

**UNIVERSIDADE FEDERAL DE ALAGOAS – UFAL**  
**CAMPUS ARAPIRACA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM AGRICULTURA E AMBIENTE - PPGAA**  
**MESTRADO EM AGRICULTURA E AMBIENTE**

**ELLEN KAROLLYNE SANTOS LOPES**

**GENOMA FUNCIONAL DE *Spondias tuberosa* Arr. Câm. (UMBUZEIRO)**

**ARAPIRACA**

**2023**

Ellen Karollyne Santos Lopes

Genoma funcional de *Spondias tuberosa* Arr. Câm. (umbuzeiro)

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Agricultura e Ambiente - PPGAA da Universidade Federal de Alagoas - UFAL, Campus Arapiraca, como requisito parcial para obtenção do grau de Mestre em Agricultura e Ambiente.

Orientador: Prof. Dr. Cícero Carlos de Souza Almeida

Arapiraca

2023



Universidade Federal de Alagoas – UFAL  
Campus Arapiraca  
Biblioteca Setorial *Campus Arapiraca* - BSCA

L864g      Lopes, Ellen Karollyne Santos  
Genoma funcional de *Spondias tuberosa* Arr. Câm. (umbuzeiro) [recurso eletrônico]  
/ Ellen Karollyne Santos Lopes. – Arapiraca, 2023.  
40 f.: il.

Orientador: Prof. Dr. Cícero Carlos de Souza Almeida.  
Dissertação (Mestrado em Agricultura e Ambiente) - Programa de Pós-Graduação  
em Agricultura e Ambiente, Universidade Federal de Alagoas, *Campus Arapiraca*,  
Arapiraca, 2023.  
Disponível em: Universidade Digital (UD) / RD- BSCA– UFAL (*Campus Arapiraca*).  
Referências: f. 33-40.

1. Genética. 2. Biologia molecular. 3. Anotação funcional. I. Almeida, Cícero Carlos  
de Souza. II. Título.

CDU 636

Bibliotecário responsável: Nestor Antonio Alves Junior  
CRB-4 / 1557

Ellen Karollyne Santos Lopes

Genoma funcional de *Spondias tuberosa* Arr. Câm. (umbuzeiro)

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Agricultura e Ambiente - PPGAA da Universidade Federal de Alagoas - UFAL, Campus Arapiraca, como requisito parcial para obtenção do grau de Mestre em Agricultura e Ambiente.

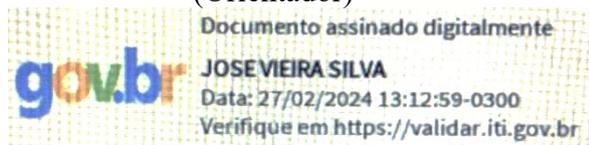
Data de aprovação: 01/08/2023.

### Banca Examinadora



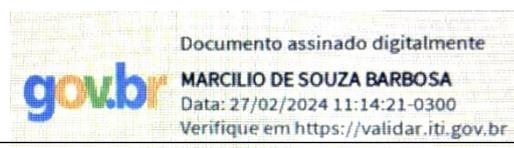
---

Prof. Dr. Cícero Carlos de Souza Almeida  
Universidade Federal de Alagoas - UFAL  
Campus de Arapiraca  
(Orientador)



---

Prof. Dr. José Vieira Silva  
Universidade Federal de Alagoas - UFAL  
Campus de Arapiraca  
(Examinador)



---

Prof. Dr. Marcílio de Souza Barbosa  
Instituto Federal de Alagoas – IFAL  
Campus de Batalha  
(Examinador)

Dedico aos meus pais, pelo empenho que tiveram - e têm ainda - em me ensinar sempre a ser uma pessoa melhor. Ao meu irmão, por ser um bom amigo.

## **AGRADECIMENTOS**

Ao meu orientador Prof<sup>o</sup> Dr. Cícero Carlos, pelos ensinamentos que me formaram nessa jornada acadêmica, pelo apoio, pela paciência e pelo tempo investido em minha formação.

Ao prof<sup>o</sup> Dr. André Seco Marques da Silva também pelos ensinamentos prestados durante minha formação no mestrado.

À FAPEAL, órgão de apoio à pesquisa que financiou este estudo.

Ao Instituto Max Planck, que também colaborou com a pesquisa.

Ao Programa de Pós-Graduação em Agricultura e Ambiente.

“A esfera de energia do planeta está aí para quem quiser vê-la e compreendê-la: um processo implacável, mas um processo mesmo assim. Há nele um espaço vazio? Então alguma coisa ocupará esse vazio. A ciência é feita de tantas coisas que parecem óbvias depois de explicadas”

(Frank Herbert, 1965)

## RESUMO

A *Spondias tuberosa* Arr. Câm. (umbuzeiro) é uma espécie de árvore conhecida no nordeste brasileiro por ser uma alternativa e recurso econômico brasileiro em regiões do semiárido. Pensando em realizar a anotação funcional do genoma da espécie, o presente estudo objetivou-se em apresentar a quantidade de genes envolvidos nos processos biológicos, nos componentes celulares e na função molecular, além disso, os códigos de enzimas e as vias metabólicas. Para isso, o DNA foi extraído pelo método de *long reads* usando Pacbio Assembly. Após detectar estruturas repetitivas, usando o GeneMask, a anotação estrutural foi montada usando preditores de genes *ab initio* (Augustus e GeneMarker) e inteligência artificial (Helixer), para a qualificação da montagem do genoma, a ferramenta BUSCO foi utilizada. A anotação funcional obteve os dados gerados pela plataforma Blast2GO, os genes foram anotados usando as categorias Gene Orthology (GO), além de enzimas e vias metabólicas. A avaliação do BUSCO na montagem gerada pelo Helixer resultou em uma precisão de 98% de completude. Os genes previstos foram distribuídos em 51.896 termos GO totais, distribuídos em três categorias principais: função molecular, processo biológico e componente celular. No presente estudo, notou-se que os genes do processo biológico envolvidos na resposta ao estresse e funções catalíticas resultaram na maior quantidade de genes para a categoria de processos biológicos e função celular, respectivamente, no genoma de *S. tuberosa*. Revelando também que a espécie apresenta um forte sistema de defesa contra condições estressoras.

**Palavras-chave:** biologia molecular; anotação funcional; genética.

## ABSTRACT

*Spondias tuberosa* Arr. Cam. (umbuzeiro) is a species of tree known in Brazil northeastern for being an alternative and Brazilian economic resource in semi-arid regions. Thinking about carrying out species genome functional annotation, the present study aimed to present the genes number involved in biological processes, cellular components and molecular function. For this, the DNA was extracted by the method of long reads using Pacbio Assembly. After detecting repetitive structures, using GeneMask, the structural annotation was assembled using *ab initio* gene predictors (Augustus and GeneMarker) and artificial intelligence (Helixer), for the genome assembly qualification, the BUSCO tool was used. Functional annotation obtained data generated by the Blast2GO platform, genes were annotated using Gene Orthology (GO) categories. BUSCO's evaluation of the assembly generated by Helixer resulted in an accuracy of 98% completeness. The predicted genes were distributed in 51,896 total GO terms, distributed in three main categories: molecular function, biological process and cellular component. In the present study, it was noted that the biological process genes involved in stress response and catalytic functions resulted in the highest number of genes for the biological process category and cell function, respectively, in the genome of *S. tuberosa*. Also revealing that the species has a strong defense system against stressful conditions.

**Keywords:** molecular biology; functional annotation; genetics.

## LISTA DE FIGURAS

Figura 1	– Imagem do fruto do umbuzeiro ( <i>S. tuberosa</i> Arr. Câm).....	14
Figura 2	– Representação gráfica do cariótipo da espécie <i>S. tuberosa</i> .....	17
Figura 3	– Dados do BUSCO para a precisão da montagem do genoma de <i>S. tuberosa</i> . dos preditores Helixer, Augustus e GeneMarker.....	25
Figura 4	– Classificação da distribuição de genes envolvidos na categoria processo biológico.....	27
Figura 5	– Classificação da distribuição de genes envolvidos na categoria função molecular.....	29
Figura 6	– Classificação da distribuição de genes envolvidos na categoria componente celular.....	30

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	10
<b>2</b>	<b>REVISÃO DE LITERATURA</b>	11
<b>2.1</b>	<b><i>Spondias tuberosa</i></b>	11
2.1.1	Família	11
2.1.2	Gênero	11
2.1.3	Localização da espécie	12
2.1.4	Morfologia	13
2.1.5	Importância econômica	14
2.1.6	Valores nutricionais	14
2.1.7	Importância no ecossistema	15
2.1.8	Variabilidade genética	15
<b>2.3</b>	<b>Transcriptoma</b>	18
<b>2.4</b>	<b>Anotação Genômica</b>	19
<b>3</b>	<b>METODOLOGIA</b>	22
<b>3.1</b>	<b>Sequenciamento de <i>Long reads</i></b>	22
3.1.1	Extração de DNA	22
3.1.2	PacBio	22
3.1.3	Omni-C	22
<b>3.3</b>	<b>Obtenção do genoma (Assemble usando PacBio)</b>	23
3.3.1	Hifiasm	23
3.3.2	Omin-c scaffolding	23
3.3.3	Análise com Busco	23
<b>3.4</b>	<b>Anotação estrutural de DNA repetitivo</b>	24
<b>3.5</b>	<b>Anotação estrutural de genes</b>	24
<b>3.6</b>	<b>Anotação funcional</b>	24
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	25
<b>4.1</b>	<b>Montagem do genoma</b>	25
<b>4.2</b>	<b>Anotação funcional</b>	25
4.2.1	Processos biológicos	26
4.2.2	Função Molecular	28
4.2.3	Componente celular	30
<b>5</b>	<b>CONCLUSÃO</b>	32
	<b>REFERÊNCIAS</b>	33

## 1 INTRODUÇÃO

O umbuzeiro, nome popular da espécie *Spondias tuberosa* Arr. Câm., é uma espécie de árvore conhecida no nordeste brasileiro por ser uma alternativa e recurso econômico brasileiro em regiões do semiárido. A espécie *S. tuberosa* ganhou destaque no mercado brasileiro e internacional devido a sua variedade de compostos bioativos, os quais possuem efeitos nutricionais, terapêuticos, propriedades físico-químicas, sensoriais, entre outros (LINS NETO; PERONI; ALBUQUERQUE, 2010; LIMA; SILVA; OLIVEIRA, 2018).

Mas ainda existem desafios a serem superados acerca dessa árvore tão importante. Por exemplo, sua comercialização limitada, consequência da produção somente no período sazonal, além disso, o fruto precisa de mais conservação após a colheita, formas melhores de processamento, melhorias nas técnicas de cultivo, de colheita, de processamento, aumento da produtividade e outros (MOREIRA *et al.*, 2021).

Para conhecer as potencialidades que a espécie *S. tuberosa* tem, a ciência busca através da genética o estudo do transcriptoma ou a fração do genoma que contém os genes. Anotações genéticas são valiosas já que a partir delas é possível compreender como as sequências do genoma e como as mudanças nessas sequências são interpretadas pelo metabolismo celular (BYRNE *et al.*, 2019; MOREIRA *et al.*, 2021; SANTOS *et al.*, 2021).

Além disso, na genética de conservação, o genoma também é usado como uma métrica para medir a diversidade. Através das informações dos genes funcionais é possível ver a diversidade adaptativa de uma espécie em um determinado ambiente ou no ambiente em que se encontra. Com genomas já conhecidos pode-se testar uma variedade de hipóteses, além de que, com o acúmulo de informações genômicas, esses dados são valiosos e podem ser viáveis para mais espécies, principalmente para espécies ameaçadas de extinção (SCHMIDT; HOBAN; JETZ, 2023).

O genoma funcional de *S. tuberosa* pode fornecer valiosas informações sobre a biologia e potenciais aplicações desta importante espécie de árvore frutífera. Até porque, dados genômicos são importantes para testar hipóteses sobre a diversidade genética, conservação e adaptação ao ambiente. Ademais, mais pesquisas sobre a espécie *S. tuberosa* são necessárias para entender sua complexidade e diversidade, tornando-se um valioso recurso genético para o desenvolvimento de novas cultivares com características melhoradas.

## 2 REVISÃO DE LITERATURA

### 2.1 *Spondias tuberosa*

#### 2.1.1 Família

A espécie *S. tuberosa* faz parte da família Anacardiaceae. Esta família é dividida entre duas subfamílias, as quais são Anacardioideae e Spondioideae, a última corresponde a espécie do estudo. A Anacardiaceae é considerada uma família importante ecologicamente e economicamente, possuindo ao todo cerca de 600 espécies em 70 gêneros presentes nos trópicos e subtropicais do mundo todo e também em zonas temperadas. Ela é diversificada tanto em sua arquitetura floral quanto pela morfologia de suas frutas, sendo então considerada uma família modelo para o estudo sobre a evolução da diversidade estrutural, já que tem correlação com a diversidade da linhagem. A diversidade de suas frutas, por exemplo, é o que faz a família Anacardiaceae ter tantos gêneros reconhecidos, como *Anacardium*, *Pistacia* e *Mangifera* (TÖLKE *et al.*, 2021; SAMEH *et al.*, 2018; SCHULZE-KAYSERS; FEUEREISEN; SCHIEBER, 2015; MITCHELL *et al.*, 2022; SALEHI *et al.*, 2019; NEZAMI; GALLEGU, 2023; NIU; GAO; LIU, 2022).

A família da Anacardiaceae inclui espécies de alta importância econômica como o caju, pistache, manga e pimenta rosa. A espécie *Mangifera indica* (mangueira) é uma árvore frutífera, considerada a mais cultivada de seu gênero e tendo mais de 1000 variedades em todo o mundo. Evidências sugerem que a *M. indica* tenha sido cultivada há mais de 4000 anos, originalmente na Ásia, se espalhando pelo mundo através da humanidade. A mangueira, por exemplo, é uma das plantas mais produzidas no Brasil, sendo considerada uma das plantas de cultura tropical. A *Pistacia vera* (pistache), assim como a mangueira, é cultivada há vários anos, além de ser originária também da Ásia. A planta pistache tem uma forte adaptação em áreas desérticas e semidesérticas em regiões temperadas e subtropicais (MANDALARI *et al.*, 2021; SHAH *et al.* 2010; YANG *et al.*, 2022; KUMAR *et al.*, 2021).

#### 2.1.2 Gênero

O gênero *Spondias* tem sua importância na família Anacardiaceae sendo o primeiro gênero a ser descrito dessa família, por Linnaeus, tendo a primeira espécie *Spondias mombin* L. (cajazeira) publicada em 1753. *Spondias* compreende árvores frutíferas com 18 espécies, possuindo apenas diferenças sutis entre elas, apresentando ainda riqueza entre a diversidade

de suas características foliares, sendo úteis para a taxonomia. Estudos sugerem que esse gênero era usado no Vale Tehuacán, no México, aproximadamente há 6500 a. C. Dentro do gênero *Spondias*, existem quatro espécies de relevância econômica na América tropical, são elas: *Spondias dulcis* L. (cajá manga), *Spondias mombin* L. (cajazeira), *S. purpurea* L. (cirigueleira) e *S. tuberosa* Arr. Câm. (umbu) (MITCHELL; DALY, 2015; MOREIRA, 2022).

Os membros de *Spondias* apresentam seu valor na medicina tradicional, sendo utilizados em tratamentos de várias doenças devido às suas substâncias antioxidantes, possuindo papel de anti-inflamatório, analgésico, anti-hipertensivo, antimicrobiano, antiviral, protetor de úlceras hepatoprotetoras, fotoprotetoras, hipoglicêmicas, entre outros. Essas características estão relacionadas aos fitoconstituintes que estão presentes no gênero, sendo eles os taninos, flavonóides, esteróis, óleos essenciais, polissacarídeos e aminoácidos (DIAS *et al.*, 2019; MOREIRA *et al.*, 2021).

### 2.1.3 Localização da espécie

A espécie *S. tuberosa* Arr. Câm, mais conhecida localmente na região semiárida brasileira por umbuzeiro, é nativa e endêmica do Brasil, presente no bioma Caatinga. Este bioma também é conhecido por apresentar várias espécies endêmicas de alto valor econômico para o país, como o umbuzeiro, mas também como fonte de renda de populações locais. Dessa forma, a exploração para o comércio ou para renda local acaba devastando o bioma e várias espécies que são pertencentes a ele, correndo perigo de extinção (BALBINO *et al.*, 2018; SANTOS *et al.*, 2021).

O bioma é exclusivamente brasileiro, conhecido por suas florestas tropicais secas em períodos sazonais, estendendo-se pelo Estado do Ceará até uma parte de Minas Gerais, sendo sua maior porção no nordeste brasileiro e, por isso, apresenta condições edafoclimáticas variadas de acordo com sua localização. É muito importante para a região Nordeste já que ocupa 54% da região com seu total de 900.000km<sup>2</sup>, correspondendo a 11% de todo o território brasileiro, sendo este território conhecido também como Polígono das Secas. O nome Caatinga é advindo do tupi Kaa e Tinga, que se traduz basicamente por “floresta branca”. Este nome foi dado devido a sua forma de floresta sem folhas verdes na maior parte do ano, conhecida também como “floresta seca” (ALVES, 2007; LIMA; SILVA; OLIVEIRA, 2018).

Entretanto, a Caatinga possui diversas formas, se separando em três categorias, que são elas a floresta seca, as caatingas arbustivas e, por fim, as estepes. Como a maior parte da

Caatinga está localizada na região do semiárido, há a perda das folhas no bioma, o que permite uma adaptação das plantas para que economizem água durante estações com menos chuva. Algumas espécies presentes nesse bioma também possuem porte lenhoso e arbustos que são ramificados no nível do solo. Como pode ser visto, a Caatinga é um bioma frágil, mesmo assim, as atividades econômicas enfrentadas neste bioma estão acompanhadas de desmatamentos, práticas de devastação e queimadas, o que traz consequências graves como o desequilíbrio do ecossistema (ALVES, 2007; LIMA; SILVA; OLIVEIRA, 2018).

Este possui uma variedade de recursos vegetais, isso possibilita que esses recursos sejam utilizados de diversas formas pela população local, principalmente de forma alimentícia. Devido a sua vasta biodiversidade, as frutas da Caatinga, assim como diversas outras frutas brasileiras, estão entre as que apresentam maior potencial alimentar, sendo estas as mais nutritivas e mais aromáticas do mundo, entretanto, a maioria delas é conhecida apenas pela comunidade local (CAMACAM; MESSIAS, 2022).

#### 2.1.4 Morfologia

A anatomia da *S. tuberosa* se destaca por um tronco vertical e cilíndrico, que pode chegar a impressionantes 10 metros de altura, com uma casca lisa e cinzenta. As folhas, compostas, alternadas e espiraladas, medem cerca de 20 a 30 cm de comprimento e 10 a 15 cm de largura. As flores, pequenas e amarelas, se reúnem em inflorescências terminais. O fruto, uma drupa globosa, tem entre 4 e 6 cm de diâmetro, com uma casca fina e enrugada, polpa suculenta e fibrosa de cor amarela ou alaranjada, e contém apenas uma semente (figura 2). O umbuzeiro carrega a característica de ter raízes tuberosas, as quais armazenam água, xerófila e é caracterizada pelo derramamento de folhas, permanecendo sem elas pelos períodos de seca, o que é característico das plantas do bioma ao qual a espécie pertence (LIMA; SILVA; OLIVEIRA, 2018; MITCHELL; DALY, 2015; SOUZA; SOUZA, 2014; SILVA; PAIVA, 2007).

Figura 1 - Imagem do fruto do umbuzeiro (*S. tuberosa* Arr. Câm).



Fonte: Moreira (2022).

### 2.1.5 Importância econômica

Assim como as várias espécies da Caatinga, a *S. tuberosa* tem importância social e econômica, pois é utilizada como fonte alimentar e de renda extra para famílias rurais, mas apenas no período de colheita, a qual é restrita em alguns meses. No somatório dos anos de 2017, 2018 e 2019, a extração do umbu chegou a um total de 23.777 toneladas, acumulando um valor de produção de R\$ 27 milhões. Fora do Brasil, a fruta do umbuzeiro atinge comercialmente lugares como a Europa, para o qual é vendido para o consumo, sendo conhecido por seu sabor exótico azedo-ácido, o qual é o motivo de tanta solicitação da fruta. Esse sabor se deve às combinações de açúcares, acidez e compostos fenólicos que são produzidos na polpa. A fruta é importante na dieta nutricional devido seu teor de vitamina C, carotenóides, minerais presentes e até nutrientes fenólicos (LIMA; SILVA; OLIVEIRA, 2018; MITCHELL; DALY, 2015; SALES, 2020).

### 2.1.6 Valores nutricionais

Sendo o teor de vitamina C muito importante, ela se apresenta com uma média de 15 mg de ácido ascórbico por 100 g de polpa. Além disso, em 100 g de parte comestível do umbu, o fruto apresenta uma composição nutricional de 37 kcal, 0,80 g de proteína, contém 9,40 g de carboidratos, fibra alimentar de 2,0 g, ainda possui alguns minerais como o Cálcio (12 mg), o Magnésio (0,03 mg), o Fósforo (13g), o Ferro (0,10 mg), o Potássio (152 mg), o Zinco (0,40 mg), vitaminas como a Tiamina (0,05 mg), a Riboflavina (0,06 mg) e Piridoxina (0,03 mg). Porém, esses valores podem ser diferentes entre os genótipos da espécie, os frutos também podem sofrer influência nos valores de pH, na relação de sólidos solúveis/acidez titulável e açúcar total. Entretanto, não foi registrado variação nos valores de pH em estádios

de maturação dos frutos (NEPA, 2011; DIAS *et al.*, 2019; IBGE, 2019; MOREIRA *et al.*, 2021).

O fruto umbu é mais conhecido por seu consumo em forma de alimento. O estudo de Paodjuenas (2018) mostrou que nas comunidades rurais do semiárido paraibano, o umbu é consumido *in natura* pela maior parte das pessoas, seguido depois de outras formas de alimento como doce, suco e picolé. Outras partes do umbu também são usadas, como o tubérculo para a produção de cocada. As folhas e frutos do umbu são usados também para forragem e sua madeira para preparo de lenha e carvão como forma de combustível. E, ainda, o umbu também é utilizado para o preparo de artefatos de carroças, para fabricação de cachimbo e de barril.

Existem algumas implicações quanto à utilização do umbu. O manejo do fruto é feito basicamente por pequenos agricultores familiares, por isso, não existe muito conhecimento e técnicas de manejo que favoreçam a produção e conservação após a colheita, o que acarreta na redução da população dessa espécie afetando, então, a vegetação nativa e a variabilidade genética que acaba se reduzindo (LINS NETO; PERONI; ALBUQUERQUE, 2010; MERTENS *et al.*, 2017; LIMA; SILVA; OLIVEIRA, 2018).

#### 2.1.7 Importância no ecossistema

A espécie *S. tuberosa* não consiste apenas em seu valor nutricional e importância comercial, mas também por sua importância no ecossistema, já que ela serve também como alimento para os polinizadores e dispersantes e também serve como forragem para os animais ruminantes domésticos. E, apesar de todas essas vantagens que a espécie traz para o agricultor local, animais e natureza, ela enfrenta ameaças provocadas pelo homem, devido à destruição do habitat. Há, então, a suspeita de que o umbuzeiro seja extinguido. Mesmo que ela ainda não seja considerada ameaçada de extinção, suspeita-se de que a espécie sofre regeneração reduzida, o que pode levar à diminuição da população. Por isso, a espécie também precisa de medidas de proteção e estratégias de conservação (MERTENS *et al.*, 2017).

#### 2.1.8 Variabilidade genética

De acordo com as características do fenótipo, a árvore umbu, enquanto sua variabilidade genética, parece estar distribuída uniformemente na região semiárida brasileira, porém, esse fato não é verdadeiro. As barreiras geográficas ou até as condições edafoclimáticas podem limitar o cruzamento e a frequência dos alelos entre as populações

(figura 2). Portanto, o umbu não é distribuído de forma uniforme no semiárido, em algumas regiões a população do umbu pode ser exclusiva, enquanto em outras elas pertencem quase ao mesmo grupo (SANTOS *et al.*, 2021; LIMA; SILVA; OLIVEIRA, 2018; MITCHELL; DALY, 2015).

A estrutura das populações e a distribuição genética dentro da espécie de *S. tuberosa* podem ter sido delineadas devido às oscilações climáticas do bioma da Caatinga, por isso, os estudos genéticos visam compreender esses eventos, tentando colaborar em sua preservação. Através de ferramentas usadas na genômica populacional é possível observar diferentes regiões existentes no genoma de um organismo. Em Alagoas, mostrado no estudo de Balbino (2018), o genótipo da espécie *S. tuberosa* era geneticamente diferente do genótipo de outras populações da Bahia, Minas Gerais e Pernambuco, sendo que as espécies dessas regiões eram claramente próximas. Esse estudo indicou a baixa diversidade genética da espécie através dos resultados oriundos do nSSR (repetições de sequência simples do núcleo), que demonstrou que dos seis loci, vinte e dois alelos foram detectados, cada *locus* variando de 2 a 5 alelos, corroborando com o estudo de Santos *et al.* (2021), o qual aponta que a diversidade genética de *S. tuberosa* não está distribuída uniformemente na região do semiárido.

A fim de aprimorar o conhecimento acerca das relações filogenéticas de *S. tuberosa* com outras espécies de plantas, bem como identificar possíveis marcadores para esforços de reprodução e conservação, a composição genética desta planta tem sido objeto de estudo. Uma pesquisa recente analisou a diversidade genética de *S. tuberosa*, valendo-se de marcadores microssatélites. Os resultados indicaram que as populações selvagens de *S. tuberosa* apresentam níveis mais elevados de diversidade genética do que as populações cultivadas, sugerindo que a domesticação pode ter resultado em perda de variabilidade genética. Outro estudo, por sua vez, utilizou técnicas de sequenciamento de DNA para investigar as relações filogenéticas entre espécies de Spondias. De acordo com os resultados obtidos, *S. tuberosa* está intimamente relacionada com *S. radlkoferi* e *S. mombin*, ambas nativas do Brasil (FRANCO *et al.*, 2015; MARTINS, 2019; SILVA JÚNIOR *et al.*, 2019).

A espécie *S. tuberosa* tem um padrão geral de habitat em quase todo o território do sertão do São Francisco, em áreas de maior adaptabilidade em decorrência dos fatores abióticos. Sua maior adaptabilidade surge em regiões do leste do Brasil, correspondendo ao norte da Bahia e ao sul de Pernambuco, em outras áreas a espécie tem menor adaptabilidade para sua ocorrência (XAVIER, 2020). Como também mostra o estudo de Santos e Nascimento (1998), a perda da variabilidade genética do umbuzeiro vem sendo questionada há décadas. Todos esses estudos sugeriram que haja uma estratégia para a preservação dessa espécie, pois

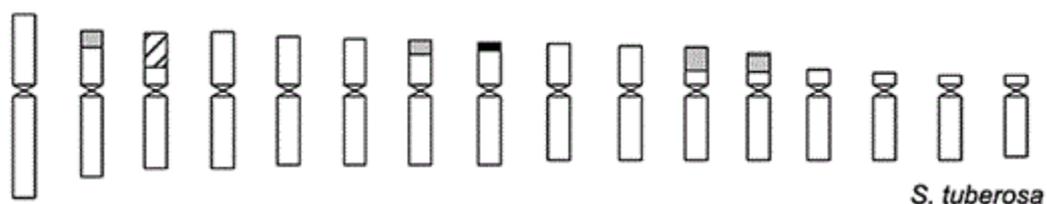
elas são muito ameaçadas de exploração, perda de habitat, expansão agrícola e atividades antrópicas, sua conservação natural também é importante, já que a baixa diversidade genética se estende por toda a Caatinga.

Além das pesquisas já realizadas, foi realizado o sequenciamento do genoma de *S. mombin*, trazendo informações valiosas sobre a estrutura genômica e evolução deste gênero. Embora não tenha sido focado especificamente em *S. tuberosa*, este estudo contribui para o conhecimento geral da espécie. Outra pesquisa utilizou marcadores RAPD para avaliar a diversidade genética de espécies de *Spondias* no nordeste do Brasil, e constatou-se que *S. tuberosa* apresenta baixos níveis de diversidade genética em comparação com outras espécies da região. Essa informação pode ter implicações importantes para a conservação da espécie. Estudos de revisão oferecem uma visão abrangente da composição fitoquímica e dos possíveis benefícios à saúde dos frutos de *Spondias*, incluindo a *S. tuberosa* (RAMOS *et al.*, 2011; PEREIRA *et al.*, 2018; MARTINS, 2019; NUNES *et al.*, 2020).

Embora a genética não seja o foco principal, é enfatizada a importância de compreender a composição genética das espécies de *Spondias* para maximizar seu potencial como fontes de nutrição e medicina. Esses estudos destacam a relevância da pesquisa genética para entender a biologia e os possíveis usos da *S. tuberosa* e outras espécies relacionadas. Em suma, este artigo oferece informações valiosas para profissionais que buscam aprofundar seus conhecimentos sobre essas espécies (RAMOS *et al.*, 2011; PEREIRA *et al.*, 2018; MARTINS, 2019; NUNES *et al.*, 2020).

O cariótipo de *S. tuberosa* apresenta 16 pares de cromossomos (Figura 2). São cromossomos pequenos no geral, apresentando apenas um par de tamanho maior. No geral, 11 pares são cromossomos metacêntricos a submetacêntricos possuindo tamanho intermediário e quatro pares são cromossomos pequenos e subtelo-cêntricos (ALMEIDA *et al.*, 2007).

Figura 2 - Representação gráfica do cariótipo da espécie *S. tuberosa*.



Fonte: Almeida *et al.* (2007).

A diversidade genética é uma das características mais importantes do genoma de *S. tuberosa*. Por meio de marcadores moleculares, como microssatélites e SNPs, diferentes

acessos da espécie foram coletados em várias regiões do Brasil e comparados. Os resultados desses estudos revelaram um alto nível de variação genética dentro da espécie, que pode ser explorada em programas de melhoramento para aumentar a produtividade, qualidade dos frutos e resistência a doenças. Esta descoberta é um passo importante para aprimorar a produção de *S. tuberosa* e garantir um futuro saudável para esta espécie valiosa (SANTOS *et al.*, 2020; MARTINS, 2019; MARTINS *et al.*, 2019).

### 2.3 Transcriptoma

O transcriptoma foi descrito por Velculescu *et al.* (1997) como a representação dos genes expressos dentro de um organismo. Segundo os autores, essa expressão pode transmitir a identidade de cada gene expresso e também o nível de expressão para uma população definida de células. Os autores também fazem uma comparação entre o genoma e o transcriptoma, sendo a diferença que o genoma é estático enquanto o transcriptoma pode ser moldado por fatores internos ou externos. Além disso, o genoma e o transcriptoma são ligados de forma dinâmica, já que compõem as características físicas de um organismo. De acordo com Ward; Ponnala e Weber (2012), o transcriptoma, mais especificamente, mostra a expressão gênica de uma célula ou tecido no momento em que o RNA total foi fornecido, representando não somente os genes que estão ativos, mas também a combinação das sequências isiformes ao todo.

A análise do transcriptoma é mais rápida do que a sequência completa do genoma, pois o transcriptoma em plantas, por exemplo, precisa apenas de um alvo pequeno a ser caracterizado, mas que compreende riqueza em informações. Através do transcriptoma é possível observar, por exemplo, a variação na expressão gênica que foi induzida devido às mudanças nas condições ambientais antes de sofrer estresse biótico ou abiótico, também os estágios de desenvolvimento de um vegetal, como o crescimento ou floração, além dos tipos de tecido, raízes ou folhas. Essas análises são feitas a partir da produção de bibliotecas de transcrição utilizando ferramentas apropriadas para produzi-las (WARD; PONNALA; WEBER, 2012; SHAW, TIAN; SHU, 2021).

A análise do transcriptoma não era tão realizada e foi lenta para se desenvolver no meio científico, porém ela se tornou mais viável devido à redução do custo do sequenciamento de alto rendimento. Agora, esta análise é possível em qualquer organismo, oferecendo métodos para a investigação da expressão gênica em tecidos específicos e são mais acessíveis (WARD; PONNALA; WEBER, 2012; SHULSE *et al.*, 2019).

Existem estudos que usa a análise do transcriptoma para identificar diferenças na expressão gênica em diferentes tratamentos, como no caso do estudo de Gamez *et al.* (2019), o qual a banana foi alvo da pesquisa e onde elas foram postas em tipos diferentes de tratamento para entender melhor as interações diretas e indiretas entre as raízes do alvo do estudo e as rizobactérias. Por fim, o estudo mostrou que os genes diferencialmente expressos eram crescentes para uma espécie de banana (*Pseudomonas fluorescens*), ao contrário do que foi visto na outra espécie (*Bacillus amyloliquefaciens*). A pesquisa é importante para entender as diferenças entre os tratamentos e como a expressão gênica pode mudar de espécie para espécie.

Essas diferenças entre tratamentos são importantes pois todo o genoma pode ser alterado, assim também pode ocorrer em caso da dosagem de nutrientes. Alterar as doses, colocando-as em abundância, por exemplo, em caso de doses abundantes de nitrogênio (N) nas raízes de *Arabidopsis*, pode perturbar as taxas de máximas de resposta transcriptômica em todo o genoma, assim como a taxa de crescimento da planta (SWIFT *et al.*, 2020). Mesmo que o nitrogênio seja um macronutriente essencial, ele limita a produção agrícola, por isso é importante estudá-lo especificamente nas plantas e como ele pode alterar o genoma através da taxa de resposta de transcriptoma e somente com essa análise é possível revelar a eficiência desse nutriente, sua utilização, assimilação e translocação (LIU *et al.*, 2022).

Espera-se para o futuro da análise de transcriptoma mais insights acerca da evolução celular das plantas, especificamente do tipo celular através do mapeamento genético, sendo possível a identificação de traços complexos que são importantes à agronomia e à economia (SHAW; TIAN; SHU, 2021).

## **2.4 Anotação Genômica**

O Projeto Genoma Humano, concluído há 20 anos, ajudou no avanço do estudo da compreensão do DNA, no recurso de manipulações genéticas, edição genômica, experimentação de mutações do DNA, decodificar as bases genéticas de várias doenças, alterações epigenéticas, criação de organismos com capacidade para a produção de remédios, entre outros. Agora é possível sequenciar os genomas de referência em larga escala facilitando esses processos, iniciando a fase da era “pós-genômica”, visando representar a biodiversidade, revolucionando a genômica da conservação (ANTUNES, 2022).

A anotação genômica é baseada em dois tipos: estrutural e funcional. O processo de anotação é caracterizado pela sequência genômica ou proteica, sendo esses processos

dependentes do conhecimento de informática dos pesquisadores e cientistas para conseguir executar as anotações. A anotação possui três caminhos, começando pela identificação de regiões repetitivas das sequências do DNA, depois consiste na realização da predição e identificação de genes codificadores de proteínas, tRNA (RNA transportador) e rRNA (RNA ribossômico); esses dois processos listados fazem parte da anotação estrutural, a próxima etapa da anotação está relacionada a atribuição das funções dos genes, sendo esta a anotação funcional (BALBINOT, 2020).

A etapa de anotação estrutural também é chamada de predição de genes, pois ela prediz a estrutura e a localização dos genes que codificam proteínas, além dos tRNA e rRNA em um genoma estudado. Existem três principais formas de predição de genes, elas são a predição *ab initio* ou também chamada de predição intrínseca; predição baseada em evidência ou método empírico ou também conhecida como predição extrínseca; por fim, a predição combinada, que se utiliza da predição intrínseca e extrínseca unidas (BALBINOT, 2020; LOPES, 2019; EJIGU; JUNG, 2020).

Em questão da análise *in silico* do material genético, a anotação funcional pode ser entendida como um dos mais importantes processos, a qual se baseia na predição da função dos genes, sendo um processo complexo por se utilizar de uma grande variedade de bancos de dados e ferramentas para o enriquecimento da anotação e garantir resultados mais precisos. Existe uma variedade de sequências em bancos de dados a serem analisadas pela anotação funcional na intenção de dar nomes aos produtos dos genes, tornando mais fácil a compreensão dos processos biológicos dos organismos e assim avançar em pesquisas (BALBINOT, 2020).

Algumas ferramentas computacionais de similaridade são usadas na anotação funcional como o *Basic Local Alignment Search Tool* (BLAST) e o HMMER (*Hidden Markov Model*), esses algoritmos são usados para realizarem comparações das sequências que estão sendo pesquisadas com outras depositadas em banco de dados de proteína ou de domínios de proteínas, objetivando-se em encontrar a semelhança, atribuindo funções e significado biológico. Ferramentas e bancos de dados como *Gene Ontology* (GO), InterProt, UniProt, *Clusters of Orthologous Groups* (COG), *KEGG Orthology* e EggNOG são utilizadas para a caracterização de domínios, famílias e busca por grupos de ortólogos, além de detectar similaridade com sequências já caracterizadas e revisadas (BALBINOT, 2020; EJIGU, JUNG, 2020; LOPES, 2019).

Essas novas tecnologias para sequenciar o DNA foram um fator importante para o aumento do conhecimento dos genomas, a bioinformática se sustenta para aprofundar mais

ainda o conhecimento do genoma dos organismos. As propostas atuais se baseiam em sequenciar, montar e anotar o genoma das espécies vegetais em todo o mundo. Por sua vez, o sequenciamento de baixa cobertura se tornou comum para medir a diversidade entre coleções e populações selvagens em muitas espécies, até mesmo para aquelas utilizadas como culturas (KERSEY, 2019).

Toda essa tecnologia avança graças aos estudos com as plantas verdes (Viridiplantae), estas que são consideradas essenciais para a vida, responsáveis pela fotossíntese que gera a maior parte da produção primária do globo terrestre, sendo utilizadas para a nutrição, ração animal, combustível, remédios, entre outros. As plantas verdes possuem aproximadamente 391 mil espécies de plantas terrestres existentes e 8 mil espécies de algas verdes. Seus genomas são diversos, variando aproximadamente em 10 Mb até mais de 100 Gb em tamanho (KERSEY, 2019).

### 3 METODOLOGIA

#### 3.1 Sequenciamento de *Long reads*

##### 3.1.1 Extração de DNA

O DNA de *S. tuberosa* obtido da folha foi isolado numa quantidade de 1,5 g de material com um kit NucleoBond HMW DNA (Macherey Nagel) para DNA de alto peso molecular. A qualidade foi avaliada com um dispositivo de pulso FEMTO (Agilent), o qual permite uma sensibilidade única, capaz de detectar ácidos nucleicos na faixa inferior do femtograma, além de sua capacidade de separar amostras de DNA e RNA. Já a quantidade de DNA foi medida com um fluorômetro Quantus™ (Promega®).

##### 3.1.2 PacBio

O sequenciamento de leituras longas foi usado para a obtenção do genoma usando a tecnologia da Pacific BioSciences (PacBio), o qual realiza o sequenciamento de molécula única em tempo real apresentando uma solução avançada para desafios não resolvidos na pesquisa genômica, transcriptômica e epigenética (RHOADS; AU, 2015).

Uma biblioteca HiFi foi então preparada de acordo com o manual "Procedimento e lista de verificação - Preparando bibliotecas HiFi SMRTbell® usando SMRTbell Express Template Prep Kit 2.0 " com uma fragmentação inicial de DNA por g-Tubes (Covaris) e classificação final do tamanho da biblioteca em frações definidas por SageELF (Sage Science). A distribuição de tamanho foi novamente controlada por pulso FEMTO (Agilent). As bibliotecas selecionadas por tamanho foram então sequenciadas em um dispositivo Sequel II com o kit Binding 2.0 e o Sequel II Sequencing Kit 2.0 por 30 h (PacBio).

##### 3.1.3 Omni-C

O kit Dovetail® Omni-C® foi escolhido já que ele fornece dados uniformes de Hi-C com cobertura por base, aproximando-se de bibliotecas com alto grau de informações de longo alcance. A biblioteca foi preparada a partir de 0,5 g de entrada de material de peso fresco. Todos os tratamentos seguiram as recomendações do fabricante do kit para plantas (Omni-C®, Dovetail®). Como etapa final, uma biblioteca compatível com Illumina foi preparada (Dovetail®) e sequenciada de 2 x 150 bp de extremidade emparelhada em um dispositivo HiSeq 3000 (Illumina) para *S. tuberosa*. Alternativamente, a biblioteca de *S.*

*tuberosa* foi sequenciada de 2 x 150 pb de extremidade emparelhada usando a tecnologia DNBseq (BGI Genomics, Hong Kong).

### 3.3 Obtenção do genoma (Assemble usando PacBio)

#### 3.3.1 Hifiasm

Para a montagem do genoma usando *long reads*, foi usado o software hifiasm (CHENG *et al.*, 2021), disponível em <https://github.com/chhyllp123/hifiasm>. Utilizando para a montagem de *S. tuberosa* com o seguinte comando:

```
hifiasm -o output.asm -t 40 reads.fq.gz.
```

As montagens preliminares foram avaliadas quanto à contiguidade e integridade com BUSCO (SEPPEY *et al.*, 2019).

#### 3.3.2 Omin-c scaffolding

As leituras Dovetail® Omni-C® foram primeiro mapeadas usando BWA (LI; DURBIN, 2009), seguindo o hic-pipeline disponível em <https://github.com/esrice/hic-pipeline>. O scaffolding Hi-C foi realizado usando SALSA2 (GHURYE *et al.*, 2019), disponível em <https://github.com/marbl/SALSA>, com parâmetros padrão. Depois de testar vários valores mínimos de qualidade de mapeamento de alinhamentos bam, o andaime final foi realizado com MAPQ10. Várias rodadas de correção de montagem guiadas por mapas de contato Hi-C e curadoria manual de andaimes foram realizadas para obter as pseudomoléculas.

#### 3.3.3 Análise com Busco

As montagens preliminares foram avaliadas quanto à contiguidade e integridade com BUSCO (SEPPEY *et al.*, 2019). BUSCO (versão 5.1.2.) (MANNI *et al.* 2021) foi usado para avaliar a completude do espaço gênico da montagem da pseudomolécula e a anotação estrutural do gene com o banco de dados 'viridiplantae\_odb10' contendo 425 genes de cópia única.

### 3.4 Anotação estrutural de DNA repetitivo

O genoma foi identificado usando a ferramenta RepeatMasker e RepeatModeler. A ferramenta RepeatMasker usa uma variedade de algoritmos para identificar repetições com base na similaridade de sequência e recursos estruturais. Depois que as repetições são identificadas, o RepeatMasker pode mascará-las substituindo a sequência de repetição por um caractere de espaço reservado, como um "N". A ferramenta RepeatModeler identifica famílias repetidas, que depois de identificadas são então refinadas e classificadas em subfamílias (TARAILO-GRAOVAC; CHEN, 2009).

### 3.5 Anotação estrutural de genes

A montagem do genoma de *S. tuberosa* foi realizada em três ferramentas diferentes usando preditores de genes *ab initio*, Augustus (HOFF; STANKE, 2019) e GeneMarker (TER-HOVHANNISYAN *et al.* 2008), e de inteligência artificial, Helixer (HOLST *et al.*, 2023). Todas as anotações foram submetidas para análise de precisão de montagem do BUSCO, que é uma ferramenta que compara uma montagem de genoma e fornece informações sobre a porcentagem de genes completos, fragmentados, ausentes e duplicados.

### 3.6 Anotação funcional

A anotação do genoma com maior integridade fornecida pela BUSCO foi escolhida para fornecer o genoma funcional sob utilização do programa Blast2GO (CONESA; GÖTZ, 2008), usando as ferramentas Blast (BORATYN, 2013), InterProScan (PAIZAN-LAFOSSE *et al.*, 2022), EggNOGG (HERNANDEZ-PLAZA *et al.*, 2023), KOALA (KANEHISA; SATO; MORISHIMA, 2016) e Pannzer (TÖRÖNEN; HOLM, 2021).

As análises estatísticas foram obtidas pela ferramenta *charts* no programa Blast2GO. Com isso, os gráficos com os resultados foram levantados pelo programa Blast2GO usando a ferramenta *graphs*, combinando os gráficos gerados pelas análises estatísticas. Os resultados para as funções dos genes foram gerados em gráficos com os termos em Genes Ortólogos (GO) e classificados em categorias e distribuição de enzimas.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Montagem do genoma

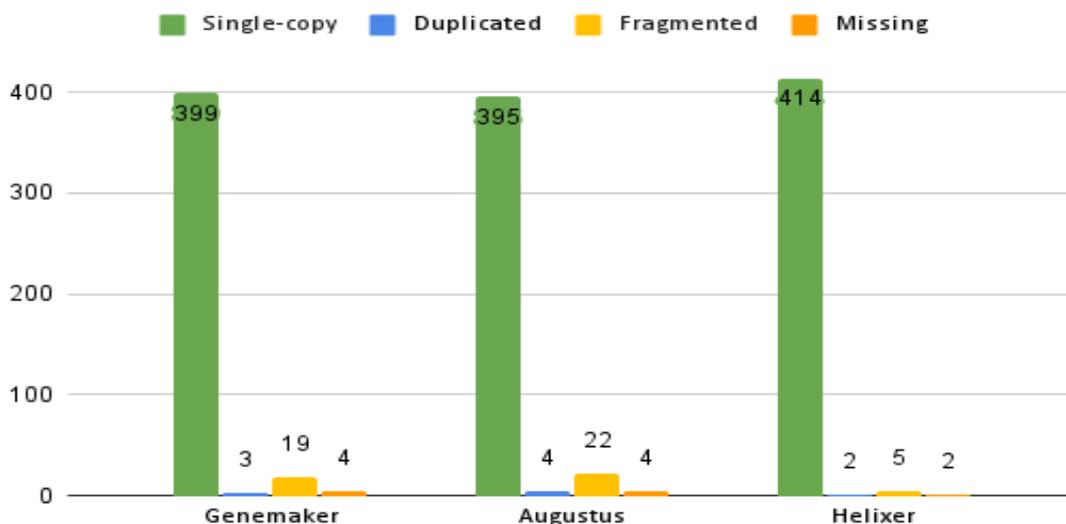
O genoma de referência foi construído em escala cromossômica para a espécie de *S. tuberosa* usando o sequenciamento PacBio HiFi e Dovetail Omni-C nos seus 16 pares de cromossomos.

Os *reads* de *S. tuberosa* tiveram uma completude de 1.800.482 *reads*, gerando um total de 29,3 G bases obtidos na plataforma Pacbio, com um comprimento máximo de leitura de 102 pb e mínimo de 50 kb, apresentando uma média de 16,3 kb de comprimento de leituras. A montagem por *long reads* gerou um tamanho de genoma para a espécie *S. tuberosa* de ~ 30,8 Mb, correspondendo a uma cobertura de 94,9%.

### 4.2 Anotação funcional

Na avaliação de precisão do BUSCO para o conjunto de dados montados do genoma do umbuzeiro usando os resultados da montagem feita na ferramenta Helixer, Augustus e GeneMarker. Em um total de 425 BUSCOs, o preditor Helixer obteve um total de 418 BUSCOs completos (incluindo cópias únicas e duplicadas), enquanto os preditores Augustus e Genemarker obtiveram 399 e 402 BUSCOs completos, respectivamente (figura 3).

Figura 3 - Dados do BUSCO para a precisão da montagem do genoma de *S. tuberosa*. dos preditores Helixer, Augustus e GeneMarker.



Fonte: BUSCO (2021).

O software que mais teve precisão na montagem foi o Helixer por apresentar 418 BUSCOs completos (98%) de 425 totais, sendo 97% de cópias únicas, 0,007% ortólogos duplicados e 1,18% de fragmentos. Além disso, uma porcentagem de 0,004% foram de BUSCOs perdidos na montagem.

O Helixer é uma ferramenta mais sensível e utiliza comprimentos de entradas mais longos, sendo melhor para a predição de genomas maiores. Em confronto, os dados de previsão de Helixer são mais concretos com os dados de RNAseq do que os dados obtidos por Augustus quando comparados (STIEHLER, 2020). Além de que modelos mais recentes de Helixer tem apresentado maior desempenho em relação a outros preditores e com versões anteriores da própria ferramenta. Ainda assim, Holst *et al.* (2023) recomendam que os pesquisadores testem outros modelos e selecionem o mais acurado para a espécie modelo, o que foi feito no presente estudo.

O conjunto de genes previstos foram anotados em termos de Ontologia Genética (GO) para classificação funcional. Os genes previstos no genoma de *S. tuberosa* foram atribuídos a 51.896 termos GO totais, devido a algumas sequências se atribuírem em vários termos GO. Os termos GO foram encontrados e divididos em três categorias principais, sendo elas: função molecular, processo biológico e componente celular.

#### 4.2.1 Processos biológicos

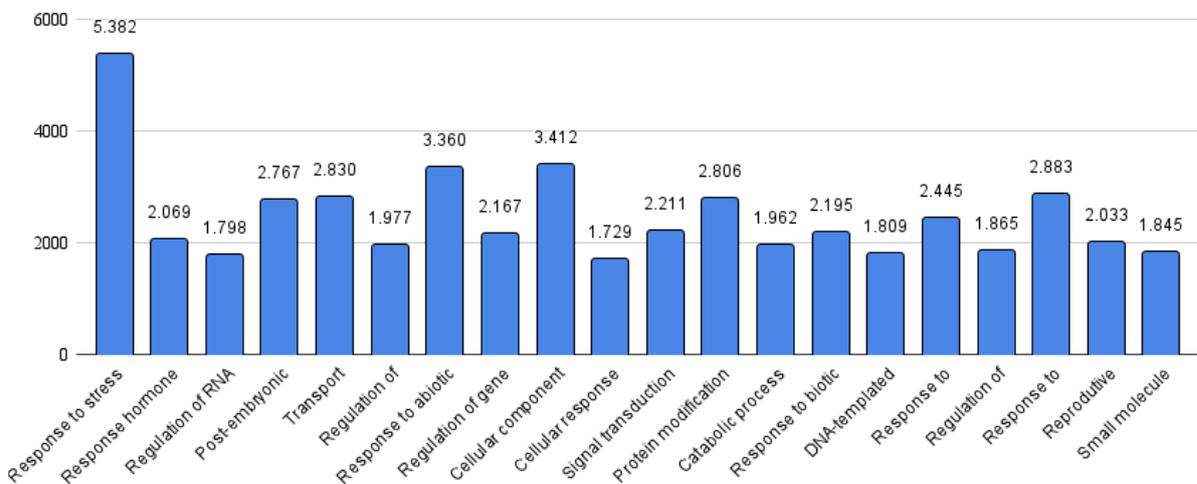
O total de genes que estão envolvidos na categoria processo biológico corresponderam a 17.213 genes, distribuídos em 21 tipos de genes para esta categoria. Os termos GO para “resposta ao estresse” estiveram em maior quantidade no genoma funcional de *S. tuberosa* para esta categoria, sendo um total de 5.382 genes, correspondendo a 10,48% de todos os genes envolvidos no processo biológico dessa espécie. Porém, quando somado com todos os genes relacionados ao estresse, sendo esses o terceiro e o décimo termos GO com maior quantidade de genes, inclui-se “resposta ao estímulo abiótico” (3.360 genes – 6,54%) e “resposta ao estímulo biótico” (2.195 genes – 4,27%), a quantidade de genes em termos GO relacionados ao estresse chega a um total de 10.937 genes, representando 21,29% de todos os genes para a categoria de processo biológico.

Em algumas plantas existem vários genes que são específicos para cada especificidade de estresse, enquanto outros genes podem ser compartilhados entre vários tipos de estresses. Os genes em resposta ao estresse abiótico e biótico são geralmente compartilhados, destacando como é complexo a resposta ao estresse e a adaptação das plantas. A tolerância ao

estresse abiótico, a exemplo, começa desde a percepção do estresse pela planta até a ativação transcricional de genes. Muitos genes ainda possuem funções desconhecidas ou não se sabe como funcionam suas interações, mas são regulados pelo estresse abiótico (MANTRI *et al.*, 2012).

O gráfico abaixo mostra os genes em termos GO e suas funções envolvidas na categoria de processo biológico (figura 4).

Figura 4 - Classificação da distribuição de genes envolvidos na categoria processo biológico.



Fonte: A autora (2023).

Outros genes de resposta também estão presentes, como: “resposta ao estímulo externo” com um total de 2.883 (5,61%) genes. “Resposta aos compostos contendo oxigênio”, tendo um total de 2.445 (4,76%). “Resposta hormonal” com 2.069 (4,03%) genes. E, por fim, “resposta celular ao estímulo químico” 1.729 (3,56%).

Os genes relacionados aos processos são: “Processo de modificação da proteína” com 2.806 (5,46%) genes. “Processo catabólico” 1.962 (3,82%). “Processo metabólico de pequenas moléculas” 1.845 (3,59%) genes.

Genes de regulação também estão presentes na categoria de processos biológicos, sendo esses: “Regulação da expressão dos genes” teve um total de 2.164 (4,22%) genes. “Regulação do processo de biossíntese celular” 1.977 (3,85%) genes. “Regulação do processo de macromolécula biossintética” 1.865 (3,63%) genes.

Duas categorias de genes estão relacionadas ao desenvolvimento, eles são: desenvolvimento pós-embriônico com um total de 2.767 (5,39%) genes e “desenvolvimento da estrutura reprodutiva”, apresentando 2.033 (3,96%) genes.

Outras categorias são: “organização da composição celular” com 3.412 (6,64%) genes totais, sendo esta a segunda maior quantidade de genes que está relacionada aos processos biológicos. “Transdução de sinal” possuindo 2.211 (4,3%) genes e “transcrição de DNA modelo” 1.809 (3,52%) genes.

No estudo de Savadi *et al.* (2022), um draft do genoma funcional de *A. occidentale*, mostrou que o termo GO “resposta ao estímulo” correspondendo à categoria processos biológicos, tem poucos genes envolvidos. Enquanto “processo celular” e “processo metabólico” possui a maior parte dos genes nesta categoria, representando mais de 60% dos genes envolvidos e quase 60%, respectivamente. Mesmo que seja uma espécie da mesma família, os genes relacionados são diferentes do que foi achado para o genoma funcional de *S. tuberosa*.

Já no estudo de Jazi *et al.* (2017), onde o estudo do transcriptoma foi feito com a espécie *P. vera*, uma espécie também pertencente à família Anacardiaceae e com estreita ligação com a espécie de estudo do presente trabalho, vários genes relacionados ao estresse foram encontrados. Isso se deve ao fato de que esta espécie em questão é vista como adaptável a condições estressoras. Ainda assim, a quantidade de genes relacionados ao estresse em comparação ao umbuzeiro é inferior.

Genes envolvidos em regulação também são importantes em espécies como *S. tuberosa*, pois esses genes auxiliam nas vias metabólicas nos diferentes compartimentos celulares para que a célula se adapte às novas condições ambientais da planta (PEIXOTO, 2020).

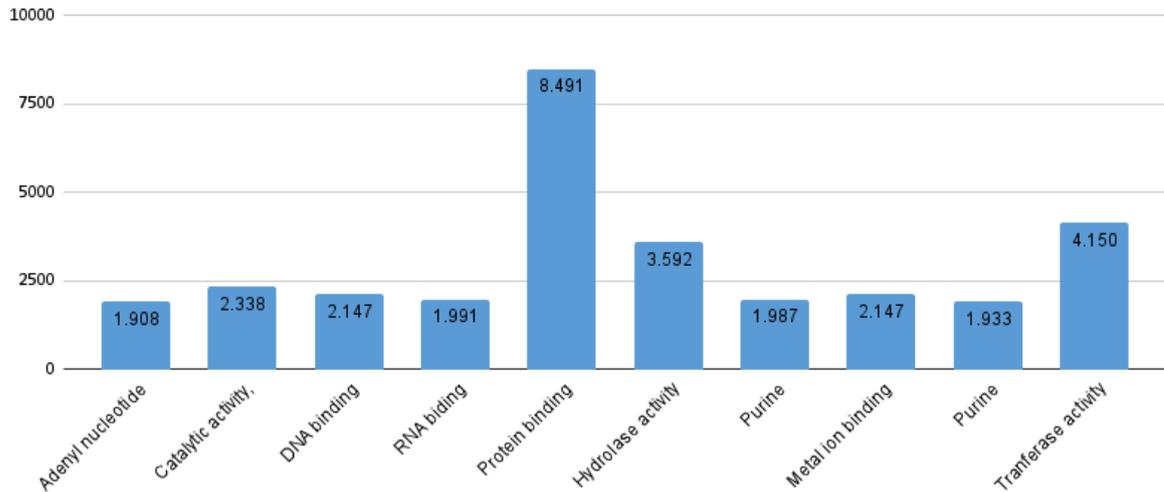
#### 4.2.2 Função Molecular

A função molecular resultou em 18.218 genes ao todo envolvidos nesta categoria, obtendo uma distribuição de 10 conjuntos de genes (figura 5). Desses conjuntos, 8.491 (27,67%) genes estão relacionados à função ligação de proteína, sendo estes o maior conjunto de genes para esta categoria. Em seguida, estão os únicos genes destas categorias envolvidos em atividade de transferase com um total 4.150 (13,52%) genes, atividade de hidrolase (3.592 genes – 11,71%) e de atividade catalítica atuando em uma proteína (3.592 genes – 7,62%).

É possível que essas atividades metabólicas do umbuzeiro possam também estar relacionadas ao estresse, visto que as funções catalíticas são responsáveis por propriedades bioquímicas de proteínas universais do estresse. Essas proteínas são permitidas para atuar em

transdutores de sinalização celular e reguladores metabólicos. Elas também proporcionam resistências ao estresse, porém essa função é pouco estudada (CHI *et al.*, 2019).

Figura 5 - Classificação da distribuição de genes envolvidos para a categoria função molecular.



Fonte: A autora (2023).

Na função molecular, a maior parte dos termos GO para esta categoria é composta basicamente por atividades de ligação, além da ligação de proteína com a maior quantidade, em ordem crescente, são eles: “ligação de íon-metálico” (2.147 genes – 7%), “ligação de DNA” (2.147 genes – 7%), “ligação de RNA” (1.991 genes – 6,49%), “ligação de ribonucleotídeo purina” (1.987 genes – 6,47%), “ligação de trifosfato de ribonucleotídeo purina” (1.933 genes – 6,3%), “ligação de nucleotídeo adenil” (1.908 genes – 6,22%).

Os resultados encontrados para a categoria função molecular foi semelhante ao de Savadi *et al.* (2017) para o genoma funcional de *A. occidentale*, onde a maior quantidade de genes é relacionada à ligação, mais de 60% dos genes, e atividade catalítica representando pouco menos de 60% dos genes.

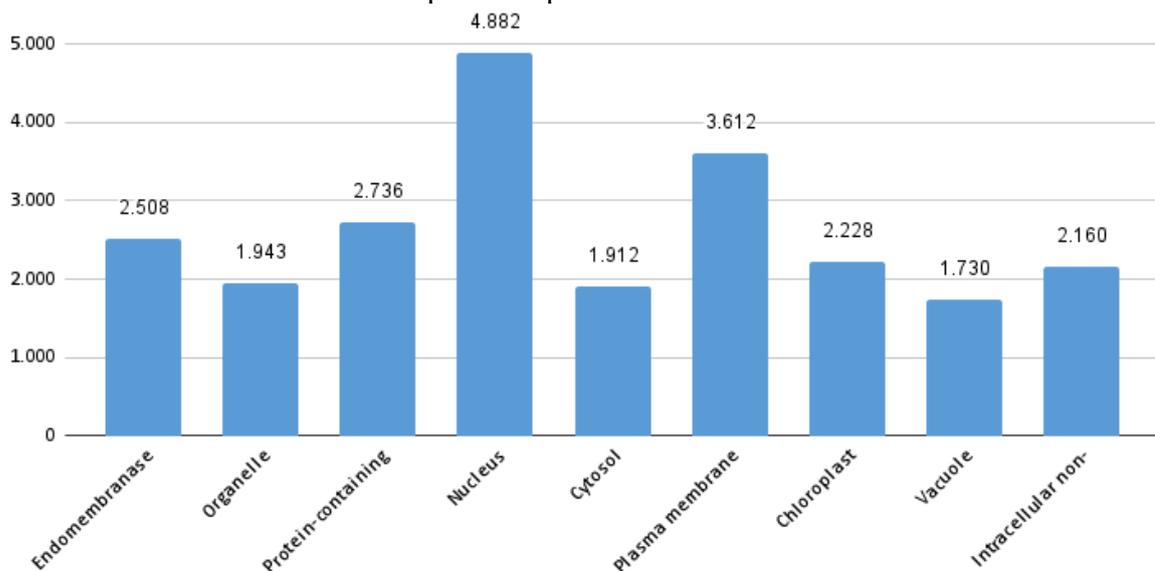
A quantidade de genes para atividade hidrolase em *S. tuberosa* também foi uma das maiores. Isso porque, no déficit hídrico, pode haver aceleração de hidrólise das proteínas para formar aminoácidos livres, os quais auxiliam também no aumento da reserva de carboidratos. Além de que, as enzimas ajudam na redução da energia de ativação, ou seja, evita a perda de energia que é importante ser mantida em uma planta que enfrenta frequentemente condições estressoras (PEIXOTO, 2020).

### 4.2.3 Componente celular

Os genes funcionais envolvidos no componente celular corresponderam a um conjunto de 16.465 genes, divididos em 9 conjuntos de genes relacionados a esta categoria (figura 7). A maior quantidade de genes funcionais está relacionada ao núcleo, sendo 4.882 (20,59%) genes, seguido pela membrana plasmática com 3.612 (15,23%) genes.

Assim como em todos os organismos vivos, o núcleo é importante por armazenar as informações genéticas que são responsáveis pela regulação do metabolismo, do crescimento e também da diferenciação celular. Não se mostrando diferente no genoma funcional de *S. tuberosa*. No núcleo acontece os processos de leitura de DNA celular para que seja transcrito como RNA mensageiro e, assim, ocorra a codificação de proteínas (TAIZ *et al.*, 2020).

Figura 6 - Gráfico demonstrando a classificação em porcentagem da distribuição de genes para componente celular.



Fonte: A autora (2023).

Outros componentes celulares anotados e classificados em termos GO envolvidos nesta categoria foram: “complexos contendo proteína”, envolvendo um total de 2.736 (11,54%) genes. “Sistema de endomembrana” envolvendo 2.508 (10,58%) genes. “Cloroplasto” com 2.228 (9,4%) genes ao todo. “Organela intracelular não limitada por membrana” com 2.160 (9,11%) genes. 1.943 (8,19%) dos genes nesta categoria se inserem no termo “membrana da organela”. Nos termos GO “citossol”, 1.912 (8,06%) genes do componente celular estão envolvidos. E, por fim, o “vacúolo” representando 1.730 (7,3%) genes.

Pode-se observar pela quantidade dos genes funcionais envolvidos em suas funções metabólicas de que o umbuzeiro é uma espécie altamente adaptada ao seu habitat. Isso é constatado também nas literaturas onde revelam a dinâmica ecofisiológica do umbuzeiro em época de seca, corroborando com os resultados do presente estudo. Por exemplo, de acordo com o que Ferri (1978) explica, ao primeiro sinal de déficit hídrico, os pequenos estômatos do umbuzeiro se fecham como um mecanismo de proteção. Como a planta tem cutículas delgadas, as transpirações observadas depois do estômato fechado são consideradas condutância cuticular.

Mesmo quando o umbuzeiro tem grande disponibilidade de água, esta árvore mantém-se sob economia, isso porque ela exerce um controle rígido sobre a transpiração, tendo estratégias para se desenvolver em seu habitat natural. Além disso das características morfoanatômicas que o ajudam a viver sob condições estressantes, o umbuzeiro possui túberos onde armazena água (LIMA FILHO; SILVA, 1988).

Assim, pode-se entender que, quando se inicia a época de seca, além da abscisão das folhas, a condutância estomática do umbuzeiro é reduzida, diminuindo também a transpiração da planta. Porém, como possui água em suas túberas, o equilíbrio é mantido. Todavia, quando se inicia a época chuvosa, esse equilíbrio é rompido, pois a superfície de transpiração aumenta e a planta não consegue manter o controle entre a absorção e transpiração (EMBRAPA, 2011).

## 5 CONCLUSÃO

O presente estudo mostrou os genes em termos de Ortologia Genética (GO) do genoma funcional da espécie *S. tuberosa*. Com isso, foi possível observar a quantidade de genes existentes em cada categoria, além de que essa espécie apresenta uma grande quantidade de genes relacionados à resposta ao estresse e atividades catalíticas que reduzem a energia de ativação para o rompimento de ligação de moléculas.

## REFERÊNCIAS

- AGUNG, P. P. *et al.* Genetic diversity of Indonesian cattle breeds based on microsatellite markers. **Asian-Australasian Journal of Animal Sciences**, v. 32, n. 4, p. 467, 2019. DOI: <https://doi.org/10.5713/ajas.18.0283>.
- ALMEIDA, C. C. de S.; CARVALHO, P. C. de L.; GUERRA, M. Karyotype differentiation among *Spondias* species and the putative hybrid Umbu-cajá (*Anacardiaceae*), **Botanical Journal of the Linnean Society**, v. 155, n. 4, p. 541–547, 2007. DOI: <https://doi.org/10.1111/j.1095-8339.2007.00721.x>.
- ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, 1990.
- ALVES, J. J. A. Geocologia da Caatinga no Semi-árido do Nordeste brasileiro. **Climatologia e Estudos da Paisagem**, v. 2, n.1, p. 58, 2007.
- ALVES, R. E.; BRITO, E. S.; CAVALCANTE, I. H. L. Composição química do umbu (*Spondias tuberosa* Arruda) in natura e processado em calda. **Revista Brasileira de Fruticultura**, v. 29, n.2, p. 345-348, 2007.
- ALVI, A. F. *et al.* The emerging key role of reactive sulfur species in abiotic stress tolerance in plants. **Physiologia Plantarum**, v. 175, n. 3, p. e13945, 2023. DOI: <https://doi.org/10.1111/ppl.13945>
- ANTUNES, A. Avanços da genômica e os desafios na decodificação da vida. **Revista Ciência Elementar**, v. 10, n. 04, p. 056, 2022.
- BALBINO, E. P. **Análise filogeográfica em *Spondias tuberosa* (Arruda Câmara)**. 2018. 52f. Dissertação (Mestrado em Agricultura e Ambiente) - Universidade Federal de Alagoas, Campus Arapiraca, Arapiraca, 2018.
- BALBINOT, E. **Workflow científico de anotação genômica funcional e curadoria manual de genomas**. 2020. 83f. Dissertação (Mestrado em Biotecnologia) - Universidade de Caxias do Sul, Caxias do Sul, 2020.
- BORATYN, G. M. *et al.* BLAST: a more efficient report with usability improvements. **Nucleic acids research**, v. 41, p. W29–W33, 2013. DOI: <https://doi.org/10.1093/nar/gkt282>.
- BYRNE, A. *et al.* Realizing the potential of full-length transcriptome sequencing. **Philosophical Transactions of the Royal Society B**, v. 374, n. 1786, p. 20190097, 2019. DOI: <https://doi.org/10.1098/rstb.2019.0097>.
- CAMACAM, B. L. M.; MESSIAS, C. M. B. de O. Potencial alimentar de frutas e plantas da caatinga: revisão integrativa. **Research, Society and Development**, v. 11, n. 9, p. e39911931997-e39911931997, 2022.
- CATLIN, N. S.; JOSEPHS, E. B. The important contribution of transposable elements to phenotypic variation and evolution. **Current Opinion in Plant Biology**, v. 65, p. 102140, 2022. DOI: <https://doi.org/10.1016/j.pbi.2021.102140>.

- CHENG, H. *et al.* Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. **Nat Methods**, v. 18, p. 170–175, 2021. DOI: <https://doi.org/10.1038/s41592-020-01056-5>.
- CHI, Y. H. *et al.* The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. **Frontiers in Plant Science**, v. 10, p. 750, 2019.
- CONESA, A.; GÖTZ, S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. **International Journal of Plant Genomics**, v. 2008, p. 619832, 2008. DOI: <https://doi.org/10.1155/2008/619832>
- DANECEK, P. *et al.* The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156–2158, 2011.
- DIAS, J. L. *et al.* Extraction of umbu (*Spondias tuberosa*) seed oil using CO<sub>2</sub>, ultrasound and conventional methods: evaluations of composition profiles and antioxidant activities. **The Journal of Supercritical Fluids**, v. 145, p. 10-18, 2019.
- EJIGU, G. F.; JUNG, J. Review on the computational genome annotation of sequences obtained by next-generation sequencing. **Biology**, v. 9, n. 9, p. 295, 2020.
- EMBRAPA. **Ecofisiologia do umbuzeiro** (*Spondias tuberosa* Arr. Cam.). José Moacir Pinheiro Lima Silva (org.). Petrolina, PE: Embrapa Semiárido, 2011.
- FESCHOTTE, C.; PRITHAM, E. J. DNA transposons and the evolution of eukaryotic genomes. **Annual Review of Genetics**, v. 41, p. 331–368, 2007. DOI: <https://doi.org/10.1146/annurev.genet.40.110405.090448>.
- FERRI, M. G.; LABOURIAU, L. G. Water balance of plants from the "Caatinga". I. Transpiration of some of the most frequent species of the "Caatinga" of Paulo Afonso (Bahia) in the rainy season. **Revista Brasileira de Biologia**, Rio de Janeiro, v.3, n.12, p.301-312, 1953.
- FLYNN, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. **Proceedings Of The National Academy of Sciences - PNAS**, v. 117, n. 7, p. 9451-9457, 2020.
- FRANCO, R. F. *et al.* Diversidade genética de populações selvagens e cultivadas de *Spondias tuberosa* (Anacardiaceae) revelada por marcadores microssatélites." **Genética e Biologia Molecular**, v. 38, n.1, p. 94-100, 2015.
- GAMEZ, R. M. *et al.* Banana (*Musa acuminata*) transcriptome profiling in response to rhizobacteria: *Bacillus amyloliquefaciens* Bs006 and *Pseudomonas fluorescens* Ps006. **BMC genomics**, v. 20, n. 1, p. 1-20, 2019. DOI: <https://doi.org/10.1186/s12864-019-5763-5>.
- HAYWARD, A.; GILBERT, C. Transposable elements. **Current Biology**, v. 32, n. 17, p. R904-R909, 2022. DOI: <https://doi.org/10.1016/j.cub.2022.07.044>.

- HOFF, K. J.; STANKE, M. Predicting genes in single genomes with AUGUSTUS. **Current Protocols in Bioinformatics**, v. 65, n. 1, p. e57, 2019. DOI: <https://doi.org/10.1002/cpbi.57>.
- HOLST, F. *et al.* Helixer – *de novo* prediction of primary eukaryotic gene models combining deep learning and a hidden markov model. **BioRxiv**, v. 1, 2023. DOI: <https://doi.org/10.1101/2023.02.06.527280>.
- HURTADO, I. Poisonous Anacardiaceae of South America. **Clinics in Dermatology**, v. 4, n. 2, p. 183-190, 1986. DOI: [https://doi.org/10.1016/0738-081X\(86\)90077-5](https://doi.org/10.1016/0738-081X(86)90077-5).
- IBGE. **Produção da extração vegetal e silvicultura 2018**. Rio de Janeiro: IBGE, 2019.
- KANEHISA, M.; SATO, Yoko; MORISHIMA, Kanae. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. **Journal of Molecular Biology**, v. 428, n. 4, p. 726-731, 2016. DOI: <https://doi.org/10.1016/j.jmb.2015.11.006>
- KAVAMURA, V. N. *et al.* Water regime influences bulk soil and rhizosphere of *Cereus jamacaru* bacterial communities in the Brazilian Caatinga biome. **PloS one**, v. 8, n. 9, p. e73606, 2013.
- KERSEY, P. J. Plant genome sequences: past, present, future. **Current Opinion in Plant biology**, v. 48, p. 1-8, 2019.
- KLEIN, C.; AGNE, S. A. A. Phosphorus: from the nutrient to pollutant!. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**, v. 8, n. 8, p. 1713–1721, 2013. DOI: 10.5902/223611706430.
- KOREN, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. **Genome Research**, v. 27, n. 5, p. 722-736, 2017.
- KUMAR, M.; RANI, K. Epigenomics in stress tolerance of plants under the climate change. **Molecular Biology Reports**, v. 50, n. 7, p. 6201–6216, 2023. DOI: <https://doi.org/10.1007/s11033-023-08539-6>.
- LANGMEAD, B.; SALZBERG, S. L. Alinhamento de leitura com intervalo rápido com Bowtie 2. **Métodos da Natureza**, v. 9, n. 4, p. 357-359, 2012.
- LIMA NETO, F. P.; SILVA JÚNIOR, J. F.; SILVA JÚNIOR, M. A. Caracterização físico-química do umbu-cajá (*Spondias spp.*) produzido no estado do Ceará e Pernambuco: uma revisão sistemática da literatura científica nacional e internacional dos últimos 15 anos. **Revista Brasileira de Produtos Agroindustriais**, v. 19, n. 2, p. 315-324, 2017.
- LIMA, M. A. C. de; SILVA, S. de M.; OLIVEIRA, V. R. de. Umbu—*Spondias tuberosa*. In: RODRIGUES, Sueli; SILVA, Ebenezer de Oliveira; BRITO, Edy Sousa de (edit.). **Exotic fruits: reference guide**. London: Academic Press, 2018. p. 427-433. ISBN: 978-0-12-803138-4. DOI: <https://doi.org/10.1016/C2014-0-02888-2>.
- LINS NETO, E. M. de F.; PERONI, N.; ALBUQUERQUE, U. P. de. Traditional knowledge and management of Umbu (*Spondias tuberosa*, Anacardiaceae): an endemic species from the

semi-arid region of Northeastern Brazil. **Economic Botany**, v. 64, n. 1, p. 11-21, 2010. DOI: <https://doi.org/10.1007/s12231-009-9106-3>.

LIU, C. *et al.* Comparative transcriptome and physiological analysis unravel proso millet (*Panicum miliaceum* L.) source leaf adaptation to nitrogen deficiency with high nitrogen use efficiency. **Environmental and Experimental Botany**, v. 199, p. 104891, 2022. DOI: <https://doi.org/10.1016/j.envexpbot.2022.104891>.

LIU, J. *et al.* O papel das proteínas 14-3-3 nas vias de sinalização celular e infecção viral. **Revista de Medicina Celular e Molecular**, v. 25, n. 9, p. 4173–4182, 2021. DOI: <https://doi.org/10.1111/jcmm.16490>

LOPES, B. T. da S. **Ab initio gene prediction combined with alignment information**. 2019. 70 p. Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2019.

MANNI, M. *et al.* BUSCO: assessing genomic data quality and beyond. **Current Protocols**, v. 1, n. 12, p. e323, 2021. DOI: <https://doi.org/10.1002/cpz1.323>

MANTRI, N. *et al.* Abiotic stress responses in plants: present and future. *In*: AHMAD, P., PRASAD, M. (ed.). **Abiotic stress responses in plants**. New York, NY: Springer, 2012. p. 1-19. DOI: [https://doi.org/10.1007/978-1-4614-0634-1\\_1](https://doi.org/10.1007/978-1-4614-0634-1_1)

MARTINS, G. *et al.* Complete mitochondrial genomes of the *Spondias tuberosa* Arr. Cam and *Spondias mombin* L. reveal highly repetitive DNA sequences. **Gene**, v. 720, p. 144026, 2019.

MARTINS, G. M. C. **Análises evolutivas em *Spondias* L.:** uma abordagem genômica do gênero. 2019. 103f. Tese (Doutorado em Agronomia) - Centro de Ciências Agrárias, Universidade Federal de Alagoas, Rio Largo, AL, 2019.

MERTENS, J. *et al.* *Spondias tuberosa* Arruda (Anacardiaceae), a threatened tree of the Brazilian Caatinga?. **Brazilian Journal of Biology**, v. 77, p. 542-552, 2017. DOI: <https://doi.org/10.1590/1519-6984.18715>.

MITCHELL, J. D.; DALY, D. C. A revision of *Spondias* L.(Anacardiaceae) in the Neotropics. **PhytoKeys**, n. 55, p. 1, 2015. DOI: <https://doi.org/10.3897/phytokeys.55.8489>.

MITCHELL, J. D. *et al.* Neotropical Anacardiaceae (cashew family). **Brazilian Journal of Botany**, v. 45, p.139-180, 2022. DOI: <https://doi.org/10.1007/s40415-022-00793-5>.

MOREIRA, M. N.; COSTA, Élia K. de C.; DONATO, S. L. R.; NARAIN, N. Phytochemical profile and antioxidant property of different genotypes of umbuzeiro fruits (*Spondias tuberosa* Arruda Câmara): a review. **Research, Society and Development**, v. 10, n. 16, p. e58101623116, 2021. DOI: 10.33448/rsd-v10i16.23116.

MOREIRA, M. N. **Caracterização e comparação de genótipos de frutos umbuzeiro (*Spondias tuberosa* Arruda Câmara) do sudoeste baiano**. 2022. 67 f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Sergipe, São

Cristóvão, 2022. Disponível em: <http://ri.ufs.br/jspui/handle/riufs/16210>. Acesso em: 28 jun. 2023.

MOREIRA, M. N. *et al.* Phytochemical profile and antioxidant property of different genotypes of umbuzeiro fruits (*Spondias tuberosa* Arruda Câmara): a review. **Research, Society and Development**, v. 10, n. 16, p. e58101623116, 2021. DOI: 10.33448/rsd-v10i16.23116.

NAITANI, S. *et al.* Plant Reactome: um recurso para caminhos de plantas e análise comparativa. **Nucleic Acids Research**, v. 45, n. D1, p. D1029–D1039, 2017. DOI: <https://doi.org/10.1093/nar/gkw932>

NEPA. **Tabela brasileira de composição de alimentos**. 4. ed. Campinas, SP: UNICAMP, 2011.

NEZAMI, E.; GALLEGO, P. P. História, filogenia, biodiversidade e novas ferramentas computacionais para micropropagação e conservação eficientes de pistache (*Pistacia* spp.) germoplasma. **Plantas (Basileia, Suíça)**, v. 12, n. 2, p. 323, 2023. DOI: <https://doi.org/10.3390/plants12020323>

NIU, Y.; GAO, C.; LIU, J. Genomas mitocondriais completos de três espécies de Mangifera, sua estrutura genômica e transferência gênica a partir de genomas de cloroplastos. **BMC Genomics**, v. 23, n. 1, p. 147, 2022. DOI: <https://doi.org/10.1186/s12864-022-08383-1>

NUNES, R. L. *et al.* O genoma de *Spondias mombin* L.: montagem, anotação e análise de RNA-seq. **BMC Genomics**, v. 21, n. 1, p. 1-12, 2020.

OLIVEIRA, E. J. *et al.* Desenvolvimento e caracterização de marcadores microssatélites para *Spondias tuberosa* Arruda (Anacardiaceae). **Genetics and Molecular Research**, v. 15, n. 2, 2016.

PAODJUNAS, R. **Conhecimento tradicional e usos do umbuzeiro (*Spondias tuberosa*) por comunidades rurais do semiárido paraibano, nordeste do Brasil**. 2018. 52 f. Dissertação (Mestrado em Desenvolvimento e Meio Ambiente) - Programa Regional de Pós-Graduação em Desenvolvimento e Meio Ambiente, Centro de Ciências Exatas e da Terra, Universidade Federal da Paraíba, João Pessoa, PB, 2018.

PEIXOTO, C. P. **Princípios de fisiologia vegetal: teoria e prática**. Rio de Janeiro: Pod Editora, 2020. ISBN: 978-65-86147-21-6.

PEREIRA, M.C. *et al.* *Spondias* spp.: uma revisão de seus usos tradicionais, fitoquímica e farmacologia. **Journal of Ethnopharmacology**, v. 224, p. 126-146, 2018.

POSSENTI, J. C. **Qualidade fisiológica de sementes de soja enriquecidas com molibdenio**. 2007. 51 f. Tese (Doutorado em Ciências) - Programa de Pós-Graduação em Ciência e Tecnologia de Sementes, Universidade Federal de Pelotas, Pelotas, RS, 2007.

PRITHAM, E. J.; PUTLIWALA, T.; FESCHOTTE, C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. **Gene**, v. 390, n. 1-2, p. 3-17, 2007. DOI: <https://doi.org/10.1016/j.gene.2006.08.008>.

RAMAKRISHNAN, M. *et al.* Transposable elements in plants: recent advancements, tools and prospects. **Plant Molecular Biology Reporter**, v. 40, p. 628-645, 2022. DOI: <https://doi.org/10.1007/s11105-022-01342-w>.

RAMOS, A. K. B. *et al.* Diversidade genética entre as espécies de *Spondias* estimada por marcadores RAPD. **Melhoramento de Culturas e Biotecnologia Aplicada**, v. 11, n. 4, p. 329-336, 2011.

RHOADS, A.; AU, K. F. Sequenciamento PacBio e suas aplicações. **Genômica Proteômica Bioinformática**, v. 13, n. 5, p. 278-289, 2015.

ROJAS-PIRELA, M. *et al.* Phosphoglycerate kinase: structural aspects and functions, with special emphasis on the enzyme from *Kinetoplastea*. **Open Biology**, v. 10, n. 11, p. 200302, 2020. DOI: <https://doi.org/10.1098/rsob.200302>.

SALES, R. P. de. **Ecologia populacional, diversidade genética e modelagem de nicho ecológico da *Spondias tuberosa***. 103 f. 2020. Dissertação (Mestrado em Ciências Florestais) - Programa de Pós-Graduação em Ciências Florestais, Universidade Federal do Rio Grande do Norte, Macaíba, RN, 2020.

SALEHI, B. *et al.* *Anacardium* plantas: química, composição nutricional e aplicações biotecnológicas. **Biomoléculas**, v. 9, n. 9, p. 465, 2019. DOI: <https://doi.org/10.3390/biom9090465>

SAMEH, S. *et al.* Genus *Spondias*: a phytochemical and pharmacological review. **Evidence-Based Complementary and Alternative Medicine**, v. 2018, 2018. DOI: <https://doi.org/10.1155/2018/5382904>.

SANTOS, C. A. F. *et al.* Diversidade genética e estrutura do germoplasma de *Spondias tuberosa* usando marcadores SNP gerados por genotipagem por sequenciamento. **Tree Genetics & Genomes**, v. 16, p. 63, 2020.

SANTOS, C. A. F.; NASCIMENTO, C. E. de S. Relação entre caracteres quantitativos do umbuzeiro (*Spondias tuberosa* A. Camara). **Pesquisa Agropecuária Brasileira**, v. 33, n. 4, p. 449-456, abr. 1998. Disponível em: <http://www.alice.cnptia.embrapa.br/alice/handle/doc/133147>. Acesso em: 28 jun. 2023.

SANTOS, V. N. *et al.* Diversity and genetic structure of *Spondias tuberosa* (Anacardiaceae) accessions based on microsatellite loci. **Revista de Biología Tropical**, v. 69, n. 2, p. 640-648, 2021. DOI: <https://doi.org/10.15517/rbt.v69i2.44194>.

SEWELAM, N. *et al.* Combined abiotic stresses repress defense and cell wall metabolic genes and render plants more susceptible to pathogen infection. **Plants**, v. 10, n. 9, p. 1946, 2021. DOI: [10.3390/plants10091946](https://doi.org/10.3390/plants10091946).

SCHMIDT, C.; HOBAN, S.; JETZ, W. Conservation macrogenetics: harnessing genetics data to meet conservation commitments. **Trends in Genetics**, v. 39, p. 816-829, nov. 2023. DOI: <https://doi.org/10.1016/j.tig.2023.08.002>

SCHULZE-KAYSERS, N.; FEUEREISEN, M. M.; SCHIEBER, A. Phenolic compounds in edible species of the Anacardiaceae family—a review. **RSC Advances**, v. 5, n. 89, p. 73301-73314, 2015. DOI: <https://doi.org/10.1039/C5RA11746A>.

SEPPEY, M.; MANNI, M.; ZDOBNOV, E. M. BUSCO: assessing genome assembly and annotation completeness. **Methods Mol Biol**, v. 1962, p. 227-245, 2019.

SHAH, K. A. *et al.* *Mangifera indica* (mango). **Pharmacognosy Reviews**, v. 4, n. 7, p. 42, 2010. DOI: <https://doi.org/10.4103/0973-7847.65325>

SHAW, R.; TIAN, X.; XU, J. Single-cell transcriptome analysis in plants: advances and challenges. **Molecular Plant**, v. 14, n. 1, p. 115-126, 2021. DOI: <https://doi.org/10.1016/j.molp.2020.10.012>.

SHULSE, C. N. *et al.* High-throughput single-cell transcriptome profiling of plant cell types. **Cell Reports**, v. 27, n. 7, p. 2241-2247. e4, 2019. DOI: <https://doi.org/10.1016/j.celrep.2019.04.054>.

SILVA, G. F.; NASCIMENTO, V. L. Umbu: características botânicas e tecnológicas. **Revista Brasileira de Fruticultura**, v. 36, n.1, p. 1-8, 2014.

SILVA, O. N.; PAIVA, J. G. A. de. Estudos morfológicos e anatômicos em folhas adultas de *Spondias tuberosa* Arruda (Anacardiaceae Lindley). **Boletín Latinoamericano y del Caribe de Plantas Medicinales y Aromáticas**, v. 6, n. 2, p. 36-43, 2007.

SIMÃO, F. A. *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics**, v. 31, n. 19, p. 3210–3212, 2015.

SOUZA, D. D. de. **Adaptações de plantas da Caatinga**. São Paulo: Oficina de Textos, 2020. eISBN: 978-65-86235-04-3.

SOUZA, M. M.; SILVA, A. S. Caracterização morfológica de frutos de umbuzeiro (*Spondias tuberosa* Arruda) cultivados em diferentes regiões do Semiárido brasileiro. **Revista Brasileira de Fruticultura**, v. 36, n. 1, p. 187-194, 2014.

STIEHLER, F. *et al.* Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. **Bioinformatics**, v. 36, n. 22-23, p. 5291-5298, 2020. DOI: <https://doi.org/10.1093/bioinformatics/btaa1044>

SWIFT, J. *et al.* Nutrient dose-responsive transcriptome changes driven by Michaelis–Menten kinetics underlie plant growth rates. **Proceedings of the National Academy of Sciences**, v. 117, n. 23, p. 12531-12540, 2020. DOI: <https://doi.org/10.1073/pnas.1918619117>.

TAIZ, L. *et al.* **Fisiologia e desenvolvimento vegetal**. 6. ed. Porto Alegre: Artmed, 2017. ISBN: 9781605352558

TAN, L. *et al.* Genomic analysis of WD40 protein family in the mango reveals a TTG1 protein enhances root growth and abiotic tolerance in Arabidopsis. **Scientific Reports**, v. 11, n. 1, p. 2266, 2021. DOI: [10.1038/s41598-021-81969-z](https://doi.org/10.1038/s41598-021-81969-z).

- TARAILO-GRAOVAC, M.; CHEN, N. Using RepeatMasker to identify repetitive elements in genomic sequences. **Current Protocols in Bioinformatics**, v. 4, p. 4.10.1–4.10.14, 2009.
- TER-HOVHANNISYAN, V. *et al.* Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. **Genome Research**, v. 18, n. 12, p. 1979-1990, 2008. DOI: <https://doi.org/10.1101/gr.081612.108>.
- TÖLKE, E. D. *et al.* Flower structure and development of *Spondias tuberosa* and *Tapirira guianensis* (Spondioideae): Implications for the evolution of the unisexual flowers and pseudomonomy in Anacardiaceae. **International Journal of Plant Sciences**, v. 182, n. 9, p. 747-762, 2021. Disponível em: <https://www.journals.uchicago.edu/doi/pdf/10.1086/716780>. Acesso em: 28 jun. 2023.
- TÖRÖNEM, P.; HOLM, L. PANNZER: a practical tool for protein function prediction. **Protein Science**, v. 31, n1, 2021. DOI: <https://doi.org/10.1002/pro.4193>.
- TROVÃO, D. M. de B. M. *et al.* Variações sazonais de aspectos fisiológicos de espécies da Caatinga. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 11, n. 3, p. 307-311, 2007.
- VAN DER AUWERA, G. A., *et al.* From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. **Current Protocols in Bioinformatics**, v. 43, n. 1110, p. 11.10.1-11.10.33, 2013.
- VELCULESCU, V. *et al.* Characterization of the yeast transcriptome. **Cell**, v. 88, n. 2, p. 243-251, 1997. DOI: [https://doi.org/10.1016/S0092-8674\(00\)81845-0](https://doi.org/10.1016/S0092-8674(00)81845-0).
- WARD, J. A.; PONNALA, L.; WEBER, C. A. Strategies for transcriptome analysis in nonmodel plants. **American Journal of Botany**, v. 99, n. 2, p. 267-276, 2012. DOI: <https://doi.org/10.3732/ajb.1100334>.
- XAVIER, G. O. **Modelagem do nicho ecológico do umbu (*Spondias tuberosa*) na região da Caatinga**. 2020. 95 f. Dissertação (Mestrado em Ciências Florestais) - Universidade de Brasília, Brasília, DF, 2020.
- YANG, E. *et al.* Taxonomy and phylogeny of novel and extant taxa in Pleosporales associated with *Mangifera indica* from Yunnan, China (Series I). **Journal of Fungi**, v. 8, n. 2, p. 152, 2022. DOI: <https://doi.org/10.3390/jof8020152>.
- ZHAO, S. *et al.* Regulation of plant responses to salt stress. **International Journal of Molecular Sciences**, v. 22, n. 9, p. 4609, 2021.
- ZHANG, H. *et al.* Respostas ao estresse abiótico em plantas. **Nature Reviews Genetics**, v. 23, p. 104–119, 2022. DOI: <https://doi.org/10.1038/s41576-021-00413-0>
- ZHU, Jian-Kang. Abiotic stress signaling and responses in plants. **Cell**, v. 167, n. 2, p. 313-324, 2016. DOI: <https://doi.org/10.1016/j.cell.2016.08.029>.
- ZIMMER, W.; MENDEL, R. Molybdenum metabolism in plants. **Plant Biology**, v. 1, p. 160-168, 1999.