



**UNIVERSIDADE FEDERAL DO CEARÁ**

***Campus* de SOBRAL**

**FACULDADE DE MEDICINA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**

**VITÓRIA VIRGÍNIA MAGALHÃES SOARES**

**IDENTIFICAÇÃO, VALIDAÇÃO E ANOTAÇÃO FUNCIONAL DE MARCADORES  
MICROSSATÉLITES EM GENÓTIPOS DE CAJUEIRO ANÃO-PRECOCE  
(*Anacardium occidentale* var. *nanum*) UTILIZANDO DADOS DE RNA-Seq.**

**SOBRAL**

**2015**

**VITÓRIA VIRGÍNIA MAGALHÃES SOARES**

**IDENTIFICAÇÃO, VALIDAÇÃO E ANOTAÇÃO FUNCIONAL DE MARCADORES  
MICROSSATÉLITES EM GENÓTIPOS DE CAJUEIRO ANÃO-PRECOCE  
(*Anacardium occidentale* var. *nanum*) UTILIZANDO DADOS DE RNA-Seq.**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia, da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Mestre em Biotecnologia. Área de Concentração: Macromoléculas.

Orientador: Prof. Dr. Rodrigo Maranguape Silva da Cunha.

**SOBRAL**

**2015**

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S657i Soares, Vitória Virgínia Magalhães.  
Identificação, validação e anotação funcional de marcadores microssatélites em genótipos de cajueiro anão-precoce (*Anacardium occidentale* var. *nanum*) utilizando dados de RNA-Seq. / Vitória Virgínia Magalhães Soares. – 2016.  
95 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Sobral, Programa de Pós-Graduação em Biotecnologia, Sobral, 2016.

Orientação: Prof. Dr. Rodrigo Maranguape Silva da Cunha .

1. Genótipo. 2. Transcriptoma. 3. Biotecnologia. I. Título.

CDD 660.6

---

VITÓRIA VIRGÍNIA MAGALHÃES SOARES

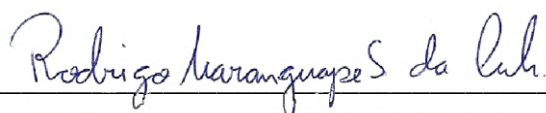
**IDENTIFICAÇÃO, VALIDAÇÃO E ANOTAÇÃO FUNCIONAL DE MARCADORES  
MICROSSATÉLITES EM GENÓTIPOS DE CAJUEIRO ANÃO-PRECOCE  
(*Anacardium occidentale* var. *nanum*) UTILIZANDO DADOS DE RNA-Seq.**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia, da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Mestre em Biotecnologia. Área de Concentração: Macromoléculas.

Orientador: Prof. Dr. Rodrigo Maranguape Silva da Cunha.

Aprovada em: 12/02/2015.

**BANCA EXAMINADORA**



---

Prof. Dr. Rodrigo Maranguape Silva da Cunha (Orientador)

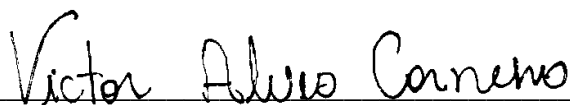
Universidade Estadual Vale do Acaraú (UVA)



---

Prof. Dr. João Garcia Alves Filho (Examinador)

Universidade Estadual Vale do Acaraú (UVA)



---

Prof. Dr. Victor Alves Carneiro (Examinador)

Universidade Federal do Ceará (UFC)

Aos meu pais, Fabiano e Vilani.

A minha irmã Waleska,

***dedico.***

## **AGRADECIMENTOS**

Finalizada essa etapa da minha vida, não poderia deixar de expressar o mais profundo agradecimento a todos àqueles que me apoiaram nesta caminhada e contribuíram para a realização deste trabalho.

Em primeiro lugar agradeço a Deus, Aquele que tornou tudo possível, me deu forças e fortaleceu minha fé em todos os momentos da minha vida.

À minha família, principalmente aos meus pais, Maria Vilani e José Fabiano e à minha irmã Waleska Virgínia, por serem minha base e que com todo carinho, dedicação, amor e compreensão me deram forças e incentivo para realizar meus sonhos.

Ao meu orientador professor Dr. Rodrigo Maranguape Silva da Cunha, pela oportunidade que me deu desde a graduação, pela contribuição em minha formação acadêmica, pela amizade, orientação e confiança durante a realização deste trabalho.

Ao meu amigo e companheiro João Garcia Alves Filho, por todo apoio e disponibilidade para a realização desse trabalho, além de amizade, carinho e dedicação.

Aos amigos do laboratório de Genética Molecular, do Núcleo de Biotecnologia de Sobral (NUBIS) que me ajudaram direta e indiretamente na realização desse trabalho. Não tenho palavras para agradecer por tanta ajuda e carinho, além da amizade e dos momentos de descontração durante o trabalho. Carlos Franciney, Pedro Cunha, Paulo de Tarso, Rafael Bastos, Erivan Alves e Flávia Muniz, meus filhos de coração, por toda a ajuda, companheirismo, amizade e por todos os momentos que passamos juntos. Aos companheiros de longas datas, os quais compartilhei muitos momentos, alegrias e tristezas antes e durante o mestrado: Aurilene Gomes e Jedson Aragão. As grandes amigas Nyanne Hardy, Mônica Valéria e Tatiana Farias, pela amizade e por todos os momentos que passamos, seja na faculdade ou no laboratório. Ao amigo Bruno Carneiro, que mesmo estando nos Estados Unidos (Mas ainda é do NUBIS), não deixou de fazer parte desse

trabalho. Aos grandes amigos Raulzito Moreira e Daniel de Brito que mesmo estando longe sempre mostraram-se disponíveis quando precisei. À técnica do NUBIS, Maria Auxiliadora que estava sempre disposta a ajudar.

Ao casal que mais amo, Annyta Frota e Ricardo Basto, por toda amizade, apoio e ajuda nos momentos que precisei.

As minhas grandes amigas Wandy Policarpo, Marianne Pontes, Iara Araújo, Dauana Mesquita e Louricélia Rodrigues pela convivência, amizade e pelos momentos de descontração na hora do café.

Aos colegas e professores do mestrado em Biotecnologia, pelo convívio e troca de conhecimentos.

As amigas do AP 102, Gessiane Martins, Ana Mikaelly, Karysia Veras e Zeneide Rodrigues, e a mais nova inquilina mas já admiro bastante Rafaele, pela amizade e compreensão durante minha ausência em casa.

À EMBRAPA Agroindústria Tropical, pela disponibilidade das amostras de folha de cajueiro.

À EMBRAPA Caprinos e Ovinos, em nome dos professores João Ricardo, Ângela Eloy e Lúcia Sider, pela disponibilidades de alguns equipamentos úteis para a realização dessa pesquisa.

Aos membros da banca examinadora, professores Victor Alves Carneiro e João Garcia Alves Filho pelas contribuições ao trabalho.

À CAPES pela bolsa concedida.

À Universidade Federal do Ceará, por contribuir na minha formação profissional.

A todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

**Muito Obrigada!**

*“A mente que se abre a uma nova ideia  
jamais voltará ao seu tamanho original”*

*(Albert Einstein)*



## RESUMO

O cajueiro (*Anacardium occidentale* L.) é uma planta nativa do Brasil com grande valor comercial. Isso contribui com a geração de milhares de empregos diretos e indiretos, especialmente na Região Nordeste, em época de estiagem. Programas de melhoramento genético vem selecionando cultivares de cajueiro melhores adaptados ao ambiente semiárido a fim de colocá-lo em um mercado cada vez mais competitivo. A busca por marcadores microssatélites pode auxiliar os programas de melhoramento no que diz respeito ao acesso à diversidade genética da espécie. O presente trabalho tem como objetivo a identificação de marcadores SSR em cajueiro com base no transcriptoma de sementes e folhas, bem como sua validação por PCR em diferentes genótipos comerciais. Utilizando ferramentas de bioinformática foram encontrados motivos SSRs do tipo di- tri- tetra- penta- e hexanucleotídeos, onde o motivo do tipo trinucleotídeo foi o mais representativo nos transcritos do cajueiro anão CCP 76 e comum variando de 60 a 65%, respectivamente. Os transcritos de cajueiro comum e anão CCP 76 compartilham um total de 298 marcadores SSR, destes, 29 foram escolhidos para amplificação por PCR, os quais 9 mostraram polimorfismo nos genótipos testados. As sequências situadas próximas aos marcadores SSR codificam proteínas, que em sua maioria pertencem a famílias gênicas. Pode-se concluir que foram encontrados regiões contendo marcadores SSRs na região transcrita de nove genótipos de cajueiro, podendo ser uma ferramenta útil nas análises genéticas e abrindo perspectivas para o papel endógeno dos SSRs na função proteica.

**Palavras-Chave:** Genótipo. SSR. Transcriptoma.

## ABSTRACT

The cashew tree is a native plant from Brazil with high market value. This contributes to the generation of thousands direct and indirect jobs, especially in the Northeast region, during dry season. Breeding programs have been selecting cashew cultivars that best adapt to semi-arid environment in order to fit it in the increasingly competitive market. The search for microsatellite markers can assist breeding programs regarding to access to genetic diversity of the species. This study aims to identify SSR markers in the cashew tree based on seeds and leaves transcriptome, as well as assess its validation by PCR technique in different commercial genotypes. Using bioinformatics tools, SSRs motifs from types di- tri- tetra- penta- and hexanucleotides were found. Trinucleotide-type motif was the most representative in transcripts both from dwarf cashew CCP 76 and common cashew, ranging from 60 to 65%, respectively. Transcripts from common cashew and dwarf cashew CCP 76 share a total of 298 SSR markers. 29 of these were selected for amplification by PCR, in which 9 showed polymorphism in genotypes tested. The sequences located near to the SSR markers encode proteins, which mostly belong to gene families. It can be concluded that regions containing SSRs markers were found in the transcribed region of nine cashew genotypes, and can be a useful tool in genetic analysis and open up prospects for endogenous role of SSRs in protein function.

**KEYWORDS:** Genotype. SSR. Transcriptome.

## LISTA DE FIGURAS

<b>Figura 1:</b> Tipos de cajueiro existentes na natureza. ....	22
<b>Figura 2:</b> Classificação dos microssatélites segundo Goldstein e Schlötterer (1999). .....	32
<b>Figura 3:</b> Diagrama de Venn demonstrando o número de SSRs totais e compartilhados entre o cajueiro comum e anão CCP 76. ....	46
<b>Figura 4:</b> Detalhe da via Metabolismo do Carbono do KEGG <i>Pathways</i> mostrando sequências do transcriptoma de semente e folha de cajueiro onde foram encontrados SSR. ....	55
<b>Figura 5:</b> Dendograma obtido por Lamas (2010), baseado nas distâncias genéticas entre cada acesso de <i>A. occidentale</i> L.....	59
<b>Figura 6:</b> Géis de eletroforese de agarose 1%, mostrando os <i>primers</i> que amplificaram 100% das amostras. ....	69
<b>Figura 7:</b> Géis de eletroforese de agarose 1%, mostrando os <i>primers</i> que não amplificaram todos as amostras. ....	70
<b>Figura 8:</b> Eletroforese em gel de agarose a 1% dos marcadores SSR mostrando o <i>primer</i> que não amplificou nenhuma das amostras. ....	71
<b>Figura 9:</b> Resultado dos produtos de PCR amplificados, mostrados em géis de agarose 3%. ....	72
<b>Figura 10:</b> Eletroforese em gel de poliacrilamida 10% corado com brometo de etídio, mostrando pares de iniciadores que amplificaram <i>loci</i> polimórficos e monomórficos. ....	75
<b>Figura 11:</b> Gel de poliacrilamida 10% mostrando resultado da PCR com problemas para alguns pares de <i>primers</i> . ....	76

## LISTA DE GRÁFICOS

<b>Gráfico 1:</b> Produção de castanha de caju em 2012, no mundo.....	26
<b>Gráfico 2:</b> Distribuição dos microssatélites no transcriptoma do cajueiro comum e anão-precoce CCP 76, utilizando o programa MISA.....	48
<b>Gráfico 3:</b> Distribuição dos microssatélites correspondendo aos motivos do tipo dinucleotídios encontrados no transcriptoma do cajueiro ( <i>A. occidentale</i> L.).....	50
<b>Gráfico 4:</b> Distribuição dos microssatélites com motivos do tipo trinucleotídios encontrados no transcriptoma do cajueiro ( <i>A. occidentale</i> L.).....	51
<b>Gráfico 5:</b> Classificação funcional das sequências contendo SSR, do transcriptoma de cajueiro, com base nos termos de <i>Gene Ontology</i> (GO) mais representativos....	53

## LISTA DE TABELAS

<b>Tabela 1:</b> Principais características dos cajueiros dos tipos comum e anão-precoce. .....	23
<b>Tabela 2:</b> Lista dos <i>primers</i> desenhados em sequências que flanqueiam SSR dos transcritos de cajueiro comum e CCP 76. ....	41
<b>Tabela 3:</b> Estatísticas de montagem do transcriptoma do cajueiro comum e anão-precoce CCP 76 usando os programas <i>Velvet</i> e <i>Oases</i> . ....	44
<b>Tabela 4:</b> Estatísticas dos SSRs identificados a partir do transcriptoma do cajueiro comum e anão-precoce CCP 76. ....	46
<b>Tabela 5:</b> Características morfológicas e agroindustriais de alguns genótipos de cajueiro anão-precoce. ....	59
<b>Tabela 6:</b> Condições térmicas da amplificação dos <i>loci</i> contendo SSR de amostras de DNA extraídas de folhas de diversos genótipos de cajueiro anão-precoce. ....	64
<b>Tabela 7:</b> Composição do gel de poli-acrilamida 10% para DNA. ....	65
<b>Tabela 8:</b> Quantificação de DNA genômico de diferentes genótipos de cajueiro anão-precoce, com valor de absorvância a 260 nm. ....	68
<b>Tabela 9:</b> Características de 28 marcadores microssatélites em <i>A. occidentale</i> var. <i>nanum</i> , mostrando o grau de polimorfismo e o número de alelos de cada <i>locus</i> . ....	74
<b>Tabela 10:</b> Lista de proteínas codificadas em sequências contendo SSR. ....	78

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

\$	Dolar
%	Porcentagem
°C	Grau Celsius
µL	Microlitro
A	Adenina
A260	Absorbância a 260 nm
A260/280	Relação entre absorbância 260 e 280
ACC	Amêndoa da Castanha do Caju
AFLP	Polimorfismo de comprimento de fragmentos amplificados ( <i>Amplified Fragment Length Polymorphism</i> )
BLAST	Ferramenta Básica de Alinhamento Local ( <i>Basic Local Alignment Search Tool</i> )
C	Citosina
CCP	Clone de Cajueiro de Pacajus
CDD	Banco de dados de Domínios Conservados ( <i>Conserved Domain Database</i> )
cDNA	Ácido desoxirribonucleico complementar
CONAB	Companhia Nacional de Abastecimento
CP	Cajueiro Pacajus
CTAB	Brometo de Cetiltrimetilamônio (Cetyl trimethylammonium bromide)
DNA	Ácido Desoxirribonucleico
dNTPs	Desoxinucleotídeos trifosfatados
EDTA	Ácido etilenodiamino tetra-acético (Ethylenediamine tetraacetic acid)
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
EST	Etiqueta de Sequência Expressa
FAO	Organização das Nações Unidas para a Alimentação e Agricultura ( <i>Food and Agriculture Organization</i> )
G	Guanina
G	Gramma
GO	Ontologia Gênica ( <i>Gene Ontology</i> )

H	Hora
HCl	Ácido Clorídrico
HP	Companhia <i>Hewlett-Packard</i>
HPS	Proteína Hidrofóbica de Soja
ISSR	Sequências Simples Repetitivas Internas ( <i>Inter Simple Sequence Repetees</i> )
KEGG	Enciclopédia de genes e genomas de Kyoto ( <i>Kyoto Encyclopedia of Genes and Genomes</i> )
Kg	Quilograma
KO	<i>Kegg Orthology</i>
LCC	Líquido da Castanha de Caju
M	Metro
M	Molar
m/v	Massa/volume
mA	Miliampere
MgCl <sub>2</sub>	Cloreto de Magnésio
Min	Minutos
MISA	MlcroSAteLLite
mL	Mililitro
mM	Milimolar
Mt	Milhões de toneladas
NaCl	Cloreto de Sódio
NAM	Sem meristema apical ( <i>No Apical Meristem</i> )
NGS	Nova Geração de Sequenciamento ( <i>Next Generation Sequencing</i> )
Nm	Nanômetro
Pb	Pares de base
PCR	Reação em cadeia da Polimerase
pH	potencial hidrogeniônico
PSA	Persulfato de amônio
RAM	Memória de acesso aleatório ( <i>Random Access Memory</i> )
RAPD	Polimorfismo de DNA amplificado ao acaso ( <i>Random Amplified Polymorphic DNA</i> )
RFLP	Polimorfismo no comprimento de fragmentos de restrição ( <i>Restriction Fragment Length Polymorphism</i> )

RGS	Reguladores da sinalização por proteína G
RNA-Seq	Sequenciamento de RNA (Ácido Ribonucleico)
Rpm	Rotações por minuto
SNP	Polimorfismo de Nucleotídeo Único ( <i>Single Nucleotide Polymorphism</i> )
SOLiD	Sequenciamento por Ligação e Detecção de Oligonucleotídeos ( <i>Sequencing by Oligonucleotide Ligation and Detection</i> )
SSAP	<i>Sequence-Specific Amplification Polymorphism</i>
SSR	Sequências Simples Repetidas ( <i>Simple Sequence Repeats</i> )
STR	Repetições Curtas em Tandem ( <i>Short Tandem Repeats</i> )
T	Timina
TBE	Tampão contendo Tris, Ácido bórico e EDTA
TE	Tampão contendo Tris e EDTA
TEMED	N,N,N',N'-tetrametil-etano-1,2-diamina
USP	Universidade de São Paulo
v/v	Volume/volume
VNTR	Sequências adjacentes que se repetem em número variável ( <i>Variable Number of Tandem Repeats</i> )
W	Wats



## SUMÁRIO

<b>INTRODUÇÃO</b> .....	<b>19</b>
<b>1 REVISÃO DE LITERATURA</b> .....	<b>21</b>
1.1 O Cajueiro .....	21
1.2 A agroindústria do cajueiro .....	24
1.3 Análises Transcriptômicas .....	26
1.4 Sequenciamento de Próxima Geração (Nova Geração de Sequenciamento - NGS) .....	27
1.5 Marcadores Moleculares .....	29
1.6 Marcadores Microsatélites .....	30
<b>2 BUSCA E CARACTERIZAÇÃO <i>in silico</i> DE MARCADORES MICROSATÉLITES NO TRANSCRIPTOMA DE SEMENTES E FOLHAS DE CAJUEIRO (<i>A. occidentale</i> L.)</b> .....	<b>34</b>
2.1 INTRODUÇÃO .....	34
2.2.1 Objetivo Geral.....	36
2.2.2 Objetivos Específicos.....	36
2.3 ESTRATÉGIA EXPERIMENTAL.....	37
2.4 METODOLOGIA .....	38
2.4.1 Obtenção das sequências do transcriptoma do cajueiro .....	38
2.4.2 Análise da qualidade das sequências e montagem <i>De novo</i> do transcriptoma de cajueiro.....	38
2.4.3 Detecção <i>in silico</i> de marcadores microsatélites .....	39
2.4.4 Anotação funcional das sequências que contém microsatélites .....	39
2.4.5 Desenho de <i>primers</i> .....	40
2.5 RESULTADOS E DISCUSSÕES.....	43
2.5.1 Montagem do transcriptoma do cajueiro comum e CCP 76 a partir de sequências de sementes e folhas.....	43

2.5.2 Busca <i>in silico</i> por marcadores microsatélites .....	44
2.5.3 Anotação funcional das sequências que contém microsatélites .....	52
2.6 CONCLUSÕES.....	56
<b>3 VALIDAÇÃO DOS MARCADORES MICROSATÉLITES EM DIFERENTES GENÓTIPOS DE CAJUEIRO ANÃO-PRECOCE (<i>A. occidentale</i> L. var. <i>nanum</i>)</b> .....	<b>58</b>
3.1 INTRODUÇÃO .....	58
3.2 OBJETIVOS.....	60
3.2.1 Objetivo Geral.....	60
3.2.2 Objetivos Específicos.....	60
3.3 ESTRATÉGIA EXPERIMENTAL.....	61
3.4 METODOLOGIA .....	62
3.4.1 Coleta do Material Biológico .....	62
3.4.2 Extração e Purificação de DNA genômico .....	62
3.4.3 Quantificação e Eletroforese do DNA .....	63
3.4.4 Amplificação dos <i>loci</i> microsatélites.....	64
3.4.5 Resolução e análise dos fragmentos amplificados .....	65
3.4.6 Identificação <i>in silico</i> das sequências amplificadas .....	66
3.5 RESULTADOS E DISCUSSÕES.....	67
3.5.1 Isolamento do DNA genômico .....	67
3.5.2 Amplificação das regiões de DNA que contém microsatélites .....	68
3.5.3 Identificação de microsatélites polimórficos .....	71
3.6 CONCLUSÕES .....	80
3.7 PERSPECTIVAS.....	81
REFERÊNCIAS BIBLIOGRÁFICAS.....	82

## INTRODUÇÃO

O cajueiro (*Anacardium occidentale* L.) é uma espécie de origem brasileira, encontrada principalmente em climas tropicais e subtropicais, e é a única espécie cultivada comercialmente do seu gênero (PAIVA; CRISÓSTOMO; BARROS, 2003). A variabilidade genética observada na cultura do cajueiro no Brasil tem sido agrupada em dois tipos bem definidos em relação ao porte, denominados de cajueiro tipo anão-precoce e cajueiro tipo comum. O cajueiro comum é o mais predominante no Nordeste do Brasil ocorrendo de forma natural sem a necessidade de plantio, enquanto que o cajueiro anão-precoce é caracterizado pelo porte baixo e precocidade (CRISÓSTOMO et al., 2001; LIMA, 1988).

O cajueiro anão-precoce vem substituindo o cajueiro comum por proporcionar maior produtividade, facilitar a colheita e a condução dos pomares em razão do seu baixo porte. Além disso, o cajueiro anão-precoce apresenta maior uniformidade da castanha, do pedúnculo e da produção, permitindo uma exploração comercial mais rentável (BARROS; CRISÓSTOMO, 1995; OLIVEIRA et al., 2002). O uso de clones representa uma forma de manejo econômico, ecológico e seguro, impedindo a invasão de pragas e doenças, além de proporcionar uma melhor utilização da variabilidade genética da espécie (PAIVA; BARROS, 2004). Uma das formas de se avaliar a variabilidade genética seria através do uso de marcadores moleculares.

Os marcadores microssatélites ou SSR possuem diversas características que os tornam marcadores moleculares de grande relevância científica e tecnológica. Os SSRs são importantes na obtenção de informações sobre a variabilidade genética, para orientar programas de melhoramento genético e na certificação de genótipos, dando uma maior confiabilidade e competitividade ao produto a ser vendido (GEORGI et al., 2012). Amiruddin e colaboradores (2012) relataram que SSRs representam uma rica fonte de marcadores moleculares devido a alta taxa de mutação nessas regiões. Ainda, acredita-se que os microssatélites são importantes para evolução dos genomas devido a estímulos de variabilidade genética e influencia na expressão de genes (LI et al., 2002; KALIA et al., 2011). Além disso, tais marcadores podem ser utilizados para estabelecer comparações evolutivas de indivíduos de uma mesma população. Dentro desse contexto o cajueiro entra como

uma espécie economicamente importante, contudo, seu potencial biotecnológico ainda é pouco explorado devido à ausência de dados genômicos.

O desenvolvimento de novas técnicas de sequenciamento tem proporcionado uma maneira mais rápida e eficiente de mapear e quantificar o transcriptoma em um método conhecido como RNA-Seq (WANG et al. 2010). A utilização da tecnologia do RNA-Seq oferece uma grande quantidade de informação sobre o transcriptoma como a avaliação dos níveis de expressão, mapeamento de genes expressos e descoberta de novos genes, sem haver a necessidade de um genoma previamente sequenciado (WANG et al. 2009).

Os dados de projetos de transcriptomas constituem uma fonte valiosa de informação molecular e podem ser analisados de diversas formas, como a descoberta das sequências dos genes transcritos, no mapeamento de rotas metabólicas ativas, na determinação de marcadores moleculares, na análise da expressão diferencial de genes de genótipos entre outros (MCGETTIGAN, 2013; MUTZ et al., 2013; TARAZONA et al., 2011). Sendo assim, a utilização de marcadores microssatélites é eficiente para caracterizar e discriminar diferentes genótipos de cajueiro e são potencialmente capazes para caracterizar e diferenciar qualquer população da espécie (PINHEIRO et al., 2008).

## 1 REVISÃO DE LITERATURA

### 1.1 O Cajueiro

O cajueiro (*Anacardium occidentale* L.) é uma cultura de grande importância para a agricultura de regiões semiáridas, contribuindo para o desenvolvimento socioeconômico de diversas regiões do mundo, incluindo a região Nordeste do Brasil (BEZERRA et al., 2008). É uma planta de origem brasileira, que possui 60 a 74 gêneros e 400 a 600 espécies de árvores, arbustos, subarbustos e trepadeiras ocorrendo principalmente em climas tropicais e subtropicais (PAIVA; CRISÓSTOMO; BARROS, 2003).

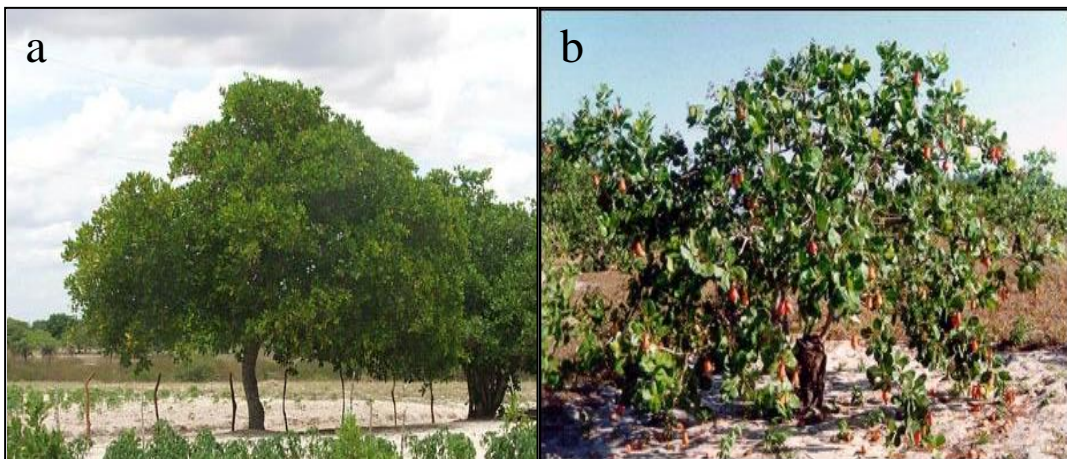
A classificação taxonômica do cajueiro é: Domínio Eukaryota, reino Plantae, divisão Magnoliophyta, classe Magnoliopsida, ordem Sapindales, família Anacardiaceae, gênero *Anacardium*, espécie *Anacardium occidentale* (\_\_\_\_\_, 2011). O gênero *Anacardium* possui 21 espécies, mas *A. occidentale* é a única espécie cultivada comercialmente.

O cajueiro é uma árvore ou arbusto (anão-precoce) apresentando folhas persistentes, simples, inteiras, alternas, de aspecto subcoriáceo, glabras e curto-pecioladas, roxo-avermelhadas quando jovem, de cor verde-amareladas quando maduras podendo cair após maturação. As flores são pequenas, curto-pediceladas, pálidas, avermelhadas, dispostas em panículas terminais, pedunculadas ramificadas e bracteadas na parte inferior. O sistema reprodutivo é predominantemente alogâmico (fecundação cruzada), com polinização predominantemente entomófila (NAMBIAR; PILLAI, 1985). O fruto (castanha) é um aquênio reniforme, formado por epicarpo, mesocarpo e endocarpo. O epicarpo é liso, coriáceo e cinzento. O mesocarpo é espesso, alveolado e cheio de LCC (líquido da castanha do caju). A amêndoa (parte comestível) é reniforme, composta por dois cotilédones brancos, carnosos, oleosos e revestida por uma película pergaminácea. O pedúnculo floral ou pseudofruto (caju) é hipertrofiado, carnoso, suculento e variável em tamanho, peso, forma e cor. O caule é normalmente tortuoso, grosso e ramificado logo a partir do nível do terreno. O sistema radicular constitui-se de uma raiz pivotante bem

desenvolvida, normalmente bifurcada, e de uma malha lateral subsuperficial (BARROS et al., 1993; CAVALCANTI et al., 2009).

Na natureza, existem basicamente dois tipos de cajueiro, o cajueiro comum (*A. occidentale* L.) e o cajueiro anão-precoce (*A. occidentale* var. *nanum*) (Figura 1). O cajueiro comum é o mais predominante no Nordeste, ocorrendo de forma natural sem a necessidade de plantio. Ele apresenta porte elevado, com altura variando entre 8 m e 15 m e extensão da copa podendo atingir até 20 m, dependendo das condições de clima, solo e sanidade (CRISÓSTOMO et al., 2001). O segundo tipo de cajueiro é conhecido como anão-precoce (cajueiro de 6 meses), e é caracterizado, principalmente pelo seu porte baixo, com altura entre 3 m e 4 m e copa atingindo até 9 m (LIMA, 1988). O tipo anão-precoce também é caracterizado pela precocidade, iniciando o florescimento entre 6 e 18 meses contra os cinco a sete anos do tipo comum. As principais características dos cajueiros dos tipos comum e anão-precoce são ilustradas na Tabela 1 (CAVALCANTI et al., 2009).

**Figura 1:** Tipos de cajueiro existentes na natureza.



**a.** Cajueiro comum (*A. occidentale* L.); **b.** cajueiro anão-precoce (*A. occidentale* var. *nanum*). **Fonte:** (PAIVA; BARROS, 2004).

**Tabela 1:** Principais características dos cajueiros dos tipos comum e anão-precoce.

<b>Características</b>	<b>Comum</b>	<b>Anão-precoce</b>
Porte (m)	Alto (8-15)	Baixo (<5)
Tamanho da copa (m)	>7	5 a 7
Primeira Florada	2 a 5 anos	6 a 18 meses
Variação do peso da castanha (g)	3 a 33	3 a 13
Variação do peso do pedúnculo (g)	20 a 500	20 a 160
Produção: castanha/planta/safra (Kg)	1 a 100	Até 43

**Fonte:** (CAVALCANTI et al., 2009).

Estudos relacionados com a determinação do número de cromossomos de *A. occidentale* são bastante divergentes. Segundo Purseglove (1968), a planta em estudo possui o número de cromossomos haplóide igual a 21, porém, a ploidia ainda não é clara.

Trabalhos realizados por Goldblatt (1981 e 1984) e por Deckers e colaboradores (2001), mostraram que o cajueiro possui um número de cromossomos polimórficos de  $2n=24$ ,  $2n=40$ ,  $2n=42$  e  $2n = 24, 30, 40, 42$ , respectivamente.

Thankamma-Pillai e Nambiar (1985) analisaram 223 células de diferentes indivíduos, e verificaram que 64-68% dessas células possuíam o cariótipo  $2n=42$ . Resultados semelhantes foram encontrados por Aliyu e Awopetu (2007), onde relataram o número, estrutura e comportamento dos cromossomos de duas populações de cajueiro que cresceram na Nigéria. Análises citológicas dessas populações revelaram que o número de cromossomos diplóide do cajueiro é de  $2n=42$  e o número haplóide  $n=21$ .

## 1.2 A agroindústria do cajueiro

Todas as partes do cajueiro são aproveitadas e podem ter importantes aplicações biotecnológicas e econômicas, além de características nutricionais e medicinais. Contudo, entre os subprodutos do cajueiro, os que se destacam como importância econômica são, o líquido da castanha do caju (LCC), o pedúnculo e a amêndoa da castanha do caju (ACC).

O líquido da castanha do caju (LCC) tem muitas aplicações como, por exemplo, no uso como aditivo para combustível para geração de calor (MOHOD; KHANDETOD; POWAR, 2008). Além disso, o LCC contém vários compostos fenólicos com diversas atividades antimicrobianas (HIMEJIMA; KUBO, 1991) e efeito moluscicida (KUBO; KOMATSU; OCHI, 1986). Mais recentemente, várias atividades dos lipídeos fenólicos presentes no LCC têm sido extensivamente estudadas do ponto de vista químico, como efeito antioxidante, antimicrobiano, antifúngicos e antiprotozoários, além de interagir com proteínas e membranas biológicas. Esses fatores certamente irão contribuir para sua aplicação na farmacologia e na medicina (STASIUK; KOZUBEK, 2010).

O pedúnculo do caju, ou falso-fruto, representa cerca de 90% do peso do fruto completo e vem se tornando importante segmento da agroindústria do caju. Entre os diversos produtos destacam-se o suco concentrado, doces, refrigerante e cajuína (PAIVA; BARROS, 2004). Além disso, é bastante utilizado como suplemento nutricional em alimentos para animais. Porém, por ser bastante perecível, o mercado valoriza mais o comércio de castanha de caju. Estudos recentes mostram que a cinza do bagaço do caju, além de ser utilizada na produção de cristais, cerâmica, fertilizante e extintores de incêndio, também tem ação antifúngica contra espécies do gênero *Fusarium* (SANTOS et al., 2011).

A amêndoa da castanha é rica em lipídeos (40 - 57%), proteínas (21%), carboidratos (20%), algumas vitaminas, como a vitamina B<sub>12</sub>, e minerais, como o cálcio, sódio, ferro, potássio, (FETUGA; BABATUNDE; OYENUGA, 1974; OLOGUNDE et al., 2011) e selênio (KANNAMKUMARATH et al., 2002), o qual desempenha um importante papel como antioxidante. Além disso, o consumo de castanhas pode reduzir o colesterol, assim como, o risco de doenças

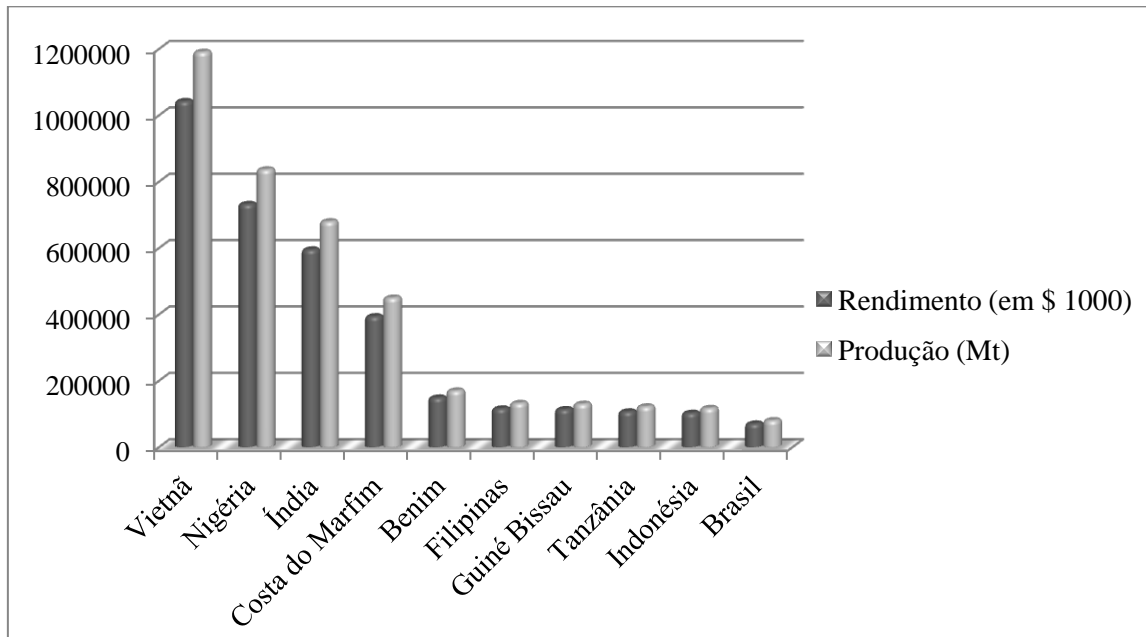


cardiovasculares e estudos têm demonstrado que alguns componentes das castanhas possuem papel cardioprotetor, antiobesidade e anticancerígeno (CHISHOLM et al., 2005).

Segundo a FAO (*Food and Agriculture Organization*), o Vietnã (1.190.900 Mt) é o maior produtor mundial de castanha de caju e o Brasil (80.630 Mt) ocupa o décimo lugar no ranking (Gráfico 1) (FAO, 2012). De acordo com a CONAB (Companhia Nacional de Abastecimento), nos períodos de 2010 a setembro de 2014 houve uma redução nas exportações da castanha de caju. Os EUA, principal importador, que em média adquiria 61% de toda castanha exportada pelo Brasil, reduziram as compras para 35,67%. Essa redução se dá em função de alguns fatores, como a crise econômica e financeira mundial e pela forte quebra da safra de castanha no Brasil, devido às adversidades do clima na região Nordeste do país (CONAB, 2014).

Apesar da importância do cajueiro para o setor agroindustrial, seu potencial biotecnológico ainda é pouco explorado devido à ausência de dados genômicos. Com os recentes avanços nas tecnologias de sequenciamento, foi possível a obtenção do transcriptoma de sementes e folhas do cajueiro anão-precoce CCP 76 (ALVES-FILHO, 2013).

**Gráfico 1:** Produção de castanha de caju em 2012, no mundo.



O rendimento é dado em dólares (\$) e a produção em milhões de toneladas (Mt). **Fonte:** FAO, 2014; adaptado pelo autor.

### 1.3 Análises Transcriptômicas

Entende-se por transcriptômica, a área científica responsável pelo estudo do transcriptoma, ou seja, o conjunto completo de transcritos em uma célula, e suas quantidades em um estágio específico do desenvolvimento ou condição fisiológica (WANG; GERSTEIN; SNYDER, 2009).

Técnicas como o RNA-Seq tem permitido o estudo de organismos que não possuem um genoma de referencia ou que apresentam poucos dados moleculares conhecidos (BRÄUTIGAM et al., 2011). Essa abordagem se baseia nas mudanças relativas da expressão em larga escala de genes, e tem sido de suma importância para o entendimento de dados moleculares em espécies não modelo, tais como *Cicer arietinum* L. (GARG et al, 2011), *Ipomoea batatas* (WANG et al, 2010) e *Anacardium occidentale* L. (ALVES FILHO, 2013).

Alves Filho (2013) utilizou a abordagem *De novo* para a montagem do primeiro esboço do transcriptoma de sementes de cajueiro (*A. occidentale* L.), utilizando dados sequenciados pela plataforma *Illumina*, que é descrita na literatura como eficiente para a verificação da expressão de genes, metilação do DNA, re-sequenciamento, bem como para o sequenciamento de transcriptoma *De novo* (PARCHMAN et al., 2010; WANG et al., 2010).

Os recentes avanços nas tecnologias de próxima geração de sequenciamento têm facilitado a obtenção de genomas e transcriptomas com rapidez e boa relação custo/benefício, representando, assim, uma alternativa atraente e eficiente para auxiliar estudos genéticos e funcionais (CHIARA et al., 2013).

#### **1.4 Sequenciamento de Próxima Geração (Nova Geração de Sequenciamento - NGS)**

O avanço do conhecimento de genomas modelo e o advento de técnicas de sequenciamento de próxima geração (*Next Generation Sequencing*) abrem a possibilidade de grande salto no conhecimento do genoma de espécies relativamente pouco conhecidas (MARTINS, 2013). As novas tecnologias de sequenciamento começaram a ser comercializadas em 2005 e estão evoluindo rapidamente. Todas essas tecnologias promovem o sequenciamento de DNA em plataformas capazes de gerar informação sobre milhões de pares de bases em uma única corrida.

Dentre as plataformas de sequenciamento de próxima geração, as mais utilizadas são as plataformas 454, Solexa e SOLiD. A plataforma 454 (Roche) foi a primeira plataforma de sequenciamento a ser comercializada e utiliza a nanotecnologia e a metodologia de pirosequenciamento. Essa plataforma utiliza como método de amplificação clonal a PCR em emulsão, onde a amplificação das moléculas de DNA ocorrem em microesferas lipídicas, e como método de detecção de nucleotídeos a DNA polimerase. (MARGULIES et al., 2005; CARVALHO; SILVA, 2010). A tecnologia SOLiD (da *Applied Biosystems*), é baseada na ligação de oligonucleotídeos e posterior detecção. Diferente da plataforma 454, emprega a

DNA ligase ao invés da DNA polimerase na reação, contudo, a amplificação clonal também ocorre por meio da PCR em emulsão (SHENDURE; JI, 2008).

O sequenciamento da plataforma Solexa (*Illumina*) é realizado por síntese e a inovação dessa plataforma consiste na clonagem *in vitro* dos fragmentos em uma plataforma sólida de vidro, processo também conhecido como PCR de fase sólida (CARVALHO; SILVA, 2010). Nessa tecnologia o DNA é fragmentado, ligado a adaptadores e em seguida aderidas a um suporte sólido. Na etapa de anelamento, o adaptador da extremidade livre encontra seu oligonucleotídeo complementar no suporte, formando uma estrutura em ponte e dando início aos ciclos de PCR (*Bridge PCR*). Na desnaturação há um aumento na temperatura desfazendo a ponte. Após vários ciclos, serão obtidos clusters de moléculas idênticas ligadas ao suporte. Para que ocorra a leitura, o *primer* de sequenciamento é anelado seguido pela adição de base (uma por vez) cada base é marcada com uma fluorescência e a leitura é feita de forma sequencial, o que permite a montagem da sequência completa de cada *cluster* (SHENDURE; JI, 2008; CARVALHO; SILVA, 2010).

As plataformas 454 da Roche e *Illumina/Solexa*, têm sido usadas na montagem *De novo* do transcriptoma. Embora a tecnologia de sequenciamento Roche possa produzir segmentos de leitura mais longos, a *Illumina* possibilita a obtenção de cobertura mais profunda e maior precisão com o mesmo custo, o que é benéfico para a descoberta de genes e marcadores moleculares (MARTINS, 2013). Estas novas tecnologias de sequenciamento usadas em conjunto com ferramentas de bioinformática, constituem um poderoso recurso para a compreensão dos genomas, especialmente de espécies pouco estudadas e de alto impacto econômico (MARTINS, 2013).

## 1.5 Marcadores Moleculares

A caracterização de genótipos de cajueiro têm sido baseada principalmente na morfologia (SWAMY; RAO; BHAT, 1998). Os marcadores morfológicos são bastante utilizados devido à sua facilidade de aplicação, contudo, a necessidade de um grande número de características, o gasto excessivo de tempo e recursos para a realização do trabalho, além da grande influencia ambiental, torna o uso desses marcadores limitado (BIANCHI et al., 2004; JESUS, 2006). Tais complicações podem ser eliminadas com o uso de marcadores moleculares.

Marcadores moleculares podem ser definidos como todo e qualquer fenótipo molecular oriundo de um gene expresso ou de um segmento de DNA (FERREIRA; GRATTAPAGLIA, 1998). Os marcadores moleculares podem ser classificados em dois grupos, de acordo com a metodologia utilizada para a identificá-los, os marcadores baseados em hidridização e os marcadores baseados na amplificação por PCR (*Polymerase Chain Reaction* - Reação em Cadeia da Polimerase).

Os marcadores identificados por hibridização são os marcadores RFLP (*Restriction Fragment Length Polymorphism* - Polimorfismo no comprimento de fragmentos de restrição) e os minissatélites ou locos VNTR (*Variable Number of Tandem Repeats* – Sequências adjacentes que se repetem em número variável) (JEFFREYS; WILSON; THEIN, 1985). Enquanto que os marcadores baseados em amplificação por PCR são os do tipo RAPD (*Random Amplified Polymorphic DNA* - Polimorfismo de DNA amplificado ao acaso) (WILLIAMS et al., 1990), AFLP (*Amplified Fragment Length Polymorphism* - Polimorfismo de comprimento de fragmentos amplificados) (ZABEAU; VOS, 1993; VOS et al., 1995), SNP (*Single Nucleotide Polymorphism* - Polimorfismo de Nucleotídeo Único), ISSR (*Inter Simple Sequence Repeats* - Sequências Simples Repetitivas Internas) (VARSHNEY et al., 2007; SEHGAL; RAINA, 2008), e Microsatélites ou SSR (*Simple Sequence Repeats* - Sequências Simples Repetidas) (HAMADA; PETRINO; KAKUNAGA, 1982; WEBER, 1990).

Marcadores moleculares podem ser derivados de qualquer tipo de dado molecular que forneça um polimorfismo detectável entre os organismos a serem comparados. Eles diferem quanto à abundância e distribuição no genoma, nível de

polimorfismo, conteúdo de informação genética, especificidade de *locus*, entre outras características (BUSO et al., 2003).

Os marcadores moleculares têm sido utilizados em análise genética com as mais diversas finalidades, tais como identificação de clones (OLIVEIRA et al., 2012), no melhoramento genético (PRASANNA et al., 2010), caracterização da diversidade genética entre cultivares (SILVA et al., 2011; MARTINS et al., 2012), paternidade (XAVIER et al., 2014), caracterização de germoplasma (SANTOS et al., 2013), construção de mapas genéticos (NAGY et al., 2012; IZZAH et al., 2014) entre outros.

Archak e colaboradores (2003) compararam três tipos de marcadores moleculares, RAPD, ISSR e AFLP, para verificar a eficiência e utilidade destes marcadores para detectar variações em dezenove acessos de cajueiro. Syed e colaboradores (2005) desenvolveram um sistema de marcador SSAP (*Sequence-Specific Amplification Polymorphism*) para o cajueiro. Segundo o autor, marcadores SSAP oferece vantagens sobre os métodos utilizados até a presente data, por apresentarem níveis mais elevados de polimorfismos para o cajueiro.

O marcador molecular utilizado como ferramenta nesse estudo foi o marcador microssatélite. Para melhor entendimento, a metodologia adotada será abordada de forma mais detalhada a seguir.

## **1.6 Marcadores Microssatélites**

Os marcadores microssatélites, também chamados de *Simple Sequence Repeat* (SSR) ou *Short Tandem Repeats* (STR), são sequências constituídas de 1 a 6 nucleotídeos repetidos em *tandem* (TAUTZ 1993; FERREIRA; GRATTAPAGLIA, 1998). Os SSRs estão ampla e uniformemente distribuídos pelo genoma dos eucariotos, eles podem ser encontrados tanto em regiões codificadoras quanto em regiões não codificadoras (ARCOT et al., 1995; WILDER; HOLLOCHER, 2001), sendo os motivos tri e hexanucleotídeos os mais comuns em regiões codificadoras, pois não alteram a leitura dos códons (TÓTH et al., 2000).

As vantagens desses marcadores sobre os outros é que são altamente informativos, polimórficos e multialélicos. São marcadores codominantes, ou seja, ambos os alelos de um indivíduo heterozigoto podem ser visualizados no gel, são baseados em PCR, amplificam regiões específicas e demandam pequena quantidade de DNA (MORGANTE; OLIVIERI, 1993; PARIDA et al., 2009).

Existem mecanismos que explicam as altas taxas de polimorfismo desses marcadores, como a inclusão de erros durante a recombinação homóloga e a ocorrência de erros da polimerase durante a replicação ou reparo do DNA (STRAND et al., 1993). Devido à sua alta mutabilidade, microssatélites são conhecidos por ter um papel significativo na evolução do genoma pela criação e manutenção da variabilidade genética (TAUTZ; TRICK; DOVER, 1986; KASHI; KING; SOLLER, 1997).

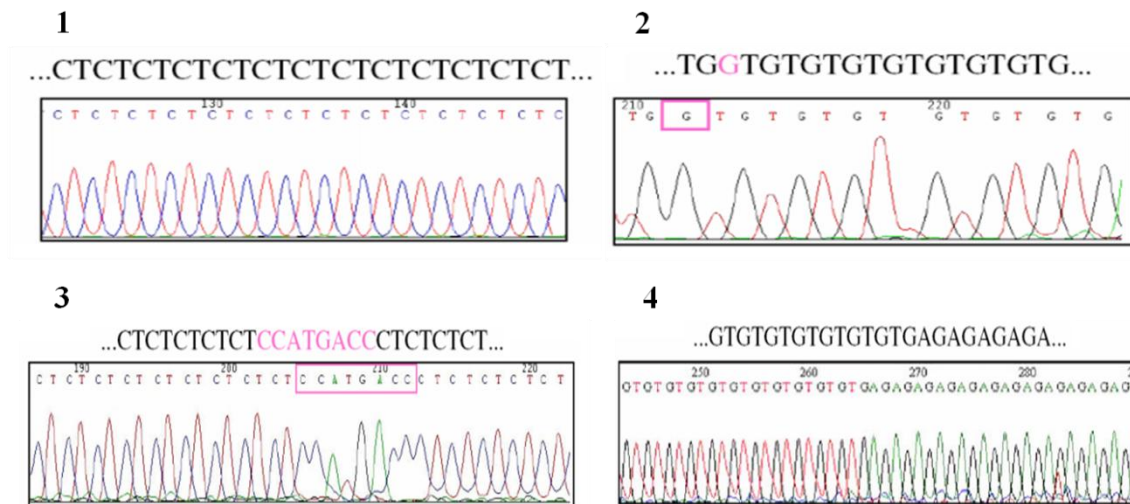
Para se amplificar regiões repetitivas do tipo microssatélite é necessário utilizar um par de *primer* específico que possua homologia com uma região única no genoma e que flanqueie uma sequência repetitiva, que pode variar, dependendo de cada indivíduo (DE FREITAS, 2005). O número de repetições de uma sequência microssatélite normalmente varia, enquanto que as regiões que a flanqueiam podem ser únicas e conservadas no genoma de diferentes indivíduos de uma mesma espécie (DE FREITAS, 2005).

Segundo Goldstein e Schlötterer (1999), os microssatélites podem ser organizados em 4 classes: Microssatélites perfeitos, onde o motivo se repete sem interrupção de outro motivo ou outra sequência; os microssatélites imperfeitos apresentam uma base que altera o padrão de repetição; nos microssatélites interrompidos há a inserção de uma pequena sequência de bases que não se encaixa no motivo de repetição e os microssatélites compostos são aqueles nos quais existe a repetição de um ou mais tipos de motivos (Figura 2).

Os SSRs têm sido aplicados em diversos estudos em plantas como, por exemplo, construção de mapas genéticos em grão de bico, tremoço, amendoim, cacau, framboesa, entre outras espécies (ZHANG et al., 2010; GAO et al., 2011; YU et al., 2012; SARGENT et al., 2012; WARD et al., 2013), nos estudos de diversidade genética em arroz, café, maçã e cana de açúcar (ZHANG et al., 2011; GELETA et al., 2012; PATZAK et al., 2012; HAMEED et al., 2012) e na identificação de

paternidade de algumas espécies, como por exemplo da cana de açúcar (XAVIER et al., 2014).

**Figura 2:** Classificação dos microssatélites segundo Goldstein e Schlötterer (1999).



**1** - Microssatélites perfeitos; **2** - Microssatélites imperfeitos; **3** - Microssatélites interrompidos; **4** - Microssatélites compostos. **Fonte:** Goldstein e Schlötterer, 1999.

As primeiras análises acerca de marcadores microssatélites para o cajueiro foram realizadas por Croxford, Robson e Wilkinson (2006), onde desenvolveram 21 marcadores microssatélites para a espécie. No ano seguinte, Cavalcanti e Wilkinson (2007) utilizaram marcadores AFLP e microssatélites na construção dos dois primeiros mapas genéticos do cajueiro, um para o cajueiro comum e outro para o anão-precoce. O mapa final incluiu apenas 11 marcadores SSR. Lamas (2010), testou 100 novos marcadores SSR para o cajueiro, destes, 14 foram caracterizados com acessos de *A. occidentale* do Banco de Germoplasma.

O número de marcadores SSR para o cajueiro ainda é limitado, sendo necessário o desenvolvimento de novos marcadores para que haja avanços no conhecimento genético da espécie.



## **CAPÍTULO I**

### **BUSCA E CARACTERIZAÇÃO *in silico* DE MARCADORES MICROSSATÉLITES NO TRANSCRIPTOMA DE SEMENTES E FOLHAS DE CAJUEIRO (*A. occidentale* L.)**

## 2 BUSCA E CARACTERIZAÇÃO *in silico* DE MARCADORES MICROSSATÉLITES NO TRANSCRIPTOMA DE SEMENTES E FOLHAS DE CAJUEIRO (*A. occidentale* L.)

### 2.1 INTRODUÇÃO

A grande quantidade de dados gerados pelas tecnologias NGS junto com erros inerentes a cada tecnologia de sequenciamento, desafiam os projetos de montagem de genomas e transcriptomas de várias espécies (MARTINS, 2013). A bioinformática tem um papel fundamental, pois os pequenos fragmentos obtidos após o sequenciamento devem ser montados para obter sequências inteiras (MARTINS, 2013).

Existem duas maneiras de se fazer a montagem de transcriptomas. A primeira é utilizando um genoma de referência e a segunda sem genoma de referência, também conhecida como montagem *De novo*. A montagem de transcriptomas com genoma de referência tem a vantagem de se poder fazer análise de expressão de genes, porém é necessário que o organismo em questão já tenha o seu genoma previamente sequenciado. Contudo, em organismos que não possuem o genoma sequenciado, é possível fazer uma montagem utilizando a estratégia *De novo*. Nesse tipo de montagem, temos a vantagem de não haver a necessidade de um sequenciamento do genoma do organismo em questão, mas a análise de expressão de genes acaba sendo inviável (MARTIN; WANG, 2011).

O processo de anotação de um genoma consiste na análise e interpretação dos dados de sequenciamento com o intuito de colocá-lo no contexto de nossa compreensão dos processos biológicos (STEIN, 2001). Uma das ferramentas utilizadas para a comparação rápida de sequências é o algoritmo BLAST (*Basic Local Alignment Search Tool* - Ferramenta Básica de Alinhamento Local) (ALTSCHUL et al., 1990).

O objetivo do BLAST é encontrar uma sequência homóloga em outro organismo que possua uma função conhecida. Torna-se então necessário a criação

de um banco de dados que associe proteínas e suas respectivas funções biológicas. Esse foi o principal objetivo do *Gene Ontology Consortium* (GO), que pode ser acessado pelo site (<http://www.geneontology.org>). Os termos do *Gene Ontology* são agrupados em três grandes categorias: Função molecular, processo biológico e componente celular (ASHBURNER et al., 2000).

Para visualização de vias metabólicas, uma das ferramentas mais utilizadas é o KEGG (*Kyoto Encyclopedia of Genes and Genomes*) *Pathways* (KANEHISA et al., 2006). Através dele é possível a visualização de vias importantes e as enzimas envolvidas no processo (KANEHISA et al., 2006).

Dado o exposto, a utilização dessas metodologias expande os conhecimentos acerca dos mecanismos moleculares utilizados pela espécie em estudo, fornecendo subsídios importantes para programas de melhoramento genético, busca por novos marcadores moleculares, análise da expressão de genes e das vias metabólicas relacionadas, dentre outros. Podendo assim, viabilizar a prospecção do seu potencial biotecnológico e econômico, bem como o fornecimento de dados para estudos posteriores.

Os marcadores moleculares adotados para a identificação individual de plantas e animais são baseados na amplificação de segmentos curtos de DNA de 1 a 6 pares de bases repetidos em *tandem*, denominados de microssatélites ou STR ou, ainda, SSR (SANSALONI, 2008). Os SSRs têm sido escolhidos como o marcador molecular para diversas aplicações em estudos de plantas devido à hipervariabilidade, codominância e extensiva cobertura no genoma (KALIA et al., 2011).

Após a identificação de uma região com motivos de microssatélite, duas outras condições tornam um *loci* propício a ser utilizado. Inicialmente é importante que as regiões flaqueadoras sejam conservadas o bastante para poderem ser utilizadas para a construção de *primers* para a reação de PCR. Além disso, é importante haver polimorfismo para o *locus*. Vale salientar que o conteúdo informativo do SSRs não está na diferença entre o número de unidade de repetição de um alelo, mas sim na frequência deste alelo em uma população. Deste modo, após a identificação, o SSR deverá ser validado, os tipos alélicos determinados e, por fim, a frequência destes alelos deve ser identificada.

## 2.2 OBJETIVOS

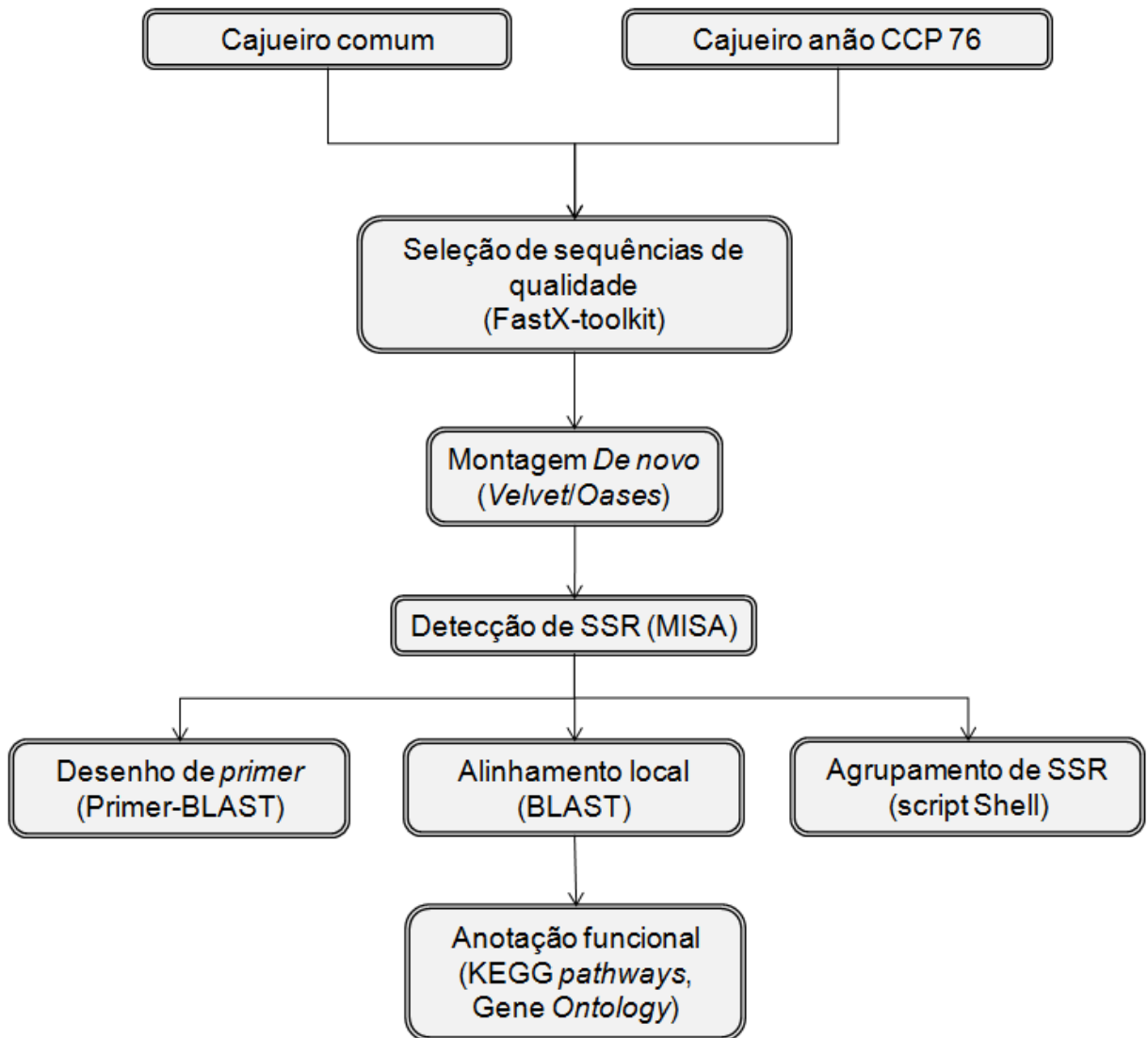
### 2.2.1 Objetivo Geral

Identificar microssatélites no transcriptoma de sementes e folhas de cajueiro comum e anão CCP 76 utilizando ferramentas *in silico* visando o desenvolvimento de novos marcadores moleculares.

### 2.2.2 Objetivos Específicos

- Obter uma montagem *De novo* do transcriptoma do cajueiro comum e anão-precoce CCP 76;
- Identificar a presença de marcadores microssatélites no transcriptoma do cajueiro comum e CCP 76;
- Avaliar a frequência dos microssatélites do tipo di, tri, tetra, penta e hexanucleotídeo;
- Identificar os microssatélites compartilhados entre o cajueiro comum e anão CCP 76;
- Atribuir anotação funcional para as sequências que contém microssatélites identificadas no transcriptoma do cajueiro comum e anão-precoce CCP 76;
- Desenhar *primers* para a amplificação das regiões que contém SSR.

## 2.3 ESTRATÉGIA EXPERIMENTAL



## 2.4 METODOLOGIA

### 2.4.1 Obtenção das sequências do transcriptoma do cajueiro

As sequências do transcriptoma do cajueiro foram previamente obtidas de trabalhos realizados por Alves-Filho (2013) e Moreira (2014). Amostras de folhas de cajueiro anão-precoce CCP 76 e sementes de cajueiro comum e anão-precoce CCP 76 foram coletadas em Itapipoca-CE. O RNA das folhas (MOREIRA, 2014) e das sementes (ALVES-FILHO, 2013) foram extraídos a partir das amostras congeladas em nitrogênio líquido, no Núcleo de Biotecnologia de Sobral e encaminhadas para sequenciamento no Laboratório de Biotecnologia Animal - Esalq-USP utilizando a plataforma *Illumina HiSeq2000*. O sequenciamento das bibliotecas de cDNA foi do tipo *paired-end* e produziu *reads* de 50 pb.

### 2.4.2 Análise da qualidade das sequências e montagem *De novo* do transcriptoma de cajueiro

O processamento dos dados foi feito em um computador HP Proliant com 8 núcleos de processamento e 16,7 Gb de memória RAM. Os dados brutos (arquivos no formato *fastq*) contendo dados de sequência e qualidade de base foram avaliados pelo programa *FastQC* (ANDREWS, 2013). As leituras (*reads*) de baixa qualidade foram removidos (trimados) utilizando a ferramenta *FastX Toolkit* (HANNON LAB, 2013) de forma a excluir nucleotídeos com qualidade *Phred* abaixo de 30 (probabilidade de um erro em 1000) e comprimento de sequência menor que 15 nucleotídeos.

A montagem *De novo* do transcriptoma de sementes e folhas do cajueiro foi feita utilizando os programas *Velvet* e *Oases* (ZERBINO; BIRNEY, 2008). Para avaliar os melhores parâmetros de montagem, utilizamos os resultados obtidos por Pontes (2013), com algumas modificações. Os principais parâmetros utilizados incluem o valor de K-mer igual a 31, cobertura esperada de 30 e corte de cobertura de 30.

Para aperfeiçoar a montagem, foi utilizada uma segunda montagem utilizando o programa Oases ajustando o corte de cobertura para 3 e o tamanho mínimo do transcrito de 300 nucleotídeos.

#### **2.4.3 Detecção *in silico* de marcadores microssatélites**

A detecção de microssatélites (SSR) foi feita utilizando o programa MISA (MicroSATellite), depositado no seguinte endereço eletrônico: <http://pgrc.ipk-gatersleben.de/misa/> (THIEL et al., 2003). As sequências dos *contigs* que contêm microssatélite foram recuperadas pelo programa *Fastacmd*. Em seguida, os *contigs* que contêm microssatélites do cajueiro anão e comum foram comparadas pelo programa BLASTn (ALTSCHUL et al., 1990). O valor E (*E-value*), que é a probabilidade do alinhamento ter sido feita ao acaso, foi abaixo de  $1 \times 10^{-5}$ .

Para a contagem dos microssatélites agrupados em categorias, foi desenvolvido um script em Shell especialmente para este propósito. O script lê o arquivo de saída do programa MISA, seleciona os SSR em di- e trinucleotídeos e os agrupa em motivos facilitando a contagem.

#### **2.4.4 Anotação funcional das sequências que contêm microssatélites**

A identificação dos genes referentes aos *contigs* que continham microssatélites foi feita pelo programa *Basic Local Alignment Search Tool* (Com a ferramenta BLASTx), através do alinhamento das sequências do transcriptoma dos cajueiros comum e anão-precoce CCP 76 contra os bancos de dados do *Swiss-Prot*. Os parâmetros foram ajustados para conter apenas sequências com o *E-value* inferior ou igual a  $1 \times 10^{-5}$ . As sequências do banco de dados foram obtidas pelo servidor *UniProt* e filtradas utilizando a palavra-chave "Plant" a fim de exibir apenas sequências referentes a proteínas vegetais.

Os números de acesso obtidos pelas identificações do BLAST foram novamente pesquisados pela ferramenta *Retrieve/ID Mapping* do *UniProt*, e os dados recuperados foram, posteriormente, convertidos em termos de GO, sendo, então, agrupados nas categorias: “Função Molecular”, “Processo Biológico” e “Componente Celular”. Além disso, destas sequências não redundantes foram obtidos os termos de KO (*Kegg Orthology*), para tornar possível a identificação das possíveis vias metabólicas que concernem às proteínas identificadas, através do servidor online *KEGG Pathways* (KANEHISA et al., 2006; CONESA et al, 2005).

#### **2.4.5 Desenho de *primers***

Cada uma das sequências compartilhadas do cajueiro anão e comum (que possuem microssatélite) juntamente com outras sequências similares pertencentes a outras plantas foram alinhadas pelo programa *ClustalW* (THOMPSON; GIBSON; HIGGINS, 2002). As 298 sequências compartilhadas entre os cajueiros comum e anão foram utilizadas para o desenho dos *primers*. Sequências consideradas pequenas (< 200 pb) ou aquelas onde os marcadores SSR estavam localizados nas extremidades não foram possíveis desenhar *primer*. A ferramenta primer-BLAST (YE et al., 2012) foi utilizada para desenhar pares de *primers* complementares a regiões flangeadoras dos microssatélites.

Foram desenhados 29 pares de *primers* para avaliar o polimorfismo em nove genótipos de cajueiro anão-precocce. Desses, 15 *primers* (V1C01 - V1C15) foram obtidos de *loci* polimórficos entre os cajueiros comum e anão CCP 76 e 14 (V2C01 - V2C14) de *loci* monomórficos (Tabela 2).



**Tabela 2:** Lista dos *primers* desenhados em sequências que flanqueiam SSR dos transcritos de cajueiro comum e CCP 76.

<b>SSR</b>		<b>Primer</b>	<b>Motivo comum</b>	<b>Motivo CCP 76</b>	<b>Tamanho amplicon</b>
V1C01	F	GCAAGAACCCACCTTCTGGA	(GA)7	(GA)8	171
	R	AACGCGTTGATCTCTCCATC			
V1C02	F	GCCATTGCTTGTGTTTGT	(TC)7	(TC)9	300
	R	CTAACATCTGAACCCACCTC			
V1C03	F	TGATGACAACTTAGGGGAG	(AT)11	(AT)10	246
	R	AACTGAAACGATACTTGAAC			
V1C04	F	AGTATCATCCCTTCTTCCCTA	(CT)11	(CT)12	388
	R	GGAAGTTGTTTCATATTCGGG			
V1C05	F	GCCATTCTTCCCCTGAGA	(AT)7	(AT)8	260
	R	TATTACATGGTTTCACTATTGGC			
V1C06	F	GGCCAATTGATGAGAAGAGA	(GGA)7	(CCT)5	122
	R	CCACCAGGACTTCCATTAC			
V1C07	F	CGAAGTACCAGATCGAATGC	(TGG)7	(TGG)6	293
	R	GGGTGTAGGGTAATTCCAG			
V1C08	F	CATATTGATGCAGACACACC	(TGA)10	(TGA)8	276
	R	CATACACTGACACCAACTTC			
V1C09	F	CGCTTGATCCTCATCAGG	(GAT)6	(GAT)5	169
	R	CCGATGTATAATTCAGTATC			
V1C10	F	CTTCCAACTGAAGGCATTATT	(ACA)8	(TGT)9	199
	R	AGTTTGATGAGTCAACTCGG			
V1C11	F	CGGGAGAATGAACTTTAACC	(ACA)10	(TGT)8	152
	R	GGGGCGATACAATTTGAAG			
V1C12	F	TAGCCAAAACAACCACCG	(CCT)8	(CCT)6	273
	R	GGACCTTGAGCAATTGAAG			
V1C13	F	AACGGTGACCCTTCAAATGC	(CTG)5	(CTG)6	104
	R	GTAAACCCTTGCCCAATCC			
V1C14	F	ATTTCAAGTGATTCCACTTGTTC	(CTT)6	(GAA)7	134
	R	ATTTCAAGTGATTCCACTTGTTC			
V1C15	F	ATGAGGCATGCTTTCAGTG	(TTC)7	(TTC)6	147
	R	AGGTCTAAGCCATCAGAAAAG			

Continua...

<b>SSR</b>		<b>Primer</b>	<b>Motivo comum</b>	<b>Motivo CCP 76</b>	<b>Tamanho amplicon</b>
V2C01	F	GGCTTGAGCGATTTGTTTGAG	(AT)7	(AT)7	224
	R	GATACTCAACCGTGAATCATATCAG			
V2C02	F	GTGGCCATTCCCAAGAGAACT	(AG)8	(AG)8	183
	R	CGAAACCTAAGACAGGTCCAAGAAT			
V2C03	F	CCCAAGGACACCAAAGGCAT	(TG)6	(TG)6	337
	R	GGTGGACCCTTCTACTCCATT			
V2C04	F	GAGCAACCTTCGGCTGTCAA	(TG)6	(TG)6	257
	R	ATCTTTCTCATCAACGGTGGG			
V2C05	F	CATCCTAACCACGGCTTTCC	(AT)7	(AT)7	205
	R	TGCTGCAGTGATGATGAGGA			
V2C06	F	TCAATCCTCAGGAGACACTCAC	(GA)6	(GA)6	190
	R	CTGAAAACGGATGGAGCCCAC			
V2C07	F	CCTTTGTCTCATCAGCAGGG	(TGG)5	(TGG)5	139
	R	GAGACGACCACCCCTTTTGC			
V2C08	F	GTGTGCAATAAGCACCAACC	(CAT)5	(CAT)5	271
	R	TGCAGATTGATTTGGCTGCTC			
V2C09	F	CACCCGTGAAAACATCTCCTC	(GCT)6	(GCT)6	236
	R	TGTGGAAAGTTGAGGGGTTCTG			
V2C10	F	GCTTCTTTGGAAGTGTGGCT	(TGC)5	(TGC)5	238
	R	CCCGAGCCTGAACCAGATAA			
V2C11	F	CGGGGCTGAATGCTCCTATT	(TCC)5	(TCC)5	400
	R	GTCCCTGGTGCAATCCTCTC			
V2C12	F	CAAGGAGGGGTTCGAAGCAG	(TTTC)6	(TTTC)6	293
	R	GTGCAACCAGTTACCCACC			
V2C13	F	GTCTATGCGCAACACGAAGTC	(GGA)5	(GGA)5	175
	R	CGACACAGACACCTCCATTCT			
V2C14	F	ACTTTCGCAACAAGAAAATAGGGA	(ATC)5	(ATC)5	232
	R	GATTGCCAAGGAATCTACAGCC			

**Fonte:** Autor.

## 2.5 RESULTADOS E DISCUSSÕES

### 2.5.1 Montagem do transcriptoma do cajueiro comum e CCP 76 a partir de sequências de sementes e folhas

Um *pool* de sequências do transcriptoma de sementes e folhas de cajueiro CCP 76, num total de 42.786.272 *reads*, foram utilizadas para a montagem do transcriptoma do cajueiro anão. Foram obtidos 68.811 transcritos (*contigs*), utilizando 17,2% dos *reads*. O tamanho do maior *contig* é de 5.382 pb e o valor de N50 é igual a 460 (Tabela 3). O transcriptoma do cajueiro comum foi obtido a partir de sequências de sementes, totalizando 16.347.083 *reads*. Após a montagem, foram obtidos um total de 24.098 transcritos, utilizando 17,7% dos *reads*. O tamanho do maior *contig* é de 3.065 pb e o valor de N50 é de 536 (Tabela 3).

O resultado da montagem do *Velvet* foi melhorada após a utilização do programa *Oases*, o qual reduziu o número de *contigs*. O programa *Oases* foi desenvolvido especificamente para a montagem de transcriptomas *De novo* usando *reads* curtos e leva em consideração a montagem feita pelo *Velvet*. Tem sido sugerido que a montagem pelo *Velvet* seguido pelo *Oases* produz os melhores *contigs/transcritos* (GARG et al., 2011).

Ashrafi e colaboradores (2012) obtiveram 68.737 *contigs* na montagem *De novo* do transcriptoma de pimentão (*Capsicum annum*), utilizando o programa *Velvet* com valor de k-mer de 31. No presente estudo, foram obtidos 68.811 *contigs* no cajueiro anão-precoce e 24.098 *contigs* no cajueiro comum, utilizando parâmetros de montagem semelhantes aos utilizados por Ashrafi e colaboradores.

**Tabela 3:** Estatísticas de montagem do transcriptoma do cajueiro comum e anão-precoce CCP 76 usando os programas *Velvet* e *Oases*.

Dados da Montagem	Cajueiro Comum	Cajueiro anão-precoce CCP 76
Número de <i>reads</i> totais	16.347.083	42.786.272
Tamanho dos <i>reads</i> (pb)	50	50
Número dos Nucleotídeos	817.354.150	2.139.313.600
Número de <i>reads</i> utilizados na montagem	2.901.318 (17,7%)	7.361.509 (17,2%)
Número de <i>contigs</i>	24.098	68.811
N50	536	460
Cobertura média	30	30
Maior <i>contig</i>	3.065	5.382

Fonte: Autor.

### 2.5.2 Busca *in silico* por marcadores microssatélites

As sequências dos transcritos obtidos no transcriptoma do cajueiro foram utilizadas para as análises de SSR através do programa MISA. Após a montagem *De novo* do transcriptoma, foram detectados, no cajueiro comum, um total de 496 SSR nas 24.098 sequências examinadas. Aproximadamente 2% das sequências analisadas continham pelo menos um microssatélite. No cajueiro anão-precoce CCP 76, foram examinadas 68.811 sequências, onde foram identificados 2363 microssatélites, o qual corresponde a 3,4% das sequências analisadas (Tabela 4). Estudos usando dados de EST (Etiqueta de Sequência Transcrita), mostram que cerca de 3-7% das sequências expressas contêm motivos SSR (THIEL et al., 2003; RAJU et al., 2010; BLANCA et al., 2011).

Os SSRs identificados no cajueiro comum estão distribuídos em 432 sequências, sendo que, 63 (14,5%) sequências contém mais de um marcador microsatélite. No cajueiro anão, os microsatélites identificados se encontram em 1.913 sequências, destas, 397 (20,7%) sequências contém mais de um SSR (Tabela 4).

Tsanakas e colaboradores (2014), identificaram no transcriptoma de pétalas de gardênia (*Gardenia jasminoides*) um total de 9.549 SSRs, sendo que 1.398 (18.3%) contém mais de um marcador microsatélite. Analisando o transcriptoma do trevo vermelho (*Trifolium pratense* L.) e *Costus pictus* foram identificados 3.127 e 8.482 SSRs, respectivamente. Mais de um SSR foi encontrado em 16% dos transcritos em *T. pratense* L. (ANNADURAI et al., 2014) e no transcriptoma de *C. pictus*, 322 (11%) sequências contém mais de um SSR (YATES et al., 2014). Assim, os resultados obtidos no presente trabalho corroboram com outros registros na literatura que utilizaram metodologias semelhantes às utilizadas.

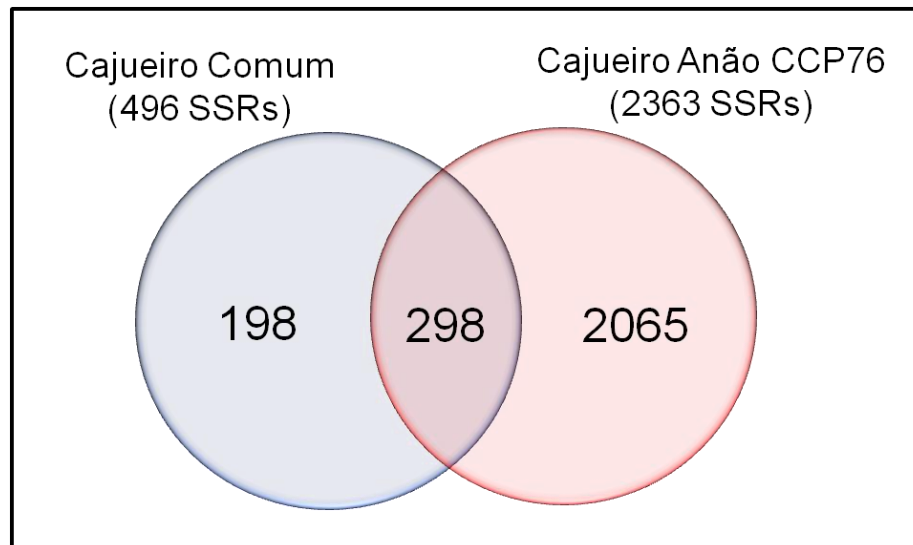
O alinhamento local entre os *contigs* de cajueiros anão e comum que contém os SSRs identificados, revelou que 298 SSRs são compartilhados (Figura 3). Destes, 96 demonstraram diferenças no número de unidades de repetição entre os genótipos avaliados, sendo este um forte indício da identificação de *locus* polimórficos.

**Tabela 4:** Estatísticas dos SSRs identificados a partir do transcriptoma do cajueiro comum e anão-precoce CCP 76.

Dados dos Marcadores Microsatélites	Cajueiro Comum		Cajueiro anão CCP 76	
	N°	%	N°	%
Número de sequências contendo SSR	432	1,8%	1.913	2,8%
Número de sequências contendo mais de 1 SSR	63	0,2%	397	0,4%
Número total de SSRs identificados	496	2,0%	2.363	3,4%

**Fonte:** Autor.

**Figura 3:** Diagrama de Venn demonstrando o número de SSRs totais e compartilhados entre o cajueiro comum e anão CCP 76.



**Fonte:** Autor.

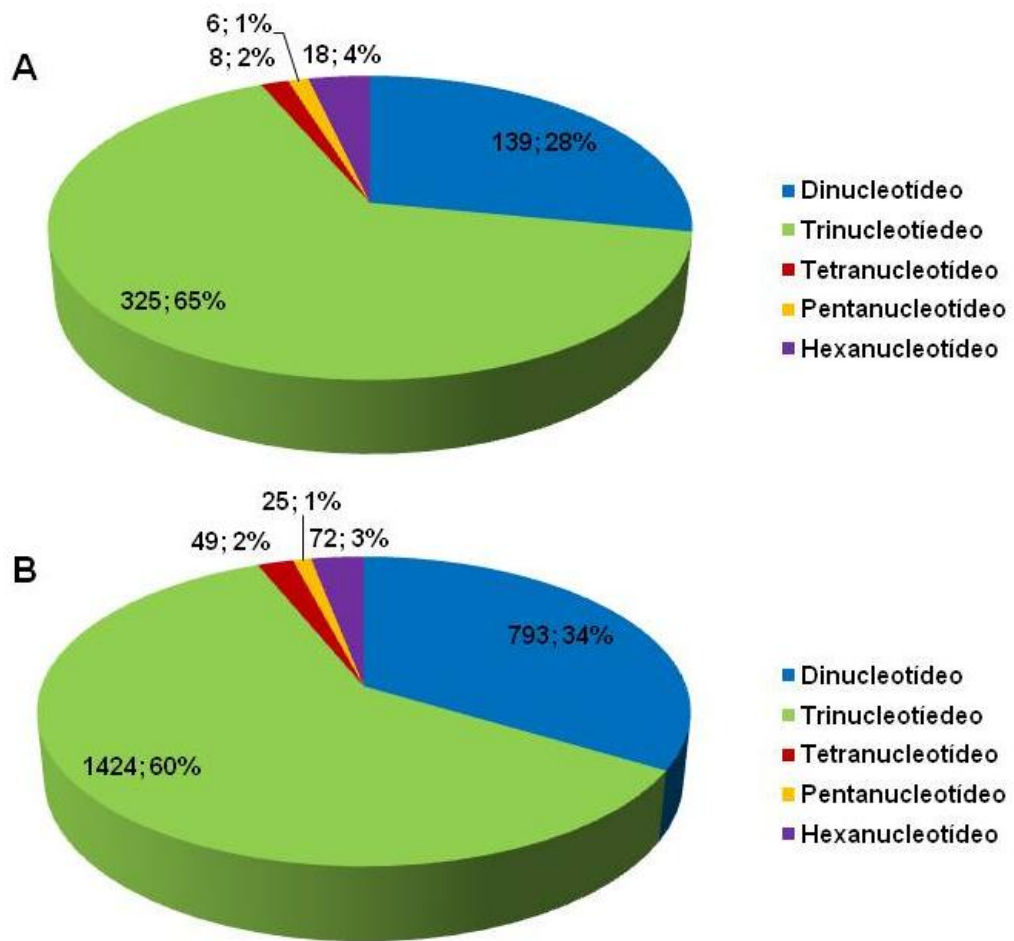
Foram identificados nas bibliotecas de transcritos do cajueiro anão e comum a presença de SSRs com motivos de repetição dos tipos, di, tri, tetra, penta e hexanucleotídeos. Tanto no cajueiro comum quanto no cajueiro anão-precoce, os motivos de repetição mais abundantes foram do tipo trinucleotídeos, seguido pelos dinucleotídeos. No cajueiro comum foi possível identificar 325 SSRs do tipo trinucleotídeo e no cajueiro anão 1424 SSRs, representando 65% e 60% do total de microsatélites, respectivamente. Quanto aos microsatélites do tipo dinucleotídeo, foram detectados 139 (28%) e 793 (34%) nos cajueiros comum e anão CCP 76, respectivamente (Gráfico 2).

Em menores proporções, também foram detectados SSRs do tipo tetra, penta e hexanucleotídeos, representando 2%, 1% e 4% dos microsatélites totais, respectivamente no cajueiro comum e 2%, 1% e 3% no cajueiro anão-precoce (Gráfico 2).

De modo geral, microsatélites em vegetais são repetições dinucleotídicas com motivos AG/CT (KALIA et al., 2011). Contudo, quando se trabalha com regiões codificadoras os motivos de repetição tendem a ser do tipo trinucleotídios (LI et al., 2002). Tal afirmação corrobora com resultados observados no presente estudo, onde se pôde observar uma maior distribuição deste tipo de motivos para ambos os genótipos, 60% dos SSRs para anão e 65% para cajueiro comum. Além disso, a presença de motivos do tipo di e tetranucleotídeos podem ser em regiões intrônicas ou nas regiões não traduzidas à montante e à jusante da região codificante (LIU et al., 1999).

Em estudos realizados com feijão-comum, foi possível identificar 331 EST-SSRs, e assim como os resultados mostrados, os motivos de repetição mais frequentes foram do tipo trinucleotídeo (45,6%), seguidos pelos dinucleotídeos (40,8%) (NETO, 2008). O mesmo pode ser observado em trabalhos feitos por Blanca e colaboradores (2011), Zhang e colaboradores (2012) e Tsanakas e colaboradores (2014) usando *Cucurbita pepo* (71,7% para trinucleotídeos e 15,3% para dinucleotídeo), *Arachis hypogaea* L. (59,6% e 36%) e *Gardenia jasminoides* (30,7% e 29,8%), respectivamente.

**Gráfico 2:** Distribuição dos microssatélites no transcriptoma do cajueiro comum e anão-precoce CCP 76, utilizando o programa MISA.



**A** - Microssatélites detectados no cajueiro comum. **B** - Microssatélites detectados no cajueiro anão-precoce CCP 76. **Fonte:** Autor.

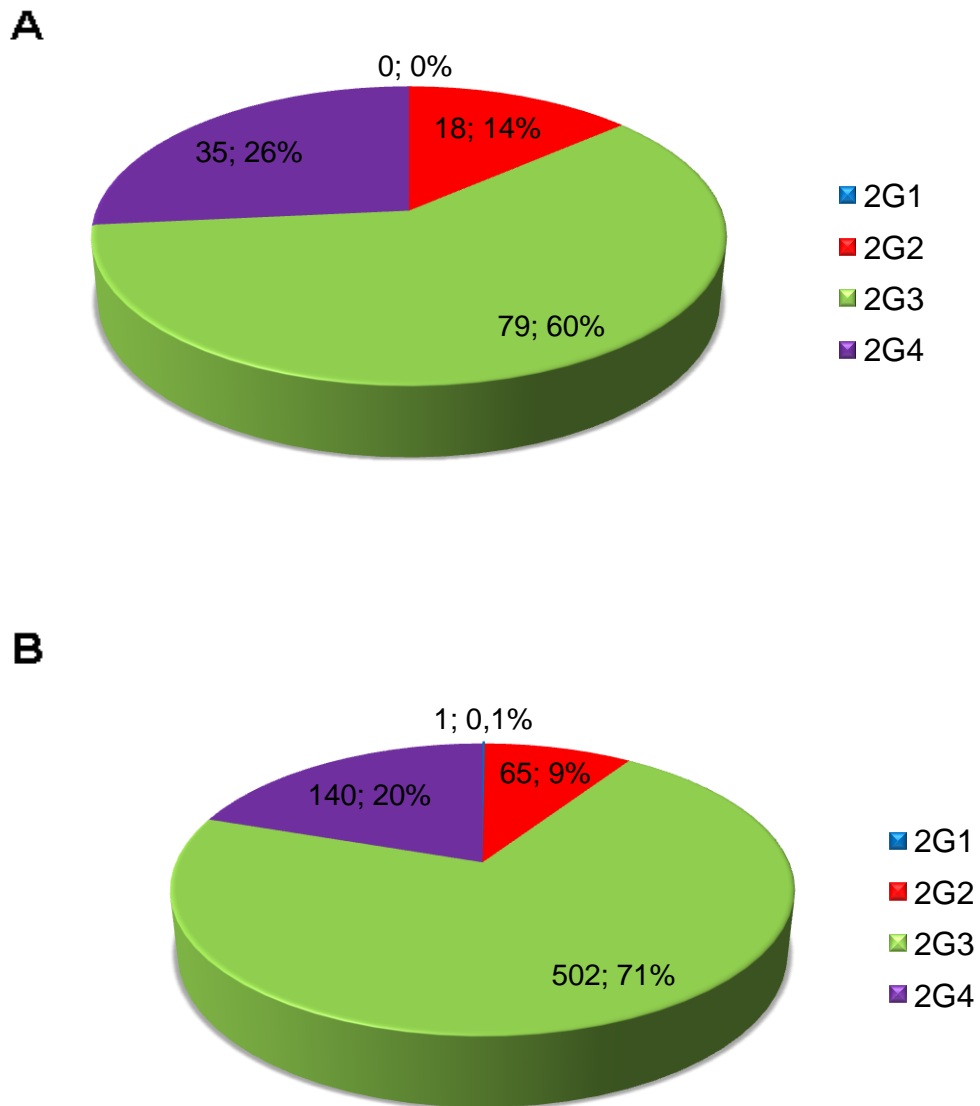


Os motivos dos microssatélites encontrados foram separados em grupos de acordo com dados de Yu e colaboradores (2011) e Zhao, Prakash e He (2012). Os dinucleotídeos foram alocados em quatro grupos: GC/CG; AC/CA/TG/GT; AG/GA/TC/CT e AT/TA, sendo observada uma maior distribuição de motivos para o grupo 3 (AG/GA/TC/CT), tanto para o cajueiro comum (79 SSRs/ 60%) quanto para o anão CCP 76 (502 SSRs/ 71%) (Gráfico 3). Vale ressaltar que não houve a presença de SSRs pertencentes ao grupo 4 (GC/CG) no cajueiro comum e apenas um (0,1%) SSR desse grupo para o cajueiro anão CCP 76. A pouca frequência (cajueiro anão) ou a ausência (cajueiro comum) de motivos de repetição do tipo GC/CG no presente estudo, corrobora com resultados encontrados por outros autores (LUO et al., 2011; BLANCA et al., 2011; WU et al., 2012).

Os motivos do tipo trinucleotídeos foram distribuídos em 10 grupos, ainda conforme proposto por Yu e colaboradores (2011) e Zhao, Prakash e He (2012). Houve uma maior abundância de marcadores SSRs pertencentes ao grupo 8 (AAG/AGA/GAA/CTT/TTC/TCT), correspondendo a 28% e 27% para os cajueiros comum e CCP 76, respectivamente (Gráfico 4).

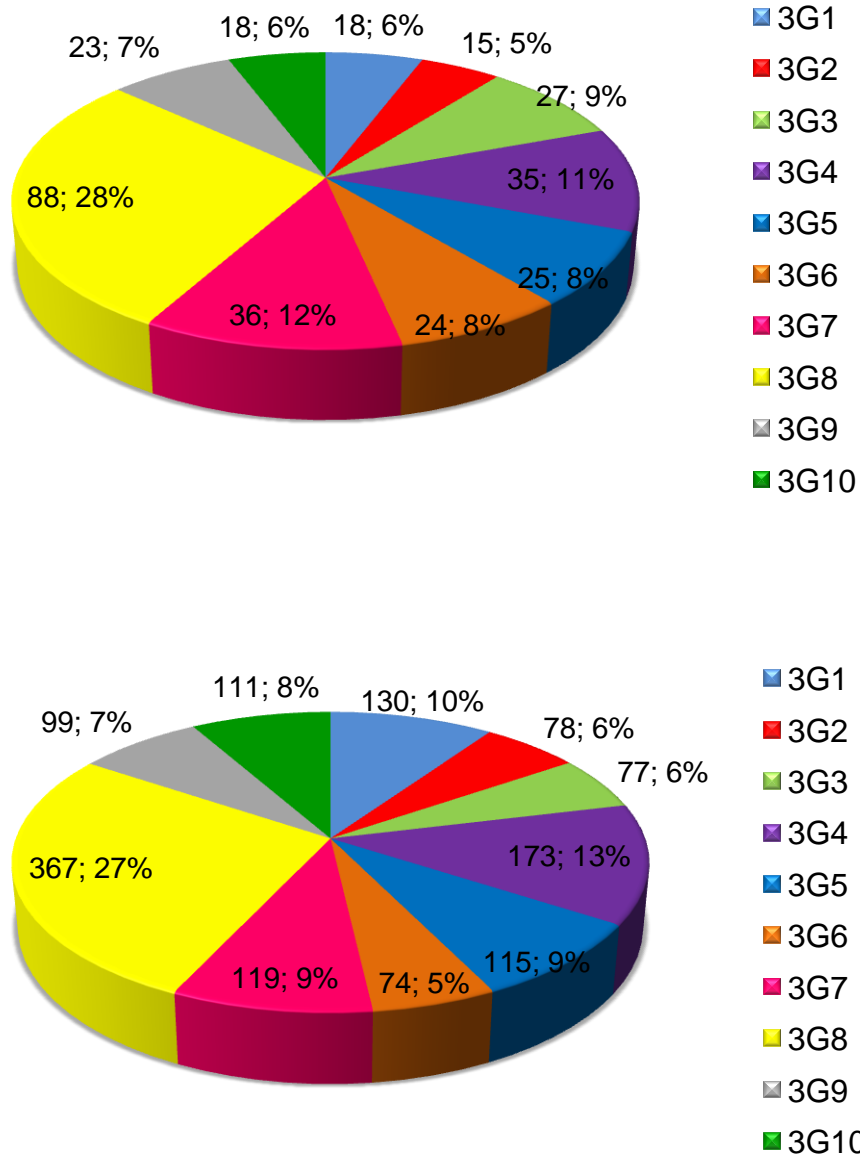
Estudos realizados com *Arachis hypogaea* L. (ZHANG et al, 2012), *Citrullus lanatus* (GUO et al., 2011), *Panax notoginseng* (LUO et al., 2011) e *Cucurbita pepo* (BLANCA et al., 2011), demonstraram que os motivos de repetição dos tipos dinucleotídeo e trinucleotídeo mais abundantes pertenciam as classes: AG/GA/CT/TC e AAG/GAA/AGA/CTT/TTC/TCT, respectivamente. Todos esses resultados corroboram com os resultados obtidos no presente estudo.

**Gráfico 3:** Distribuição dos microssatélites correspondendo aos motivos do tipo dinucleotídios encontrados no transcriptoma do cajueiro (*A. occidentale* L.).



Distribuição dos microssatélites do tipo dinucleotídios no cajueiro comum (**A**) e no cajueiro anão-precoce CCP 76 (**B**). **2G1** - Dinucleotídios pertencentes ao grupo 1 (GC/CG); **2G2** - Dinucleotídios pertencentes ao grupo 2 (CA/AC/GT/TG); **2G3** - Dinucleotídios pertencentes ao grupo 3 (CT/TC/GA/AG); **2G4** - Dinucleotídios pertencentes ao grupo 4 (AT/TA). **Fonte:** Autor.

**Gráfico 4:** Distribuição dos microssatélites com motivos do tipo trinucleotídios encontrados no transcriptoma do cajueiro (*A. occidentale* L.).



Distribuição dos microssatélites do tipo dinucleotídios no cajueiro comum **(A)** e no cajueiro anão-precoce CCP 76 **(B)**.

**3G1** - Trinucleotídios pertencentes ao grupo 1 (GGC/GCG/CGG/CCG/CGC/GCC);

**3G2** - Trinucleotídios pertencentes ao grupo 2 (ACG/CGA/GAC/TGC/GCT/CTG);

**3G3** - Trinucleotídios pertencentes ao grupo 3 (AGC/GCA/CAG/TCG/CGT/GTC);

**3G4** - Trinucleotídios pertencentes ao grupo 4 (ACC/CAC/CCA/TGG/GTG/GGT);

**3G5** - Trinucleotídios pertencentes ao grupo 5 (AGG/GGA/GAG/TCC/CCT/CTC);

**3G6** - Trinucleotídios pertencentes ao grupo 6 (AGT/GTATAG/ACT/CAT/ATC);

**3G7** - Trinucleotídios pertencentes ao grupo 7 (ATG/TGA/GAT/TAC/ACT/CTA);

**3G8** - Trinucleotídios pertencentes ao grupo 8 (AAG/AGA/GAA/CTT/TTC/TCT);

**3G9** - Trinucleotídios pertencentes ao grupo 9 (AAC/ACA/CAA/TTG/TGT/GTT);

**3G10** - Trinucleotídios pertencentes ao grupo 10 (AAT/ATA/TAA/ATT/TTA/TAT).

Fonte: Autor.

Os microssatélites possuem diversas características que os tornam marcadores moleculares de extrema relevância científica e tecnológica. Amiruddin e colaboradores (2012) relataram que SSRs representam uma rica fonte de marcadores moleculares devido à alta taxa de mutação causada ao longo da fita de DNA nessas regiões. Ainda, acredita-se que os microssatélites são importantes para evolução dos genomas devido a estímulos de variabilidade genética e influência na expressão de genes (LI et al., 2002; KALIA et al., 2011). Os microssatélites são importantes para o desenvolvimento de estratégias para conservação da espécie (KALIA et al., 2011). Assim, a identificação de SSRs para a espécie estudada representa importante fonte de dados para o desenvolvimento de estratégias de identificação, conservação e relações evolutivas entre genótipos.

### **2.5.3 Anotação funcional das sequências que contém microssatélites**

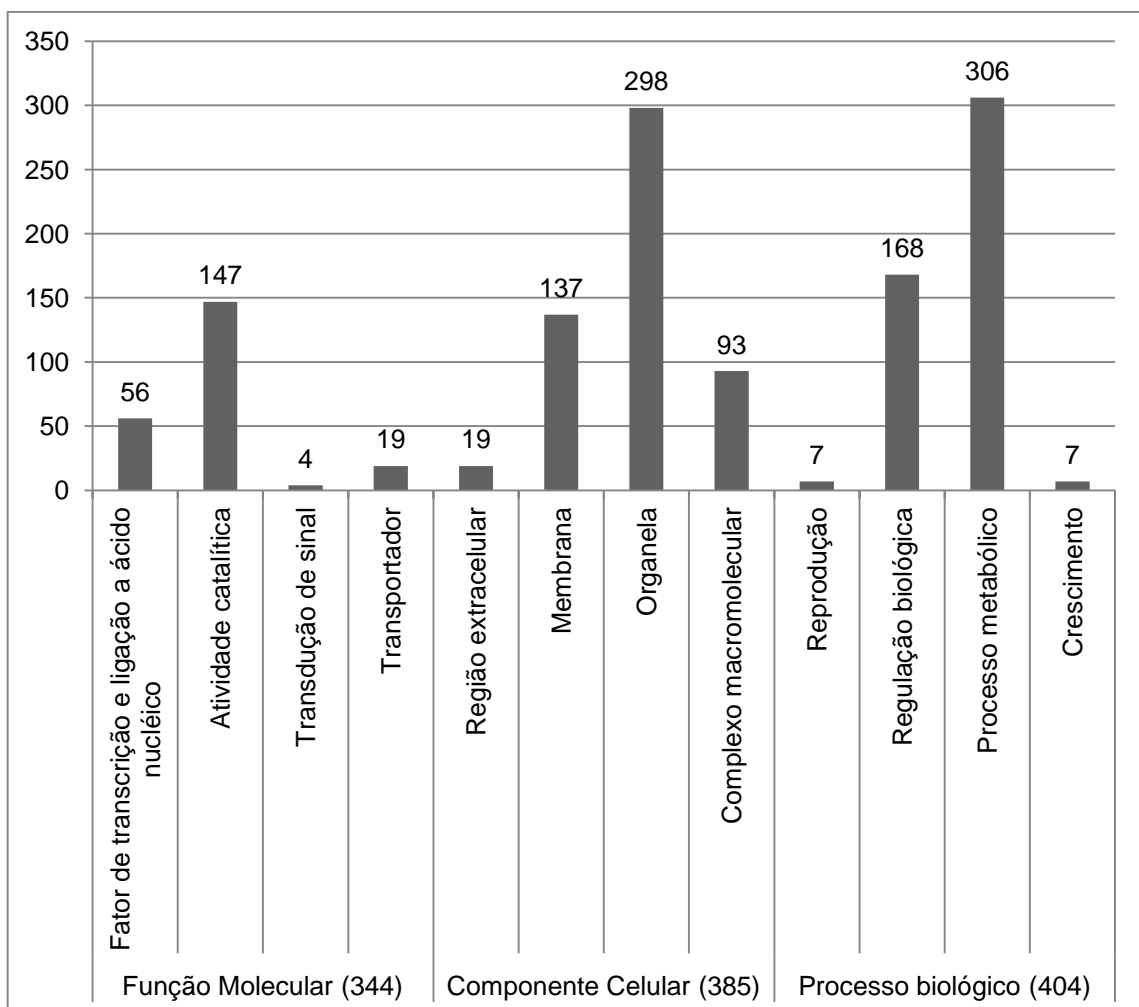
Para avaliar os genes que se localizam nas sequências onde foram encontrados os microssatélites, as sequências de cajueiro comum e CCP 76 com SSR detectado foram reunidos e utilizados para anotação funcional.

A identificação da função dos transcritos contendo microssatélites foi feita com base em similaridade de sequência utilizando a ferramenta BLAST contra o banco de dados não redundantes do *Swiss-Prot* disponível na base de dados do *Uniprot*. O banco de dados do *Swiss-Prot*, acessado em janeiro de 2015, contém 547.357 sequências cadastradas.

Um total de 930 (37,8%) transcritos de cajueiro mostraram similaridade com sequências do banco de dados do *Swiss-Prot*. Com base na identificação pelo BLAST, foi feita uma anotação funcional pelo *Gene Ontology* (GO). Foram reportados 477 termos de GO para as sequências do transcriptoma de castanha e folha de cajueiro. O gráfico 5 apresenta a distribuição dos transcritos contendo regiões microssatélites, os quais foram classificados em três categorias baseados nos termos de GO, a saber: processo biológico (404), componente celular (385) e função molecular (344).

Na categoria função molecular, as subcategorias mais representativas foram atividade catalítica (42,7%) e Fator de transcrição e ligação a ácido nucleico (16,3%). Em termos da classificação componente celular, a maior parte dos transcritos foi distribuído nas subclasses Organela (77,4%) e Membrana (137 35,6%). Enquanto que a categoria processo biológico é composta principalmente por proteínas envolvidas no processo metabólico (75,7%) e na regulação biológica (41,6%).

**Gráfico 5:** Classificação funcional das sequências contendo SSR, do transcriptoma de cajueiro, com base nos termos de *Gene Ontology* (GO) mais representativos.



**Fonte:** Autor.

Xie e colaboradores (2013), analisando o transcriptoma de *Pyropia haitanensis*, observaram que na categoria função molecular (FM) a subcategoria mais representativa foi atividade catalítica. Além disso, as subcategorias organela e processo metabólico, estão entre as cinco mais representativas das categorias componente celular (CC) e processo biológico (PB), respectivamente. Em estudos realizados com amendoim (*Arachis hypogaea* L.) e *Macrotyloma uniflorum*, observou-se que na categoria PB, a subclasse processo metabólico foi a mais representativa, com 30,42% e 18,86%, respectivamente. A subcategoria atividade catalítica foi a segunda mais representativa, para a categoria função molecular, para ambas as espécies, enquanto que na categoria CC, diferente dos resultados obtidos no presente trabalho, as subclasses mais representativas são célula e parte da célula para o amendoim (ZHANG et al., 2012) e para *Macrotyloma uniflorum* (BHARDWAJ et al., 2013), essa categoria não foi analisada.

A lista de números de acesso pertencente a proteínas identificadas pelo BLAST foi convertida em uma lista de número de acesso do banco de dados do KEGG *Orthology* (KO) utilizando ferramentas online do *Uniprot*.

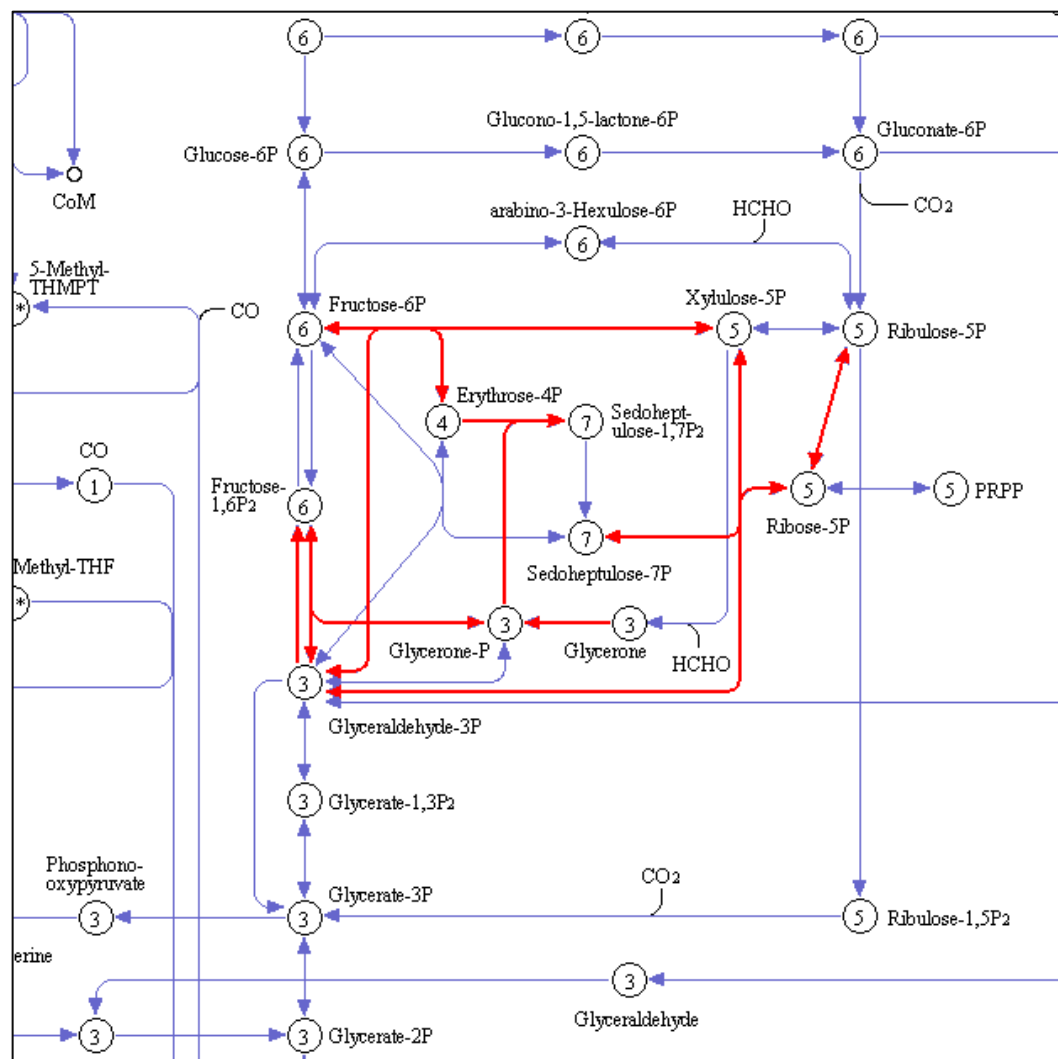
Para as sequências contendo SSR do transcriptoma do cajueiro, foram encontrados 153 termos de KO, podendo estar presente em 136 mapas metabólicos KEGG. Entre as vias identificadas, destacam-se a Biossíntese de metabólitos secundários (KO011110), onde foram encontrados 11 enzimas como a aldeído desidrogenase (K00128), transcetolase (K00615), mio-inositol-1 monofosfatase (K01610), aldolase (K01807) e cinamoil-CoA redutase (K09753).

Também foram encontrados 8 enzimas envolvidas na transdução de sinal de hormônios vegetais (KO04075) como proteína IAA responsiva a auxina (K14432), proteína contendo domínio de ligação ao jasmonato (K13464) e proteína fosfatase 2C (K14497). Seis enzimas envolvidas no metabolismo do carbono (Ko01200) também foram encontrados e estão mostrados na figura 4.

Muitos relatos têm demonstrado que uma parcela significativa dos microsatélites está localizada em regiões transcritas, incluindo genes codificadores de proteínas (MORGANTE et al., 2002; HAMMOCK; YOUNG, 2004; NICOT et al., 2004). Em um ensaio pioneiro, Marcott e colaboradores (1999) analisaram 70.882

proteínas presentes no banco de dados *Swiss-Prot* e estimaram que 14% do total continham seqüências repetidas, sendo tal abundancia três vezes maior em eucariotos que em procariotos. A presença desses marcadores em seqüências ESTs e em proteínas sugere que algumas dessas repetições podem estar envolvidas em processos metabólicos e de evolução gênica (LI et al., 2004).

**Figura 4:** Detalhe da via Metabolismo do Carbono do KEGG *Pathways* mostrando seqüências do transcriptoma de semente e folha de cajueiro onde foram encontrados SSR.



As setas vermelhas indicam as reações realizadas pelas enzimas relacionadas às seqüências contendo SSR que foram identificadas nesse trabalho. As setas azuis representam as enzimas que não foram identificadas no presente estudo. **Fonte:** Autor.

## 2.6 CONCLUSÕES

Ao analisar o transcriptoma de sementes e folhas do cajueiro, foram detectados mais de dois mil marcadores SSR, havendo compartilhamento de quase 300 SSRs no cajueiro comum e anão CCP 76. Entre os motivos de repetição avaliados, os do tipo trinucleotídeos foram os mais abundantes em ambos os tipos de cajueiro. Além disso, foi possível desenhar pares de *primers* para a amplificação das regiões que contém microssatélites e identificar prováveis funções e vias metabólicas relacionadas aos transcritos contendo microssatélites.



## **CAPÍTULO II**

### **VALIDAÇÃO DOS MARCADORES MICROSSATÉLITES EM DIFERENTES GENÓTIPOS DE CAJUEIRO ANÃO-PRECOCE (*A. occidentale* L. var *nanum*)**

### **3 VALIDAÇÃO DOS MARCADORES MICROSSATÉLITES EM DIFERENTES GENÓTIPOS DE CAJUEIRO ANÃO-PRECOCE (*A. occidentale* L. var. *nanum*)**

#### **3.1 INTRODUÇÃO**

Uma das mais importantes contribuições do melhoramento genético de plantas tem sido o desenvolvimento de variedades para diferentes ambientes. O uso de clones resistentes representa uma forma de manejo econômico, ecológico e seguro, impedindo a invasão de pragas e doenças, além de proporcionar uma melhor utilização da variabilidade genética da espécie (PAIVA; BARROS, 2004).

O melhoramento genético do cajueiro-anão precoce no Brasil, iniciou-se em 1956, com a introdução de plantas no Campo Experimental de Pacajus. Os primeiros clones comerciais foram lançados nos anos 80, a saber: CCP 06 e CCP 76, em 1983, e CCP 09 e CCP 1001, em 1987. Eles foram obtidos a partir das plantas matriz CP 06, CP 76, CP 09 e CP 1001, respectivamente. Esses clones foram lançados visando uma melhoria da qualidade de castanhas e ainda são considerados os principais clones comerciais disponíveis (BARROS et al., 1984; BARROS, 1988; ALMEIDA; ARAÚJO; LOPES, 1993).

Posteriormente, foram implantados outros clones, em 1996 pôde-se obter os clones Embrapa 50, Embrapa 51 (BARROS et al., 2000). O primeiro teve origem do cruzamento entre o cajueiro anão-precoce CCP 06 e a planta matriz CP 07, enquanto que o outro surgiu a partir da planta matriz P 500E (PAIVA; BARROS, 2004). O clone BRS 189 originou-se do cruzamento entre os clones de cajueiro-anão precoce CCP 1001 e CCP 76 e foi lançado para o plantio em 2000 (PAIVA; BARROS, 2004). Devido ao alto teor de vitamina C e baixa quantidade de taninos, o pedúnculo desse clone tem sido aproveitado para o mercado de mesa (BARROS et al., 2002). Em 2002, o clone BRS 226 foi lançado para o plantio comercial, esse clone foi obtido da planta matriz de cajueiro-anão número 42 (MAP-42). Por ter um dos maiores pesos de castanha é bastante recomendado para o mercado da amêndoa (PAIVA et al., 2002). Os indicadores agroindustriais para os cajueiros anão-precoce estão listados na Tabela 5.

Lamas (2010), utilizou 14 marcadores SSR par avaliar as relações de vínculo entre 36 acessos de *A. occidentale*, por meio de uma análise de similaridade genética, agrupando-os na forma de um dendograma. Através dessa análise foi possível distinguir os diferentes acessos de cajueiro utilizando os marcadores microssatélites empregados no estudo. A figura 5 mostra o dendograma obtido por Lamas, com algumas modificações, sendo selecionado somente os genótipos utilizados no presente estudo.

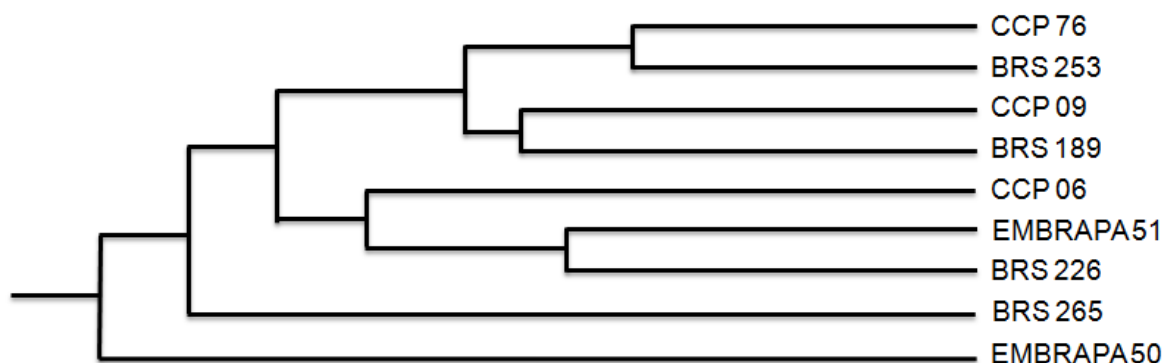
A utilização de marcadores microssatélites é eficiente para caracterizar e discriminar diferentes genótipos de cajueiro e são potencialmente capazes para caracterizar e diferenciar qualquer população de cajueiro (PINHEIRO et al., 2008).

**Tabela 5:** Características morfológicas e agroindustriais de alguns genótipos de cajueiro anão-precoce.

Genótipo	Altura média	Diâmetro médio da copa	Peso médio da amêndoa	Peso médio do pedúnculo
CCP 06	2,11 m	4,52 m	1,6 g	76,5 g
CCP 09	2,15 m	4,65 m	2,1 g	87 g
CCP 76	2,68 m	4,98 m	1,80 g	135 g
Embrapa 50	3,41 m	7,67 m	2,9 g	111 g
Embrapa 51	3,52 m	7,79 m	2,6 g	104 g
BRS 189	3,16 m	5,9 m	2,1 g	155,4 g
BRS 226	1,24 m	2,20 m	2,72 g	102,6

Fonte: (PAIVA; BARROS, 2004), adaptado pelo autor.

**Figura 5:** Dendograma obtido por Lamas (2010), baseado nas distâncias genéticas entre cada acesso de *A. occidentale* L.



## **3.2 OBJETIVOS**

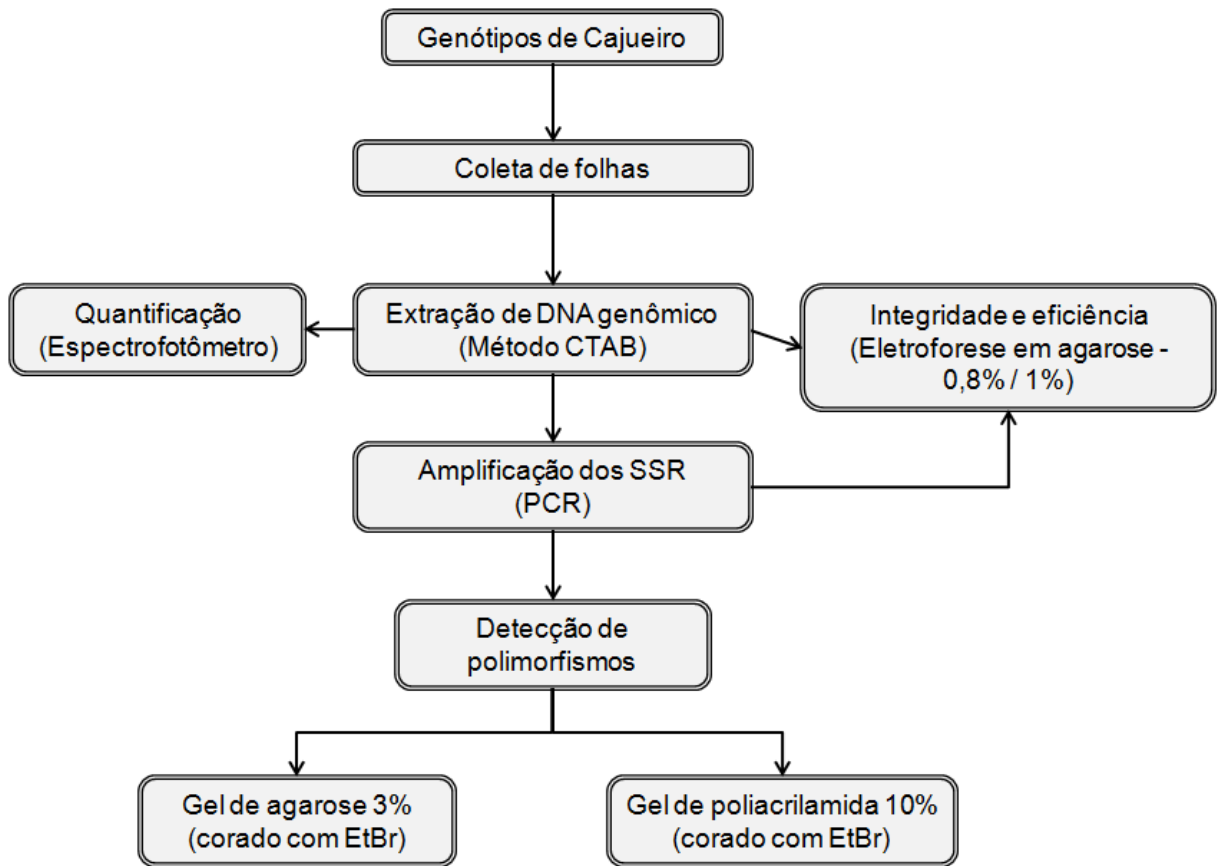
### **3.2.1 Objetivo Geral**

Validar e caracterizar marcadores moleculares em genótipos de cajueiro utilizando microssatélites presentes em regiões transcritas.

### **3.2.2 Objetivos Específicos**

- Amplificar regiões genômicas genótipos de cajueiro que contém microssatélites;
- Analisar a presença de alelos polimórficos nas regiões amplificadas dos nove genótipos de cajueiro anão-precoce estudados.
- Caracterizar a distribuição das frequências alélicas dos microssatélites desenvolvidos.
- Identificar sequências codificantes nos transcritos que contém microssatélite.

### 3.3 ESTRATÉGIA EXPERIMENTAL



### **3.4 METODOLOGIA**

#### **3.4.1 Coleta do Material Biológico**

Folhas de cajueiro anão-precoce (CCP 76, Embrapa 50, BRS 189, CCP 09, Embrapa 51, CCP 06, BRS 265, BRS 253, BRS 226) foram adquiridas na Embrapa Agroindústria Tropical (Fazenda Experimental de Pacajús, Ceará). As amostras ficaram armazenadas em freezer -20 °C até a etapa de extração de DNA.

#### **3.4.2 Extração e Purificação de DNA genômico**

Para a validação dos marcadores SSR foi extraído DNA genômico de folhas dos diferentes genótipos de cajueiro em estudo, utilizando o protocolo de Doyle e Doyle (1987), com adaptações. Aproximadamente 0,6 g de limbo de folhas, de cada genótipo, foram pesados e macerados com nitrogênio líquido. O pó das folhas foi colocado em tubos de centrifugação de 15 mL, contendo 6 mL de tampão de extração CTAB (Tris-HCl 110 mM pH8,0; NaCl 1,4 M; EDTA 20 mM; CTAB 2%) acrescido de 12 µL de β-mercaptoetanol 0,2%, previamente aquecido, em banho-maria, a 65 °C. Em seguida, as amostras foram incubadas a 60 °C por 16 h, em banho-maria. Após o período de incubação, foi adicionado 1 volume (6 mL) de clorofórmio:álcool isoamílico (24:1 v/v). As amostras ficaram em repouso, a temperatura ambiente, por 10 min e, em seguida, centrifugadas a 5.000 rpm por 10 min, a temperatura ambiente. O sobrenadante foi transferido para um novo tubo e repetiu-se a etapa de adição de clorofórmio:álcool isoamílico (24:1 v/v). Os ácidos nucleicos foram precipitados pela adição de 2/3 do volume de isopropanol 100% gelado, as amostras ficaram incubadas a 4 °C por 2 h e posteriormente centrifugadas (5.000 rpm/5 min/ temperatura ambiente). O sobrenadante foi descartado e o precipitado solubilizado em 2 mL de NaCl 1 M. As amostras foram novamente precipitadas pela adição de 5 mL de etanol 100% gelado e incubadas a 4 °C por 16 h. Após esse período, foram centrifugadas (5.000 rpm/5 min/ temperatura ambiente) e o sobrenadante descartado. O precipitado foi lavado com etanol 70%,

seguido por centrifugação a 2.500 rpm por 5 min. O sobrenadante foi descartado e o precipitado seco em capela por aproximadamente 5 min, para posterior solubilização em 500 µL de tampão TE (Tris-HCl 10 mM pH 8,0; EDTA 1 mM).

### 3.4.3 Quantificação e Eletroforese do DNA

A concentração do DNA dos nove genótipos de cajueiro anão-precoce foi estimada a partir dos valores de absorvância a 260 nm (A<sub>260</sub>) com a utilização de um espectrofotômetro *GenQuant pro* (GE Healthcare). A razão A<sub>260</sub>/280 foi usada como parâmetro para se estimar a pureza de cada preparação (SAMBROOK; FRITSCH; MANIATIS, 1989).

A eletroforese do DNA foi realizada em gel de agarose 0,8%, segundo protocolo descrito por Sambrook, Fritsch e Maniatis (1989). Para o preparo do gel foi utilizado 2 g de agarose e 250 mL de tampão TBE 0,5X (Tris 45 mM; Ácido Bórico 45 mM; EDTA 1 mM; pH 8,0). Após a completa dissolução da agarose, a mistura foi resfriada até uma temperatura aproximada de 60 °C e adicionado brometo de etídio 0,25 mg/mL. O gel permaneceu em repouso à temperatura ambiente até solidificar. As amostras a serem submetidas a corrida eletroforética foram previamente misturadas com tampão de amostra (azul de bromofenol 0,25% (m/v), glicerol 30% (v/v), preparada em tampão TE). Em seguida, o gel foi submerso em tampão TBE 0,5X, as amostras foram aplicadas e a corrida eletroforética seguiu-se em duas etapas: Primeiro a 70 V, 200 mA e 10 min, e em seguida, a 90 V, 200 mA e 40 min. O gel foi revelado em um sistema de fotodocumentação L-Pix contendo transluminador e uma câmera Canon A620 ligado diretamente a um computador onde as fotodocumentações ficam armazenadas.

### 3.4.4 Amplificação dos *loci* microssatélites

Para a amplificação dos *loci* microssatélites, realizou-se a reação em cadeia da polimerase (PCR). O volume total da reação foi de 25  $\mu$ L, sendo utilizado 300 ng de DNA genômico, 0,4  $\mu$ M de *primers* (F/R), 0,2 mM de dNTPs, 1,75 unidade de Taq DNA polimerase, tampão 10X (contendo  $MgCl_2$ ) e o volume foi completado com água estéril tipo 1. Foi realizada uma reação adicional contendo todos os reagentes, com exceção de DNA genômico (controle negativo). Os pares de *primers* desenhados, a partir das sequências compartilhadas entre os cajueiros comum e anão CCP 76 (ver tópico "Detecção *in silico* de marcadores microssatélites e desenho de *primer* para PCR"), foram testados em cada genótipo de cajueiro anão-precoce descrito anteriormente.

As condições térmicas utilizadas foram as seguintes: desnaturação do DNA a 95 °C durante 5 minutos, seguido de 40 ciclos de amplificação (95 °C durante 30 segundos, 50 °C durante 30 segundos e 72 °C durante 30 segundos), seguido por extensão final a 72 °C durante 5 minutos (Tabela 6). As reações de amplificação foram feitas em termociclador *ThermalCyclers (Amplitherm)*. Após a reação de PCR, os amplicons foram analisados em gel de agarose 1% para verificar a eficiência da amplificação.

**Tabela 6:** Condições térmicas da amplificação dos *loci* contendo SSR de amostras de DNA extraídas de folhas de diversos genótipos de cajueiro anão-precoce.

Etapas		Temperatura	Tempo
Desnaturação inicial		95 °C	5 min
Desnaturação	40X	95 °C	30 seg
Anelamento		50 °C	30 seg
Extensão		72 °C	30 seg
Extensão final		72 °C	5 min

Fonte: Autor



### 3.4.5 Resolução e análise dos fragmentos amplificados

Os produtos amplificados foram analisados em gel de agarose 3% e em gel de poliacrilamida 10% para a detecção dos possíveis *loci* polimórficos. A eletroforese em gel de agarose 3% foi realizada seguindo o protocolo descrito por Sambrook, Fritsch e Maniatis (1989), como descrito no tópico "Quantificação e Eletroforese do DNA".

Para a preparação do gel de poliacrilamida 10%, empregou-se a metodologia de Rickwood e Hames (1990), com adaptações. O método consiste no preparo de um gel de separação contendo poliacrilamida 10% e um gel de empilhamento (concentração) contendo poliacrilamida 5% (Tabela 7). O sistema de eletroforese utilizado foi o Ruby™ da GE Healthcare com placas de 15 cm de comprimento e 1,5 mm de espessura. O preparo do gel de poliacrilamida foi feito conforme descrito na tabela 8 com volume ajustado para 50 mL para o gel de separação e 15 mL para o gel de empilhamento. Foi adicionado 2 µL de tampão de amostra (Glicerol 20%, TE e Azul de Bromofenol 0,02%) em 6 µL do produto de PCR, e em seguida, aplicados no gel. O tampão de corrida utilizado era composto por Tris/HCl 0,3% pH8,8, Glicina 1,44% e SDS 0,1% e a eletroforese foi conduzida com diferença de potencial de 100 V por aproximadamente 2 h.

**Tabela 7:** Composição do gel de poliacrilamida 10% para DNA.

Constituintes	Gel de Separação	Gel de Empilhamento
Acrilamida/Bisacrilamida	10%	5%
Tris/HCl	0,4 M (pH 8,8)	0,1 M (pH 6,8)
SDS	0,1 M	0,1 M
PSA	0,13%	0,1%
TEMED	0,06%	0,1%

Para a revelação dos fragmentos amplificados foi utilizado uma solução de brometo de etídio 0,5 µg/mL, por 1 h. Para digitalização, os géis foram levados para um sistema de fotodocumentação em gel L-Pix™ contendo transluminador e uma câmera Canon™ A620 ligado diretamente a um computador onde as imagens ficam armazenadas.

O grau de polimorfismo de cada marcador microssatélite foi avaliado de acordo com a presença de um determinado *locus* SSR e os alelos de todos os *loci* microssatélites foram identificados por ordem crescente de tamanho, definida pela migração eletroforética.

#### **3.4.6 Identificação *in silico* das sequências amplificadas**

Os transcritos utilizados para desenho de *primer* também foram identificados com base em homologia de sequência utilizando a ferramenta BLAST. Foram utilizados nas identificações basicamente três bancos de dados, *Conserved Domain Database* (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), *Swiss-prot database* ([www.uniprot.org](http://www.uniprot.org)) e o *GenBank* (<http://www.ncbi.nlm.nih.gov/refseq/>). Foram consideradas válidas as identificações com melhores valores de *E-value* (probabilidade de alinhamento ser feito ao acaso) no CDD (Conserved Domain Database) marcadas como "*trusted*". As sequências não identificadas pelo CDD foram analisadas pelo BLAST feito em terminal por linha de comando restringindo as identificações para o valor-E inferiores a  $1 \times 10^{-5}$ . Sequências não identificadas no *Swiss-prot* foram analisadas pelo BLAST usando o banco de dados do *GenBank* (NCBI) considerando alinhamentos com identidade superior a 75%.

### 3.5 RESULTADOS E DISCUSSÕES

#### 3.5.1 Isolamento do DNA genômico

A análise do DNA genômico dos diferentes genótipos de cajueiro anão-precoce, mostrou concentrações que variaram de 150 ng/μL (CCP 76) a 793 ng/μL (Embrapa 51) e relação  $A_{260}/A_{280}$  entre 1,35 (CCP 09) e 2,33 (CCP 76) (Tabela 8). Segundo Sambrook, Fritsch e Maniatis (1989), a relação entre as absorvâncias a 260 nm e 280 nm deve ser maior que 1,75 para que uma amostra seja considerada de boa qualidade e livre de contaminações por proteínas. Contudo, outros pesquisadores apontam que amostras com relação entre 1,6 e 2,0 possuem qualidade satisfatória, e valores acima de 2,0 podem indicar contaminação por fenol ou outros alcoóis (ROMANO, 1998).

A eletroforese em gel de agarose 0,8% revelou bandas uniformes com tamanho maior de 23.000 pb (Dados não mostrados). Os fragmentos de DNA obtidos não apresentaram degradação aparente, possuindo tamanho e qualidade necessários para a realização de estudos posteriores.

**Tabela 8:** Quantificação de DNA genômico de diferentes genótipos de cajueiro anão-precoce, com valor de absorvância a 260 nm.

Amostra	Concentração do DNA (ng/ $\mu$ L)	Relação ( $A_{260}/A_{280}$ )
1 - CCP 76	150	2,33
2 - Embrapa 50	157	1,93
3 - BRS 189	215	2,00
4 - CCP 09	463	1,35
5 - Embrapa 51	793	2,15
6 - CCP 06	424	2,10
7 - BRS 265	349	2,21
8 - BRS 253	373	1,88
9 - BRS 226	368	2,12

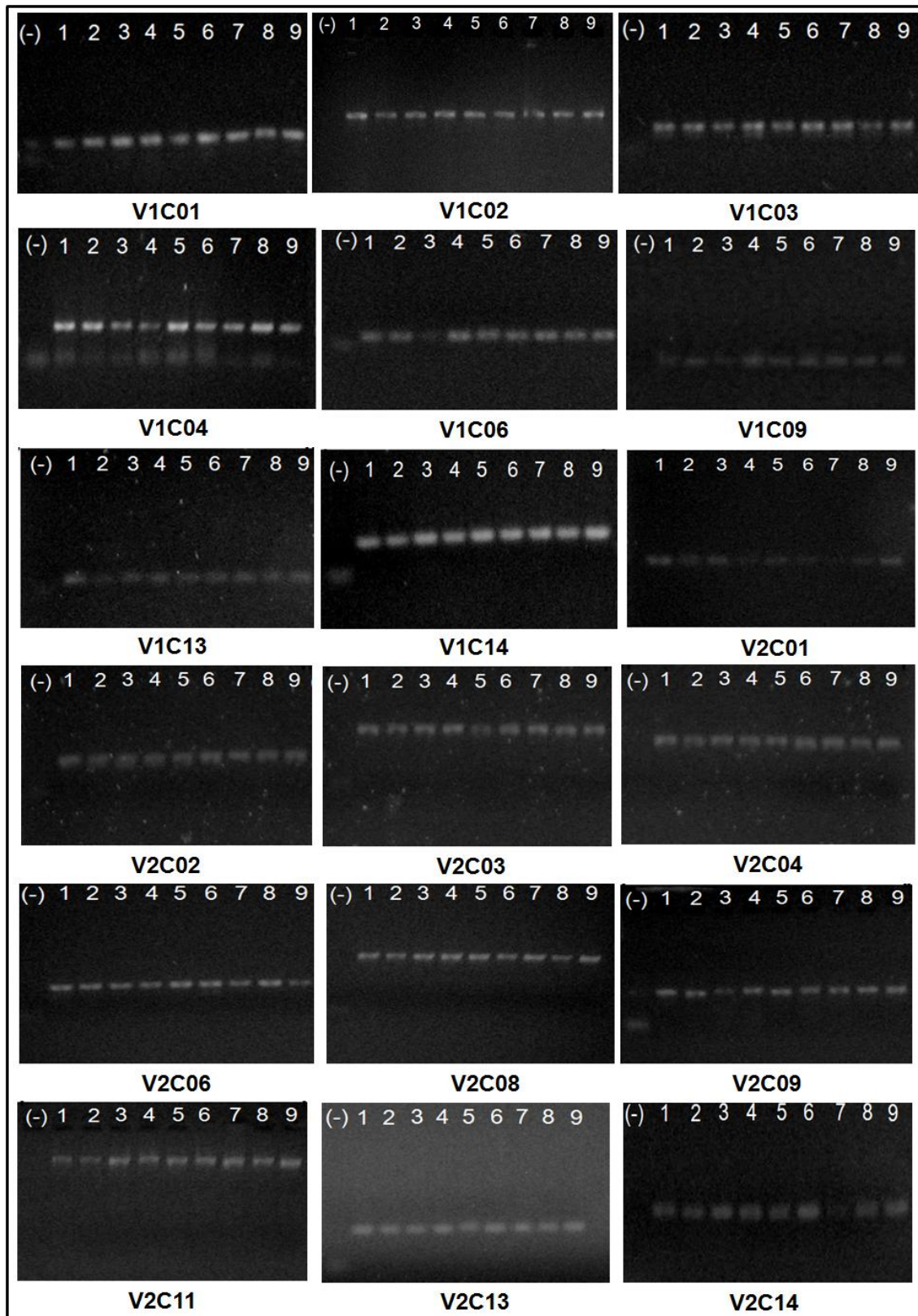
A concentração do DNA foi dada em nanograma por microlitro (ng/ $\mu$ L). A relação entre as absorvâncias ( $A_{260}/A_{280}$ ) devem variar de 1,5 a 2,0. **Fonte:** Autor.

### 3.5.2 Amplificação das regiões de DNA que contém microssatélites

Os produtos de PCR amplificados foram analisados em gel de agarose 1% para verificar a eficiência da amplificação. Dentre os 29 pares de *primers*, 19 (66 %) amplificaram com sucesso 100% das amostras (Figura 6). Enquanto que 9 (31%) pares de *primers* observou-se a ocorrência de alelos nulos, isto é, não houve amplificação em todas as amostras (Figura 7) e apenas 1 (3%) par de *primer* não amplificou nenhuma das amostras (Figura 8). Dados obtidos da análise dos genótipos de *Coffea canephora* mostraram que dos 21 marcadores utilizados, 14 (66,7%) apresentaram bandas bem definidas e sete (33,3%) não apresentaram produtos de amplificação (MOTTA et al., 2011).

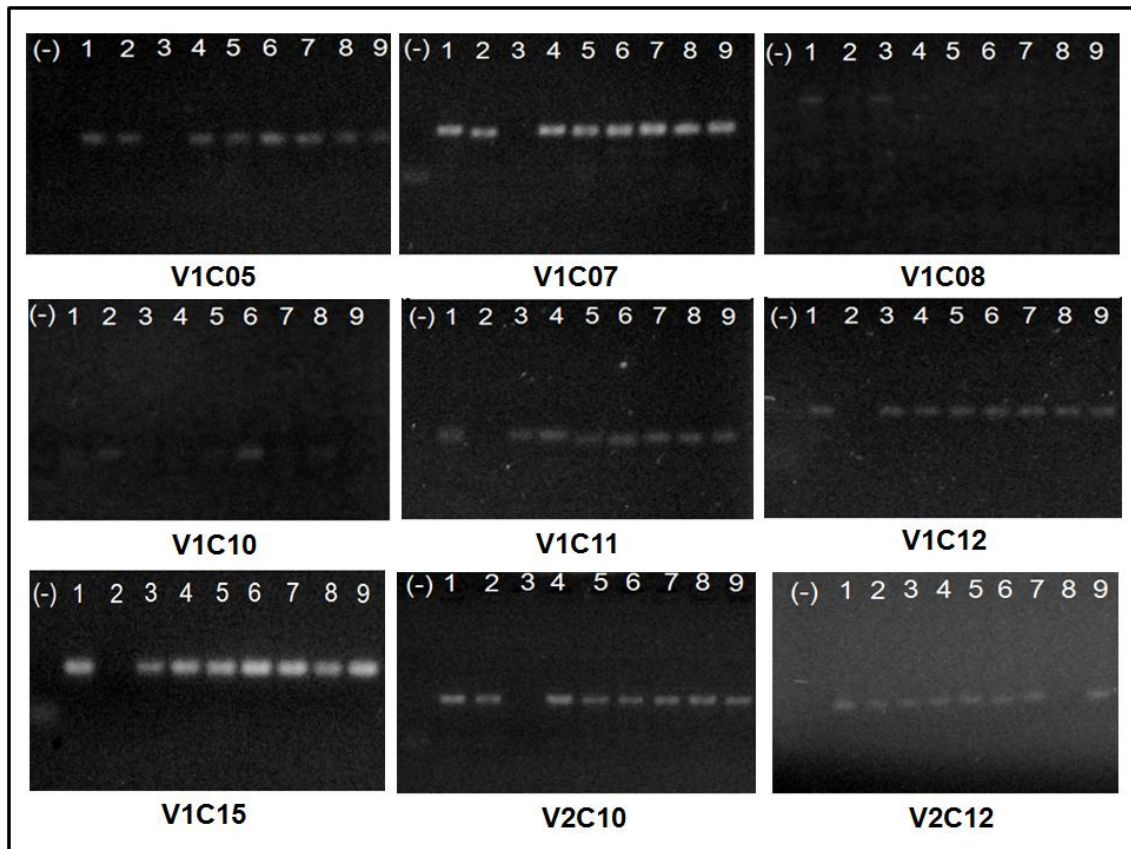
Santos (2010), ao trabalhar na identificação e caracterização de marcadores microssatélites de aveia, revelou que 100% dos pares de *primers* utilizados amplificaram, contudo 30% destes, apresentaram falhas na amplificação de alguns indivíduos. A ausência de amplificação em alguns genótipos pode demonstrar diferenças nas sequências que flanqueiam os microssatélites, produzindo alelos nulos, os quais podem ser detectados após várias repetições do experimento, para garantir que não ocorreram falhas na reação (SANTOS, 2010).

**Figura 6:** Géis de eletroforese de agarose 1%, mostrando os *primers* que amplificaram 100% das amostras.



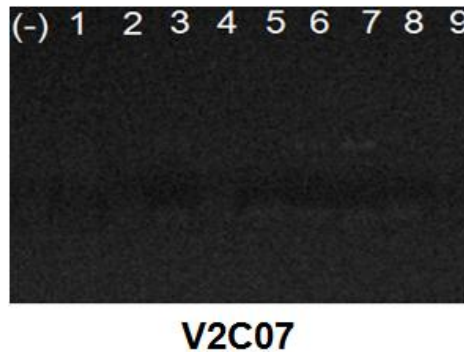
(-) Negativo. Raias de 1 - 9: Produtos de PCR referentes as amostras 1: CCP 76; 2: Embrapa 50; 3: BRS 189; 4: CCP 09; 5: Embrapa 51; 6: CCP 06; 7: BRS 265; 8: BRS 253; 9: BRS 226. Fonte: Autor.

**Figura 7:** Géis de eletroforese de agarose 1%, mostrando os *primers* que não amplificaram todos as amostras.



(-) Negativo. **Raias de 1 - 9:** Produtos de PCR referentes as amostras **1:** CCP 76; **2:** Embrapa 50; **3:** BRS 189; **4:** CCP 09; **5:** Embrapa 51; **6:** CCP 06; **7:** BRS 265; **8:** BRS 253; **9:** BRS 226. Fonte: Autor.

**Figura 8:** Eletroforese em gel de agarose a 1% dos marcadores SSR mostrando o *primer* que não amplificou nenhuma das amostras.



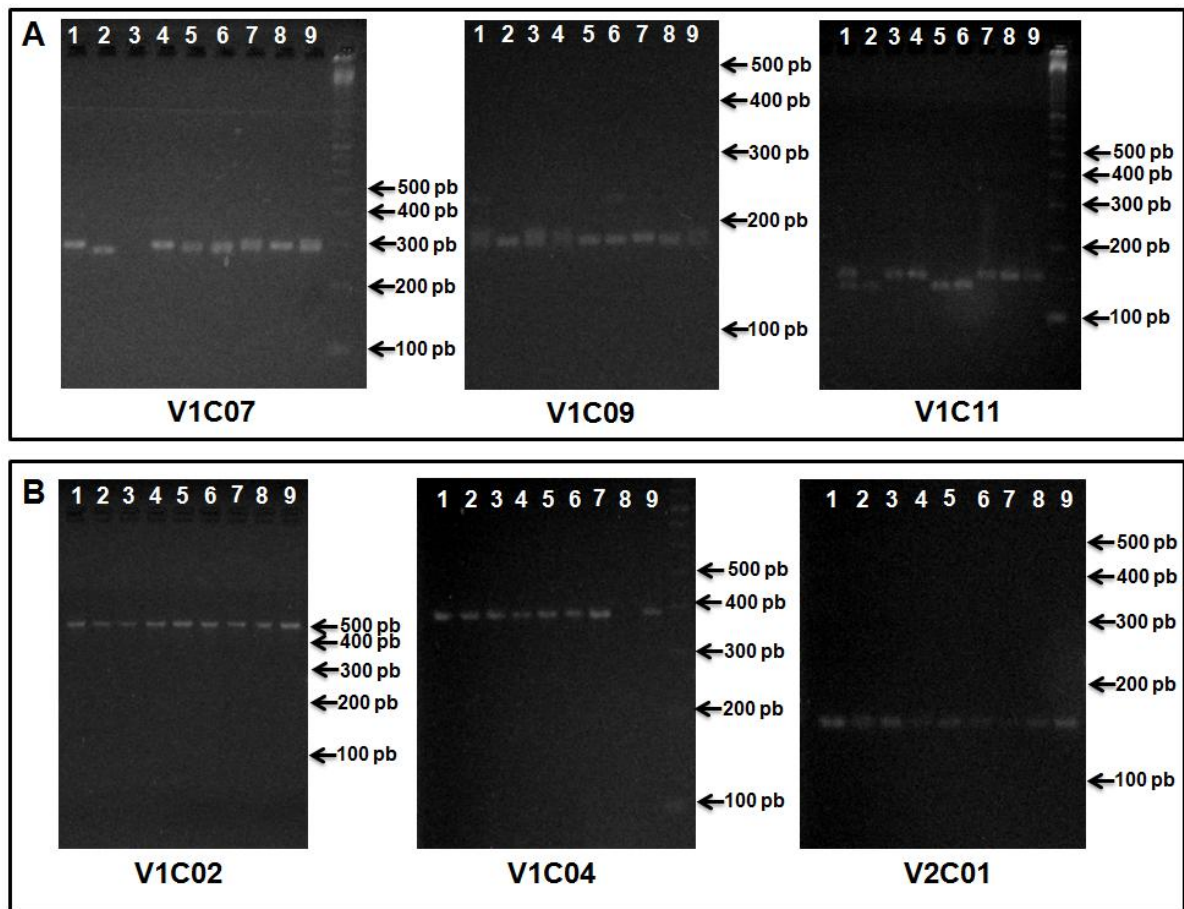
Os produtos de PCR mostrados acima foram amplificados pelo *primer* V2C07. (-) Negativo. **Raias de 1 - 9:** Produtos de PCR referentes as amostras **1:** CCP 76; **2:** Embrapa 50; **3:** BRS 189; **4:** CCP 09; **5:** Embrapa 51; **6:** CCP 06; **7:** BRS 265; **8:** BRS 253; **9:** BRS 226. Fonte: Autor.

### 3.5.3 Identificação de microssatélites polimórficos

Para a detecção dos possíveis *loci* polimórficos entre os genótipos de cajueiro anão-precoce, os produtos de PCR com amplificação confirmada em gel a 1% (total de 28 amplificações positivas) foram analisados em géis de agarose 3% e poliacrilamida 10%.

Na primeira etapa da triagem, utilizando gel de agarose 3%, todos os pares de *primers* que amplificaram pelo menos uma das amostras, foram testados nos nove indivíduos em estudo. Dentro deste grupo, um total de 8 *loci* (28,6%) apresentaram possíveis polimorfismos (Figura 9A), enquanto que nos 20 (71,4%) marcadores restantes não foi possível detectar polimorfismo em gel de agarose a 3% (Figura 9B). Entretanto, neste nível de resolução é difícil declarar com certeza a inexistência de polimorfismo, o que demanda uma segunda análise para uma avaliação mais precisa do potencial de cada par de iniciadores. Além disso, a resolução de fragmentos em géis de agarose 3% não permite distinguir heterozigotos com precisão ou alelos próximos em tamanho (SANSALONI, 2008).

**Figura 9:** Resultado dos produtos de PCR amplificados, mostrados em géis de agarose 3%.



**A** - Géis de agarose 3% mostrando produtos amplificados por 3 marcadores (V1C07, V1C09 e V1C11) que apresentaram possíveis *loci* polimórficos. **B** - Géis de agarose 3% mostrando produtos amplificados por 3 marcadores (V1C02, V1C04 e V2C01) que não foi possível detectar polimorfismo. **Raias de 1 - 9:** Produtos de PCR referentes as amostras 1: CCP 76; 2: Embrapa 50; 3: BRS 189; 4: CCP 09; 5: Embrapa 51; 6: CCP 06; 7: BRS 265; 8: BRS 253; 9: BRS 226. **Fonte:** Autor.



Para averiguar a efetiva ocorrência de polimorfismo, todos os marcadores foram novamente analisados na segunda etapa de triagem em géis de poliacrilamida 10%. Dos 28 pares de iniciadores avaliados, pôde-se observar com precisão que 9 (32%) amplificaram *loci* polimórficos e 12 (43%) eram monomórficos (Tabela 9). Além disso, foi observado também a presença de 7 (25%) marcadores que produziram bandas de difícil interpretação, sendo necessário mais análises para afirmar o grau de polimorfismo entre as amostras estudadas (Tabela 9). A Figura 10 mostra três marcadores com alelos polimórficos (V1C07, V1C11 e V2C09) e três com alelos monomórficos (V2C04, V2C12 e V2C05).

O número médio de alelos obtidos por *locus* foi 3,5, variando de 1 a 6 alelos. O *locus* V1C03 apresentou a maior diversidade alélica, com 6 alelos. A menor diversidade alélica foi evidenciada em 10 marcadores estudados, a saber: V1C06, V1C10, V1C12, V1C13, V1C14, V1C15, V2C04, V2C08, V2C12 e V2C13, os quais apresentaram somente 1 alelo (Tabela 9). Resultados semelhantes foram obtidos por Chen e colaboradores (2014), que ao desenvolverem 90 marcadores SSR para 131 acessos de feijão comum (*Phaseolus vulgaris* L.), obtiveram uma média de 3,56 alelos por *locus*, variando de dois a sete alelos. Segundo os autores, o número médio de alelos observado é menor do que as estimativas anteriores, considerando o número de genótipos avaliados.

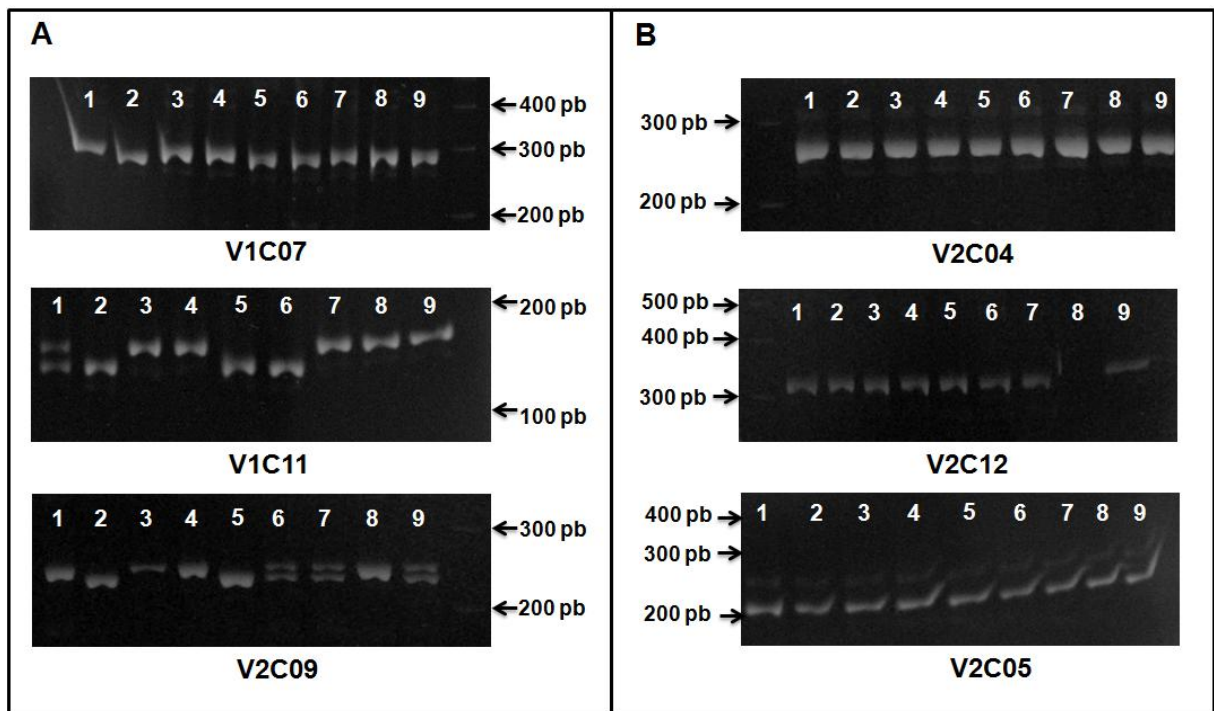
De acordo com Paiva e Barros (2004), o genótipo Embrapa 50 descende da matriz CP 06, a mesma que também deu origem ao genótipo CCP 06. As análises do padrão de bandas nos géis mostraram semelhança para ambos os genótipos em todos os SSR amplificados, exceto para os marcadores V1C03 e V2C14. O motivo para esta diferença no padrão de bandas pode ser devido a fatores como mutações nos gametas da planta matriz ou distorções na migração eletroforética.

**Tabela 9:** Características de 28 marcadores microssatélites em *A. occidentale* var. *nanum*, mostrando o grau de polimorfismo e o número de alelos de cada *locus*.

<b>Marcador SSR</b>	<b>Grau de Polimorfismo</b>	<b>Nº de alelos</b>
V1C01	Polimórfico	4
V1C03	Polimórfico	6
V1C07	Polimórfico	2
V1C09	Polimórfico	2
V1C11	Polimórfico	2
V2C02	Polimórfico	2
V2C09	Polimórfico	2
V2C10	Polimórfico	2
V2C14	Polimórfico	2
V1C06	Monomórfico	1
V1C10	Monomórfico	1
V1C12	Monomórfico	1
V1C13	Monomórfico	1
V1C14	Monomórfico	1
V1C15	Monomórfico	1
V2C04	Monomórfico	1
V2C05	Monomórfico	2
V2C06	Monomórfico	2
V2C08	Monomórfico	1
V2C12	Monomórfico	1
V2C13	Monomórfico	1
V1C02	Ininterpretável	-
V1C04	Ininterpretável	-
V1C05	Ininterpretável	-
V1C08	Ininterpretável	-
V2C01	Ininterpretável	-
V2C03	Ininterpretável	-
V2C11	Ininterpretável	-

**Fonte:** Autor.

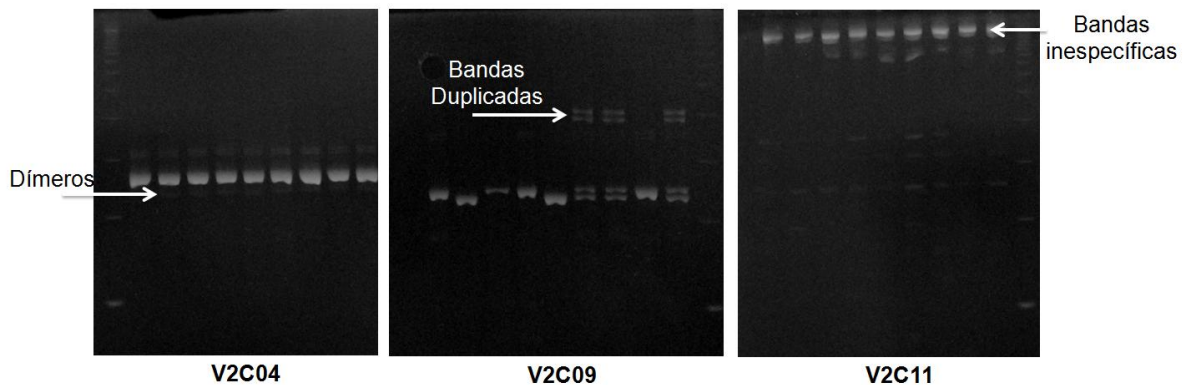
**Figura 10:** Eletroforese em gel de poliacrilamida 10% corado com brometo de etídio, mostrando pares de iniciadores que amplificaram *loci* polimórficos e monomórficos.



**A** - Pares de *primers* (V1C07, V1C11 e V2C09) que amplificaram *loci* polimórficos; **B** - Pares de *primers* (V2C04, V2C12 e V2C05) que amplificaram *loci* monomórficos. **Raias de 1 - 9:** Produtos de PCR referentes as amostras **1:** CCP 76; **2:** Embrapa 50; **3:** BRS 189; **4:** CCP 09; **5:** Embrapa 51; **6:** CCP 06; **7:** BRS 265; **8:** BRS 253; **9:** BRS 226. **Fonte:** Autor.

O gel de poliacrilamida tem melhor resolução, quando comparado à géis de agarose, contudo, a presença de bandas duplicadas, de dímeros e a amplificação de produtos de forma inespecífica, foram detectados nos géis de poliacrilamida (Figura 11). Em *Petunia*, 100% dos *loci* foram amplificados com sucesso, sendo que 7,7% destes apresentam um padrão de bandas ininterpretável. Em *Passiflora*, 58,3% dos *loci* amplificaram com sucesso, sendo que 33,3% destes apresentam bandas múltiplas (KRIEDT, 2009).

**Figura 11:** Gel de poliacrilamida 10% mostrando resultado da PCR com problemas para alguns pares de *primers*.



Os *primers* V2C04, V2C09, V2C11, apresentaram alguns problemas durante a eletroforese em gel de poliacrilamida 10%, como a presença de dímeros, bandas duplicadas e inespecíficas. **Fonte:** Autor.

Wang e colaboradores (2009), ao analisarem 12 cultivares de *Cornus florida* L., usando dois conjuntos de *primers* e cinco métodos de eletroforese, observaram que o gel de agarose tem resolução inferior ao gel de poliacrilamida, no entanto, devido a maior sensibilidade da coloração com prata foi detectado alguns ruídos de fundo e a presença de bandas não específicas. Em estudos de caracterização e mapeamento de microssatélites em *Eucalyptus spp.*, foi possível observar tanto em géis de agarose 2,5%, quanto em poliacrilamida a presença de *loci* duplicados e produtos amplificados de forma inespecífica (SANSALONI, 2008).

Analisando marcadores SSR de jaboticabeira (*Plinia sp.*), Martins (2013), verificou que um dos pares de iniciadores mostrou problemas na amplificação, pois houve excesso na amplificação de bandas inespecíficas e a presença de dímeros. Segundo o autor, tal problema foi minimizado com a redução da quantidade de  $MgCl_2$  para eliminar a amplificação das bandas inespecíficas e reduzindo a quantidade de cada *primer*, impedia o aparecimento de dímeros na amplificação das amostras.

A detecção de produtos secundários e não específicos aumentam a dificuldade de identificação dos alelos com precisão. Além disso, é difícil calcular com exatidão o tamanho dos alelos pelos métodos tradicionais de eletroforese em gel (agarose e poliacrilamida), devido às diferenças entre as bandas durante a migração (WANG et al., 2009).

As 29 sequências contendo SSR, utilizadas para o desenho de *primer*, foram identificadas com base em similaridade de sequência utilizando a ferramenta BLAST. Destas, 15 foram identificadas utilizando o banco de dados de domínios conservados (CDD), 9 usando o *Swiss-Prot* e 5 utilizando o *GenBank* (Tabela 10).

Dos 9 marcadores SSR onde foram detectados polimorfismos nos genótipos de cajueiro, 5 (55,6%) fazem parte de famílias gênicas de acordo com o banco de dados do CDD, uma codifica fator de transcrição responsivo ao etileno, identificada pelo *GenBank* e três foram identificadas pelo *Swiss-Prot*.

O termo famílias gênicas, ou famílias multigênicas, é usado para incluir um conjunto de genes de um mesmo organismo que codificam proteínas com notável similaridade estrutural, quanto ao número e organização dos pares de bases nitrogenadas, embora eles possam exibir diferentes funções (FARAH, 2007). Acredita-se que as famílias multigênicas tenham surgido por uma série de eventos de duplicação durante a evolução e que o acúmulo de mutações ocorridas ao longo do tempo é responsável pelas pequenas diferenças observadas hoje entre esses genes (FARAH, 2007).

O microssatélite V1C01, se localiza próximo ao gene que codifica uma proteína hidrofóbica de soja (HPS). Essa proteína é sintetizada no endocarpo e, subsequentemente, depositada na superfície da semente, além disso, é considerada a principal proteína alergênica de soja (SANTOS, 2007). A hidrofobicidade e a topografia da superfície de variedades de soja expressando altos níveis de HPS podem afetar a fixação e penetração de patógenos, influenciar as propriedades de absorção de água e/ou mediar a fixação do endocarpo para a superfície da semente (ODANI et al., 1987; SWANSON et al., 1991; GONZALEZ et al., 1995; GIJZEN et al., 1999).

**Tabela 10:** Lista de proteínas codificadas em sequências contendo SSR.

SSR	Num. Acesso	Banco de dados	Descrição	Valor-E
V1C01	cd01958	CDD	Subfamília tipo Proteína Hidrofóbica de Soja (HPS)	3.32e-35
V1C02	cd05283	CDD	Cinamil álcool desidrogenase (CAD)	1.74e-124
V1C03	cd00051	CDD	Motivo de ligação ao cálcio, braço-EF	9.89e-29
V1C04	pfam09179	CDD	Domínio de ligação ao substrato TiIS, encontrado no tRNA(Ile) lisidina sintetase (TiIS)	6.14e-03
V1C05	pfam03937	CDD	Succinato desidrogenase	5.59e-22
V1C06	cl20177	CDD	Regulador de fissão mitocondrial	5.46e-03
V1C07	cl18461	CDD	Regulador de fissão mitocondrial	5.93e-04
V1C08	smart00769	CDD	Resposta hipersensitativa ao estresse hídrico	2.64e-25
V1C09	cd12118	CDD	Acil-CoA graxo sintetase	0e+00
V1C10	cl21161	CDD	Família de inibidor MyoD	1.75e-04
V1C11	cd08746	CDD	Regulador da proteína G sinalizadora (RGS)	1.54e-03
V1C12	COG0225	CDD	Metionina sulfoxido redutase	1.87e-82
V1C13	PLN03237	CDD	DNA topoisomerase 2	1.18e-03
V1C14	pfam02365	CDD	Proteína do meristema apical (NAM)	3.50e-75
V1C15	TIGR00879	CDD	Proteína transportadora de açúcar (SP)	2.86e-11
V2C01	P51823	Swiss-Prot	Fator de ribosilação de ADP	1,00e-005
V2C02	Q42676	Swiss-Prot	Transcetolase, cloroplastica (TK)	1,00e-005
V2C03	XM_002523459.1	GenBank	Apolipoproteína D, putativa	1,00e-005
V2C04	Q9STX5	Swiss-Prot	Homólogo a endoplasmína (regulada por glicose)	1,00e-005
V2C05	Q7XZU1	Swiss-Prot	Fosfoinositida fosfatase SAC4	1,00e-005
V2C06	XM_006420228.1	GenBank	Proteína hipotética	1,00e-005
V2C07	XM_010455196.1	GenBank	Proteína tolerante a metal	1,00e-005
V2C08	P54766	Swiss-Prot	Proteína nuclear ligante de GTP	1,00e-005
V2C09	XM_006474129.1	GenBank	Fator de transcrição responsivo ao etileno	1,00e-005
V2C10	A3BDI8	Swiss-Prot	Proteína associada ao estresse, zinc finger	1,00e-005
V2C11	Q13435	Swiss-Prot	Fator de splicing 3B subunidade 2	1,00e-005
V2C12	Q9AXJ4	Swiss-Prot	Fator de iniciação da transcrição eucariótica 5A	1,00e-005
V2C13	XM_007883109.1	GenBank	Proteína hipotética	1,00e-005
V2C14	Q01859	Swiss-Prot	ATP sintase mitocondrial, subunidade beta	1,00e-005

**Fonte:** Autor.

O SSR V1C11 está localizado em uma sequência pertencente a família das proteínas reguladoras da sinalização por proteína G (RGS). As proteínas RGS, são um grupo diverso de proteínas multifuncionais que regulam os eventos de sinalização celular que ocorrem à jusante aos receptores acoplados a proteína-G (WATSON et al., 1996). Compreendem mais de 20 proteínas diferentes que têm sido classificadas em subfamílias com base na homologia estrutural (NUNN et al., 2006).

O marcador SSR V2C10, está localizado próximo ao gene que codifica a proteína dedos de zinco (*zinc finger*). Essas proteínas possuem diversas funções e geralmente estão associadas ao estresse. Em *Arabidopsis*, por exemplo, uma proteína *Zinc finger* (Zat12) possui expressão aumentada quando submetida ao estresse oxidativo (DAVLETONA et al., 2005).

Recentemente, tem aumentado o número de detecção de SSR em regiões codificantes e regiões não transcritas (UTR). Esses dados oferecem uma importante informação para o estudo da função endógena dos SSR. Existem fortes evidências que os SSR não estão distribuídos aleatoriamente em regiões codificantes de proteínas, UTRs e introns. Dados substanciais indicam que a expansão dos SSR em regiões codificantes podem levar ao ganho ou perda de função dos genes via mudança no frame de leitura, por exemplo. Variações dos SSR em regiões 5'-UTR podem afetar a regulação da expressão gênica via transcrição e tradução. Sendo assim, os SSR podem fornecer uma base molecular para compreender a rápida adaptação dos organismos às mudanças ambientais (LI et al., 2004).

### **3.6 CONCLUSÕES**

Dado o exposto, percebemos que foi possível validar e caracterizar marcadores microssatélites em genótipos de cajueiro anão-precoce. Observou-se também a presença de polimorfismos nas regiões amplificadas dos genótipos de cajueiro, podendo ser úteis em estudos de diversidade genética. Além disso, foi possível identificar as proteínas codificadas nas sequências que contém SSR. A identificação de SSRs polimórficos presentes em sequências expressas do cajueiro também poderão auxiliar na compreensão do papel dos SSR na evolução dos genes.



### 3.7 PERSPECTIVAS

A detecção *in silico* de SSR em sequências de transcritos de cajueiro seguida pela sua validação *in vitro* possibilitou a detecção de polimorfismos entre nove genótipos testados. O sequenciamento dos produtos amplificados irá fornecer uma caracterização mais detalhada dos alelos encontrados possibilitando a determinação da frequência alélica, que é importante para a genotipagem do cajueiro. Esses estudos irão contribuir com a cajucultura no sentido de reconhecer os genótipos mais adequados para determinadas regiões aumentando a eficiência de produção e vantagens no mercado de exportação.

Por outro lado, a detecção de SSR em sequências codificantes podem fornecer informações para a melhor compreensão da evolução de genes e proteínas. Um estudo mais detalhado envolvendo o sequenciamento dos genes portadores de SSR em genótipos do cajueiro, correlacionando diversidade de alelos contendo SSR com variação fenotípica pode elucidar sobre o papel dos SSR em sequências codificantes relacionados ao ganho ou perda da função dos genes.

## REFERÊNCIAS BIBLIOGRÁFICAS

\_\_\_\_\_. **Noções elementares de botânica do cajueiro**. 2011. Disponível em: <<http://www.cajucultura.com/botanica.html>>, acesso em: 18 de setembro de 2014.

ALIYU, O. M.; AWOPETU, J. A. Chromosome studies in Cashew (*Anacardium occidentale* L.). **African Journal of Biotechnology**, v. 6, n. 2, p. 131-136, 2007.

ALMEIDA, J. I. L.; ARAÚJO, F. E.; LOPES, J. G. V. **Evolução do cajueiro anão precoce na Estação Experimental de Pacajus**, Ceará. Fortaleza: EPACE, 1993. 17p.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.

ALVES-FILHO, J. G. **Análises transcriptômica e proteômica de sementes de cajueiro (*Anacardium occidentale* L) visando aplicações biotecnológicas**. 2013. 198 p. Tese. (Doutorado em Bioquímica). Universidade Federal do Ceará. Fortaleza, 2013.

AMIRUDDIN, N. et al. Characterisation of full-length cDNA sequences provides insights into the *Eimeria tenella* transcriptome. **BMC Genomics**, v. 13, p. 21, 2012.

ANDREWS, S. **FastQC**: a quality control tool for high throughput sequence data. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Acesso em: 30 de setembro de 2013.

ANNADURAI, R. S. et al. Next generation sequencing and *de novo* transcriptome analysis of *Costus pictus* D. Don, a non-model plant with potent anti-diabetic properties. **BMC Genomics**, v. 13, n. 663, p. 1-15, 2014.

ARCHAK, S. et al. Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. **Genome**, n. 46, p. 362-369, 2003.

ARCOT, S. S. et al. Alu repeats: A source for the genesis of primate microsatellites. **Genomics**, v. 29, p. 136-144, 1995.

ASHBURNER, M. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, p. 25-29, 2000.

ASHRAFI, H. et al. De novo assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. **BMC Genomics**, v. 13, n. 571, p. 1-15, 2012.

BARRIL, P; NATES, S. **Introduction to Agarose and Polyacrylamide Gel Electrophoresis Matrices with Respect to Their Detection Sensitivities**. Gel Electrophoresis – Principles and Basics. 2012. 346 p.

BARROS, L. de M. et al. BRS 189 dwarf cashew clone cultivar. **Crop Breeding and Applied Biotechnology**, v.2, n.1, p.157-158, 2002.

BARROS, L. de M. et al. Seleção de clones de cajueiro anão para o plantio comercial no Estado do Ceará. **Pesquisa Agropecuária Brasileira**, v.35, n.11, p. 2197-2204, 2000.

BARROS, L. de M. **Melhoramento**. In: LIMA, V. P. M. S. (Ed.) A cultura do cajueiro no Nordeste do Brasil. Fortaleza: BNB/ETENE, 1988. p. 321-356.

BARROS, L. de M.; et al. **A cultura do cajueiro anão**. Fortaleza: EPACE, 1984. 67p.

BARROS, L. M. et al. **Recomendações técnicas para a cultura do cajueiro anão-precoce**. Fortaleza: Embrapa, 1993. 64 p.

BARROS, L. M.; CRISÓSTOMO, J. R. Melhoramento genético do cajueiro. In: ARAÚJO, J. P. P.; SILVA, V. V. (Org.). **Cajucultura: modernas técnicas de produção**. Fortaleza: Embrapa Agroindústria Tropical, 1995. p. 73-93.

BEZERRA, M. A. et al. Physiology of cashew plants grown under adverse conditions. **Brazilian Journal of Plant Physiology**, v. 19, n. 4, p. 449-461, 2008.

BHARDWAJ, J. et al. Comprehensive transcriptomic study on horse gram (*Macrotyloma uniflorum*): De novo assembly, functional characterization and comparative analysis in relation to drought stress. **BMC Genomics**, v. 14, p. 1-18, 2013.

- BIANCHI, V. J. et al. Caracterização molecular de cultivares de pessegueiro e nectarineira com microsátélites. **Revista Brasileira de Fruticultura**, v. 26, n. 3, p. 490-493, 2004.
- BLANCA, J. et al. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). **BMC Genomics**, v. 12, n. 104, p. 1-15, 2011.
- BRÄUTIGAM, A. et al. Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C3 and C4 species. **Journal of Experimental Botany**, v. 62, p. 3093-3102, 2011.
- BUSO, G. S. C. et al. Marcadores Microsátélites em Espécies Vegetais. **Revista Biotecnologia Ciência e Desenvolvimento**, v. 30, p. 46-50, 2003.
- CARVALHO, M. C. da C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, n. 3, p. 735-744, 2010.
- CAVALCANTI, J. J. V. et al. Avanços, desafios, e novas estratégias do melhoramento genético do cajueiro no Brasil. In: **I Simpósio Nordestino de Genética e Melhoramento de Plantas**. Fortaleza: EMBRAPA/CNPAT, 2009. 210 p.
- CAVALCANTI, J. J. V; WILKINSON, E. M. J. The first genetic maps of cashew (*Anacardium occidentale* L.). **Euphytica**, v. 157, p. 131–143, 2007.
- CHEN, M. et al. Development of mapped simple sequence repeat markers from common bean (*Phaseolus vulgaris* L.) based on genome sequences of a Chinese landrace and diversity evaluation. **Mol Breeding**, v. 33, p. 489–496, 2014.
- CHIARA M. et al. *De Novo* Assembly of the Transcriptome of the Non-Model Plant *Streptocarpus rexii* Employing a Novel Heuristic Recover Locus-Specific Transcript Clusters. **PLoS ONE**, v. 8, p. 1-10, 2013.
- CHISHOLM, A. et al. Cholesterol lowering effects of nuts compared with a canola oil enriched cereal of similar fat composition. **Nutrition, Metabolism & Cardiovascular Diseases**, v. 15, p. 284–292, 2005.

CONAB. **Companhia Nacional de Abastecimento**, 2014. Disponível em: <[http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14\\_10\\_16\\_08\\_20\\_47\\_castanhadecajusetembro2014.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14_10_16_08_20_47_castanhadecajusetembro2014.pdf)> Acesso em: 2014.

CONESA, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674-3676, 2005.

CRISÓSTOMO, L. A. et al. **Cultivo do cajueiro anão precoce: Aspectos fitotécnicos com ênfase na adubação e na irrigação**. Fortaleza: Embrapa Agroindústria Tropical, 2001. p. 1-19.

CROXFORD, A. E.; ROBSON, M.; WILKINSON, M. J . Characterization and PCR multiplexing of polymorphic microsatellite *loci* in cashew (*Anacardium occidentale* L.) and their cross-species utilization. **Molecular Ecology Notes**, v. 6, p. 249– 251, 2006.

DAVLETOVA, S. et al. The Zinc-Finger Protein Zat12 Plays a Central Role in Reactive Oxygen and Abiotic Stress Signaling in *Arabidopsis*. **Plant Physiology**, v. 139, p. 847-856, 2005

DE FREITAS, P. D. **Estudos de diversidade genética em camarões utilizando marcadores moleculares**. Manual Prático. São Paulo. vol. 2. 2005.

DECKERS, J. et al. **Cashew Crop Production in Tropical Africa**. Romain H. Reamaekers (Ed.) Published by Directorate General for International Cooperation (DGIC), Brussels, Belgium. p. 236-238, 2001.

DOYLE, J. J. T.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13- 15, 1987.

FAO. Food and Agricultural commodities production. **Food and Agriculture Organization**, 2012. Disponível em: <<http://faostat.fao.org>> Acesso em: outubro de 2014.

FARAH S. B. Decifrando o genoma humano. In: **DNA Segredos e Mistérios**. 2007.

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

FETUGA, B.; BABATUNDE, G.; OYENUGA, V. Composition and nutritive value of cashew nut to the rat. **Journal of Agricultural Food Chemistry**, v. 22, p. 678-682, 1974.

GAO, L-L. et al. Development of genomic resources for the narrow-leafed lupin (*Lupinus angustifolius*): construction of a bacterial artificial chromosome (BAC) library and BAC-end sequencing. **BMC Genomics**, v. 12, n. 521, p. 1-15, 2011.

GARG, R. et al. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. **DNA research**, v. 18, p. 53-63, 2011.

GELETA, M. et al. Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers. **The Scientific World Journal**. v. 2012, 11.p., 2012.

GEORGI, L. et al. Cranberry microsatellite marker development from assembled next-generation genomic sequence. **Molecular breeding**, v. 30, p. 227-237, 2012.

GIJZEN, M. et al. Hydrophobic protein synthesized in the pod endocarp adheres to the seed surface. **Plant Physiology**, v. 120, p. 951-959, 1999.

GOLDBLATT, P. Index to Plant Chromosome Numbers 1975–1978. **Monographs in Systematic Botany from the Missouri Botanical Garden**, v. 6, p. 1–553, 1981.

GOLDBLATT, P. Index to Plant Chromosome Numbers 1979–1981. **Monographs in Systematic Botany from the Missouri Botanical Garden**, v. 8, p. 1–427, 1984.

GOLDSTEIN, D. B.; SCHLÖTTERER, C. **Microsatellites: Evolution and Applications**. Oxford University Press, New York, 368p. 1999.

GONZALEZ, R. et al. Soybean hydrophobic protein and soybean hull allergy. **Lancet**, v. 346, p. 48-49, 1995.

GUO, S. et al. Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. **BMC Genomics**, v. 12, p. 1-13, 2011.

HAMADA, H. M.; PETRINO, M. G.; KAKUNAGA, T. A novel repeated element with Z-DNAforming potential is widely found in evolutionarily diverse eukaryotic genomes. **National Academy of Sciences**, v. 79, p. 6465–6469, 1982.

HAMEED, U. et al. Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (*Saccharum* spp) cultivars resistant and susceptible to red rot. **Genetic and Molecular Research**. n. 2, p. 1195-1204, 2012.

HAMMOCK, E. A. D.; YOUNG, L. J. Functional microsatellite polymorphism associated with divergent social structure in voles species. **Molecular Biology and Evolution**, v. 21, p. 1057-1063, 2004.

HANNON LAB. **FASTX Toolkit**. Disponível em: <[http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)>. Acesso em: 02 de outubro de 2013.

HIMEJIMA, M.; KUBO, I. Antibacterial agents from the cashew *Anacardium occidentale* (Anacardiaceae) nut shell oil. **Journal of Agricultural and Food Chemistry**, v. 39, p. 418-421, 1991.

IZZAH, N. K. et al. Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. *capitata* L.) and construction of an EST-based genetic map. **BMC genomics**, v. 15, n. 149, p. 1-13, 2014.

JEFFREYS, A.; WILSON, V.; THEIN, S. Hypervariable 'minisatellite' regions in human DNA. **Nature**, v. 314, p. 67–73, 1985.

JESUS, O. N. de. **Caracterização morfológica e molecular de cultivares de bananeira**. 2006. 83 f. Dissertação. (Mestrado em Agronomia). Universidade Federal Rural de Pernambuco. Recife, Pernambuco. 2006.

KALIA, R. K. et al. Microsatellite markers: an overview of the recent progress in plants. **Euphytica**, v. 177, p. 309-334, 2011.

KANEHISA, M. et al. From genomics to chemical genomics: new developments in KEGG. **Nucleic Acids Research**, v. 34, p. D354-D357, 2006.

KANNAMKUMARATH, S. S. et al. HPLC-ICP-MS determination of selenium distribution and speciation in different types of nut. **Analytical and Bioanalytical Chemistry**, v. 373, p. 454-460, 2002.

KASHI, Y.; KING, D.; SOLLER, M. Simple sequence repeats as a source of quantitative genetic variation. **Trends in Genetics**, v. 13, p. 74-78, 1997.

KRIEDT, R. A. **Isolamento de marcadores microssatélites para *Petúnia integrifolia* subesp. *depauperata* (Solanaceae) e *Passiflora ovalis* (Passifloraceae)**. 2009. 57 f. Monografia (Graduação em Ciências Biológicas) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

KUBO, I.; KOMATSU, S.; OCHI, M. Molluscicides from the cashew *Anacardium occidentale* and their large-scale isolation. **Journal of Agricultural and Food Chemistry**, v. 34, p. 970-973, 1986.

LAMAS, N. da S. **Desenvolvimento, caracterização e uso de marcadores microssatélites no mapeamento genético de características agrônômicas de cajueiro (*Anacardium occidentale* L.)**. 2010. 144f. Programa de Pós-Graduação em Biologia Molecular, UNIVERSIDADE DE BRASÍLIA – UnB, Dissertação.

LI, Y. et al. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. **Molecular Ecology**, v. 11, p. 2453-2465, 2002.

LI, Y-C. et al. Microsatellites Within Genes: Structure, Function, and Evolution. **Molecular Biology and Evolution**, v. 21, n. 6, p. 991-1007, 2004.

LIMA, V. P. M. S. A cultura do cajueiro no nordeste do Brasil. In: BARROS, L. M. **Melhoramento**. Fortaleza: Etene, 1988. 35 p.

LIU, Z. et al. Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. **Biophysical Research Communication**, v. 259, p. 190-194, 1999.



LUO, H. et al. Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. **BMC Genomics**, v. 12, p. 1-15, 2011.

MARCOTTE, E. M. et al. A census of protein repeats. **Journal of Molecular Biology**, v. 293, p. 151-160, 1999.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437, p. 376-380, 2005.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature reviews**, v. 12, p. 671-682, 2011.

MARTINS, A. 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. 198 f. Tese (Doutorado em Biologia Molecular) - Universidade de Brasília, Brasília. 2013.

MARTINS, T. B. et al. Validação da Metodologia de Caracterização Molecular de Cultivares de Soja por Meio de Marcadores Microssatélites. In: REUNIÃO BRASILEIRA DE FERTILIDADE DO SOLO E NUTRIÇÃO DE PLANTAS, 30. 2012, Maceió/Alagoas, **Anais...** Alagoas, 2012. p. 4.

MCGETTIGAN, P. A. Transcriptomics in the RNA-seq era. **Current Opinion in Chemical Biology**, v. 17, p. 4-11, 2013.

MOHOD, A. G.; KHANDETOD, Y. P.; POWAR, A. G. Processed cashew shell waste as fuel supplement for heat generation. **Energy for Sustainable Development**, v. 12, p. 73-76, 2008.

MOREIRA, R. F. **Alteração no perfil de expressão dos genes de folhas de cajueiro CCP76 (*Anacardium occidentale* L.) em resposta ao estresse salino**. 2014. 60 f. Dissertação. (Mestrado em Biotecnologia). Universidade Federal de Ceará. Sobral, 2014.

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellite are preferentially associated with nonrepetitive DNA in plants genomes. **Nature Genetics**, v. 30, p. 194-200, 2002.

- MORGANTE, M.; OLIVIERI, A. M. PCR-amplified microsatellites as markers in plant genetics. **The Plant Journal**, v.3, n.1, p.175-182, 1993.
- MOTTA, L. B. et al. Validação de marcadores microssatélites em genótipos de café (*Coffea arabica* e *Coffea canephora*). In. Encontro Latino Americano de Iniciação Científica e Encontro Latino Americano de Pós-Graduação, 15 e 11., 2011, Espírito Santo. **Resumos...** Espírito Santo: Universidade do Vale do Paraíba, 2011. p. 1 - 4.
- MUÑOZ-MÉRIDA, A. et al. De novo assembly and functional annotation of the olive (*Olea europaea*) transcriptome. **DNA Research**, v. 7, p. 1-16, 2013.
- MUTZ, K. O. et al. Transcriptome analysis using next-generation sequencing. **Current Opinion in Biotechnology**, v. 24, p. 22–30, 2013.
- NAGY, E. D. et al. A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. **BMC Genomics**, v. 13, n. 469, p. 1-11, 2012.
- NAMBIAR, M. C.; PILLAI, P. K. T. Cashew. In: BOSE, T. K. **Fruits of India: Tropical and subtropical**. Calcutta, India, 1985. p. 409-438.
- NETO, J. R. C. F. **Análise *in silico* de EST-SSR em *Phaseolus vulgaris* e *Glycine max* e transferibilidade de marcadores para *Vigna unguiculata***. Dissertação. (Mestrado em Genética). Universidade Federal de Pernambuco. Recife, 2008.
- NICOT, N. et al. Study of simple sequence repeat (SSR) markers from wheat expressed tags (ESTs). **Theoretical and Applied Genetics**, v. 109, p. 800-805, 2004.
- NUNN, C. et al. RGS17/RGSZ2 and the RZ/A family of regulators of G-protein signaling. **Seminars in Cell & Developmental Biology**, v. 17, p. 390–399, 2006.
- ODANI, S. et al. Soybean hydrophobic protein: isolation, partial characterization and the complete primary structure. **European Journal of Biochemistry**. v. 162, p. 485-491, 1987.
- OLIVEIRA, M. V. C. et al. Caracterização de clones de mandioca utilizando marcadores microssatélites. **Revista Ciência Agronômica**, v. 43, n. 1, p. 170-176, 2012.

- OLIVEIRA, V. H.; et al. **Cultivo do cajueiro anão precoce**. Fortaleza: Embrapa Agroindústria Tropical. 2002.
- OLOGUNDE, M. O. et al. Preliminary nutritional evaluation of cashew nuts from different locations in Nigeria. **Continental Journal Food Science and Technology**, v. 5, n. 2, p. 32-36, 2011.
- PAIVA, J. R. de; et al. **Clone de cajueiro-anão precoce BRS 226 ou Planalto**: nova alternativa para o plantio na Região Semi-Árida do Nordeste. Fortaleza: Embrapa Agroindústria Tropical, 2002. 4p.
- PAIVA, J. R.; BARROS, L. de M. **Clones de Cajueiro**: Obtenção, Características e Perspectivas. 1ª edição. Fortaleza: Embrapa Agroindústria Tropical, 2004. 25p.
- PAIVA, J. R.; CRISÓSTOMO, J. R.; BARROS, L. M. **Recursos genéticos do cajueiro: coleta, caracterização e utilização**. Fortaleza: Embrapa, 2003. 43p.
- PARCHMAN, T. L.; et al. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. **BMC genomics**, 11, 180, 2010.
- PARIDA, S. K. et al. Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. **Theoretical and Applied Genetics**, v. 118, p. 327–338, 2009.
- PATZAK, J. et al. Comparison of genetic diversity structure analyses of SSR molecular marker data within apple (*Malus × domestica*) genetic resources. **Genome**, v. 55, n. 9, p. 647-665, 2012.
- PINHEIRO, C. R. et al. **Caracterização de genótipos de cajueiro por meio de marcadores microssatélites**. XX Congresso Brasileiro de Fruticultura XX Congresso Brasileiro de Fruticultura. Vitória: [s.n.]. 2008.
- PONTES, N. H. L. **Análises comparativas das montagens do transcriptoma de sementes em maturação de cajueiro (*A. occidentale* var. *nanum*) com diferentes valores de sobreposição de sequências (k-mers) utilizando a abordagem do grafo de Bruijn**. 2013. 36 f. Monografia (Graduação em Ciências Biológicas). Universidade Estadual Vale do Acaraú. Sobral, 2013.

PRASANNA, B. M. et al. Molecular marker-assisted breeding options for maize improvement in Asia. **Molecular Breeding**, v. 26, n. 2. p. 339-356, 2010.

PURSEGLOVE, J. W. **Tropical crops: dicotyledons**. Reino Unido: Londres, 1968. 719 p.

RAJU, N. L. et al. The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). **BMC Plant Biology**, v.10, p. 45, 2010.

RICKWOOD, D.; HAMES, B. D. **Gel Electrophoresis of Nucleic Acids: A Practical Approach**, 2ª edição, IRL Press, 1990.

ROMANO, E. **Extração de DNA de tecidos vegetais**. In: BRASILEIRO, A. C. M.; CARNEIRO, V. T. C. Manual de transformação genética de plantas. Brasília: Embrapa, 1998. p. 163-177.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS. **Molecular cloning**. A laboratory manual. Second edition. ed. [S.l.]: Cold Spring Harbor Laboratory Press, 1989.

SANSALONI, C. P. **Desenvolvimento, caracterização e mapeamento de microssatélites de tetra e pentanucleotídeos em *Eucalyptus spp.*** 2008. 114 p. Dissertação (Mestrado em Biologia Molecular) - Universidade de Brasília, Brasília, 2008.

SANTOS, F. F. dos. **Validação de marcadores microssatélites derivados de regiões funcionais do genoma de aveia**. 2010. 54 f. Dissertação (Mestrado em Biotecnologia) - Universidade Federal de Pelotas, Rio Grande do Sul, 2010.

SANTOS, P. de O. **Tegumentos de sementes como barreiras químicas de defesa contra o ataque do inseto *Callosobruchus maculatus* e fungos fitopatogênicos**. 2007. 104 p. Dissertação (Mestrado Biociências e Biotecnologia). Universidade Estadual do Norte Fluminense Darcy Ribeiro. Campos dos Goytacazes, Rio de Janeiro. 2007.

SANTOS, P. S. et al. Caracterização molecular de germoplasma de patchouli (*Pogostemon* sp.) por marcadores RAPD. **Scientia Plena**, v, 9, n. 5, p. 1-7, 2013.

SANTOS, R. P. et al. Compositional analysis of cashew (*Anacardium occidentale* L.) peduncle bagasse ash and its in vitro antifungal activity against *Fusarium* species. **Revista Brasileira de Biociências**. v. 9, p. 200-205, 2011.

SARGENT, D. J. et al. A microsatellite linkage map for the cultivated strawberry (*Fragaria 3 ananassa*) suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection. **Theoretical and Applied Genetics**. v. 124, p. 1229–1240, 2012.

SEHGAL, D.; RAINA, S. N. **DNA markers and germplasm resource diagnostics: new perspectives in crop improvement and conservation strategies**. In: ARYA, I. D.; ARYA, S. Utilization of biotechnology in plant sciences. Dehradun: [s.n.], 2008. p. 39–54.

SHENDURE, J.; JI, H. Next-generation DNA sequencing. **Nature Biotechnology**, v. 28, n. 10, p. 1135-1145, 2008.

SILVA, K. V. P. et al. Variabilidade genética entre acessos do gênero *Manihot* por meio de marcadores moleculares ISSR. **Pesquisa Agropecuária Brasileira**, v.46, n.9, p.1082-1088, 2011.

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, p. 1-9, 2013.

SOUER, E. et al. The No Apical Meristem Gene of *Petunia* Is Required for Pattern Formation in Embryos and Flowers and Is Expressed at Meristem and Primordia Boundaries. **Cell**, v. 85, p. 159–170, 1996.

STASIUK, M.; KOZUBEK, A. Biological activity of phenolic lipids. **Cellular and Molecular Life Sciences**., v. 67, p. 841-860, 2010.

STEIN, L. Genome annotation: from sequence to biology. **Nature**, v. 4, p. 493-503, 2001.

STRAND, M. et al. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. **Nature**, v. 365, p. 274-276, 1993.

SWAMY, K. R. M.; RAO, E. V. V. B.; BHAT, M. G. Catalogue of minimum descriptors of cashew (*Anacardium occidentale* L.) germplasm accessions-II. **National Research Centre for Cashew**, Puttur, Karnataka, India, 1998.

SWANSON, M. C. et al. Source of the aeroallergen of soybean dust: a low molecular mass glycopeptide from the soybean tela. **Journal of Allergy and Clinical Immunology**, v. 87, p. 783-788, 1991.

SYED, N. H. et al. Ty1-copiaretrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). **Theoretical and Applied Genetics**, v. 110, p. 1195–1202, 2005.

TARAZONA, S. et al. Differential expression in RNA-seq: A matter of depth. **Genome Research**, v. 21, p. 2213-2223, 2011.

TAUTZ, D. Notes on the defunction and nomenclature of tandemly repetitive DNA sequences. **EXS**, v. 67, p. 21-28, 1993.

TAUTZ, D.; TRICK, M.; DOVER, G. A. Cryptic simplicity in DNA is a major source of genetic variation. **Nature** v. 322, p. 652-656, 1986.

THANKAMMA-PILLAI, P. K.; NAMBIAR, M. C. Study of microsporogenesis in cashew. **Acta Horticulturae**, n. 108, p. 280-281, 1985.

THIEL, T. et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, v. 106, p. 411-422, 2003.

THOMPSON, J. D.; GIBSON, T. J.; HIGGINS, D. G. Multiple Sequence Alignment Using ClustalW and ClustalX. **Current Protocols in Bioinformatics**, 2002.

TÓTH, G.; GÁSPARI, Z.; JURKA, J. Microsatellites in different eukaryotic genomes: Survey and analysis. **Genome Research**, v. 10, p. 967-981, 2000.

TSANAKAS, G. F. et al. *De novo* transcriptome analysis of petal senescence in *Gardenia jasminoides* Ellis. **BMC Genomics**, v. 15, n. 554, p. 1-16, 2014.

VARSHNEY, R. K. et al. **Genic molecular markers in plants: development and applications**. In: VARSHNEY, R. K.; TUBEROSA, R. *Genomics assisted crop improvement: genomics approaches and*. [S.l.]: Springer, 2007. p. 13–29.

VOS, P. et al. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Res.**, v. 21, 1995.

WANG , X. W. et al. De novo characterization of a white fly transcriptome and analysis of its gene expression during development. **BMC Genomics**, v. 11, p.400, 2010.

WANG, X. et al. A new electrophoresis technique to separate microsatellite alleles. **African Journal of Biotechnology**, v. 8, n. 11, p. 2432-2436, 2009.

WANG, Z. et al. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). **BMC Genomics**, v. 11, p. 726-740, 2010.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews**, v. 10, p. 57-63, 2009.

WARD, J. A. et al. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. **BMC Genomics**, v. 14, n. 2, p. 1-14, 2013.

WATSON, N. et al. RGS family members: GTPase-activating proteins for heterotrimeric G-protein alpha-subunits. **Nature**, v. 383, p. 172-175, 1996.

WEBER, J. L. Informativeness os human (dC-dA)<sub>n</sub> (dG-dT)<sub>n</sub> polymorphism. **Genomics**, v. 7, p. 524-530, 1990.

WILDER, J.; HOLLOCHER, H. Mobile elements and the genesis of microsatellites in dipterans. **Molecular Biology and Evolution**, v. 18, p. 384-392, 2001.

WILLIAMS, J. G. et al. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Res.** , v. 18, p. 6531–6535, 1990.

- WU, W. et al. Development of simple sequence repeat (SSR) markers that are polymorphic between cultivars in *Brassica rapa* subsp. *rapa*. **African Journal of Biotechnology**, v. 11, p. 2654-2660, 2012.
- XAVIER, M. A. et al. Paternity identification in sugarcane polycrosses by using microsatellite markers. **Genetics and Molecular Research**, v. 13, n. 1, p. 2268-2277, 2014.
- XIE, C. et al. Characterization of the global transcriptome for *Pyropia haitanensis* (Bangiales, Rhodophyta) and development of cSSR markers. **BMC genomics**, v. 14, p. 1-13, 2013.
- YATES, S. A. et al. *De novo* assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. **BMC Genomics**, v. 15, n. 453, p. 1-15, 2014.
- YE, J. et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. **BMC Bioinformatics**, v. 13, n. 134, p. 1-11, 2012.
- YU, J. K. et al. A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. **Genes Genomes Genetics**, v. 2, n. 1, p. 43–58, 2012.
- YU, J. N. et al. Fast and Cost-Effective Mining of Microsatellite Markers Using NGS Technology: An Example of a Korean Water Deer *Hydropotes inermis argyropus*. **PLoS ONE**, v. 6, 2011.
- ZABEAU, M.; VOS, P. **Selective restriction fragment amplification**: a general method for DNA fingerprinting. 0 534 858 A1, 1993.
- ZERBINO, D. R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, v. 18, p. 821-829, 2008.
- ZHANG, J. et al. De novo assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). **BMC Genomics**, v. 13, n. 90, p. 1-6, 2012.



ZHANG, P. et al. Population Structure and Genetic Diversity in a Rice Core Collection (*Oryza sativa* L.) Investigated with SSR Markers. **PLoS ONE**, v. 6, n. 12, 2011.

ZHANG, X. et al. A BAC/BIBAC-based physical map of chickpea, *Cicer arietinum* L. **BMC Genomics**, v. 11, n. 501, p. 1-10, 2010.

ZHAO, Y.; PRAKASH, C. S.; HE, G. Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. **BMC Research Notes**, v. 5, p. 362, 2012.