetições.

Assim, um total de dez sequências genômicas montadas para o genoma cloroplastidial de *A. cearensis* apresentaram regiões microssatélites. Uma sequência apresentou cinquenta regiões microssatélites, uma sequência apresentou vinte regiões, uma sequência apresentou quinze regiões, uma sequência apresentou doze regiões, uma sequência apresentou dez regiões, uma sequência apresentou seis regiões, duas sequências apresentaram quatro regiões e uma sequência apresentou duas regiões microssatélites. Totalizando 123 regiões microssatélites.

A distribuição dos microssatélites cloroplastidiais por tamanho de motivo de repetição foi de 85 regiões microssatélites com tamanho dinucleotídeo, 28 regiões com tamanho trinucleotídeo e 8 regiões com tamanho tetranucleotídeo (Figura 5). A variação entre o número de repetições foi de 4-8 para dinucleotídeo, 4-5 para trinucleotídeo, 3-4 e para tetranucleotídeo.

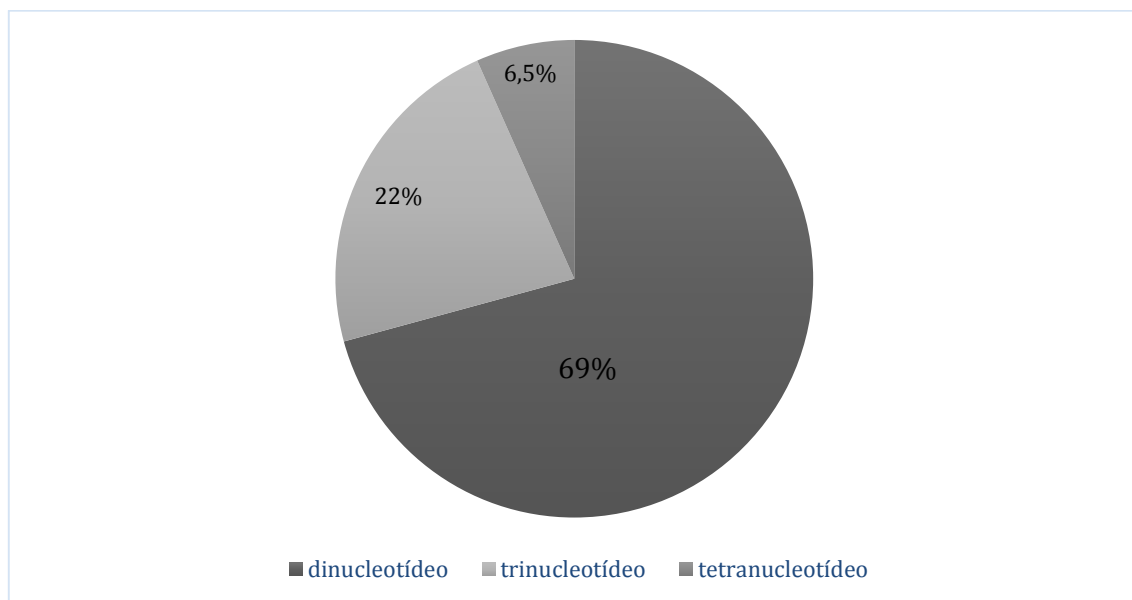


Figura 6: Distribuição dos microssatélites por tamanho de motivo de repetição no genoma

cloroplastidial de *Amburana cearensis*.

Enquanto para o genoma mitocondrial de *A. cearensis*, um 117 sequências genômicas apresentaram regiões microssatélites. Duas sequências apresentaram dez regiões microssatélites, uma sequência apresentou nove regiões, uma sequência apresentou oito regiões, uma sequência apresentou sete regiões, uma sequência apresentou seis regiões, cinco sequências apresentaram cinco regiões, doze sequências apresentaram quatro regiões, nove sequências apresentaram três regiões, 57 sequências apresentaram uma região microssatélite. Totalizando 270 regiões microssatélites.

A distribuição dos microssatélites por tamanho de motivo de repetição para o genoma mitocondrial foi de 200 regiões microssatélites com tamanho dinucleotídeo, 13 regiões com tamanho trinucleotídeo, 49 regiões com tamanho tetranucleotídeo, 6 regiões com tamanho pentanucleotídeo e 2 regiões com tamanho hexanucleotídeo. A variação entre o número de repetições foi de 3-8 para dinucleotídeo, 3-5 para trinucleotídeo, 3-6 para tetranucleotídeo e 3-4 para pentanucleotídeo e 3 para hexanucleotídeo (Figura 6).

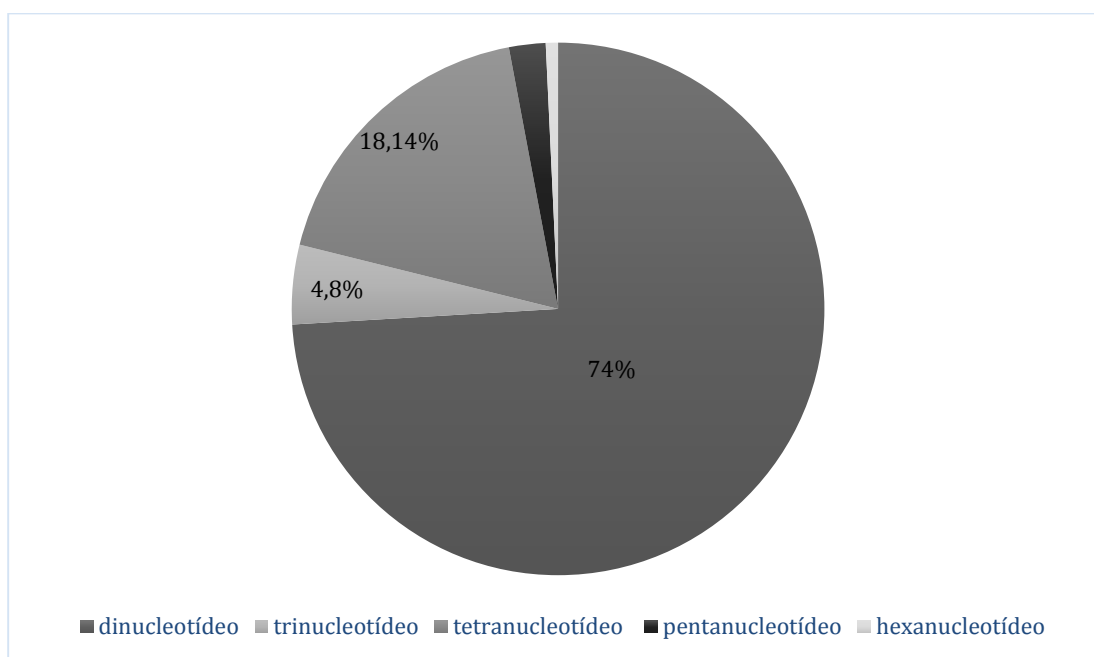


Figura 7: Distribuição dos microssatélites por tamanho de motivo de repetição no genoma mitocondrial.

Considerando o total de regiões microssatélites identificadas no genoma cloroplastidial e todos os tamanhos de motivos, os mais frequentes foram, para dinucleotídeos: AT (83 %) e AG (3,5 %), para trinucleotídeos: TTA (21%) e TAT (21%), tetranucleotídeos: TTAT (25 %) e AATA (12 %). Desse modo, o tamanho de motivo mais frequente foi dinucleotídeo e o motivo de repetição mais frequente foi AT.

Para o genoma mitocondrial, considerando o total de regiões microssatélites identificadas e todos os tamanhos de motivos, os mais frequentes foram para dinucleotídeos: AT (37 %) e AG (14,5 %); para trinucleotídeos: TTA (23%) e AAG (15%); tetranucleotídeos: CTTT (8 %) e TTCA (4%); pentanucleotídeos: CTAA (16%), AGAGA (16%) e GAATT (16%) e hexanucleotídeos (0,74%). Desse modo, o tamanho de motivo mais frequente foi o dinucleotídeo e o motivo de repetição mais frequente foi AT.

Foram desenhados três marcadores microssatélites para o genoma cloroplastidial de *A. cearensis* através do *software QDD* (Tabela 5). Todos os marcadores possuem motivo de repetição AT, sendo assim, possuem fins demonstrativos, pois não são considerados potenciais marcadores, visto que seu motivo de repetição é indesejado no desenvolvimento de marcadores microssatélites devido à formação de *hairpins*.

Tabela 5: Lista dos marcadores microssatélites desenvolvidos para o genoma cloroplastidial de *Amburana cearensis*.

<i>Primer</i>	<i>Sequência do primer</i>	Motivo de repetição
ACE_CP_01	F: AAACATGCACAGATCTAACTCC R: GTGCTTCAATTTCTTTGACAAT	(AT) ₇
ACE_CP_02	F: CAGATCATAATGAAACATTCCC R: TCTATTATTTCTGGATTCGCAA	(AT) ₅
ACE_CP_03	F: GCAGATCCAAGATGTATTCCTA R: TAATTGGTCGTGTATTAGCAGA	(AT) ₅

Inicialmente, foram desenhados 24 pares de *primers* para o genoma mitocondrial de *A. cearensis*, cujas regiões microssatélites possuem motivo de repetição do tipo dinucleotídico. Considerando o total de regiões microssatélites identificadas, as mais frequentes foram: AT (70 %), AG (25 %) e AC (4 %). Levando em consideração os critérios adotados para seleção de

potenciais marcadores, todas as regiões contendo o motivo AT foram excluídas, restando sete marcadores microssatélites selecionados (Tabela 6), com algumas observações. O número de repetições dos marcadores destacados é considerado baixo para o desenvolvimento de marcadores do tipo microssatélite, para maior confiabilidade é desejável uma repetição *em tandem* acima de 10.

Tabela 6: Seleção de marcadores microssatélites para o genoma mitocondrial de *A. cearensis*

<i>Primer</i>	<i>Sequência do primer</i>	Motivo de repetição
ACE_mit_01	F: CTTATTAGATCCAGAACCCTCC R: AGAGGCACTACTAGGAATAGGG	(AC) ₅
ACE_mit_02	F: GTTGACCCGAAATCTCATTC R: CACATCAAAGAGGTCAAGGA	(AG) ₅
ACE_mit_03	F: CACGACTTCCTATGCTCTAAAC R: CTTTACCACGCCTAACTAAATG	(AG) ₅
ACE_mit_04	F: AAATCCAAACTAAGGAAAGCAT R: TCTCTTTCAAGAAAGGATTGAA	(AG) ₅
ACE_mit_05	F: CACGACATTAGTGGCTTAGTTC R: TTTCTTTTCAGTCAAGGTCTAA	(AG) ₅
ACE_mit_06	F: TAGTAGCGTAAAGCCGTAAGTC R: GTTGGGAAGAAAGATGGAGT	(AG) ₅
ACE_mit_07	F: TGTGAGCTAGTTCCTTTGCTAT R: GAACTTTCTTCACTTCAGCTTG	(AG) ₅

6. Discussão

Foi possível realizar de forma satisfatória, a montagem das sequências genômicas para o genoma nuclear, cloroplastidial e mitocondrial de *A. cearensis*. O tamanho do genoma nuclear foi de 285 Mb, enquanto que os genomas cloroplastidial e mitocondrial apresentaram tamanho de 136 Kb e 551 Kb, respectivamente. Por meio deste volume de dados foi possível acessar uma quantidade considerável de regiões microssatélites da espécie. Tais resultados demonstram que não é necessário o sequenciamento completo do genoma de uma espécie para seleção de regiões microssatélites promissoras, conforme o observado em outros trabalhos: MARTINS (2013), SILVA (2016) e SANTOS (2018).

Com base nos rascunhos dos três genomas analisados, foi observado que o tamanho de motivo mais frequente correspondeu aos dinucleotídeos e o motivo de repetição ao tipo AT. Apesar do motivo AT ser constante em genomas de plantas, seu uso não é indicado para o desenho de marcadores microssatélites devido à possibilidade de geração de *hairpin*, que pode diminuir a eficiência da amplificação durante a PCR (ANTUNES, 2017; FAN et al., 2019). Em função disso, as regiões AT não foram priorizadas para o desenho de marcadores neste trabalho. As regiões com motivos mono, tri e hexanucleotídeos também foram excluídas. Os mononucleotídicos apresentam maior susceptibilidade a erros de deslizamento durante a etapa de PCR. Já os tri e hexanucleotídeos possuem presença frequente em regiões gênicas, acarretando marcadores microssatélites com baixo polimorfismo (XU et al., 2013; ANTUNES, 2017; FAN et al., 2019).

Um critério relevante adotado para a escolha das regiões microssatélites foi a preferência por regiões com maior número de repetição em *tandem*. Isso pode ser justificado por ZALAPA et al. (2012), no qual relata que regiões microssatélites longas e perfeitas possuem maior probabilidade de serem polimórficas. Quanto mais longa e perfeita a repetição microssatélite, maior a frequência de mutação, enquanto que as repetições curtas e imperfeitas, possuem frequência de mutação mais baixa (VIEIRA et al., 2016). Segundo NADEEM et al. (2017), para uma PCR bem sucedida, o desenvolvimento de um marcador é o parâmetro mais importante. Se todos os pontos estiverem equilibrados, exceto o marcador, levará a falhas no funcionamento do protocolo de PCR.

Este trabalho permitiu a seleção de regiões microssatélites de alta qualidade para os

três genomas analisados, principalmente para o genoma nuclear, por possuir maior tamanho genômico e, conseqüentemente, maior quantidade de informações a serem extraídas. Foram desenhados 28.978 marcadores microssatélites para 9.374 regiões microssatélites nucleares. Assim, uma quantidade significativa de regiões foi submetida a filtros de qualidade rigorosos a fim de selecionar as regiões mais promissoras e favoráveis para o desenvolvimento de marcadores microssatélites.

Uma tabela com trinta marcadores microssatélites nucleares promissores para a análise de diversidade genética da espécie *A. cearensis* foi disponibilizada neste trabalho. Essa quantidade de marcadores microssatélites é interessante para estudos de genética de populações pois gera uma boa margem para testes. Comparando com a literatura, normalmente os autores desenvolvem números inferiores de marcadores microssatélites: BARATELI (2018) desenhou 20 marcadores microssatélites para o genoma de *Stryphnodendron adstringens*, sendo 16 polimórficos; SHERMAN et al. (2012) desenhou 23 marcadores microssatélites para a espécie *Zostera muelleri*, com amplificação total e LIU et al. (2020) desenhou 16 marcadores microssatélites para *Angelica dahurica*, sendo 14 polimórficos. Ambos autores utilizando a tecnologia de Sequenciamento de Nova Geração.

Além da vantagem do número alto de marcadores nucleares disponibilizados, os trinta marcadores foram desenhados para amplificação em PCR *multiplex*, uma alternativa extremamente positiva em estudos com espécies florestais com grande potencial de uso, que é o caso *A. cearensis*. Em estudos de genética de populações com espécies florestais geralmente realizam-se análises em larga escala, utilizando quantidade significativa de indivíduos. A multiplexação envolve a amplificação simultânea de vários loci em uma única reação de PCR, possibilitando um maior acesso de informações genéticas a um menor custo, pois é possível a redução do consumo de reagentes de alto valor financeiro sem comprometer a utilidade do teste, além da redução do trabalho laboratorial. Todos os marcadores microssatélites de *A. cearensis* foram projetados para serem amplificados em condições de temperatura e quantidade de reagentes semelhantes na PCR *multiplex*, (RENSHAW e MARK et al., 2007; GUICHOUX, 2011; LI et al., 2018; GUERRERO-CÓZAR et al., 2020).

Considerando o processo de seleção de regiões e de desenho de marcadores microssatélites, a transferibilidade de marcadores de uma espécie para outra é bastante oportuna, principalmente para espécies do mesmo gênero. Os marcadores nucleares desenhados para *A. cearensis* podem ser testados também em outras espécies do gênero, *Amburana acreana*

e *Amburana eritrosperma*.

Marcadores microssatélites organelares também foram disponibilizados neste trabalho, porém com algumas observações quanto às suas características, pois mesmo com a flexibilização adotada, foram encontradas poucas regiões microssatélites para esses genomas. Foram desenhados três pares de *primers* para amplificação de regiões microssatélites no genoma cloroplastidial de *A. cearensis*. Os marcadores cloroplastidiais foram desenhados com a ressalva de que não possuem todos os atributos necessários para um marcador promissor, principalmente porque os três marcadores possuem motivo de repetição AT e baixo número de repetição em *tandem*. Como discutido anteriormente, marcadores microssatélites com repetições AT tem possibilidade de amplificação com geração de *hairpin*, conhecido também como “grampo de cabelo”, que pode diminuir a eficiência da PCR (XU et al., 2013; ANTUNES, 2017). Além disso, o baixo número de repetições em *tandem* está associado à baixa possibilidade de polimorfismo: SINGH et al. (2017) desenvolveram marcadores microssatélites cloroplastidiais com motivo AT para *Gladiolus*, porém estes marcadores não foram selecionados para caracterizar e estimar a diversidade genética dos cultivares. XU et al. (2015) desenvolveram marcadores cloroplastidiais com motivo AT para *Ginkgo biloba*, no entanto esses marcadores não apresentaram polimorfismos. Porém, mesmo que o motivo AT possa apresentar problemas para a qualidade dos marcadores microssatélites, na literatura há trabalhos que utilizam marcadores cloroplastidiais com esse tipo de motivo de repetição (PARK et al., 2016; BEYER et al., 2017; DENG et al., 2017; SINGH et al., 2017). Logo, sugere-se que sejam realizados testes para a verificação da qualidade e polimorfismo dos pares de *primers* para microssatélites cloroplastidiais, disponibilizados no presente estudo.

Foram desenhados sete pares de *primers* para a amplificação de regiões microssatélites do genoma mitocondrial de *A. cearensis*. Todos os marcadores disponibilizados possuem tamanho de motivo dinucleotídico (AC e AG) com 5 repetições em *tandem*. É desejável que marcadores microssatélites possuam pelo menos 10 repetições em *tandem* para motivos de repetição do tipo dinucleotídico. No entanto, na literatura há trabalhos desenvolvidos para marcadores mitocondriais que utilizam marcadores com baixo número de repetições: KHERA et al. (2015) desenvolveram mtSSR para feijão *bóer* com motivo (AG)₆ que apresentou regiões polimórficas, sendo úteis para distinguir os genótipos da espécie em estudos evolutivos e filogenéticos. WANG et al. (2017) também desenvolveram e testaram marcadores mtSSR em *Millettia pinnata* com motivos (GA)₆ e (CT)₆ que apresentaram polimorfismos e se mostraram

úteis na avaliação da diversidade genética da espécie.

Em suma, mesmo que tais marcadores não preencham todos os requisitos estipulados neste trabalho, eles ainda preenchem a maioria dos requisitos, tais como: foram desenhados a partir de regiões com microsatélites perfeitos, não possuem motivos de repetição AT, não possuem muitas bases iguais (ex: AAAG) e não estão associados a elementos transponíveis. Desse modo, possuem competência para serem investigados.

7. Conclusão

Foram desenhados trinta pares de *primers* para regiões microsatélites nucleares de *A. cearensis*, desenvolvidos para amplificação em *PCR multiplex* e com potencial para discriminação individual e transferibilidade para as outras espécies do gênero *Amburana*. Além disso, foram desenhados três pares de *primers* para regiões microsatélites cloroplastidiais e sete para microsatélites mitocondriais, com competência para serem investigados quanto ao polimorfismo. Os resultados do presente estudo tornam possível a disponibilização de novos marcadores microsatélites que podem facilitar o estudo da diversidade e estrutura genética de *A. cearensis* e do gênero *Amburana*.

8. Apêndice

Tabela 7: Apêndice com cem marcadores microssatélites nucleares extras para a espécie *A. cearensis*.

Marcador ID	Sequência do <i>primer</i> (F+R)	Motivo de repetição	Temperatura de anelamento C	Tamanho do fragmento
ACE_01	F: GTTAAGTTCAGAGGCAGAATTG R: TGTCATGGTAATGGTCTTCTTT	(AC) ²⁹	55	380
ACE_02	F: GCCTTTCACATAATCAACTTTC R: TTTGAAGAGGAGAAAGAAATGA	(AC) ²⁹	55	224
ACE_03	F: CACTTGTATAGCTTGCGTATGA R: TTTGAAGAGGAGAAAGAAATGA	(AC) ²⁹	55	318
ACE_04	F: TATCTCCAAGCGAAGTTAAGAA R: GAGTGTAGCCAGTGATCTGAAT	(AG) ²⁶	55	379
ACE_05	F: CTAGCATTAGGAATATGCAAGC R: TTGTTGACAGTGATGTCTCCTA	(AG) ²⁶	55	226
ACE_06	F: AGCCAATCTTTATGGGATACTT R: GAGTGTAGCCAGTGATCTGAAT	(AG) ²⁶	56	326
ACE_07	F: GTGAGAGTGAGAGTGTGTCCTC R: TGTCCACTTCAACATTCAAAG	(AG) ²⁶	56	158
ACE_08	F: TCGTCTGTATTATTCTTCGGAT R: ATGCAGCACACATATTACATTG	(AC) ²⁵	56	321

ACE_09	F: TCGTCTGTATTATTCTTCGGAT R: ATGCAGCACACATATTACATTG	(AC) ²⁵	56	380
ACE_10	F: ACTAAATAGCACACTTGCATCC R: GGGAATAGTAACAACAGCGTTA	(AC) ²⁵	55	290
ACE_11	F: TGGTTTCTGGAACTAAATATGC R: GGGAATAGTAACAACAGCGTTA	(AC) ²⁵	55	233
ACE_12	F: ACTAAATAGCACACTTGCATCC R: GCATTAATGGTCTAATGCGTAG	(AC) ²⁵	55	380
ACE_13	F: GGAAAGTGCAAGAAACACAA R: TTCTTCTTCTGCTTCTCTTCTG	(AG) ²³	56	191
ACE_14	F: AACGATGAACTCGGTTATTATG R: TTCTTCTTCTGCTTCTCTTCTG	(AG) ²³	56	319
ACE_15	F: ATGGCCATGTTGTGATTAGTAT R: AAATGAATATGCCTGTGAAATC	(AG) ²³	56	255
ACE_16	F: AATCTAGGGTGAAGACCTGAAT R: AAATGAATATGCCTGTGAAATC	(AG) ²³	56	392
ACE_17	F: CACTAAACATTGCATTTGAACA R: AAATGAATATGCCTGTGAAATC	(AG) ²³	56	192
ACE_18	F: AGTACAAGCATTTCATACACACG R: TTAGCGAATATGAGGATTGTTT	(AG) ²³	56	237
ACE_19	F: AGTACAAGCATTTCATACACACG R: AGCTTCTACTAGTGGTCAGCTCT	(AG) ²³	55	260
ACE_20	F: AGTACAAGCATTTCATACACACG R: AATGCTTATTGATCACGACTTT	(AG) ²³	55	361

ACE_21	F: CCGCTTAAATAATGTGACTAGC R:TTTCGTCCAGAGTGATAAAGAG	(AG) ²²	55	219
ACE_22	F:AACTAAGACTCTCTTACCAAACACA R:TTTCGTCCAGAGTGATAAAGAG	(AG) ²²	55	384
ACE_23	F: CCGCTTAAATAATGTGACTAGC R:GAGAAAGGAAGGGTGTTTAGTT	(AG) ²²	55	267
ACE_24	F: AGTTGAGCTTATTGGTCAATTT R:TTTAGACTCGGCTCATCAATA	(AG) ²²	54	384
ACE_25	F: CCTTCTCCATAGTAACCATCA R:TAAGCCAATGAAACAAATTACC	(AC) ²²	54	357
ACE_26	F: CTTCTCCTCATCATCCTAGACA R:CGATATAGCAAAGAACAGAAGG	(AC) ²²	54	265
ACE_27	F: ATGGATGCTCAAATAAGACAAC R:GATAAAGACCTAATTGCCACAA	(AC) ²²	55	398
ACE_28	F: ATGGATGCTCAAATAAGACAAC R:CCCAAATAAGCTTAATCAACAA	(AC) ²²	55	285
ACE_29	F:TAGAGTAGCAGCGAAAGCTAAT R:CCCAAATAAGCTTAATCAACAA	(AC) ²²	55	203
ACE_30	F:ATTTGGTTAGTCTCCCAACTGT R:AGAAATAAGAGGAATAACGGGA	(AG) ²¹	56	211
ACE_31	F: CTGTTCTTTCCACTTCATTGTT R:GAAGGAACAACAGGCTCTAGTA	(AG) ²¹	56	358
ACE_32	F: CTGTTCTTTCCACTTCATTGTT R:AGTACAAATGGCATCTTAGCTC	(AG) ²¹	56	322

ACE_33	F:TACTTGGGTTTGGTAGAGAATG R:GATAAGCGTGTGTCTCTGATTT	(AG) ²¹	56	223
ACE_34	F: ATGTTTATTAGCACGTAAACGG R:GATAAGCGTGTGTCTCTGATTT	(AG) ²¹	56	329
ACE_35	F: ATGTTTATTAGCACGTAAACGG R:GAAAGTCAGCTTTGAGGTACAA	(AG) ²¹	56	398
ACE_36	F:CAATATGCCTTGTCTTCTCTTT R:AACGAGTTATGCCTTCAATTT	(AG) ²¹	55	316
ACE_37	F: CAATATGCCTTGTCTTCTCTTT R:GGCAAACATTAATCACAATGA	(AG) ²¹	55	350
ACE_38	F: CAATATGCCTTGTCTTCTCTTT R:GAAATACCAATTCTGCCATAAA	(AG) ²¹	55	173
ACE_39	F: CACTCCAGTAATTTGAGCTACG R:TGATTGCACTCTATCATTTGAA	(AG) ²¹	55	301
ACE_40	F: CACTCCAGTAATTTGAGCTACG R:AAAGTCACATTAATCACAATGC	(AG) ²¹	55	248
ACE_41	F:AACCTGGGTTCCCTAATAAACTT R:AAAGTCACATTAATCACAATGC	(AG) ²¹	55	358
ACE_42	F: TCTACAGCATTGTATCATTCG R:ATAGGGTTTAAGAGGAAGCAAG	(AG) ²¹	55	240
ACE_43	F:TCAGACATAGGAATGAAGAAGC R:ATAGGGTTTAAGAGGAAGCAAG	(AG) ²¹	56	340
ACE_44	F:AAGGTTGAAGAATGAGATCAGA R:ATAGGGTTTAAGAGGAAGCAAG	(AG) ²¹	56	357

ACE_45	F:ATTCTGACTTGTCTCCTCCG R:GAGGTGAGATTGGAGACGAG	(AG) ²¹	56	355
ACE_46	F:GTTGTCCTCCCATTATCTTCTT R:GAGGTGAGATTGGAGACGAG	(AG) ²¹	56	336
ACE_47	F:GCTTCTTCTATCTTCCTCAGTTT R:ACATGAATTGGTGAGTGAGAAT	(AG) ²¹	56	161
ACE_48	F:TTGATCTCTCTCTTGTCCAATC R:AATTAATTCCAGGTTTCATGTG	(AG) ²¹	56	383
ACE_49	F:TTGACTTCCCTAAACTTGACTG R:TTCCTAAAGCTGATGAAACTGT	(AC) ²⁰	56	205
ACE_50	F:CAATTATCATTGAGTCTCGGT R:ATGAGTACGTTTGTGAGCAAAT	(AC) ²⁰	56	305
ACE_51	F:CTTGAACATGGTTTGTCAAAGT R:TGAGAAATGATTTAGTGTGACTTG	(AC) ²⁰	56	189
ACE_52	F:CTTGAACATGGTTTGTCAAAGT R:TACCTCATGGATGAGTACGTTT	(AC) ²⁰	56	359
ACE_53	F:CTTTCATTGCTTTCATCTCATC R:GTCTCTTTCACAGACAAGCAAG	(AG) ²⁰	56	349
ACE_54	F:CTTTCATTGCTTTCATCTCATC R:ATTAGGACTACAGGATCGGTCT	(AG) ²⁰	56	363
ACE_55	F:CTTTCATTGCTTTCATCTCATC R:TCAGGAGTAGTGAAATGAGGAC	(AG) ²⁰	56	318
ACE_56	F:AACTCTTCCAGAAATTGAAACA R:TTGTACCCTTAACAATACACCC	(AG) ²⁰	55	392

ACE_57	F:AACTGGGAACAGGAATATCAAT R:TTGTACCCTTAACAATACACCC	(AG) ₂₀	55	242
ACE_58	F:CTGTGAAAGTGAAGATGGAGAT R:TTGTACCCTTAACAATACACCC	(AG) ₂₀	55	324
ACE_59	F:TATTGACTACTCACCACAACCA R:GGTTGGATGTAATTCAGTATGG	(AG) ₂₀	55	246
ACE_60	F:GAAGATCAACGGAGTTCAGTAA R:CAACGTTGAATTATTAGGGAAG	(AG) ₂₀	55	264
ACE_61	F:GAAGATCAACGGAGTTCAGTAA R:GGATGTAATTCAGTATGGTTGG	(AG) ₂₀	55	355
ACE_62	F:AGCTCAGATACAAGCTCCATAA R:ACAGAATATGCTTTGCTCTGTT	(AG) ₂₀	55	240
ACE_63	F:AGGGAAAGAGAGTATTTGAGGA R:ATAACGATGAGGGAGTCAGTC	(AG) ₂₀	56	400
ACE_64	F:AGCTCAGATACAAGCTCCATAA R:CCTTTCCTCACTCTTCTTCT	(AG) ₂₀	56	325
ACE_65	F:GCCATGTATGAATAATTTAGCC R:AAATACCTTCTCAATCAGTTCG	(AG) ₂₀	55	332
ACE_66	F:GCCATGTATGAATAATTTAGCC R:TACAAAGAGAATAACAACGCAG	(AG) ₂₀	55	183
ACE_67	F:ACCTGCAACAATATTCCAATTA R:ACATATTGGTTGTTCCAACTTC	(AG) ₂₀	55	394
ACE_68	F:TCAAATCACTAATGTGTTGCAG R:GAGGTAAGGAATTGGGAGACTTC	(AG) ₂₀	55	261

ACE_69	F:CTATAAATTCCACAACCACCAC R:GAGGTAAGGAATTGGAGACTTC	(AG) ₂₀	56	361
ACE_70	F:CTATAAATTCCACAACCACCAC R:TGCCATGATCACAAGTATCTTA	(AG) ₂₀	56	223
ACE_71	F:CATATGCACACATGCAACTATT R:AAGATTCAGAGAAGGTTGGAAA	(AG) ₂₀	56	342
ACE_72	F:CATATGCACACATGCAACTATT R:ACTTGTACAGAGCAAGCTCATC	(AG) ₂₀	56	376
ACE_73	F:CATATGCACACATGCAACTATT R:GGGTTTAGTTTCATCAGCTACTT	(AG) ₂₀	56	244
ACE_74	F:CCAATTATGCATGAATCTCACT R:CATAGTTTCCCATCATCTTCAA	(AG) ₂₀	56	292
ACE_75	F:CCAATTATGCATGAATCTCACT R:ATCCTCATCCTCCAAGTACCTA	(AG) ₂₀	55	176
ACE_76	F:CCAATTATGCATGAATCTCACT R:AAGAGCAACACTCCTCTCAAC	(AG) ₂₀	55	350
ACE_77	F:GAATTTCTTCACGCATAGGTAA R:ATACATTCGCGTATTTCTACGA	(AG) ₂₀	55	239
ACE_78	F:CCTATCAATAGCAGTGTCGTTT R:ACTTGTACAGAGCAAGCTCATC	(AG) ₂₀	55	338
ACE_79	F:CATATGCACACATGCAACTATT R:GGGTTTAGTTTCATCAGCTACTT	(AG) ₂₀	56	357
ACE_80	F:CCAATTATGCATGAATCTCACT R:CATAGTTTCCCATCATCTTCAA	(AG) ₂₀	56	376

ACE_81	F:CCAATTATGCATGAATCTCACT R:ATCCTCATCCTCCAAGTACCTA	(AG) ²⁰	56	219
ACE_82	F:CCAATTATGCATGAATCTCACT R:AAGAGCAACACTCCTCTCAAC	(AG) ²⁰	56	336
ACE_83	F:GAAATTGGACCTTGACTTTGTA R:TTACATTGTGAATTGTGATGCT	(AC) ¹⁹	55	196
ACE_84	F:TTTGTTATGGCTGTTATGATGA R:CAAAGACAAGAAGAAAGACGAC	(AC) ¹⁹	55	337
ACE_85	F:GAAATTGGACCTTGACTTTGTA R:GAAAGACGACCAATGAAGAGTA	(AC) ¹⁹	55	395
ACE_86	F:TTTAGATGAAAGATATTGCGTG R:TTAAATCACTTTGAGATGGTCC	(AC) ¹⁹	55	174
ACE_87	F:TTAGATGAAAGATATTGCGTGG R:CGCCTAAGTTTCAATTCTGTAA	(AC) ¹⁹	57	390
ACE_88	F:TTTAGATGAAAGATATTGCGTG R:TAAATTAAACACAACCCAATCC	(AC) ¹⁹	54	305
ACE_89	F:TACAAATCAGCCAATCTTTCTC R:CAGCCTTGTAACCTTTCATTAG	(AC) ¹⁹	56	347
ACE_90	F:ATCATATAGAGGCCAGACATTG R:ATCATTCCCTCGTACTTCACTCA	(AC) ¹⁹	55	203
ACE_91	F:ATCATATAGAGGCCAGACATTG R:CAGCCTTGTAACCTTTCATTAG	(AC) ¹⁹	55	380
ACE_92	F:CCATCCTCTTAATTGTTGATGT R:TTGGGTGAAGTGAAGAATTTAT	(AC) ¹⁹	55	226

ACE_93	F:CCATCCTCTTAATTGTTGATGT R: CGAAAGGGTTAGAGGATAAGAT	(AC) ¹⁹	55	388
ACE_94	F: AATCACCCACCGTATCTATCTC R:CTTCCTCTTTGAGCACCATT	(AAGG) ⁸	57	352
ACE_95	F: GAGCAATTATGTTCTTCTTCCA R:GGTTCTCAAGTTCATTCTCTCA	(AAACC) ⁸	57	173
ACE_96	F:AGATGTCTCGGTAATTCACAAT R:GTTTCATGGAAGAGTGAATTTGT	(AACC) ⁷	55	282
ACE_97	F:TGGAACTTGGTATAGGTATGG R:CCTCTGAGAACTAGAGAGGACC	(AATT) ⁷	56	233
ACE_98	F: TGATTTGAACCTACACCCTTAC R:ATCCATGTTTCATTCATGTGTCT	(ACAT) ⁷	56	208
ACE_99	F:GAGAGAGAATGGGTAGTGATGA R:GTGACGAGCTAAAGTTTCTGAG	(AGAGG) ⁶	55	320
ACE_100	F:CTTCTTCAACAATCTCTCCGT R:CAATCCCATGTTTGATACCTAC	(AGCCG) ⁵	56	376

9. Referências

- ADHIKARI, S. et al. Application of molecular markers in plant genome analysis: a review. **The Nucleus**, v. 60, n. 3, p. 283-297, 2017.
- ALLEN, John F. Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. **Proceedings of the National Academy of Sciences**, v. 112, n. 33, p. 10231-10238, 2015.
- ALMEIDA, J. P. N. et al. Germination and development of *Amburana cearensis* seedlings as influenced by seed weight, light and temperature. **Acta Scientiarum. Agronomy**, v. 39, n. 4, p. 525-533, 2017.
- ALVIM, B. F. M. et al. In vitro conservation of *Amburana cearensis* (Fabaceae). **Ciência Rural**, v. 50, n. 7, 2020.
- ANDREWS, S. et al. **FastQC**: a quality control tool for high throughput sequence data. 2010. Disponível em: < <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.>. Acesso em: 20 de janeiro de 2021.
- ANTONIO, R. P. et al. Identificação e mapeamento de QTLs de resistência do meloeiro à podridão do colo causada por *Fusarium solani* f. sp. cucurbitae. **Embrapa Semiárido-Artigo em periódico indexado (ALICE)**, 2019.
- ANTUNES, A. M. A. **Tamanho, montagem de novo e anotação do genoma de *Dipteryx alata* (Leguminosae)**. 2017. Tese de doutorado. Goiânia, Goiás.
- ARAÚJO, J. R. da S.. **Avaliação do potencial antioxidante, citotóxico e genotóxico do extrato foliar de *Amburana Cearensis* (Allemao) AC SM**. 2018. Dissertação de Mestrado. Recife, Pernambuco.
- BARATELI, L. O.. **Desenvolvimento de marcadores microssatélites para *Stryphnodendron adstringens* (barbatimão-Fabaceae)**. 2018. Dissertação de Mestrado. Goiânia, Goiás.
- BARRAL, E. C. **Biologia reprodutiva, ecologia de populações e filogeografia de *Amburana cearensis*, leguminosae ameaçada de extinção**. 2018. Tese de doutorado. Recife, Pernambuco.
- BASTÍAS, A. et al. Draft genome sequence data of maqui (*Aristotelia chilensis*) and identification of SSR markers. **Data in brief**, v. 27, p. 104545, 2019.

BHANDARI, H. R. et al. Assessment of genetic diversity in crop plants-an overview. **Adv. Plants Agric. Res**, v. 7, n. 3, p. 00255, 2017.

BHATTARAI, G.; MEHLENBACHER, S. A. In silico development and characterization of tri-nucleotide simple sequence repeat markers in hazelnut (*Corylus avellana* L.). **PloS one**, v. 12, n. 5, 2017.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, 2014.

BRAGA, R. S. **Genética geográfica de *Hymenaea stigonocarpa* (FABACEAE)**. 2015. Dissertação de mestrado. Goiânia, Goiás.

BUSO, GSC et al. Transferibilidade de Primers Microsatélites Desenhados de Sequências Expressas de *Brachiaria brizantha* para *B. decumbens* e *B. humidicola*. **Embrapa Recursos Genéticos e Biotecnologia-Comunicado Técnico (INFOTECA-E)**, 2016.

CANUTO, K. M.; SILVEIRA, E. R. Constituintes químicos da casca do caule de *Amburana cearensis* AC Smith. **Química Nova**, v. 29, n. 6, p. 1241, 2006.

CARVALHO, P.E. R. Cerejeira-da-Amazônia-Amburana acreana. **Embrapa Florestas-Circular Técnica (INFOTECA-E)**, Paraná, 2007.

CARVALHO, M. C. da C. G.; SILVA, D. C. G da. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, n. 3, p. 735-744, 2010.

CNCFlora. *Amburana cearensis* in Lista Vermelha da flora brasileira versão 2012.2 Centro Nacional de Conservação da Flora. Disponível em <[http://cncflora.jbrj.gov.br/portal/pt-br/profile/Amburana cearensis](http://cncflora.jbrj.gov.br/portal/pt-br/profile/Amburana%20cearensis)>. Acesso em 9 julho 2021.

DANIELL, H. et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. **Genome biology**, v. 17, n. 1, p. 1-29, 2016.

DAR, A. A.; MAHAJAN, R.; SHARMA, S.. Molecular markers for the characterization and conservation of plant genetic resources. **Indian J Agric Sci**, v. 89, p. 1755-1763, 2019.

D'HAENE, B. et al. Accurate and objective copy number profiling using real-time quantitative PCR. **Methods**, v. 50, n. 4, p. 262-270, 2010.

ELLEGREN, H. Microsatellite mutations in the germline: implications for evolutionary inference. **Trends in genetics**, v. 16, n. 12, p. 551-558, 2000.

FAN, H. et al. Effects of secondary structures of DNA templates on the quantification of qPCR. **Journal of Biomolecular Structure and Dynamics**, v. 37, n. 11, p. 2867-2874, 2019.

FACHIM, E.; GUARIM, V. L. M. S. Conservação da biodiversidade: espécies da flora de Mato Grosso. **Acta Botanica Brasilica**, v. 9, n. 2, p. 281-287, 1995.

FILICHKIN, S. A. et al. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. **Genome research**, v. 20, n. 1, p. 45-58, 2010.

Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB128477>>. Acesso em: 26 jul. 2021

GALTIER, N.. The intriguing evolutionary dynamics of plant mitochondrial DNA. **BMC biology**, v. 9, n. 1, p. 1-3, 2011.

GARRIGAN, Daniel; HAMMER, Michael F. Reconstructing human origins in the genomic era. **Nature Reviews Genetics**, v. 7, n. 9, p. 669-680, 2006.

GUERRERO-CÓZAR, I. et al. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. **Scientific reports**, v. 10, n. 1, p. 1-16, 2020.

GUIMARÃES, R. A. **Variabilidade genética de *Dipteryx alata* Vogel (Leguminosae): novos marcadores microssatélites, fluxo de pólen in situ e ex situ e genética da paisagem em escala local**. 2018. Tese de doutorado. Goiânia, Goiás.

GRATTAPAGLIA, D.; FERREIRA, M. E. Introdução ao uso de marcadores moleculares em análise genética. **EMBRAPA, Brasília**, v. 3, 1998.

GIULIETTI, A.M.; RAPINI, A.; ANDRADE, M.J.G.; QUEIROZ, L.P.; SILVA, J.M.C. **Plantas raras do Brasil**, 496p, 2009.

GÓMEZ-RODRÍGUEZ, M. V. et al. Identification of an olive (*Olea europaea* L.) core collection with a new set of SSR markers. **Genetic Resources and Crop Evolution**, v. 68, n. 1, p. 117-133, 2021.

GUICHOUX, E. et al. Current trends in microsatellite genotyping. **Molecular ecology resources**, v. 11, n. 4, p. 591-611, 2011.

GUIMARÃES, R. A. **Variabilidade genética de *Dipteryx alata* Vogel (Leguminosae): novos marcadores microssatélites, fluxo de pólen in situ e ex situ e genética da paisagem em escala local**. 2018. Tese de doutorado. Goiânia, Goiás.

HAVIRD, J. C. et al. Causes and consequences of rapidly evolving mtDNA in a plant lineage. **Genome biology and evolution**, v. 9, n. 2, p. 323, 2017.

- HE, Q. et al. Primers are decisive for sensitivity of PCR. **Biotechniques**, v. 17, n. 1, p. 82, 84, 86-7, 1994.
- HILMARSSON, H. S. et al. Population genetic analysis of a global collection of *Fragaria vesca* using microsatellite markers. **PloS one**, v. 12, n. 8, 2017.
- IDREES, M.; IRSHAD, M. Molecular markers in plants for analysis of genetic diversity: a review. **European Academic Research**, v. 2, n. 1, p. 1513-1540, 2014.
- JONES, N. et al. Markers and mapping revisited: finding your gene. **New Phytologist**, v. 183, n. 4, p. 935-966, 2009.
- KACZMARCZYK, D. Techniques based on the polymorphism of microsatellite DNA as tools for conservation of endangered populations. **Applied Ecology and Environmental Research**, v. 17, n. 2, p. 1599-1615, 2019.
- KALIA, R. K. et al. Microsatellite markers: an overview of the recent progress in plants. **Euphytica**, v. 177, n. 3, p. 309-334, 2011.
- KAPIL, A. et al. Data mining to detect common, unique, and polymorphic simple sequence repeats. In: **Bioinformatics: Sequences, Structures, Phylogeny**. Springer, Singapore, 2018. p. 141-154.
- KHERA, P. et al. Mitochondrial SSRs and their utility in distinguishing wild species, CMS lines and maintainer lines in pigeonpea (*Cajanus cajan* L.). **Euphytica**, v. 206, n. 3, p. 737-746, 2015.
- KUMAR, B.; KUMAR, Umesh; YADAV, Hemant Kumar. Identification of EST-SSRs and molecular diversity analysis in *Mentha piperita*. **The Crop Journal**, v. 3, n. 4, p. 335-342, 2015.
- KUMAR, D et al. Mining and Characterizing the SSR Markers for Black Rice using the Illumina Sequencing Platform. **Preprints**, 2020.
- LATALIZA, A. A. B. et al. Antioxidant, cytotoxic, antiproliferative and acetylcholinesterase inhibition properties of the extract from *Amburana cearensis*. **Journal of Multidisciplinary Engineering Science and Technology**, v. 6, p. 9336-9343, 2019.
- LEITE, V. G. et al. Floral development of the early-branching papilionoid legume *Amburana cearensis* (Leguminosae) reveals rare and novel characters. **International Journal of Plant Sciences**, v. 176, n. 1, p. 94-106, 2015.
- LI, Da et al. A multiplex microsatellite PCR method for evaluating genetic diversity in grass carp (*Ctenopharyngodon idellus*). **Aquaculture and fisheries**, v. 3, n. 6, p. 238-245, 2018.
- LIU, H. et al. Sex-specific markers developed by next-generation sequencing confirmed an

XX/XY sex determination system in bighead carp (*Hypophthalmichthys nobilis*) and silver carp (*Hypophthalmichthys molitrix*). **Dna Research**, v. 25, n. 3, p. 257-264, 2018.

LIU, Q. et al. Development and characterization of 16 novel microsatellite markers by Transcriptome sequencing for *Angelica dahurica* and test for cross-species amplification. **BMC plant biology**, v. 20, n. 1, p. 1-8, 2020.

LORENZI, H. **Arvores brasileiras- manual de identificação e cultivo de plantas arbóreas do Brasil**, Nova Odessa, 2 ed., v.2, 1998.

MA, S. et al. An RNA sequencing transcriptome analysis and development of EST-SSR markers in Chinese hawthorn through Illumina sequencing. **Forests**, v. 10, n. 2, pág. 82, 2019.

MARDIS, E. R. Next-generation DNA sequencing methods. **Annual Review Genomics Human Genetics**, v. 9, p. 387-402, 2008.

MARDIS, E. R. Next-generation sequencing platforms. **Annual review of analytical chemistry**, v. 6, p. 287-303, 2013.

MARKOULATOS, P.; SIAFAKAS, N.; MONCANY, M. Multiplex polymerase chain reaction: a practical approach. **Journal of clinical laboratory analysis**, v. 16, n. 1, p. 47-51, 2002.

MARTINS, Al. M. **Sequenciamento de DNA, montagem de novo do genoma e desenvolvimento de marcadores microssatélites, indels e SNPs para uso em análise genética de *Brachiaria ruziziensis***. 2013. Tese de Doutorado. Brasília, Distrito Federal.

MELO, P. Z. **Tamanho genômico e marcadores microssatélites para *Pterodon pubecens* e *Pterodon emarginatus* (Leguminosae)**. 2018. Dissertação de mestrado. Goiânia, Goiás.

MIRANDA, F. D. et al. Tópicos especiais em genética e melhoramento II.: **CAUFES**, Universidade do Espírito Santo, 2018.

MYBURG, A. A.; GRATTAPAGLIA, D.; TUSKAN, G. A.; HELLSTEN, U.; HAYES, R. D.; GRIMWOOD, J.; et al. The genome of *Eucalyptus grandis*. **Nature**, v. 5010, n. 7505, p. 356–362, jun. 2014.

MESQUITA, A. G. G., GUIMARÃES, C. T., PARENTONI, S. N., & PAIVA, E. Recuperação do genitor recorrente em milho utilizando retrocruzamento assistido por marcadores microssatélites. **Revista Brasileira de Milho e Sorgo**, 2005.

MEGLÉCZ, E. et al. QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. **Molecular Ecology Resources**, v. 14, n. 6, p. 1302-1313, 2014.

MOKHTAR, M. M.; ATIA, M. A. M. SSRome: an integrated database and pipelines for exploring microsatellites in all organisms. **Nucleic acids research**, v. 47, n. D1, 2019.

NADEEM, M.A. et al. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biotechnology & Biotechnological Equipment**, v. 32, n. 2, p. 261-285, 2018.

NASCIMENTO, W. DE M. C. **Xarope de chambá (*Jsticia pectoralis* jacq.) No tratamento da tosse e sintomas respiratórios: um ensaio clínico randomizado**. 2018. Dissertação de Mestrado. Sobral, Ceará.

OCHOA, I. et al. QualComp: a new lossy compressor for quality scores based on rate distortion theory. **BMC bioinformatics**, v. 14, n. 1, p. 187, 2013.

OLIVEIRA, M. T. A. et al. Serine protease inhibition and modulatory-antibiotic activity of the proteic extract and fractions from *Amburana cearensis*. **Food and Chemical Toxicology**, v. 135, p. 110946, 2020.

PAN, L. et al. Development of 12 chloroplast microsatellite markers in *Vigna unguiculata* (Fabaceae) and amplification in *Phaseolus vulgaris*. **Applications in plant sciences**, v. 2, n. 3, p. 1300075, 2014.

PARIDA, S. K. et al. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. **Theoretical and Applied Genetics**, v. 118, n. 2, p. 327-338, 2009.

PARK, H. et al. Development of chloroplast microsatellite markers for the endangered *Maianthemum bicolor* (Asparagaceae sl). **Applications in plant sciences**, v. 4, n. 8, p. 1600032, 2016.

PATIL, P. G. et al. Genome-wide characterization and development of simple sequence repeat markers for genetic studies in pomegranate (*Punica granatum* L.). **Trees**, v. 34, n. 4, p. 987-998, 2020.

PEREIRA, E. P. L. et al. *Amburana cearensis* seed extracts protect PC-12 cells against toxicity induced by glutamate. **Revista Brasileira de Farmacognosia**, v. 27, p. 199-205, 2017.

PEREIRA, E. P. L. **Mercadores bioquímicos da atividade antioxidante em sementes de amburana cearensis (Fr. Allemão) AC Smith submetidas a estresse hídrico**. Dissertação de mestrado, 2010. Salvador, Bahia.

PERVAIZ, Tariq et al. Association between Chloroplast and Mitochondrial DNA sequences in Chinese *Prunus* genotypes (*Prunus persica*, *Prunus domestica*, and *Prunus avium*). **BMC plant biology**, v. 15, n. 1, p. 1-10, 2015.

QUEIROZ, L. P. 2009. Leguminosas da Caatinga. 1. ed. **Feira de Santana: Universidade Estadual de Feira de Santana**. Pp 243-244.

RAZA, Khalid; AHMAD, Sabahuddin. Recent advancement in next-generation sequencing techniques and its computational analysis. **International Journal of Bioinformatics Research and Applications**, v. 15, n. 3, p. 191-220, 2019.

RENSHAW, M. A. et al. Microsatellite multiplex panels for genetic studies of gray snapper (*Lutjanus griseus*) and lane snapper (*Lutjanus synagris*). **Fishery Bulletin**, v. 105, n. 3, p. 436-439, 2007.

ROSA, AJ de M.; PAIVA, Samuel Rezende. Marcadores moleculares e suas aplicações em estudos populacionais de espécies de interesse zootécnico. **Embrapa Cerrados-Documents (INFOTECA-E)**, 2009.

SAFONOVA, Y.; BANKEVICH, A.; PEVZNER, P. A. dipSPAdes: assembler for highly polymorphic diploid genomes. **Journal of Computational Biology**, v. 22, n. 6, p. 528-545, 2015.

SANTOS, C. A. F. et al. Variabilidade genética, com base em marcadores RAPD, de três espécies arbóreas ameaçadas de extinção no semi-árido brasileiro. **Embrapa Semiárido-Artigo em periódico indexado**, 2007.

SANTOS, S. K. S. et al. Transferibilidade de marcadores SSR, diversidade e estrutura genética em *Syagrus oleracea*. **Brazilian Journal of Development**, v. 6, n. 8, p. 59931-59947, 2020.

SANTOS, R. DE P. Desenvolvimento bioinformático de marcadores moleculares para a conservação de *Brycon orbignyanus* (Characiformes: Bryconidae), uma espécie de importância socioambiental. 2018. Dissertação de Mestrado. São João Del Rei, Minas Gerais.

SELEME, E. P. **Amburana Schwacke & Taub.(Leguminosae, Papilionoideae): estudos taxonômicos e sistemáticos**. Tese de doutorado. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro. Rio de Janeiro, 2014.

SELEME, E.P. et al. A taxonomic review and a new species of the South American woody genus *Amburana* (Leguminosae, Papilionoideae). **Phytotaxa**, v. 212, n. 4, p. 249-263, 2015.

SILVA, C. da. **Validação in silico de marcadores SSR de eugenia uniflora e transferibilidade para sete espécies de eucalyptus**. 2016. Trabalho de Conclusão de Curso. São Gabriel, Rio Grande do Sul.

SINGH, N. et al. Development of cpSSR markers for analysis of genetic diversity in *Gladiolus* cultivars. **Plant Gene**, v. 10, p. 31-36, 2017.

SHEN, Qi et al. Large scale identification of SSR markers in *perilla* by next generation

sequencing. **Oil Crop Science**, v. 4, n. 2, p. 100-109, 2019.

SHERMAN, Craig DH et al. Development of twenty-three novel microsatellite markers for the seagrass, *Zostera muelleri* from Australia. **Conservation Genetics Resources**, v. 4, n. 3, p. 689-693, 2012.

SILVA, P. I.T. et al. Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. **Bmc Genomics**, v. 14, n. 1, p. 1-9, 2013.

SILVEIRA, E. S. **Desenvolvimento, avaliação da atividade leishmanicida e toxicológica de nanosistema de cumarina (1,2-benzopirona)**. 2020. Tese de doutorado. Fortaleza, Ceará.

SINGH, V.K. et al. Next-Generation Sequencing (NGS) Tools and Impact in Plant Breeding. In *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*; Springer: Cham, Switzerland. 563–612, 2015.

SOARES, T. N. et al. Patterns of genetic variability in central and peripheral populations of *Dipteryx alata* (Fabaceae) in the Brazilian Cerrado. **Plant Systematics and Evolution**, v. 301, n. 5, p. 1315-1324, 2015.

SOUSA, R. C. DE B. **CARACTERIZAÇÃO DE MICROSSATÉLITES E DESENVOLVIMENTO DE MARCADORES SSR EM *Khaya grandifoliola* C. DC.** 2019. Dissertação de mestrado. Goiânia, Goiás.

SUNG, W. et al. Simple sequence repeat variation in the *Daphnia pulex* genome. **BMC genomics**, v. 11, n. 1, p. 1-11, 2010.

TAHERI, S. et al. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. **Molecules**, v. 23, n. 2, p. 399, 2018.

TAKAHASHI, Yu; YOKOI, Suji; TAKAHATA, Yoshihito. Genetic divergence of turnip (*Brassica rapa* L. em. Metzg. subsp. *rapa*) inferred from simple sequence repeats in chloroplast and nuclear genomes and morphology. **Genetic resources and crop evolution**, v. 63, n. 5, p. 869-879, 2016.

TELLES, M. P. C. et al. Discovery and characterization of new microsatellite locus in *Dipteryx alata* vogel (Fabaceae) using next-generation sequencing data. **Genetics and Molecular Research**, v. 16, n. 2, 2017.

TINIO, C. E. et al. Development of microsatellite markers from genomic DNA of *Parashorea malaanonan* (Dipterocarpaceae) using next-generation sequencing. **Silvae Genetica**, v. 68, n. 1, p. 22-25, 2019.

- TOMAR, Ram Sewak S. et al. Development of chloroplast-specific microsatellite markers for molecular characterization of alloplasmic lines and phylogenetic analysis in wheat. **Plant Breeding**, v. 133, n. 1, p. 12-18, 2014.
- SANTOS, S. S., et al. Multi-objective optimization in systematic conservation planning and the representation of genetic variability among populations. **Genetics and Molecular Research**, Ribeirão Preto, v. 14, n. 2, p. 6744-6761, 2015.
- VIEIRA, M. L. C. et al. Microsatellite markers: what they mean and why they are so useful. **Genet Mol. Biol.**, Ribeirão Preto, v. 39, n. 3, p. 312-328, 2016.
- WANG, P. et al. Characterization and development of EST-SSR markers from a cold-stressed transcriptome of Centipedegrass by Illumina paired-end sequencing. **Plant molecular biology reporter**, v. 35, n. 2, p. 215-223, 2017.
- WANG, M. et al. Genome-wide microsatellite characterisation and marker development in *Chenopodium quinoa*. **Annals of Applied Biology**, v. 175, n. 3, p. 415-423, 2019.
- WANG, Y. et al. Chloroplast and mitochondrial microsatellites for *Millettia pinnata* (Fabaceae) and cross-amplification in related species. **Applications in plant sciences**, v. 5, n. 5, p. 1700034, 2017.
- XU, J. et al. Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. **DNA research**, v. 20, n. 5, p. 497-509, jun. 2013.
- XU, M. et al. Development of novel chloroplast microsatellite markers for *Ginkgo biloba*. **Genet. Mol. Res.**, v. 14, p. 7715-7720, 2015.
- XUE, S. et al. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. **Horticulture research**, v. 6, n. 1, p. 1-13, 2019.
- YASODHA, R. et al. Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): insights into SSR diversity, phylogeny and conservation. **DNA Research**, v. 25, n. 4, p. 409-419, 2018.
- ZALAPA, J. E. et al. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) locus in the plant sciences. **American journal of botany**, v. 99, n. 2, p. 193-208, 2012.
- ZANE, L.; BARGELLONI, L.; PATARNELLO, T. Strategies for microsatellite isolation: a review. **Molecular ecology**, v. 11, n. 1, p. 1-16, 2002.
- ZHANG, Y. et al. De novo assembly of transcriptome and development of novel EST-SSR markers in *Rhododendron rex* Lévl. through illumina sequencing. **Frontiers in plant science**,

v. 8, p. 1664, 2017.

ZHONG, C. et al. Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel Phytophthora resistance gene, RpsHC18, in soybean. **Theoretical and applied genetics**, v. 131, n. 3, p. 525-538, 2018.

ZOLET, A. C. T. et al. Marcadores moleculares na era genômica: metodologias e aplicações. **Sociedade Brasileira de Genética**, 2017.