



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

GLADYANE SANTOS DA SILVA

**ABORDAGEM METABOLÔMICA BASEADA EM RMN PARA ESTUDO DE
CASCA E FOLHAS DO BANCO ATIVO DE GERMOPLASMA DE *Spondias mombin***

FORTALEZA

2021

GLADYANE SANTOS DA SILVA

ABORDAGEM METABOLÔMICA BASEADA EM RMN PARA ESTUDO DE CASCAS E
FOLHAS DO BANCO ATIVO DE GERMOPLASMA DE *Spondias mombin*

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Química. Área de concentração: Química.

Orientador: Profa. Dra. Dávila Zampieri.

Coorientador: Dr. Guilherme Julião Zocolo

FORTALEZA

2021

Página reservada para ficha catalográfica.

Utilize a ferramenta *online* Catalog! para elaborar a ficha catalográfica de seu trabalho acadêmico, gerando-a em arquivo PDF, disponível para download e/ou impressão.

(<http://www.fichacatalografica.ufc.br/>)

GLADYANE SANTOS DA SILVA

ABORDAGEM METABOLÔMICA BASEADA EM RMN PARA ESTUDO DE CASCAS E
FOLHAS DO BANCO ATIVO DE GERMOPLASMA DE *Spondias mombin*

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Química. Área de concentração: Química.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Profa. Dra. Dávila Zampieri (Orientador)
Universidade Federal do Ceará (UFC)

Dr. Guilherme Julião Zocolo (Coorientador)
Empresa Brasileira de Pesquisa Agropecuária (Embrapa)

Profa. Dra. Gisele Simone Lopes
Universidade Federal do Ceará (UFC)

Dr. Edy Sousa de Brito
Empresa Brasileira de Pesquisa Agropecuária (Embrapa)

À Deus.

Aos meus pais, Ana Lúcia e Beneval, que me incentivaram durante toda a vida.

AGRADECIMENTOS

À Deus, a qual sou grata pela vida, pela perseverança dada para enfrentar todos os empecilhos que surgiram no decorrer desse trabalho e por ser refúgio em todos os momentos.

À minha família, que sempre me deu o melhor, mesmo em momentos de dificuldade. Obrigada por todo apoio, motivação e demonstração de orgulho por cada passo da minha vida acadêmica!

Ao meu namorado, Ícaro Marques, por toda ajuda, paciência, incentivo e pôr sempre tornar tudo isso mais fácil e mais leve. Obrigada por sempre estar presente e por todo amor dado!

À todos os meus amigos por toda paciência nos meus momentos de estresse e por entender a minha ausência, na maior parte das vezes. Obrigada por tudo!

Aos amigos que a pós-graduação reaproximou cada vez mais, Cristine, Rodolfo e Robson. Obrigada pela amizade, pelo apoio e por tornar esse mestrado mais divertido e reconfortante.

Aos colegas do laboratório, em especial ao Jhonyson, Oiram, Priscila, Francilene, Sheila, Ana Clara, Jorge e Yandra que me ajudaram tremendamente em várias etapas desse trabalho.

À minha orientadora, Dra. Dávila Zampieri, por toda paciência e ensinamento compartilhado. Obrigada por sempre tornar as coisas menos complexas e nos tratar de forma tão amável!

Ao meu coorientador, Dr. Guilherme Zocolo, por todo suporte, preocupação e ensinamento. Obrigada por sempre está presente ainda que em outro estado e por sempre nos motivar, nos desafiar e acreditar em cada aluno seu!

Ao grupo BioMass, do qual faço parte, por todo apoio, suporte, aprendizado e momentos de descontração.

À Embrapa, por todo suporte oferecido para a realização desse trabalho, em especial, aos membros do LMQPN, Paulo, Lorena, Dr. Edy e Dr. Kirley, por todo apoio e contribuição.

À CAPES e ao CNPQ, pela concessão da bolsa de mestrado e por subsidiarem o projeto de pesquisa.

Ao Programa de Pós-Graduação em Química da UFC, ao qual devo minha formação, por toda ajuda e oferta de cursos/disciplinas que contribuíram significativamente para meu conhecimento.

À Universidade Federal do Ceará, minha segunda casa, por tornar meu sonho real!

RESUMO

A exploração de frutíferas na região Nordeste do Brasil ocorre, em sua grande maioria, de forma extrativista, utilizando técnicas de manejo inadequadas que podem levar à perda de recursos genéticos e genótipos contendo compostos bioativos de importância farmacológica. Essa prática se repete com a cajazeira, cujos cultivos em pomares domésticos e comerciais ainda são escassos. Por se tratar de uma espécie ainda em domesticação, citogeneticamente pouco conhecida e com alto potencial produtivo nas regiões de origem, é de enorme relevância o estabelecimento de estratégias de conservação de sua variabilidade genética a fim de evitar o risco de erosão genética e mantê-la disponível para uso em programas de melhoramento de plantas. Assim, no presente trabalho, foi realizado um estudo do perfil químico do banco ativo de germoplasma (BAG) contendo 22 acessos de cajazeira (separados entre folhas e cascas) divididos entre clones e progênies, visando contribuir com a geração de conhecimento e agregação de valor. Para tanto, a técnica de metabolômica não-alvo foi aplicada por meio de uma abordagem baseada em Ressonância Magnética Nuclear para elucidar os perfis químicos dos acessos e genótipos. Observou-se a presença de metabólitos de interesse químico/farmacológico com possíveis aplicações técnicas, medicinais e nutricionais, entre os quais a quercetina, o ácido elágico e o ácido gálico. Moléculas de açúcar e aminoácidos, já relatados como marcadores de resistência a condições de estresse hídrico, também foram encontrados em abundância nas folhas e na casca, informação valiosa para fins de melhoramento genético. A análise multivariada foi utilizada para investigar a variabilidade química entre os acessos e correlacionar os metabólitos descritos com o local e horário de coleta. Assim, os resultados obtidos justificaram o estabelecimento de uma base química associada ao BAG, que pode ser útil para subsidiar programas de melhoramento genético da cajazeira, fundamental para aumentar a produtividade agrícola e a qualidade nutricional das lavouras em face das condições adversas a que estão sujeitos.

Palavras-chave: Conservação de recursos genéticos. Elucidação estrutural. Análise multivariada. Melhoramento de plantas.

ABSTRACT

The exploitation of fruit trees in the Northeast region of Brazil occurs in the great majority, in an extractive manner, using inadequate management techniques that lead to the loss of genetic resources and genotypes containing bioactive compounds of pharmacological importance. This practice is repeated with the cajazeira tree, whose cultivations in domestic and commercial orchards are few. As it is a species still in domestication, cytogenetically little known, and with high productive potential in the regions of origin, it is hugely significant to establish strategies for conserving its genetic variability in order to avoid the risk of genetic erosion and maintain it available for use in plant breeding programs. Thus, in the present work, a chemical profile study of the active germplasm bank (AGB) was carried out containing 22 accessions of cajazeira (separated between leaves and barks) divided between clones and progenies. To contribute to generating knowledge and adding value to the AGB, the non-target metabolomics technique was applied through an approach based on Nuclear Magnetic Resonance to elucidate the accessions and genotypes' chemical profiles. It was observed the presence of metabolites of chemical/pharmacological interest with possible technical, medicinal, and nutritional applications, among which quercetin, ellagic acid, and gallic acid can be mentioned. Sugar molecules and amino acids already reported as markers of resistance to stress conditions were also found in abundance in the leaves and bark. Multivariate analysis was used to investigate the chemical variability between accessions and correlate the described metabolites with the place and time of collection. Thus, the results obtained justified the establishment of a chemical base associated with the active germplasm bank, which can be useful to subsidize programs of genetic improvement of the cajazeira tree, essential to increase the agricultural productivity and nutritional quality of the crops in the face of adverse conditions to which they are subjected.

Keywords: Conservation of genetic resources. Structural elucidation. Multivariate analysis. Plant breeding.

LISTA DE FIGURAS

Figura 1 - Relações biossintéticas entre o metabolismo basal e especial. As rotas biossintéticas para a síntese de metabólitos especiais ocorrem pelas seguintes vias: (1) via dos aminoácidos, (2) via do chiquimato, (3) via dos policetídeos, (4) via do ácido mevalônico e (5) via do metileritritol fosfato (MEP)	19
Figura 2 - Mapa de distribuição geográfica de <i>Spondias mombin</i> no Brasil	27
Figura 3 - Categorias e subcategorias do estudo metabolômico	34
Figura 4 - Esquema resumido do fluxograma de trabalho envolvido na abordagem metabolômica alvo	35
Figura 5 - Esquema resumido do fluxograma de trabalho envolvido na abordagem metabolômica não-alvo	36
Figura 6 – Fluxograma do estudo metabolômico em plantas. *Algumas etapas de preparo de amostra podem ser alteradas conforme o método analítico utilizado	37
Figura 7 - Procedimento envolvido na Ressonância Magnética Nuclear	40
Figura 8 - Estudo metabolômico envolvendo Ressonância Magnética Nuclear	42
Figura 9 - Etapas de pré-processamento que podem ocorrer no conjunto de dados estudado em metabolômica	43
Figura 10 - Faseamento de linhas espectrais de RMN, visando a obtenção de fases puras ..	46
Figura 11 - Comportamento da matriz de dados frente aos diferentes métodos de centralização e escalonamento: (a) sem tratamento; (b) centrado na média; (c) auto-escalonamento e (d) Pareto; (e) escalonamento por escala de alcance; (f) o escalonamento por estabilidade de variável	48
Figura 12 - Gráficos de <i>scores</i> e <i>loadings</i> , respectivamente, inspecionados simultaneamente a fim de entender as relações e tendências entre grupos observados nas amostras e quais variáveis (metabólitos) são responsáveis	51
Figura 13 - Dendrograma HCA para algoritmo aglomerativo (A) e fluxo de agrupamento de algoritmo divisivo (B)	52

Figura 14 - Gráficos de <i>scores</i> OPLS (esquerda) e <i>loadings</i> (direita) para o componente preditivo 1 <i>versus</i> componente ortogonal 1, com pontuações coloridas pela variável y.....	53
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Capítulo 1. Untargeted NMR-Based metabolomics applied to study of leaves and bark of the active germplasm bank of *Spondias mombin* L. using multivariate analysis tools

Figure 1 - Representative ¹ H NMR spectra from extracts of the barks (a) and leaves (b) of <i>S. mombin</i>	65
------------------------------------------------------------------------------------------------------------------------------	----

Figure 2 - PCA modeling using the ¹ H NMR dataset of barks from different genotypes of <i>S. mombin</i> : a) bidimensional scores coordinate system (PC1 × PC2), where “C” and “P” mean clone and progenies, respectively; b) PC1 loadings plotted in lines form with the relevant identified compounds	68
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Figure 3 - PCA modeling using the ¹ H NMR dataset of leaves from different genotypes of <i>S. mombin</i> : a) bidimensional scores coordinate system (PC1 × PC2), where “C” and “P” mean clone and progenies, respectively; b) PC1 loadings plotted in lines form with the relevant identified compounds	71
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

LISTA DE TABELAS

Tabela 1 - Cronologia da descoberta de fármacos a partir de plantas	15
Tabela 2 - Principais vias metabólicas do metabolismo basal	17
Tabela 3 - Número de gênero e espécies que compõem as tribos de Anacardiaceae	22
Tabela 4 - Espécies de <i>Spondias mombin</i> L. e sua distribuição geográfica	23
Tabela 5 - Metabólitos encontrados em <i>Spondias</i> spp.	24
Capítulo 1. Untargeted NMR-Based metabolomics applied to study of leaves and bark of the active germplasm bank of <i>Spondias mombin</i> L. using multivariate analysis tools	
Table 1 - Coding and procedence of clones and progenies of leaves and barks of <i>Spondias mombin</i> L.	61
Table 2 - Primary and secondary metabolites identified in <i>Spondias mombin</i> L. leaves	96
Table 3 - Primary and secondary metabolites identified in <i>Spondias mombin</i> L. barks ..	98

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	15
2.1	Biodiversidade como fonte de moléculas bioativas	15
2.2	Metabolismo vegetal	16
2.2.1	<i>Metabolismo primário</i>	17
2.2.2	<i>Metabolismo secundário</i>	18
2.3	A árvore cajazeira	21
2.3.1	<i>A Família Anacardiaceae</i>	21
2.3.2	<i>O gênero Spondias</i>	22
2.3.3	<i>Spondias mombin L.</i>	27
2.4	Banco ativo de germoplasma (BAG)	31
2.5	Metabolômica	33
2.5.1	<i>Preparo da amostra</i>	36
2.5.2	<i>Aquisição de dados</i>	38
2.5.2.1	<i>Ressonância magnética nuclear (RMN)</i>	39
2.5.3	<i>Tratamento de dados</i>	42
2.5.3.1	<i>Pré-processamento de dados</i>	42
2.5.3.2	<i>Pré-tratamento de dados</i>	46
2.5.3.3	<i>Métodos quimiométricos em metabolômica</i>	49
3	OBJETIVOS	54
3.1	Objetivo geral	54
3.2	Objetivos específicos	54
4	Capítulo 1. Untargeted NMR-Based metabolomics applied to study of leaves and bark of the active germplasm bank of <i>Spondias mombin L.</i> using multivariate analysis tools	56
5	CONCLUSÃO	85
	REFERÊNCIAS	86
	APÊNDICE A – SUPPLEMENTARY DATA	95

1 INTRODUÇÃO

Devido as várias diferenças climáticas e, conseqüentemente, ao predomínio de diversos biomas, o Brasil é um dos maiores responsáveis pela diversidade de plantas do mundo, caracterizando-se como uma fonte abundante de metabólitos candidatos à potenciais agentes medicinais (GIULIETTI; HARLEY, 2005; NEWMAN; CRAGG, 2016; SIMÕES et al., 2017).

Esses compostos são oriundos de vias metabólicas resultantes da evolução e adaptação do vegetal ao ambiente e/ou produzidos para atuar como possíveis mecanismos de defesa, podendo, ainda, ter funções associadas ao crescimento e ao desenvolvimento (DIAS; URBAN; ROESSNER, 2012; LEWINSOHN, 2017; SIMÕES et al., 2017).

A riqueza em metabólitos de diferentes classes pode ser encontrada na frutífera *Spondias mombin*, conhecida por seus inúmeros efeitos medicinais, como propriedades antioxidantes, antivirais, anti-inflamatórias, efeito antimicrobiano e adstringente, atividade antidiabética, antipsicótica, bem como atividade antifúngica (COSTA et al., 2013; DOS SANTOS et al., 2010; PESSOA et al., 2006; REMILA et al., 2015)

Compostos encontrados nas folhas dessa árvore, como quercetina e ácido elágico, pertencentes à classe dos flavonoides e taninos, respectivamente, demonstraram ser opções preventivas e terapêuticas contra diferentes enfermidades, dentre as quais estão o câncer (MATTIETTO; MATTA, 2011; RAUF et al., 2018; SILVA et al., 2011; WANG et al., 2016), vírus da herpes tipo 1 (SIQUEIRA et al., 2020), depressão e ansiedade (INGRET et al., 2018), doenças relacionadas ao fígado (NWIDU et al., 2018), entre outras.

Além de sua importância na indústria farmacêutica, o elevado consumo dos frutos da cajazeira é de extrema importância para o agronegócio das regiões Norte e Nordeste do país, as quais detêm a maior ocorrência dessa frutífera (RAMIRES, 2016). Sua exploração, no entanto, ocorre, na maioria das vezes, de forma extrativista, utilizando práticas favoráveis à perda dos recursos genéticos e de genótipos com compostos de importância medicinal (RAMIRES, 2016).

Assim, para otimizar o uso dessa espécie e impedir que o processo de erosão genética venha a ocorrer, torna-se necessário a criação de bancos ativos de germoplasma, mantendo-a disponível para uso em programas de melhoramento genético e assegurando a variabilidade genética existente (MACHADO et al., 2015; SPEHAR, 2012). Das cinco coleções de cajá encontradas nos centros de pesquisa do país, nenhuma apresenta, ainda, informações relacionadas ao perfil químico de seus acessos (FONSECA et al., 2018).

A análise dos perfis metabólicos da cajazeira, bem como a prospecção de moléculas candidatas à fármacos, têm sido particularmente desafiadora, devido a complexa e diversificada

natureza química inerente às plantas (DIAS; URBAN; ROESSNER, 2012). Dessa forma, a metabolômica tem se destacado na pesquisa de produtos naturais no que concerne à otimização biotecnológica da descoberta e validação de metabólitos secundários farmacologicamente ativos, abrangendo uma gama de abordagens que vão do holístico ao reducionista (TRENERRY; ROCHFORT, 2010). Para esse fim, técnicas de análise robustas e de alta resolução, juntamente com a análise multivariada, têm sido amplamente utilizadas (FORCISI et al., 2013; FRANCILENE et al., 2019; GAD; EL-AHMADY; ABOU-SHOER, 2012).

Com base nas informações descritas, o presente trabalho visa realizar um estudo metabolômico com as folhas e cascas da espécie *Spondias mombin* L. com base em um banco ativo de germoplasma, envolvendo o estudo de variabilidade química entre acessos de cajazeira. No decorrer do trabalho, a técnica de Ressonância Magnética Nuclear (RMN) será aliada à análise multivariada da matriz estudada.

2 REFERENCIAL TEÓRICO

2.1 Biodiversidade como fonte de moléculas bioativas

De acordo com o Fórum Nacional de 1986 (Washington, EUA), a Biodiversidade – ou diversidade biológica - é entendida como a variabilidade de organismos vivos de todas as origens, compreendendo os ecossistemas terrestres, marinhos e outros ecossistemas aquáticos, bem como os complexos ecológicos de que fazem parte; incluindo também a diversidade dentro de espécies, entre espécies e ecossistemas (SIMÕES et al., 2017).

Entre os países detentores da megadiversidade do planeta, o Brasil é a principal e maior fonte de riqueza em plantas do mundo, reconhecendo a existência de 34916 espécies (SIMÕES et al., 2017). A variedade e complexidade de seus biomas, bem como a posse da maior cobertura de florestas tropicais do mundo, concentrada especialmente na Floresta Amazônica, são os principais fatores responsáveis por essa classificação (PIMENTEL et al., 2015).

Esse imenso patrimônio biológico tem valor econômico-estratégico inestimável em várias atividades, como segmentos do agronegócio, têxtil e energia, entretanto sua maior potencialidade se encontra na descoberta de novos agentes bioativos, embora apenas um pequeno percentual (15%) tenha sido submetido a algum estudo científico para investigar seus efeitos farmacoterapêuticos (CALIXTO, 2003; SIMÕES et al., 2017).

No que tange à descoberta de novos produtos, o impacto da pesquisa na área de produtos naturais, principalmente nos séculos XVIII e XIX, contribuiu de forma significativa para o estabelecimento de vários medicamentos com base na medicina popular (Tabela 1) (MGL et al., 2010; SIMÕES et al., 2017).

Tabela 1 – Cronologia da descoberta de fármacos a partir de plantas

Fármaco	Espécie	Ano	Recurso terapêutico
Digitoxina	<i>Digitalis</i>	1785-1875	Cardiotônico
Morfina	<i>Papaver</i>	1805	Hipnoanalgésico
Quinina	<i>Cinchona</i>	1820	Antimalárico
Atropina	<i>Atropa</i>	1833	Anticolinérgico
Fisostigmina	<i>Physostigma</i>	1864	Anticolinesterásico
Pilocarpina	<i>Pilocarpus</i>	1875	Colinérgico
Efedrina	<i>Ephedra</i>	1887	Adrenérgico

Cocaína	<i>Erythroxylon</i>	1895	Anestésico local
Tubocurarina	<i>Chondrodendron</i>	1895	Bloqueador neuromuscular
Dicumarol	<i>Melilotus</i>	1941	Anticoagulante
Reserpina	<i>Rauwolfia</i>	1952	Neuroléptico

Fonte: (SIMÕES et al., 2017).

Nesse contexto, centrado na obtenção e desenvolvimento de fármacos, o mercado de fitomedicamentos continua crescendo consideravelmente e despertando o interesse de muitas indústrias farmacêuticas multinacionais, assegurando uma fonte de oportunidades para o Brasil, em especial na pesquisa científica e tecnológica de compostos bioativos produzidos pelas plantas (CALIXTO, 2003; SIMÕES et al., 2017).

2.2 Metabolismo vegetal

O termo “metabolismo” refere-se à totalidade de processos químicos que ocorrem em um organismo vivo, incluindo reações simples e complexas, mediadas por uma infinidade de enzimas. De forma geral, o metabolismo permite o crescimento, desenvolvimento e reprodução do vegetal, possibilitando, assim, a perpetuação das espécies (ALMEIDA, 2017; SIMÕES et al., 2017).

A partir da captura de energia e substâncias químicas de ambientes externos, os organismos vivos sintetizam e degradam uma infinidade de compostos capazes de atender às necessidades fisiológicas distintas de sobrevivência e adequação do hospedeiro em ambientes sob constante mudança, os quais são denominados metabólitos, geralmente restrito a compostos com peso molecular de até 1.500 Da (AHARONI; GALILI, 2011; WENG, 2014). Em relação a rede metabólica de plantas, em específico, acredita-se que ela seja de longe mais extensa do que na maioria de outros organismos (AHARONI; GALILI, 2011; WENG, 2014).

Os produtos do metabolismo são divididos em metabólitos primários, ou basais, que englobam reações e caminhos de importância vital, e metabólitos secundários, ou especiais, responsáveis pelo desempenho de diversas funções importantes para o crescimento e desenvolvimento, incluindo a interação planta-meio ambiente (DEBORDE et al., 2017; FANG; FERNIE; LUO, 2019; FUMAGALI et al., 2008; POTT; OSORIO; VALLARINO, 2019).

2.2.1 Metabolismo primário

O metabolismo primário, ou basal, é destinado principalmente à manutenção de vida e reprodução. Seus compostos são essenciais à todos os seres vivos e possuem funções bem definidas, como divisão celular, crescimento, respiração, armazenamento e reprodução. Nesse grupo, estão inclusos aminoácidos, açúcares, nucleotídeos, vitaminas e ácidos graxos (A. HUSSEIN; A. EL-ANSSARY, 2019; SIMÕES et al., 2017).

Os metabólitos primários da maioria das espécies são conservados em grande parte, apresentando grande semelhança, o qual pode ser evidenciado pelo padrão de fornecimento e armazenamento de energia, especialmente, moléculas de ATP; síntese e armazenamento de substâncias similares, como gorduras, carboidratos e proteínas, por vias metabólicas semelhantes; o uso de enzimas para a catálise de reações metabólicas; código genético idêntico para os organismos; dentre outros (Tabela 2) (DEBORDE et al., 2017; FANG; FERNIE; LUO, 2019; SIMÕES et al., 2017; WENG, 2014).

Tabela 2 – Principais vias metabólicas do metabolismo basal

Metabolismo	Biossíntese	Degradação
Carboidratos	Fotossíntese, ciclo de Calvin, ciclo dos ácidos C-4, gliconeogênese	Clivagem hidrolítica de carboidratos, glicólise, ciclo da pentose-fosfato
Gorduras	Síntese de lipídeos, complexo-ácido graxo-síntase, acilglicerídeos, fosfolipídeos, glicolipídeos, carotenoides, esteróis	Clivagem hidrolítica de lipídeos, β -oxidação de ácidos graxos
Proteínas	Biossíntese de aminoácidos e proteínas (processo de tradução)	Clivagem hidrolítica de proteínas, degradação e conversão dos aminoácidos
Acetil-coenzima A	Carboidratos, ácidos graxos, aminoácidos, cetogênese, terpenos, esteroides	Ciclo de Krebs, cadeia respiratória, ciclo do glicoxilato
Ácidos nucleicos	Biossíntese de nucleotídeos do RNA a partir de bases púricas e pirimídicas, replicação de DNA, formação, formação de flavinas e pteridinas a partir de GTP	Clivagem de DNA e RNA

Fonte: Adaptado de (SIMÕES et al., 2017).

A partir de quatro vias metabólicas principais com precursores provenientes do metabolismo primário, as quais envolvem síntese, degradação e conversão, são produzidos o segundo grupo de metabólitos a serem discutidos mais adiante, os metabólitos secundários (DOS SANTOS, 2015; GALDEANO; KLEINGESINDS, 2016).

2.2.2 Metabolismo secundário

Os produtos do metabolismo secundário, ou ainda especializado, visto que esses são de distribuição restrita, encontrados somente em determinadas espécies, não estão diretamente relacionados aos processos primários das plantas, como crescimento e desenvolvimento (DEBORDE et al., 2017; FANG; FERNIE; LUO, 2019; GALDEANO; KLEINGESINDS, 2016; SIMÕES et al., 2017).

Durante muito tempo, acreditou-se que os metabólitos especiais eram apenas subprodutos do metabolismo primário, sem papel específico (SIMÕES et al., 2017). Entretanto, no decorrer do tempo, ficou claro a função vital dessas substâncias para as plantas em virtude das mesmas utilizarem rotas biossintéticas elaboradas, com alto consumo de energia na síntese de compostos (SIMÕES et al., 2017).

Ao contrário do metabolismo basal, o metabolismo especializado é caracterizado por diferenças que podem ser constatadas pelas distintas vias metabólicas dos organismos, lhes conferindo alta variabilidade e destaque com a descoberta de suas múltiplas funções, ainda que seus produtos sejam encontrados em concentrações relativamente baixas (DOS SANTOS, 2015; SIMÕES et al., 2017).

Esses metabólitos atuam primeiramente na defesa do vegetal, protegendo a planta de predadores e patógenos e preservando o fruto até o crescimento da semente. Uma das principais estratégias é a promoção do sabor amargo que algumas substâncias costumam apresentar, levando os herbívoros a reconhecerem e evitarem as plantas que as contêm (SIMÕES et al., 2017). Por outro lado, alguns metabólitos secundários atuam de forma contrária, atraindo animais responsáveis por causarem efeitos benéficos à planta, como a polinização (SIMÕES et al., 2017).

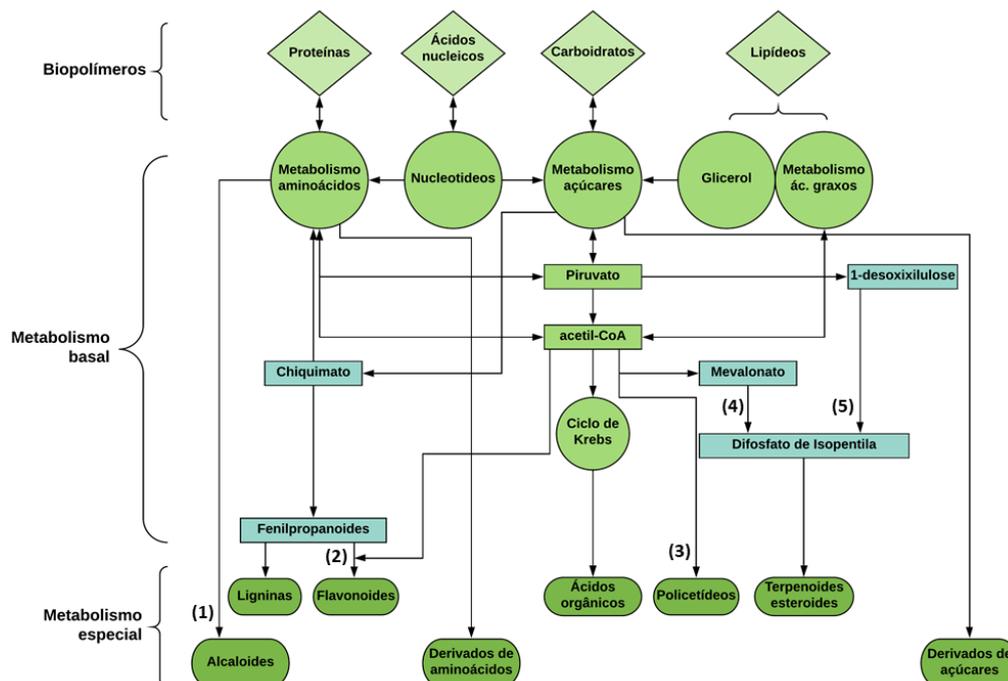
Alguns desses compostos podem estar envolvidos, ainda, na reprodução vegetal, na proteção de plantas contra raios ultravioleta e espécies reativas de oxigênio (ERO), bem como em efeitos alelopáticos, onde há a competição entre os vegetais como resposta à estresses bióticos e abióticos. Para muitos dos metabólitos especializados, no entanto, ainda não se conhece a real função, uma vez que esses são muito diversos, já sendo descritos, em

angiospermas, mais de 50.000 (DEBORDE et al., 2017; EVA; WEBER, 2010; SIMÕES et al., 2017).

Como já mencionado, todos os metabólitos especiais são derivados de componentes do metabolismo basal, os quais são aproveitados fora do processo de geração de energia da fotossíntese, glicólise e ciclo de Krebs para o fornecimento de intermediários biossintéticos. Os blocos de construção (precursores) mais importantes para a biossíntese desses são derivados da acetil coenzima A, ácido chiquímico, ácido mevalônico e 1-desóxi-D-xilulose-5-fosfato (DOS SANTOS, 2015; KABERA, 2014).

Ainda que sejam conhecidos cerca de 100.000 metabólitos especiais advindos de plantas, somente algumas rotas biossintéticas são utilizadas para sua síntese, os quais envolvem 5 principais vias: (1) via dos aminoácidos, (2) via do chiquimato, (3) via dos policetídeos, (4) via do ácido mevalônico e (5) via do metileritritol fosfato (MEP) (Figura 1) (SIMÕES et al., 2017).

Figura 1 – Relações biossintéticas entre o metabolismo basal e especializado. As rotas biossintéticas para a síntese de metabólitos especializados ocorrem pelas seguintes vias: (1) via dos aminoácidos, (2) via do chiquimato, (3) via dos policetídeos, (4) via do ácido mevalônico e (5) via do metileritritol fosfato (MEP)



Fonte: Adaptado de (SIMÕES et al., 2017).

Através das vias biossintéticas descritas na Figura 1, são formados os três principais grupos de metabólitos especializados, organizados com base em suas estruturas químicas, levando em conta aspectos biogenéticos: terpenóides, alcaloides e compostos fenólicos (SIMÕES et al., 2017).

Por meio das vias do ácido mevalônico (4) e do MEP (5), são produzidos os terpenos. A organização desses metabólitos segue a regra do isopreno (unidades pentacarbonadas, C-5), o qual baseia-se em observações de química orgânica. O bloco de construção de todos os terpenos utiliza como matéria-prima o difosfato de isopentila (IPP), que é um isopreno na forma ativa em equilíbrio com seu isômero, o difosfato de dimetilalila (DMAPP) (DOS SANTOS, 2015; SIMÕES et al., 2017).

Pela via do ácido mevalônico, para biossíntese dos terpenóides, esses dois compostos são obtidos pela condensação de três moléculas de acetil-CoA, com a participação do fosfato, e descarboxilação subsequente. A biossíntese pela rota do MEP, uma rota alternativa, leva a formação de dois intermediários-chave a partir da glicose, o 1-desóxi-D-xilulose (DOX) e 2-metileritritol fosfato, que, em última instância, vão originar o IPP e DMAPP. A partir dessa via, são obtidos os blocos de construção de hemiterpenos (C-5), monoterpenos (C-10), diterpenos (C-20) e tetraterpenos (C-40) (DOS SANTOS, 2015; SIMÕES et al., 2017).

A biossíntese dos alcaloides, compostos geralmente caracterizados pela ocorrência de um átomo de nitrogênio no estado oxidativo em um anel heterocíclico, é obtida pela via dos aminoácidos (1). Para a maioria desses compostos, a regra dos aminoácidos é válida, o qual se baseia na afirmação de que todos os alcaloides verdadeiros são derivados destes (OLIVOTO et al., 2017; SIMÕES et al., 2017).

O grupo amínico dos aminoácidos origina o átomo de nitrogênio do heterociclo de um alcaloide. Mediante a descarboxilação inicial da molécula de um aminoácido, é formada a amina biogênica, que vai promover a biossíntese dos alcaloides. É importante ressaltar ainda que nem todos os compostos desse grupo seguem a regra dos aminoácidos, como os alcaloides acridônicos, esteroidais, imidazólicos e púricos (OLIVOTO et al., 2017; SIMÕES et al., 2017).

Os compostos fenólicos das plantas, um dos maiores grupos de constituintes vegetais secundários sintetizados, são formados pela via biossintética do ácido chiquímico (2). A maior parte das substâncias desse grupo são derivadas da *L*-fenilalanina ou, raramente, da *L*-tirosina. Esses últimos, ao sofrerem desaminação oxidativa, catalisados pela fenilalanina-liase na biossíntese do ácido cinâmico, viabilizam, por meio da degradação oxidativa de sua cadeia lateral, a formação de ácidos fenólicos e fenóis simples (COHEN; KENNEDY, 2010; KABERA, 2014; SIMÕES et al., 2017).

Em função dos processos coevolutivos que resultaram em uma sofisticada maquinaria biossintética e uma alta capacidade de interação de produtos do metabolismo de plantas com proteínas e outras biomoléculas presentes nos seres vivos, os metabólitos secundários, dado à sua grande diversidade estrutural, apresentam em sua grande maioria, atividades biológicas/farmacológicas relevantes (SIMÕES et al., 2017).

Esses compostos, assim chamados de produtos naturais, representam uma empolgante biblioteca de compostos bioativos filtrados por seleção natural, podendo ser empregados *in natura*, em preparações galênicas simples, em formulações farmacêuticas com maior valor agregado, ou como substâncias puras, os ingredientes ativos dos medicamentos (SIMÕES et al., 2017; WINK, 2015).

Estes podem, ainda, ser a base para a obtenção de moléculas farmacologicamente ativas por meio de processos de síntese parcial, visando incrementar propriedades biofarmacêuticas ou servir como modelos moleculares para o planejamento e síntese total de novos fármacos (SIMÕES et al., 2017; WINK, 2015).

Muitos desses metabólitos especializados são extraídos de plantas já utilizadas na medicina popular com aplicações estabelecidas, como é o caso da cajazeira, árvore típica da região Nordeste do Brasil, utilizada no tratamento de diversas infecções (CORADIN; CAMILLO; PAREYN, 2018; WINK, 2015).

2.3 A árvore Cajazeira

2.3.1 A Família Anacardiaceae

Os principais centros de diversidade dessa família estão no México, América do Sul, África Meridional e equatorial, Madagascar, Indochina e Malásia. No Brasil, esta é representada por muitos gêneros de árvores de importância econômica, como *Anacardium*, *Lithraea*, *Schinus*, *Spondias* e *Tapirira* (LACCHIA et al., 2016; PELL et al., 2010).

Os membros de Anacardiaceae são bem conhecidas por seus frutos e sementes comestíveis (manga, caju, cajá, entre outros), sua madeira valiosa e produção de verniz natural. Os metabólitos secundários de espécies deste grupo também têm sido amplamente investigados, sendo destaque estudos com foco na atividade medicinal de fenólicos, ésteres e taninos. (LACCHIA et al., 2016; PELL et al., 2010).

A característica mais significativa dos membros das espécies de Anacardiaceae é a presença de condutos resinosos no córtex e no lenho, com conseqüente formação de resina

nessas partes, embora a exsudação ocorra, de forma semelhante, nas folhas, flores e frutos (FONSECA et al., 2018).

A família Anacardiaceae compreende, aproximadamente, 81 gêneros e 800 espécies, sendo subdividida em cinco tribos: Anacardieae, Rhoeeae, Spondiadeae, Semecarpeae e Dobineae, uma classificação baseada em características morfológicas. O número de gêneros e espécies alocadas a essas tribos são mostradas na Tabela 3 (SCHULZE-KAYSERS; FEUEREISEN; SCHIEBER, 2015).

Tabela 3 – Número de gêneros e espécies que compõem as tribos de Anacardiaceae

Tribo	Gêneros	Espécies
Dobineae	2	3
Semecarpeae	5	80
Anacardieae	8	120
Spondiadeae	17	140
Rhoeeae	40	300

Fonte: (SCHULZE-KAYSERS; FEUEREISEN; SCHIEBER, 2015).

A partir desta, pode-se observar que as tribos Anacardieae, Spondiadeae e Rhoeeae são de longe as tribos mais importantes em relação ao número de gêneros e espécies, abrangendo praticamente todas as culturas economicamente relevantes da família Anacardiaceae, incluindo o gênero *Spondias*, um dos gêneros mais comumente encontrados no país e altamente valorizados no que tange à investigação sobre fitoterapia de plantas medicinais, pertencente a tribo Spondiadeae (SAMEH et al., 2018; SCHULZE-KAYSERS; FEUEREISEN; SCHIEBER, 2015).

2.3.2 O gênero *Spondias*

O gênero *Spondias* compreende 19 espécies, a saber *S. acida* Blume, *S. admirabilis* J. D. Mitch. & Daly, *S. chinensis* (Merr) F. P. Metcalf, *S. dulcis* Parkinson, *S. expeditionaria* Hook. f., *S. globosa* J. D. Mitch. & Daly, *S. macrocarpa* Engl., *S. malayana* Kosterm., *S. mombin* L., *S. novoguineensis* Kosterm., *S. pinnata* (L. f.) Kurz, *S. purpurea* L., *S. radlkoferi* Donn. Sm., *S. tefyi* J. D. Mitch., *S. testudinis* J. D. Mitch. & Daly, *S. tuberosa* Arruda, *S. venulosa* (Engl.) Engl., *S. xerophila* Kosterm. e *S. bahienses* P, um híbrido entre *S. tuberosa*

e *S. mombin*. A distribuição de cada táxon encontra-se registrada na Tabela 4 (MITCHELL; DALY, 2015; SAMEH et al., 2018).

Tabela 4 – Espécies de *Spondias mombin* L. e sua distribuição geográfica

Espécie	Distribuição geográfica
<i>S. acida</i> Blume	Península Malaia
<i>S. admirabilis</i> J. D. Mitch. & Daly	Mata Atlântica do Rio de Janeiro
<i>S. chinensis</i> (Merr) F. P. Metcalf	Ásia
<i>S. dulcis</i> Parkinson	Pacífico, amplamente cultivado nos trópicos
<i>S. expeditionaria</i> J.D. Mitch. & Daly	Mata Atlântica do Espírito Santo e Minas Gerais
<i>S. globosa</i> J. D. Mitch. & Daly	Oeste da Amazônia, atípico em Zulia (Venezuela)
<i>S. macrocarpa</i> Engl.	Mata Atlântica brasileira
<i>S. malayana</i> Kosterm.	Malásia, Filipinas
<i>S. mombin</i> L.	México à Bolívia e Brasil, amplamente cultivado em trópicos úmidos
<i>S. novoguineensis</i> Kosterm.	Nova Guiné e Ilhas Salomão
<i>S. pinnata</i> (L. f.) Kurz	Subcontinente indiano, Indochina e Sul da China
<i>S. purpurea</i> L.	Florestas tropicais secas do México e América Central
<i>S. radlkoferi</i> Donn. Sm.	México à América Central e Noroeste da Venezuela e Colômbia
<i>S. tefyi</i> J. D. Mitch.	Madagascar
<i>S. testudinis</i> J. D. Mitch. & Daly	Sudoeste da Amazônia
<i>S. tuberosa</i> Arruda	Nordeste do Brasil, caatinga
<i>S. venulosa</i> (Engl.) Engl.	Caatinga e Mata Atlântica brasileira
<i>S. xerophila</i> Kosterm.	Sri Lanka
<i>S. bahienses</i> P.	Endêmica do Brasil, com ocorrência restrita à Região Nordeste

Fonte: Adaptado de (BATISTA et al., 2015; MITCHELL; DALY, 2015).

Os membros de *Spondias* spp. são amplamente estudados na etnofarmacologia, sendo utilizados no tratamento de inúmeras enfermidades, incluindo dor de estômago, diarreia, diabetes, demência, anemia, disenteria, dentre outros (SAMEH et al., 2018).

Os frutos de *S. dulcis* são utilizados pela população rural do Bangladesh para aumentar visão e prevenir infecções oculares, enquanto as de *S. tuberosa* são consumidas pelas comunidades rurais no Brasil devido à seu alto valor nutricional. Os frutos de *S. mombin* são usados na Nigéria como diuréticos. O pó do fruto maduro de *S. pinnata* é usado na Índia como antídoto para flechas venenosas (SAMEH et al., 2018).

Em relação às estruturas foliares de espécies desse gênero, uma infusão de folhas frescas de *S. purpurea* é usada no México para o tratamento de dor de estômago e flatulência, assim como a casca da mesma espécie, usada em casos de anemia, diarreia, disenteria e infecções de pele. Em Belize, a decocção foliar de *S. mombin* é usada para tratar diarreia e disenteria, bem como pelas populações da Nigéria, Benin e Togo, que a utilizam para a manutenção da boa memória (SAMEH et al., 2018).

Na Índia, a goma produzida a partir de *S. pinnata* é usada no tratamento de disenteria, úlceras, diarreia e doenças de pele, bem como demulcente em casos de bronquite e a casca é utilizada no tratamento de articulações dolorosas, bem como na prevenção de vômitos e tratamento de diarreia e disenteria (SAMEH et al., 2018).

No Brasil, especialmente nas regiões Norte e Nordeste, as principais frutíferas de importância socioeconômica são o umbuzeiro (*Spondias tuberosa* Arruda), a cajazeira (*Spondias mombin* L.), a serigueleira (*Spondias purpurea* L.), a cajaraneira (*Spondias dulcis* Parkinson) e os híbridos naturais entre esses, todas exploradas de forma extrativista ou em pomares domésticos (FONSECA et al., 2018).

Em relação aos metabólitos primários e secundários encontrados no gênero, rico em diferentes classes dos mesmos, estão inclusos compostos fenólicos, esteróis, triterpenos, saponinas, óleos essenciais, aminoácidos e polissacarídeos (Tabela 5).

Tabela 5 – Metabólitos encontrados em *Spondias* spp.

Composto	Espécie	Parte encontrada (tipo de extrato)
Taninos e Pseudotaninos		
Geranina	<i>S. mombin</i> L.	Folhas e caules (80% EtOH)
Galoil-geranina	<i>S. mombin</i> L.	Folhas e caules (80% EtOH)
Galoil-glicose	<i>S. purpurea</i> L.	Frutos (Acetona)
Flavonoides		
Ramnetina	<i>S. purpurea</i> L.	Frutos (Acetona)

Isoramnetina	<i>S. purpurea</i> L.	Frutos (Acetona)
Canferol	<i>S. purpurea</i> L.	Frutos (Acetona)
Canferide	<i>S. purpurea</i> L.	Frutos (Acetona)
Astragalina	<i>S. purpurea</i> L.	Frutos (Acetona)
Isoquercetina	<i>S. purpurea</i> L.	Frutos (Acetona)
Quercetina dihidratada	<i>S. purpurea</i> L.	Frutos (Acetona)
Rutina	<i>S. venulosa</i> (Engl.) Engl.	Folhas (80% MeOH)
Ramnetina-3- <i>O</i> - β -D-rutinosídeo	<i>S. venulosa</i> (Engl.) Engl.	Folhas (80% MeOH)
Quercetina-3- <i>O</i> -[α -ramnopiranosil-(1 \rightarrow 2)]- α -ramnopiranosil-(1 \rightarrow 6)- β -glicopiranosídeo	<i>S. venulosa</i> (Engl.) Engl.	Folhas (80% MeOH)
Fenólicos e derivados		
Ácido gálico	<i>S. purpurea</i> L.	Frutos (Acetona)
3-cafeoil ácido quínico	<i>S. purpurea</i> L.	Frutos (Acetona)
Galato de metila	<i>S. pinnata</i> (L. f.) Kurz	Casca (70% MeOH)
Esteróis e terpenóides		
β -amirina	<i>S. pinnata</i> (L. f.) Kurz	Fruto (MeOH)
Ácido oleanólico	<i>S. pinnata</i> (L. f.) Kurz	Fruto (MeOH)
24-metilenocicloartanona	<i>S. pinnata</i> (L. f.) Kurz	Partes aéreas (EtOH)
Estigmast-4-en-3-ona	<i>S. pinnata</i> (L. f.) Kurz	Partes aéreas (EtOH)
β -sitosterol	<i>S. pinnata</i> (L. f.) Kurz	Partes aéreas (EtOH)
Ácido lignocérico	<i>S. pinnata</i> (L. f.) Kurz	Partes aéreas (EtOH)
β -sitosterol- β -D-glicosídeo	<i>S. pinnata</i> (L. f.) Kurz	Partes aéreas (EtOH)
Estigmast-9-en-3,6,7-triol	<i>S. mombin</i> L.	Casca (MeOH)
3-hidróxi-22-epóxiestigmastano	<i>S. mombin</i> L.	Casca (MeOH)
Ergosteril triterpeno 1	<i>S. pinnata</i> (L. f.) Kurz	Casca (CHCl ₃ /MeOH)
Ergosteril triterpeno 2	<i>S. pinnata</i> (L. f.) Kurz	Casca (CHCl ₃ /MeOH)
Lupeol	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)
Ácido equinocístico-3- <i>O</i> - β -D-galactopiranosil-(1 \rightarrow 5)- <i>O</i> - β -D-xilofuranosídeo	<i>S. pinnata</i> (L. f.) Kurz	Raízes (EtOH)
α -pineno	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)
β -pineno	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)

Cariofileno	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)
Humuleno	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)
Indeno	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)
Cadineno	<i>S. mombin</i> L. e <i>S. purpurea</i> L.	Folhas (Hidrodestilação)
Aminoácidos e carboidratos		
Glicina	<i>S. pinnata</i> (L. f.) Kurz	Frutos
Cisteína	<i>S. pinnata</i> (L. f.) Kurz	Frutos
Serina	<i>S. pinnata</i> (L. f.) Kurz	Frutos
Alanina	<i>S. pinnata</i> (L. f.) Kurz	Frutos
D-galactose	<i>S. pinnata</i> (L. f.) Kurz	Frutos
D-xilose	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
L-arabinose	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
2,3,4,6-tetra- <i>O</i> -metilglicose	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
2,3,6-tri- <i>O</i> -metilglicose	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
2,3-di- <i>O</i> -metilglicose	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
3- <i>O</i> -metilglicose	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
Propan-1,2-dióico	<i>S. pinnata</i> (L. f.) Kurz	Frutos (Aquoso)
Ácido 3-carboxil- β -D-glicopiranosil-(6' \rightarrow 1'')- β -D-glicofuranosídeo	<i>S. pinnata</i> (L. f.) Kurz	Frutos (EtOH)

Fonte: Adaptado de (SAMEH et al., 2018).

Devido a riqueza de metabólitos secundários, como visto na Tabela 4, as espécies de *Spondias* spp. apresentam inúmeros efeitos farmacológicos já relatados, como atividade citotóxica, demonstrado pelo extrato etanólico das cascas de *S. pinnata* contra linhagens de adenocarcinoma pulmonar humano (A549) e adenocarcinoma de mama humano (MCF-7), bem como linhagens de glioblastoma (U87), sendo esse último testado com compostos isolados da fração acetato de etila da mesma parte da planta (SAMEH et al., 2018).

O extrato metanólico (70% v/v) das cascas de *S. pinnata* e o extrato etanólico desse fruto são descritos por apresentarem potenciais efeitos antioxidantes. O extrato metanólico de *Spondias purpurea* demonstrou o mesmo efeito. Atividades protetoras de ulceração gástrica foram relatadas com o extrato metanólico das cascas de *S. pinnata* e extrato aquoso das folhas de *Spondias mombin*. Efeitos hepatoprotetores, fotoprotetores, anti-inflamatórios, antiartrítico, analgésico, antipirético, trombolítico, antimicrobial, anti-hipertensivo, antifertilidade,

antihemíntico, diurético, laxante, antiepiléptico e antipsicótico também foram evidenciados em *Spondias* spp. (SAMEH et al., 2018).

A maior parte desses efeitos farmacológicos podem ser evidenciados na espécie *Spondias mombin*, o qual será visto com mais detalhes adiante.

2.3.3 *Spondias mombin* L.

Árvore nativa do sul do México até o Sudeste do Brasil, conhecida popularmente como cajazeira. É encontrada nas regiões tropicais da América, África e Ásia, onde cresce abundantemente no Brasil, nas regiões Norte, Nordeste, Centro-Oeste e Sudeste, ocorrendo naturalmente em grupos ou de forma isolada (Figura 2). As regiões geográficas que concentram a diversidade genética da cajazeira englobam a Mata Atlântica e a Amazônia Ocidental do Estado do Acre, Brasil, e regiões limítrofes do Peru e da Bolívia (BOFF et al., 2020; CABRAL et al., 2016; CORADIN; CAMILLO; PAREYN, 2018; FONSECA et al., 2018; MATTIETTO; MATTA, 2011; SILVA et al., 2017).

Figura 2 – Mapa de distribuição geográfica de *Spondias mombin* no Brasil



Fonte: Adaptado de (CORADIN; CAMILLO; PAREYN, 2018).

A cajazeira exibe elevada adaptação às regiões de condições climáticas distintas, ocorrendo na Amazônia, Mata Atlântica, serras e litoral do Nordeste brasileiro; zonas úmidas e subúmidas, especialmente nas regiões costeiras onde há maior incidência de chuva; nos limites mais úmidos do Agreste e nas regiões de encostas de serras do Ceará e do Rio Grande do Norte; e nas caatingas do Semiárido quando cultivada (CORADIN; CAMILLO; PAREYN, 2018).

As áreas de maior ocorrência dessa frutífera arbórea são a região do Brejo Paraibano, no Estado da Paraíba; a zona litorânea, próxima a Fortaleza, e as serras de Guaramiranga,

Meruoca, Baturité e Ibiapaba, no Ceará; bem como a região sul da Bahia (Figura 2). Populações nativas de cajazeira ocorrem em florestas úmidas tropicais a semidecíduas, como também em florestas de galerias e ilhas florestais em savanas, porém é menos frequente em florestas de planícies inundadas (CORADIN; CAMILLO; PAREYN, 2018).

Nas regiões de ocorrência natural dessa espécie, principalmente nas regiões Norte e Nordeste do país, o cajá tem considerável importância social e econômica, sendo um dos produtos mais valorizados nos mercados, devido seu sabor exótico, aroma intenso e propriedades nutricionais (CORADIN; CAMILLO; PAREYN, 2018; MATTIETTO; MATTA, 2011; TIBURSKI et al., 2011).

Este é consumido fresco e, na maior parte das vezes, de forma processada na fabricação de sucos, picolés, sorvetes, néctares e geleias, devido sua alta acidez e perecibilidade, não requerendo tecnologias avançadas e, conseqüentemente, grandes investimentos (BRITO et al., 2018a; CORADIN; CAMILLO; PAREYN, 2018; MATTIETTO; MATTA, 2011; TIBURSKI et al., 2011).

Ainda que a exploração do cajá, nas regiões brasileiras já mencionadas, seja de forma quase exclusivamente extrativista, sendo poucos os pomares domésticos e cultivos comerciais, seu potencial produtivo é alto, não atendendo, muitas vezes, a demanda dos consumidores. Essa demanda crescente tem despertado o interesse de vários setores do agronegócio, desempenhando um papel importante na geração de renda e empregos durante o período de colheita em todas as áreas que ocorre (BRITO et al., 2018b, 2018a; CORADIN; CAMILLO; PAREYN, 2018; FONSECA et al., 2018; MATTIETTO; MATTA, 2011).

A cajazeira também é utilizada para extração de madeira, a qual é leve, mole, fácil de trabalhar e de média durabilidade. A casca presta-se à modelagem, xilogravura e fonte de substâncias adstringentes. As árvores, cujo porte pode alcançar, aproximadamente, 20 m, são utilizadas para diversos fins, como cerca-viva, árvores de sombra e ornamentais, bem como alimento para animais da região (CORADIN; CAMILLO; PAREYN, 2018; SACRAMENTO; SOUZA, 2009).

Assim como a maioria dos frutos tropicais, o cajá tem bom valor nutricional, possuindo altos níveis de potássio, cálcio, magnésio, fósforo, cobre, ferro, manganês, açúcares, vitaminas A e C e alguns carotenoides, os quais contribuem para a coloração amarelada da fruta e derivados de sua polpa (BAILÃO et al., 2015; FONSECA et al., 2018; MATTIETTO; MATTA, 2011; SILVA et al., 2017).

Compostos fenólicos presentes na polpa de cajá também foram amplamente estudados em termos de seus efeitos antioxidantes e seus possíveis benefícios à saúde humana,

evidenciando ações anti-inflamatórias, expectorantes e febrífugas, bem como efeitos inibidores de suas proteínas contra algumas enzimas digestivas de mamíferos, como a α -amilase. Embora a importância desses inibidores na nutrição humana ainda não seja tão clara, alguns estudos sugerem que o uso medicinal dos mesmos pode influir na digestão de carboidratos de pacientes diabéticos ou obesos (ADEDOKUN et al., 2010; MATTIETTO; MATTA, 2011).

Efeitos biológicos de outras partes da árvore que não sejam os frutos, a saber folhas, ramos, caule e cascas, também foram relatados na literatura, apresentando fitoquímicos como taninos elágicos, ácido fenólicos, flavonoides, saponinas, triterpenos, resinas, esteróis e glicosídeos de antraquinona, os quais representam potenciais fontes de novos medicamentos farmacêuticos e nutracêuticos à base de plantas. Por meio do estudo de composição volátil das folhas foi possível identificar, ainda, a presença de como o β -cariofileno e γ -cadineno (BRITO et al., 2018b; CABRAL et al., 2016; CORADIN; CAMILLO; PAREYN, 2018; DA SILVA et al., 2012; LINS et al., 2020; MATTIETTO; MATTA, 2011; OJO et al., 2018; SABIU et al., 2016).

Na medicina popular, o chá das folhas de *S. mombin* é usado na cicatrização de feridas, demonstrando ação anti-inflamatória, contra distúrbios gastrointestinais, uretrite, cistite, prostatite, no tratamento de diabetes *mellitus* e anemia, bem como inflamações nos olhos, boca e garganta. Essas também são utilizadas na maioria dos sistemas de fitoterapia do mundo para atuar como antidepressivos, ansiolíticos e nos problemas do trato reprodutivo feminino, sendo usado por parteiras para ajudar a induzir o parto, reduzir sangramentos e dores durante e depois (ASUQUO et al., 2013; BRITO et al., 2018b; CABRAL et al., 2016; DA SILVA et al., 2012; J. A. O OLUGBUYIRO; J. O MOODY; M. T HAMANN, 2013; SABIU et al., 2016).

Estudos científicos relataram a atividade antiviral pronunciada contra o vírus Herpes (tipo 1) (BOFF et al., 2020) e o vírus Coxsackie a partir dos extratos de folhas de cajazeira e do consequente isolamento de dois elagitaninos, geranina e galoil-geranina (BOFF et al., 2020). Outros estudos citaram e/ou confirmaram sua potencial atividade abortiva, atividade inibidora de β -lactamase, anti-inflamatória, antioxidante, leishmanicida, antiedematogênica, antimicrobiana e adstringente, efeito antipsicótico com uso potencial em transtornos psiquiátricos e atividade antidiabética usando a atividade inibidora da α -amilase do 3- β -olean-12-en-3-il (9Z)-hexadec-9-enoato das folhas de *S. mombin* L. (LINS et al., 2020; MATTIETTO; MATTA, 2011; SABIU et al., 2016).

A casca do caule da cajazeira foi descrita por apresentar propriedades antituberculares e suas folhas como fonte de um potencial antidiabético ativo, também relatando efeitos hipoglicêmicos, bem como as alterações bioquímicas e histológicas do extrato etanólico de

Spondias mombin em ratos diabéticos induzidos por aloxana (J. A. O OLUGBUYIRO; J. O MOODY; M. T HAMANN, 2013; OJO et al., 2018).

O conhecimento das atividades biológicas enfatiza a importância dessa árvore não apenas como uma espécie agroflorestal, mas também como um poderoso recurso econômico, sendo possivelmente necessário a futura expansão do cultivo desta, devido à falta de plantações comerciais e sazonalidade (ADEDOKUN et al., 2010; MATTIETTO; MATTA, 2011). Assim, para que isso seja possível, devem ser estudados os métodos de melhoramento, propagação e produção de cajá.

Para essa espécie, a propagação sexuada é realizada por meio da semente ou caroço, onde muitos dos caroços são estéreis e cujas sementes, que podem chegar até 5, têm dormência, bem como germinação lenta, baixa e desuniforme. A existência de dormência e mais de uma semente na maioria dos caroços é importante para a perpetuação da espécie, podendo ser vantajosa para a propagação sexual, contanto que sejam solucionados os problemas de germinação e de retirada das sementes (CORADIN; CAMILLO; PAREYN, 2018).

A reprodução da cajazeira por sementes é necessária e importante para multiplicar novos genótipos e, principalmente, para a formação de porta-enxertos e clonagem, já que mudas de cajazeira podem ser usadas como porta-enxertos interespecíficos para outras espécies de *Spondias* (CORADIN; CAMILLO; PAREYN, 2018; RAMIRES, 2016).

Na propagação assexuada, são empregados a estaquia e a enxertia, por serem os processos vegetativos que propiciam as maiores taxas de formação de mudas clonadas. Para a estaquia, utiliza-se um pedaço de estaca de caule ou de raiz como propágulo, no entanto há relatos de que a propagação vegetativa por esse método usando a tecnologia existente tem limitações significativas e a tecnologia para produção comercial de mudas é inexistente (CORADIN; CAMILLO; PAREYN, 2018; MATTIETTO; MATTA, 2011; RAMIRES, 2016).

Dessa forma, a enxertia tem sido utilizada com sucesso na fruticultura do cajá, ainda que o melhor método de enxertia ainda seja objeto de debate entre os cientistas. Alguns autores recomendam o enxerto de fissura, enquanto outros indicam que a emenda e o enxerto lateral são melhores métodos. Essa divergência é explicada pelos pesquisadores como uma consequência da grande variabilidade genética das árvores frutíferas (CORADIN; CAMILLO; PAREYN, 2018; MATTIETTO; MATTA, 2011; RAMIRES, 2016).

Em se tratando de variabilidade genética, a cajazeira é citogeneticamente pouco conhecida, assim como a maior parte das frutíferas exploradas de forma extrativista, cujo manejo é realizado com práticas desfavoráveis ao meio ambiente, responsáveis pela perda de recursos genéticos (FONSECA et al., 2018; RAMIRES, 2016).

Dados de alta relevância para a descrição e caracterização dos diversos genótipos existentes nessa espécie, relacionados à sua morfologia, produção, características fisiológicas e fenologia são, ainda, incipientes. A perda de recursos genéticos com compostos bioativos de valor farmacológico também é consequência de uma área devastada, fazendo-se necessário, portanto, a otimização do uso destas espécies por meio da conservação dos recursos genéticos a partir da implantação de bancos de germoplasma (RAMIRES, 2016).

2.4 Banco ativo de germoplasma (BAG)

A conservação de recursos genéticos está associada à manutenção da biodiversidade na Terra, levando em consideração a importância da caracterização das diferentes espécies de seres vivos, bem como o desenvolvimento destas em seus respectivos habitats e suas respectivas implicações (SPEHAR; CLAUDIO; KARIA, 2012).

Esses constituem-se como parte fundamental da diversidade biológica utilizada pelo homem para promover o desenvolvimento sustentável agrícola, tratando da variação genética das diferentes espécies de plantas com importância socioeconômica para os locais que em ocorre naturalmente (RAMIRES, 2016). De forma geral, estão inclusos nesse conceito as espécies silvestres e parentais, raças locais e suas variedades, linhagens melhoradas e com atributos específicos, indivíduos submetidos à experimentos, etc. (FERREIRA, 2011; ROA et al., 2016).

Os recursos genéticos, que compreendem a variabilidade genética das diferentes espécies como de extrema relevância para uso atual e futuro, são representados pelo germoplasma, que se constitui como a coleção genética transmitida entre gerações por células reprodutivas, abrangendo toda estrutura com eficiência regenerativa sobre um organismo vivo, com fins de utilização para pesquisa (FERREIRA, 2011; POOLS et al., 2020; SPEHAR; CLAUDIO; KARIA, 2012).

A conservação do germoplasma, principalmente os de origem vegetal, é uma prioridade consolidada a nível mundial, tendo como principal estratégia a criação de bancos ativos, o qual é caracterizado por armazenar a diversidade genética de uma ou mais espécies (BARATA et al., 2011; FERREIRA, 2011).

Geralmente, esses se constituem como uma base física, mantidos em centros de pesquisa ou instituições públicas e privadas que colecionam o germoplasma de forma *ex situ*, ou seja, fora do habitat natural, sob a forma de sementes, *in vitro*, criopreservação e em campo (FERREIRA, 2011). Esse banco é caracterizado por conter um número de acessos muito maior

que o da coleção de trabalho, mantendo a variabilidade genética das espécies com propósito de preservação, para uso e intercâmbio, sem efetuar descartes (BARATA et al., 2011; BRAVERMAN, 2014; FERREIRA, 2011; RAMIRES, 2016; SPEHAR; CLAUDIO; KARIA, 2012).

Os BAGs existentes em toda região brasileira são avaliados em mais de 250, sendo somente 5 referentes à espécie de cajazeira, conservadas em empresas ou instituições como a Embrapa - Centro de Pesquisa Agropecuária do Trópico Úmido (CPATU - AM), com 35 acessos; a Embrapa Meio Norte (CPAMN - PI), com 38 acessos; a Empresa Baiana de Desenvolvimento Agrícola (EBDA - BA), com 2 acessos; a Empresa Estadual de Pesquisa Agropecuária da Paraíba (EMEPA - PB), com 34 acessos e o Instituto Agrônomo de Pernambuco (IPA - PE), com 33 acessos (FONSECA et al., 2018; SPEHAR, 2012).

A conservação de germoplasma já foi recomendada por inúmeros autores por diversos motivos, dentre eles a oportunidade de uso futuro dos acessos, em virtude do risco de erosão genética e da possível extinção de espécies que habitam territórios ameaçados por impactos ambientais e humanos; a chance de servirem como subsídios à programas de melhoramento, pesquisas aplicadas, recuperação de áreas, ou plantio comercial, ambos de interesse nacional e internacional; a perspectiva de expandir o conhecimento técnico-científico; a necessidade de preservar espécies parentais silvestres de plantas cultivadas; a possibilidade de aumentar o número de espécies presentes nas coleções *ex situ* já existentes, propondo novos fins para espécies não tão conhecidas ou de uso restrito com valor potencial (SPEHAR, 2012).

Após a conservação do germoplasma, a caracterização é uma das etapas mais importantes, demonstrando a distinção morfológica, fenológica, reprodutiva, molecular, bioquímica e genética entre os acessos estudados, além de ser uma importante etapa na agregação de valor (RAMIRES, 2016). A caracterização química, em específico, pode ser útil para distinguir os acessos conservados de uma mesma espécie quanto à presença de compostos de valor biológico, bem como a concentração destes na matriz, auxiliando e/ou intensificando a aplicação dos mesmos em diversas áreas de estudo (RAMIRES, 2016).

Entretanto, nenhum dos BAGs de cajazeira já mencionados possuem informações relacionadas à caracterização química de seus constituintes, bem como a análise multivariada dos mesmos, relacionando-os aos tipos de acessos e distribuição geográfica, tornando fundamental a busca por novos modelos que visem a valorização e diversificação do material conservado. Assim, o estudo da variação química nos acessos de interesse é uma abordagem extremamente rica e inovadora na construção do BAGs (ALMEIDA et al., 2012; ARANHA et al., 2017; DA SILVA et al., 2012; MUKHERJEE, 2019; RAMIRES, 2016).

2.5 Metabolômica

As ciências ômicas têm como principal finalidade compreender o funcionamento do sistema celular de organismos, bem como suas respectivas alterações biológicas. A metabolômica, cujo termo foi introduzido por Steven Oliver em 1998, faz parte desse conjunto de ciências através do estudo das alterações dos metabólitos, assim como a genômica (estudo da alteração dos genes), a transcriptômica (estudo das alterações dos transcritos) e a proteômica (estudo das alterações das proteínas) (ALSEEKH; FERNIE, 2018; CANUTO et al., 2018a).

A metabolômica centraliza seus estudos na comparação de perfis metabólicos de indivíduos sujeitos à diferentes condições genéticas, ambientais ou patológicas (PILON et al., 2020). Na metabolômica de plantas, em específico, esse perfil metabólico está relacionado à um grupo de sinais cromatográficos e/ou espectroscópicos de uma amostra bruta ou de um extrato, os quais evidenciam características químicas intrínsecas à determinadas classes de produtos naturais. Os dados analíticos obtidos devem representar o maior número de metabólitos detectáveis possíveis presentes na matriz de estudo (FUNARI et al., 2013).

A metabolômica possui um enfoque muito versátil com aplicações em vários âmbitos, como análises clínicas, alimentícia e nutrição, esportiva, ambiental, toxicologia forense, microbiologia, dentre outras (CANUTO et al., 2018b). Devido a ampla gama de aplicações, essa ciência vem crescendo de forma exponencial na química de produtos naturais, a julgar pela quantidade de trabalhos publicados nessa área (FUNARI et al., 2013).

Inicialmente, os estudos metabolômicos eram divididos em duas categorias, classificados de acordo com a qualidade dos dados e dos metabólitos a serem detectados em estudos alvo (do inglês, *targeted*) e estudos não-alvo (do inglês, *untargeted*). O primeiro avalia os compostos que tenham associação à rotas metabólicas particulares ou compostos pré-definidos de classes de produtos naturais específicas, como resposta a um dado estímulo, já o segundo leva em consideração todos os metabólitos de modo indiscriminado (CANUTO et al., 2018b; PILON et al., 2020).

Atualmente, existem outras subcategorias que exprimem as diferentes finalidades e especificidades dessa ciência, como *metabolic fingerprint* e *plant metabolomics*, como abordagens não-alvo, bem como *metabolic profiling* e *target analysis*, como abordagens alvo (Figura 3) (CANUTO et al., 2018a; KRASTANOV, 2010; PILON et al., 2020).

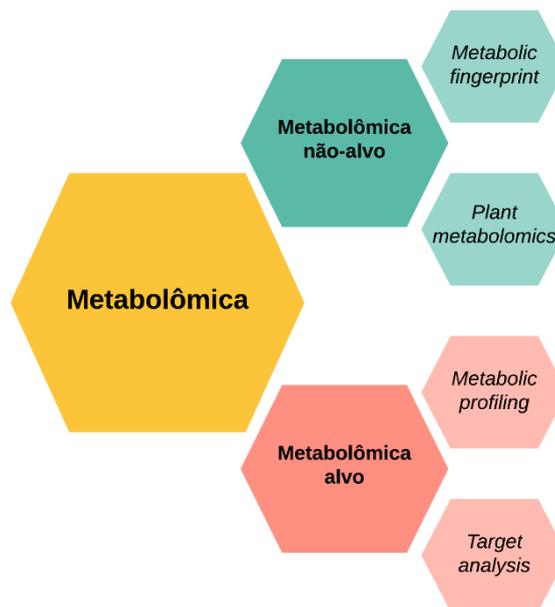
Através da abordagem não-alvo *metabolic fingerprint*, é possível analisar o perfil químico de um indivíduo. De modo usual, a quantificação é relacionada entre amostras controle

e tratada, com o objetivo de compará-las e realizar análise discriminatória. Esta classifica e/ou agrupa dados referentes aos compostos, identificados ou não, sujeitos à condições experimentais diversas (PILON et al., 2020).

O estudo metabolômico por *plant metabolomics* considera a tentativa não enviesada de determinar e quantificar a maior quantidade possível de metabólitos de órgãos vegetais, tecidos ou na planta como um todo (PILON et al., 2020).

O *metabolic profiling*, diferentemente do *metabolic fingerprint*, tem como objetivo a identificação e quantificação de classes metabólicas encarregadas de fazer a distinção entre grupos tratados e controle, os quais são capazes de diferenciá-los geneticamente e fenotipicamente. Por fim, o estudo metabolômico direcionado à *target analysis* tem a finalidade de detectar ou descrever de substâncias ou grupos específicos que possuem propriedades químicas ou uma rota metabólica semelhantes, considerando, usualmente, dados quantitativos e qualitativos (PILON et al., 2020).

Figura 3 – Categorias e subcategorias do estudo metabolômico



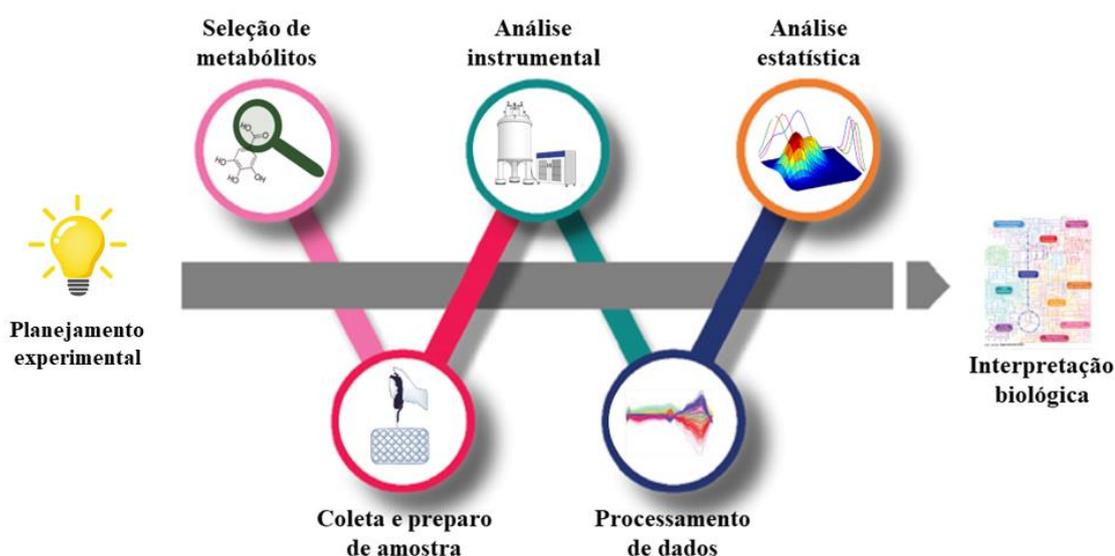
Fonte: Adaptado de (PILON et al., 2020).

Como os estudos utilizando abordagens metabolômicas estava adquirindo proporções crescentes, a Sociedade de Metabolômica criou em 2005, o *Metabolomics Standards Initiative* (MSI), o qual estabelece uma padronização nos trabalhos publicados nessa área e nos procedimentos utilizado para tais estudos, a fim de proporcionar uma descrição clara da matriz biológica estudada e todos os componentes do estudo, bem como possibilitar a aplicação,

compartilhamento e reutilização efetiva dos dados (CANUTO et al., 2018b). Assim, para o desenvolvimento de estudos metabolômicos, alvo e não-alvo, é recomendado que estes sejam conduzidos por uma série de passos já definidos (Figuras 4 e 5) (CANUTO et al., 2018a).

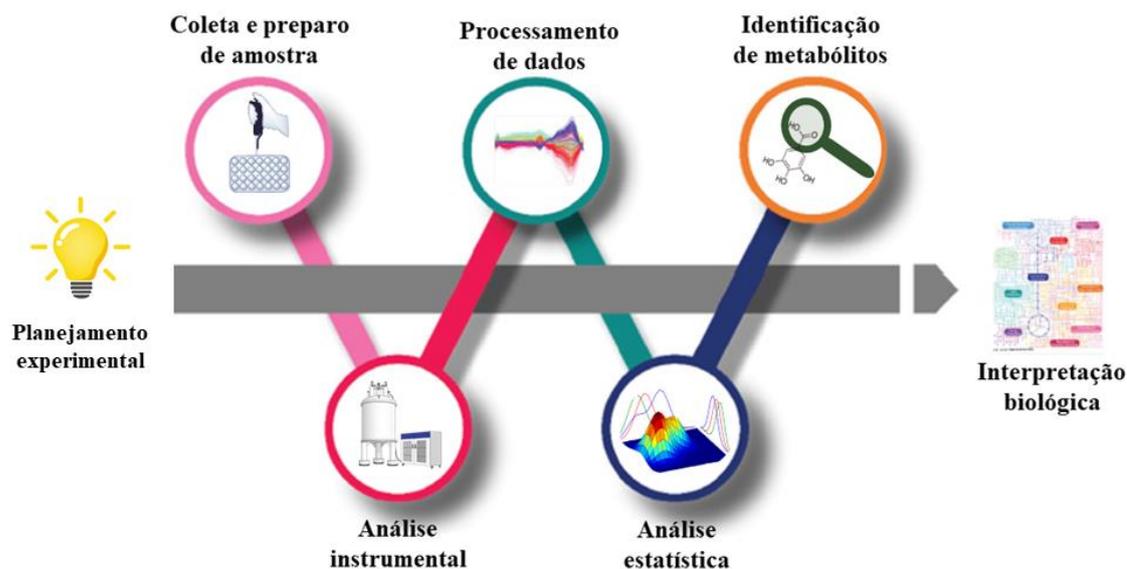
O objetivo não é descrever como realizar experimentos metabolômicos, mas formular um conjunto mínimo de padrões que descrevam os experimentos e métodos de análise associados. Em geral, é exigido informações sobre o planejamento experimental, tais como, métodos de coleta, preparo de amostra e extração de metabólitos, caso tenha ocorrido, além dos métodos analíticos que foram utilizados e a forma que os dados obtidos por tais métodos foram processados (CUBERO-LEON; PEÑALVER; MAQUET, 2014; PILON et al., 2020).

Figura 4 – Esquema resumido do fluxograma de trabalho envolvido na abordagem metabolômica alvo



Fonte: (CANUTO et al., 2018a; PILON et al., 2020).

Figura 5 - Esquema resumido do fluxograma de trabalho envolvido na abordagem metabolômica não-alvo



Fonte: Adaptado de (CANUTO et al., 2018a; PILON et al., 2020).

De início, para que um estudo metabolômico seja desenvolvido, uma questão biológica deve ser previamente definida, a qual deve ser elucidada no fim do mesmo. Outro fator a ser levado em consideração é o tipo de abordagem a ser aplicada (global ou alvo), pois, a partir desta são definidas as etapas de planejamento experimental e análise de dados, as quais podem atingir de forma direta o resultado do trabalho (CANUTO et al., 2018b).

Uma vez que o problema biológico foi definido, é preciso escolher o tipo de amostra a ser trabalhada, a quantidade que será avaliada em cada grupo, bem como a forma que estas serão coletadas e, por seguinte, transportadas e armazenadas, sendo esta etapa de planejamento experimental uma das mais importantes do estudo (CANUTO et al., 2018b).

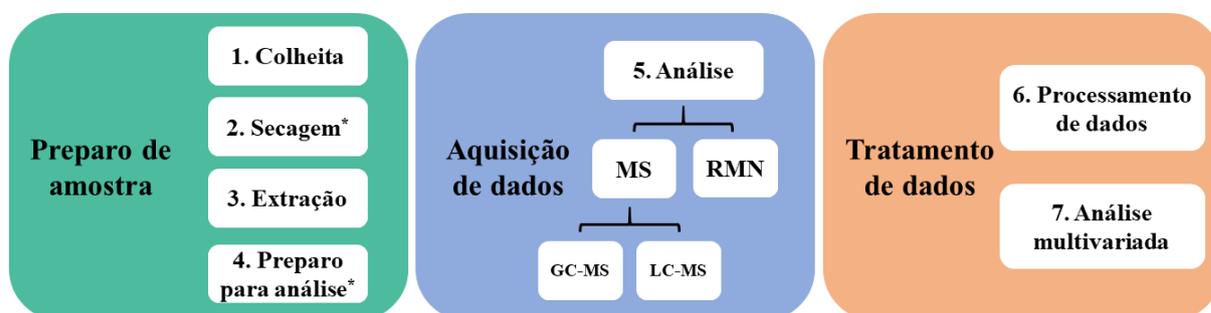
2.5.1 Preparo da amostra

Uma das questões mais importantes na metabolômica é a escolha do método de preparo de amostra, visto que as amostras biológicas utilizadas apresentam grande complexidade, heterogeneidade e ampla faixa de concentração dos metabólitos presentes (CANUTO et al., 2018b). Apesar dos grandes avanços na automação e utilização de sistemas robóticos, grande parte desse trabalho é ainda manual, levando essa etapa a ser, provavelmente, a mais suscetível a erros, consumindo boa parte do tempo e recursos da maioria das análises químicas, com importantes consequências para os compostos isolados e precisão dos resultados (GIKA et al., 2014; KIM; VERPOORTE, 2010).

Dessa forma, todas as considerações práticas sobre esse processo devem ser levadas em consideração desde o início, garantindo a reprodutibilidade do experimento e minimizando todas as fontes de variação possíveis, sem que a composição dos metabólitos presentes na amostra seja alterada. É importante, ainda, manter o procedimento o mais simples e rápido possível, visto que a metabolômica trabalha com um grande número de amostras simultaneamente (GIKA et al., 2014; KIM; VERPOORTE, 2010; KRASTANOV, 2010; LIU; LOCASALE, 2017; SCHRIPSEMA, 2010; WOLFENDER et al., 2013).

Em se tratando do estudo metabolômico de plantas, a maioria dos laboratórios usa seu próprio método específico de preparo de amostras, o qual envolve, de forma geral, quatro etapas: colheita, secagem, extração do material vegetal e preparo da amostra para a análise, embora algumas dessas etapas, como extração ou secagem, possam ser evitadas, dependendo do método analítico a ser usado e das características da amostra em questão (Figura 6) (KIM; VERPOORTE, 2010; SCHRIPSEMA, 2010).

Figura 6 – Fluxograma do estudo metabolômico em plantas. * Algumas etapas de preparo de amostra podem ser alteradas conforme o método analítico utilizado



Fonte: Adaptado de (KIM; VERPOORTE, 2010).

No ato da coleta, a padronização do procedimento é um dos pilares da metabolômica, devendo ser considerado fatores como a parte da planta a ser utilizada (folhas, frutos, cascas), sazonalidade, horários de coleta e os estágios ontogenéticos (PILON et al., 2020).

Em casos em que as coletas foram executadas em locais diferentes, é recomendado que haja padronização em conformidade ao estágio de crescimento vegetativo. Uma vez coletada a planta, o procedimento de retirada de água dessas matrizes é de extrema importância, visto que sua presença afeta de forma negativa várias etapas, prejudicando o processo de extração por solventes orgânicos por meio da alteração na proporção dos mesmos, o armazenamento da amostra e, conseqüentemente, as análises instrumentais. O processo de secagem pode ser

realizado em estufas de ar circulante à 40°C, em ambientes à vácuo sob temperatura ambiente e através de processos de sublimação (PILON et al., 2020).

A seleção de um procedimento de extração ideal deve ser investigada em conformidade com a plataforma analítica escolhida, considerando a polaridade e solubilidade dos compostos a serem detectados. Aspectos como características do solvente, razão solvente e amostra, duração do experimento e temperatura também devem ser considerados (CANUTO et al., 2018a; KIM; VERPOORTE, 2010; KRASTANOV, 2010; PILON et al., 2020; WOLFENDER et al., 2013).

Diferentes procedimentos de extração têm sido avaliados com o objetivo de se obter perfis metabólicos representativos, podendo ser mencionados a maceração, extração por Soxhlet, fluído supercrítico, líquidos iônicos e extrações por solvente à alta temperatura e pressão. Após a obtenção dos extratos, a fase seguinte envolve a seleção do método analítico, o qual deve se tratar de uma técnica reproduzível, simples e rápida que permita determinar o maior número possível de metabólitos (PILON et al., 2020).

2.5.2 Aquisição de dados

Seguindo a etapa de preparo de amostra e, em alguns casos, de procedimentos pré-analíticos, de modo que as amostras evidenciem de forma significativa o sistema biológico a ser estudado, a aquisição de dados requer o uso de técnicas analíticas avançadas, as quais devem ser capazes de detectar objetivamente uma ampla variedade de metabólitos em uma ampla faixa dinâmica, com alta sensibilidade, resolução e repetibilidade (BADDINI; NETTO; TEIXEIRA, 2014; ERNST et al., 2014; TUGIZIMANA; PIATER; DUBERY, 2013; WOLFENDER et al., 2013).

É improvável que única técnica analítica consiga cobrir todo o metaboloma de uma matriz vegetal, devido à grande variedade de estruturas químicas e concentrações nas quais os metabólitos estão presentes e na capacidade das plataformas analíticas, visto que estas possuem suas próprias vantagens e limitações, tanto em seletividade quanto em sensibilidade estudado (CANUTO et al., 2018b; CUBERO-LEON; PEÑALVER; MAQUET, 2014; FUNARI et al., 2013; PILON et al., 2020).

Esse obstáculo requer o uso de multiplataformas de análise no estabelecimento do perfil metabólico das amostras, minimizando as deficiências individuais das ferramentas analíticas e proporcionando, assim, um maior número de metabólitos detectados e, por consequência, uma

maior compreensão do sistema biológico estudado (CANUTO et al., 2018a; CUBERO-LEON; PEÑALVER; MAQUET, 2014; FUNARI et al., 2013; PILON et al., 2020).

A escolha da plataforma analítica depende, principalmente, do estudo em questão, levando em consideração a classe de compostos, suas propriedades químicas e físicas e seus níveis de concentração. Como tal, dentre as técnicas analíticas mais comumente utilizadas, destacam-se a cromatografia gasosa e líquida acoplada à espectrometria de massas (CG-EM ou CLAE-EM, respectivamente), a injeção direta em espectrometria de massas (ID-EM) e a ressonância magnética nuclear (PILON et al., 2020; TUGIZIMANA; PIATER; DUBERY, 2013). Nesse contexto, serão abordados com mais detalhes a técnica de ressonância magnética nuclear.

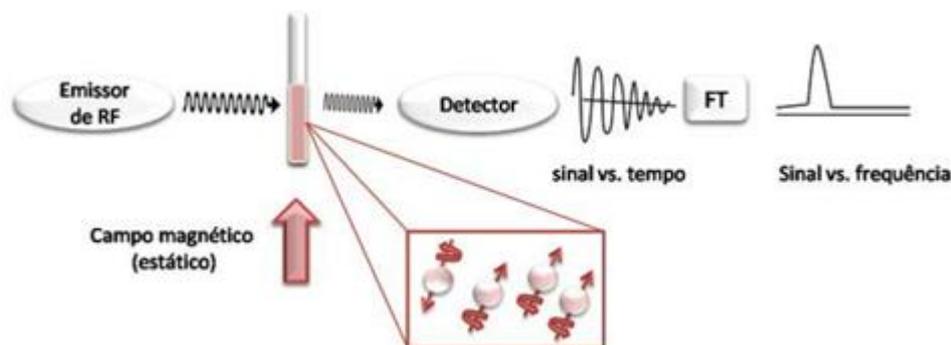
2.5.2.1 Ressonância magnética nuclear (RMN)

A RMN é uma técnica analítica que possibilita a obtenção de informações estruturais e dinâmicas de moléculas, se baseando na detecção das propriedades magnéticas dos núcleos, as quais podem ser comparadas à pequenos piões com carga que giram em torno do seu eixo, produzindo o seu próprio campo magnético, como o ^1H ou o ^{13}C , os quais se comportam como pequenos magnetes (LUZYANIN; ABRANTES, 2010).

Conforme descrito, em um experimento de RMN, quando as amostras são colocadas sob a ação de um campo magnético e irradiadas com uma radiofrequência, a absorção de energia permite que os núcleos sejam promovidos dos estados de rotação de baixa energia para alta energia (TUGIZIMANA; PIATER; DUBERY, 2013).

A emissão subsequente de radiação durante o processo de relaxamento gera sinais registrados em um espectro de RMN como desvios químicos, representando frequências de todos os núcleos visíveis na amostra, em relação à de um hidrogênio de referência presente em um determinado composto (Figura 7) (TUGIZIMANA; PIATER; DUBERY, 2013).

Figura 7 – Procedimento envolvido na Ressonância Magnética Nuclear



Fonte: (LUZYANIN; ABRANTES, 2010).

A amostra a ser analisada é colocada em tubos que podem variar entre 1 à 5 mm de diâmetro e diluída com solvente deuterado, que permite minimizar os sinais que provém desse mesmo solvente. Em relação ao magneto, em estudos metabolômicos geralmente são utilizados instrumentos de 500 ou 600 MHz, porque esses campos são de fácil acesso, sendo recomendável trabalhar no campo magnético mais alto disponível, uma vez que o número de metabólitos observados depende em grande parte da força desse campo magnético (PILON et al., 2020; SMOLINSKA et al., 2012).

O uso da RMN marcou o início da metabolômica como uma ferramenta na bioquímica e na análise fitoquímica, testemunhando um grande número de avanços em estudos fundamentais, desenvolvimento de métodos e aplicações em uma ampla variedade de estudos, incluindo ciências biomédicas, ciências de plantas, alimentos, nutrição, desenvolvimento de medicamentos, ciências energéticas e ambientais (NAGANA GOWDA; RAFTERY, 2017; TUGIZIMANA; PIATER; DUBERY, 2013).

A RMN contribuiu significativamente para o campo da metabolômica e apesar de sua menor sensibilidade em comparação à espectrometria de massa, ela demonstrou ter várias vantagens sobre outras técnicas, fornecendo uma análise eficiente, confiável, robusta e não destrutiva de extratos ou amostras brutas, com reprodutibilidade e repetibilidade primordiais, tornando a técnica um candidato ideal para o estabelecimento de um banco de dados público (SMOLINSKA et al., 2012; WALKER, 2013; WOLFENDER et al., 2013).

A utilização da espectroscopia de RMN está muito associada à determinação da estrutura molecular de pequenas e médias moléculas, incluindo metabólitos inéditos, sendo empregada predominantemente à estudos metabolômicos ao propiciar a análise concomitante de diversos grupos de metabólitos primários e secundários. Essa técnica pode ser aplicada,

ainda, no estudo de macromoléculas, como proteínas e ácidos nucleicos (LUZYANIN; ABRANTES, 2010; PILON et al., 2020).

O método é simples, rápido e possui uma alta capacidade de processamento, exigindo pouco ou nenhum preparo de amostra, diminuindo a probabilidade de perda amostral ou a inserção de alguma variação no procedimento extrativo, além de ser uma técnica não seletiva, ou seja, não depender das características químicas dos metabólitos analisados (CANUTO et al., 2018a; PILON et al., 2020; TUGIZIMANA; PIATER; DUBERY, 2013; WOLFENDER et al., 2013).

Como a RMN é aplicada rotineiramente em análises metabolômicas sem purificação da amostra, a identificação abrangente e precisa dos metabólitos pode oferecer enormes desafios, como resultado da sobreposição de sinais e efeitos de blindagem pelos elétrons vizinhos (GRIFFIN; MERZABAN, 2013; HOOFT, 2016; TUGIZIMANA; PIATER; DUBERY, 2013; WOLFENDER et al., 2013).

Uma das maneiras de superar esse problema consiste no uso da espectroscopia de RMN bidimensional (2D), rotineiramente usada em análises metabolômicas para isolamento de sinais e atribuição inequívoca de metabólitos, uma vez que é útil para aumentar a dispersão do sinal e elucidar a conectividade entre os mesmos, possibilitando o espalhamento do conteúdo espectral em um plano bidimensional e permitindo que os compostos possam ser melhor observados (GRIFFIN; MERZABAN, 2013; HOOFT, 2016; TUGIZIMANA; PIATER; DUBERY, 2013; WOLFENDER et al., 2013).

Além disso, a baixa sensibilidade inerente à RMN consiste em uma das principais limitações da técnica, sendo consideravelmente aumentada com o uso de sondas de refrigeração criogênica, microsondas, aumento da força do campo magnético, diminuição da amostragem com o uso de tubos de até 1,6 mm e, mais recentemente, com a introdução da polarização nuclear dinâmica (DNP) (GRIFFIN; MERZABAN, 2013; PILON et al., 2020).

Uma vantagem muito importante da RMN é a quantificação absoluta, sem o uso de curvas de calibração para cada metabólito detectado. A resposta linear do experimento de RMN é a chave-benefício dessa técnica sobre outros métodos analíticos (PILON et al., 2020; SMOLINSKA et al., 2012; WALKER, 2013; WOLFENDER et al., 2013).

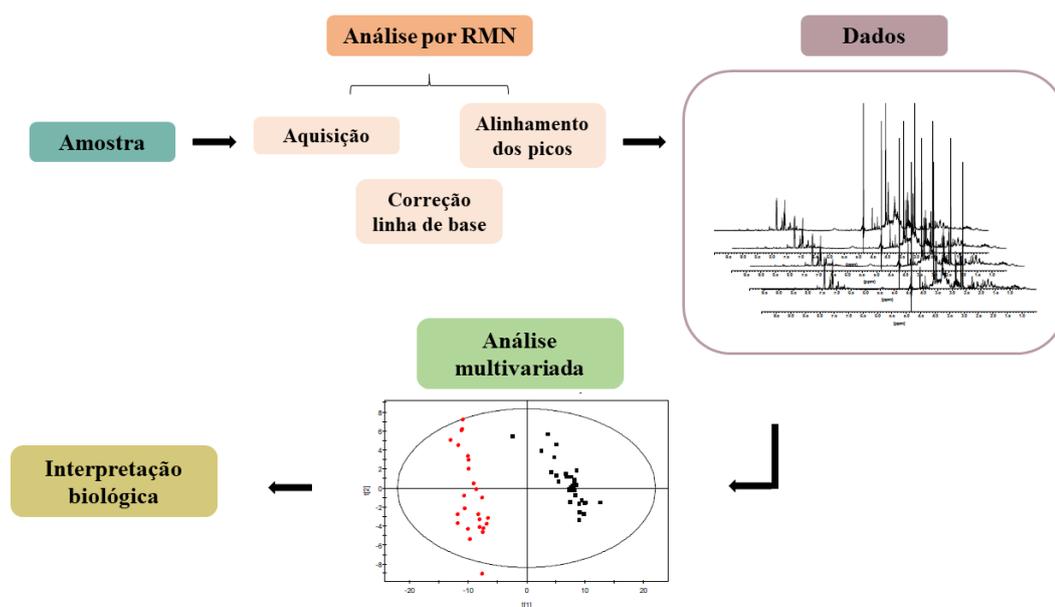
As intensidades dos sinais observados no espectro são diretamente proporcionais à concentração molar, sendo usado, geralmente, a adição de padrões internos para a obtenção da concentração absoluta dos compostos, o qual não pode ser instável no solvente, não ser reativo ou possuir sinais residuais de água, passar facilmente para o estado gasoso e exibir sinais junto aos sinais amostrais. Essas características são fundamentais para assegurar a seletividade,

precisão e acurácia do método, e, conseqüentemente, sua reprodutibilidade e confiabilidade (PILON et al., 2020; SMOLINSKA et al., 2012; WALKER, 2013; WOLFENDER et al., 2013).

Em geral, a metabolômica baseada em RMN utiliza uma combinação de coleta espectroscópica e análise multivariada de dados para procurar padrões dentro de grupos de espectros e monitorar comportamentos metabólicos. Uma vez que os dados são adquiridos, o primeiro passo é transformá-los em um formato adequado para posterior tratamento estatístico (Figura 8) (BRENNAN, 2014; GRIFFIN; MERZABAN, 2013).

Em geral, algumas etapas de pré-processamento são necessárias, as quais incluem correção da linha de base, alinhamento, classificação, normalização e escala, descritas com mais detalhes na subseção 2.5.3 (BRENNAN, 2014; GRIFFIN; MERZABAN, 2013).

Figura 8 – Estudo metabolômico envolvendo Ressonância Magnética Nuclear



Fonte: Adaptado de (BRENNAN, 2014).

2.5.3 Tratamento de dados

2.5.3.1 Pré-processamento de dados

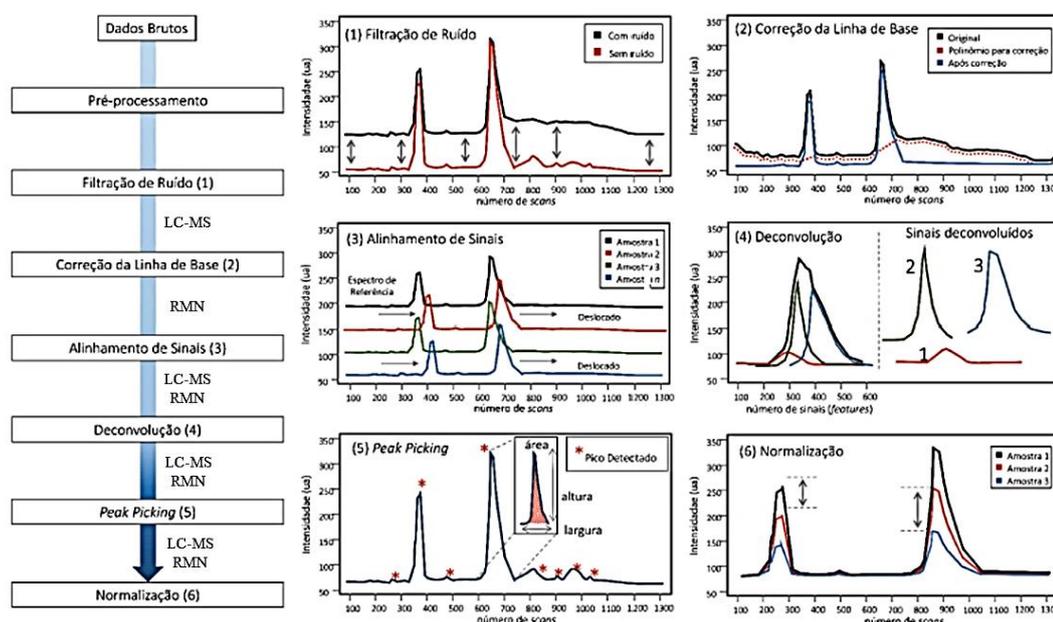
O pré-processamento é a etapa que fica no intermédio da obtenção de dados pela plataforma analítica utilizada e a análise estatística. Nesse momento, a melhoria da qualidade

dos sinais e a redução da interferência de ruídos são os principais objetivos, levando o conjunto de dados a ser comparável entre si (PILON et al., 2020).

Esse procedimento é necessário para assegurar a otimização dos resultados, garantindo que os picos sejam detectados e quantificados com alta qualidade. Este consiste na transformação dos dados em uma matriz robusta, na qual é possível descartar alterações que possam vir a ocorrer em virtude de variações presentes na amostra e no equipamento, como divergência na concentração de sais, temperatura e pH (PILON et al., 2020).

Algumas funções de pré-processamento só podem ser efetuadas em técnicas analíticas específicas, como, por exemplo, o faseamento e referenciamento, característico dos espectros de RMN. Outras, de forma contrárias, podem ser realizadas em todos os tipos de matrizes de dados, as quais incluem filtração de ruído, correção da linha de base, alinhamento de sinais, deconvolução, *peak picking* e normalização (Figura 9) (CUBERO-LEON; PEÑALVER; MAQUET, 2014; PILON et al., 2020).

Figura 9 - Etapas de pré-processamento que podem ocorrer no conjunto de dados estudado em metabolômica



Fonte: Adaptado de (PILON et al., 2020).

A etapa de filtração pode ser executada em sinais advindos de técnicas cromatográficas e tem como objetivo atenuar o ruído de interferências instrumentais, utilizando, para esse propósito a filtração combinada ou a filtração pela mediana em uma janela de tamanho

determinado. É recomendado que essa função se baseie em um espectro de massas em que não haja sinais de interesse, impedindo que os sinais referentes às moléculas minoritárias ou fracamente ionizáveis sejam removidas (PILON et al., 2020).

Em relação aos espectros obtidos por RMN, em especial os espectros unidimensionais, distorções da linha de base são frequentemente obtidas a partir das mudanças dos pontos primários do FID, os quais acrescentam modulações de baixa frequência por toda extensão da faixa espectral. A correção dessas distorções é uma etapa primordial, uma vez que a oscilação das intensidades pode gerar resultados estatísticos e quantitativos de forma errônea, prejudicando a análise biológica (PILON et al., 2020).

Em metabolômica, a correção da linha de base é realizada diretamente no espectro de RMN, levando em consideração as características da matriz quando os algoritmos a serem utilizados são escolhidos. Estes constroem uma linha de base teórica e subtraem os espectros reais para minimizar as distorções, exibindo modificações apenas na forma em que os cálculos e a modelagem da linha base são realizados (PILON et al., 2020).

O alinhamento dos picos detectados visa remover as mudanças entre as amostras de um determinado sinal para garantir a extração de informações úteis. Durante a aquisição dos dados de LC-MS, variáveis como pH, mudanças na temperatura, vazão ou concentração da fase móvel, presença de sais, efeito de matrizes, degradação da coluna cromatográfica entre outras, podem acarretar em variações não-lineares quando um mesmo analito for eluído, afetando de modo negativo as etapas subsequentes (PILON et al., 2020; YI et al., 2016).

Essas variações podem ser minimizadas por algoritmos de alinhamento que têm como base valores definidos de tempos de retenção para um composto específico, ou no momento da detecção dos picos, presumindo a variação dos sinais iguais por todo o cromatograma do sistema biológico (PILON et al., 2020; YI et al., 2016).

Em RMN, o alinhamento é uma das primeiras etapas no estágio de pré-processamento de dados, visto que mudanças de pico são comuns nos espectros de RMN de amostras biológicas, podendo impedir a descoberta de padrões diferenciais em um conjunto de dados. Essas variações, assim como na LC-MS, acontecem devido variações nas amostras, como pH e efeitos de matrizes, assim como da própria técnica, prejudicando que tendências entre os espectros sejam encontradas (BRENNAN, 2014; PILON et al., 2020; VU; LAUKENS, 2013).

Assim, os espectros são inicialmente alinhados com um padrão interno em cada espectro, útil na remoção de mudanças globais do conjunto de dados. Em seguida, um alinhamento fino ou local é realizado, utilizando algoritmos que se baseiam, principalmente, em métodos de distorção (BRENNAN, 2014; PILON et al., 2020; VU; LAUKENS, 2013).

A deconvolução é entendida como uma resolução matemática de picos ou sinais sobrepostos, decompondo um pico em dois ou mais modelos de pico pré-determinados. Essa ferramenta amplia a capacidade experimental, economizando tempo, mão de obra e dinheiro para adquirir informações sobre os analitos-alvo (PIERCE et al., 2012; ZENG et al., 2014).

A detecção de picos pelo processo de *peak picking* possibilita identificar um sinal em tempo de retenção específico ou valor de m/z , demonstrando a intensidade máxima dos pontos e a soma respectiva soma destes, equivalente à sua área (PILON et al., 2020).

O principal objetivo quanto da detecção de um sinal é a extração de dados para a caracterização e quantificação dos metabólitos presentes em um sistema biológico, devendo identificar sinais verdadeiros e evitar falsos positivos para reduzir a dimensão dos dados e tornar a análise mais viável (PILON et al., 2020).

Por fim, a normalização é aplicada para corrigir variações que são inevitavelmente introduzidas durante o preparo e injeção da amostra, as quais intervêm na interpretação química e biológica na análise estatística. A supressão dessas variações possibilita, por exemplo, comparar sinais entre duas amostras com diferentes concentrações, sendo fundamental o uso de um sinal de referência para atingir tal objetivo (PIERCE et al., 2012; PILON et al., 2020).

No que tange à normalização dos sinais de RMN, a técnica padrão é a normalização por área total, podendo ser empregado outros métodos de acordo com a complexidade das matrizes, como a normalização por padrão interno, por quociente probabilístico, por histograma ou por agrupamento de *clusters* (PILON et al., 2020).

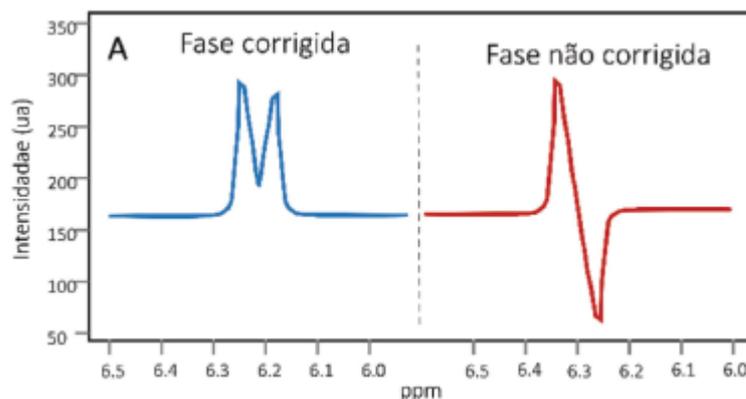
Como mencionado anteriormente, duas etapas podem ser acrescentadas ao pré-processamento de dados obtidos por RMN, sendo elas o referenciamento e o faseamento. Esse último é realizado após o cálculo da transformada de Fourier (FT), característico de análises metabolômicas que envolvem dados de RMN, convertendo os dados adquiridos no domínio tempo em domínio de frequência (PILON et al., 2020).

O referenciamento (ou calibração) busca resolver a variabilidade global dos deslocamentos químicos, possibilitando a analogia precisa entre os dados analíticos de diferentes amostras, frequentemente conseguida pelo deslocamento químico do sinal residual do solvente deuterado utilizado, bem como ao se adicionar um padrão interno de deslocamento conhecido à solução, como o tetrametilsilano (TMS) e hexametilsiloxano (HMDSO) (PILON et al., 2020).

O espectro também deve ser faseado para garantir que as linhas dos espectros sejam visualizadas com fases puras. A diferença de fases em um espectro ocorre, em especial, em

virtude da variabilidade do ponto de origem do FID, sendo corrigida pela alteração manual ou automática das fases de primeira e segunda ordem (Figura 10) (PILON et al., 2020).

Figura 10 - Faseamento de linhas espectrais de RMN, visando a obtenção de fases puras



Fonte: (PILON et al., 2020).

Quando os dados metabolômicos limpos e normalizados estiverem prontos, é importante usar o método de pré-tratamento de dados apropriado antes de iniciar a análise estatística, devendo ser reduzido os efeitos dos erros técnicos e de medição, enquanto é realizado o aprimoramento das variações biológicas relevantes (SETHI et al., 2017).

2.5.3.2 Pré-tratamento de dados

A escolha do método de pré-tratamento dos dados depende da questão científica e do método de análise a ser utilizado, desempenhando um papel importante na obtenção e interpretação dos resultados quando, por exemplo, métodos de análise multivariada são considerados (SETHI et al., 2017).

A análise de dados de RMN e LC-MS através de métodos de projeção de variáveis latentes, discutidas mais adiante, leva em conta o espectro ou perfil médio e a variedade biológica nos dados, seja ela qual for, o qual ocorre, frequentemente, de forma atenuada. Dessa forma, os métodos de centralização e escalonamento contemplam diferentes formas de analisar os dados, sendo estratégica frequentes nos estudos metabolômicos (PILON et al., 2020).

A centralização pretende-se remover o deslocamento dos dados e focar na variação biológica, bem como semelhanças/dissimilaridades entre as amostras nos dados. Os metabólitos mais abundantes exibem altos valores na tabela de dados e, subsequentemente, mostram

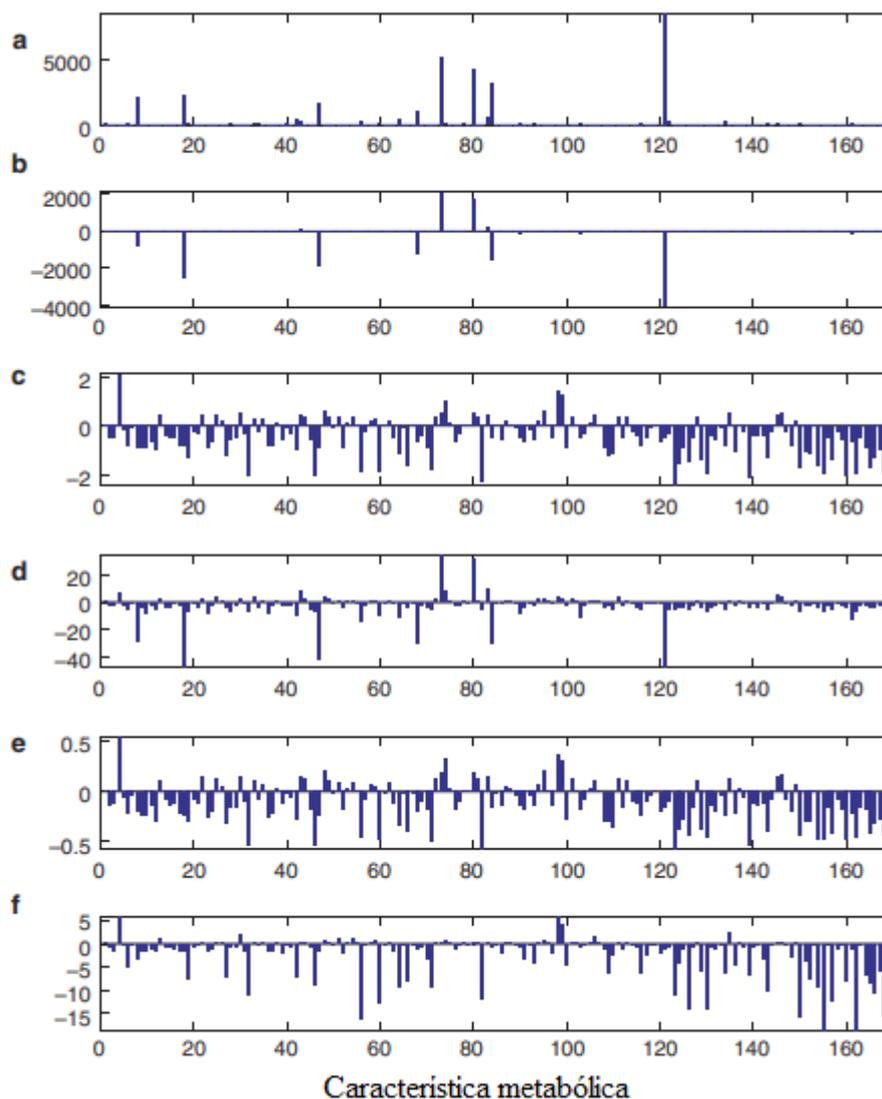
grandes diferenças entre as amostras em comparação com os metabólitos de baixa abundância (SETHI et al., 2017).

Entretanto, centralizar os dados por si só pode não ser suficiente para encontrar biomarcadores, uma vez que os metabólitos altamente abundantes contribuirão predominantemente para o modelo, podendo mascarar os metabólitos biologicamente importantes, mas com baixa abundância, e, assim, tornar tendencioso os resultados da análise estatística (SETHI et al., 2017).

Consequentemente, o dimensionamento de cada recurso da tabela de dados, que potencialmente corresponde a um metabólito, precisa ser cuidadosamente considerado. Assim, o escalonamento é fundamental para extrair informações adequadas, se destacando as estratégias de auto-escalonamento, escalonamento por Pareto, escalonamento por estabilidade de variável e escalonamento por escala de alcance (PILON et al., 2020; SETHI et al., 2017).

A Figura 11 mostra o efeito da centralização e do escalonamento para um conjunto de dados de LC-MS, onde em (a), a maioria dos picos parece ter pouca abundância, existindo alguns muito abundantes. Em (b), a centralização da média conduziu os picos à uma distribuição próxima à zero, informação relevante quando conjuntos de dados de diferentes plataformas analíticas são consideradas (MS e RMN juntos). Isso também é visível em (d), após escalonamento por Pareto, embora os picos de baixa abundância tenham sido “inflados” até certo ponto. Por outro lado, os picos parecem ser mais comparáveis entre si após o auto-escalonamento e escalonamento por escala de alcance, respectivamente Figuras 11c e 11e. Em (f), o escalonamento por estabilidade de variável parece mostrar dados mais comparáveis em relação aos dados centralizados na média e em escala de Pareto (PILON et al., 2020; SETHI et al., 2017).

Figura 11 - Comportamento da matriz de dados frente aos diferentes métodos de centralização e escalonamento: (a) sem tratamento; (b) centrado na média; (c) auto-escalonamento e (d) Pareto; (e) escalonamento por escala de alcance; (f) o escalonamento por estabilidade de variável



Fonte: (SETHI et al., 2017).

Após as etapas de pré-processamento e pré-tratamento de dados, o grande volume de dados produzidos é então abordado pelos métodos de análise quimiométrica, usadas para projetar, processar, visualizar, explorar e analisar os dados metabômicos, extraindo informações relevantes através de métodos de análise multivariada e univariada (CANUTO et al., 2018a; PILON et al., 2020; SETHI et al., 2017).

2.5.3.3 Métodos quimiométricos em metabolômica

A quimiometria pode ser definida como “a disciplina química que utiliza métodos matemáticos e estatísticos que empregam lógica formal para projetar ou selecionar procedimentos e experimentos de medição ideais, bem como para fornecer informações relevantes a partir da análise de dados químicos” (SETHI et al., 2017).

Por ser uma ferramenta dinâmica, com novos algoritmos criados e/ou alterados frequentemente, dispositivos tecnológicos com softwares voltados para o cálculo numérico são amplamente empregados, tendo como constituinte básico uma matriz de dados (SOUZA; RONEI JESUS POPPI, 2012).

Como os métodos analíticos geram uma grande quantidade de dados, a quimiometria tem sido fundamental para o desenvolvimento da metabolômica, uma vez que esta fornece métodos capazes de reduzir o dimensionamento do conjunto de dados sem que haja perda da qualidade das informações geradas (CORNEJO-BÁEZ et al., 2020; KUMAR; SHARMA, 2018; SETHI et al., 2017).

Para este fim, são empregados diferentes métodos multivariados de análise para a modelagem dos dados, levando em consideração a classificação e discriminação de variáveis de estudo responsáveis pela diferenciação entre os grupos amostrais, por meio da análise da matriz obtida (CANUTO et al., 2018b; PILON et al., 2020). De forma geral, os métodos multivariados projetam as amostras em um espaço de menor dimensão, representando-as como pontos (CANUTO et al., 2018a; PILON et al., 2020).

Diferentes técnicas de análise multivariada são utilizadas para a modelagem de dados, podendo estes serem não-supervisionados, os quais oferecem uma visão global do comportamento amostral, ou supervisionados, onde as variáveis são definidas em classes ou relacionadas a um valor específico (PILON et al., 2020).

Métodos não-supervisionados incluem análise de componentes principais (PCA, do inglês *Principal Component Analysis*) e HCA (do inglês, *Hierarchical Cluster Analysis*), já em relação aos métodos supervisionados, fazem parte a análise discriminante por mínimos quadrados parciais (PLS-DA, do inglês, *Partial Least Squares Discriminant Analysis*) e projeções ortogonais para estruturas latentes (OPLS-DA, do inglês, *Orthogonal Partial Least Squares Discriminant Analysis*) (CANUTO et al., 2018a; PILON et al., 2020; SOUZA; RONEI JESUS POPPI, 2012).

A PCA é a ferramenta não-supervisionada mais empregada como ferramenta de exploração em metabolômica, podendo ser visto como base para outros métodos multivariados.

Como não contém suposições sobre os dados, é usado como uma ferramenta de visualização e exploração no início de qualquer análise, a fim de detectar tendências, grupos e discrepâncias (SETHI et al., 2017).

Durante uma modelagem de PCA, a dimensionalidade dos dados é algebricamente reduzida em um novo sistema de eixos, denominados componentes principais (PCs), mantendo as informações presentes nos dados originais o máximo possível (SOUZA; RONEI JESUS POPPI, 2012).

A redução dos dados permite que uma matriz multidimensional seja projetada em um espaço de baixa dimensão, proporcionando visualização interpretável dos dados complexos e destacando, como consequência, possíveis semelhanças ou diferenças. No entanto, não é um método de classificação, mas permite que os usuários entendam a estrutura e os padrões em um determinado conjunto de dados (CORNEJO-BÁEZ et al., 2020; KUMAR; SHARMA, 2018; LUBES; GOODARZI, 2017; PILON et al., 2020; SETHI et al., 2017).

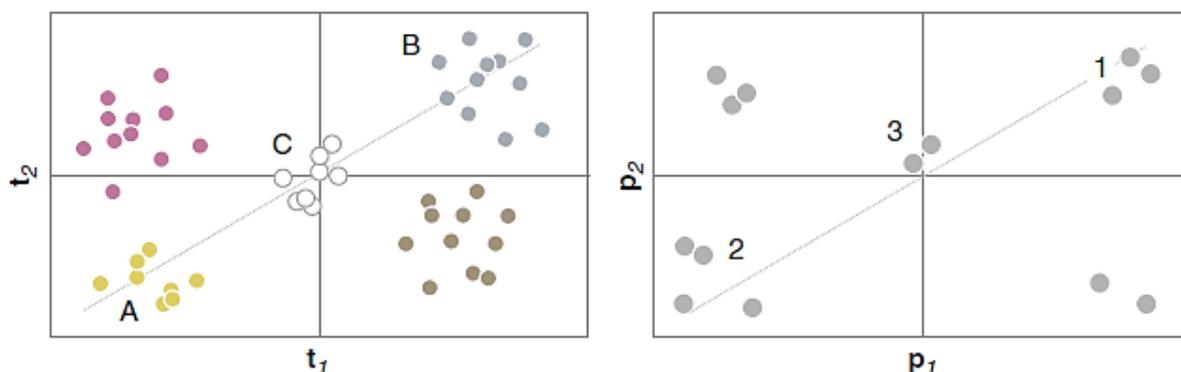
Na PCA, a matriz de dados original é convertida em duas matrizes distintas, *scores* e *loadings*. Os *scores* representam as coordenadas das amostras no sistema de eixos formados pelos componentes principais, já os *loadings* refletem a maneira que estas variáveis estão combinadas nas PCs, demonstrando o quanto cada variável original está contribuindo para a formação desta (SOUZA; RONEI JESUS POPPI, 2012).

A distribuição espacial das amostras pode ser observada em um gráfico de *scores*, onde amostras semelhantes são agrupadas e amostras diferentes se separam, enquanto o gráfico de *loadings* ajuda a detectar as variáveis responsáveis pela formação do agrupamento de um determinado conjunto de dados (CORNEJO-BÁEZ et al., 2020; LUBES; GOODARZI, 2017; SOUZA; RONEI JESUS POPPI, 2012).

Os eixos Y e X de um gráfico de *scores* são denominados PC1 e PC2, respectivamente, e ilustram a variação dentro e entre os grupos. O primeiro componente principal (PC1) é o que melhor representa os dados, enquanto o segundo (PC2) geralmente contém a segunda variação máxima de dados. Às vezes, no entanto, é necessário observar um espaço tridimensional para evitar uma decisão baseada em apenas dois PCs, devido a sobreposição da amostra (LUBES; GOODARZI, 2017).

O gráfico de *loadings*, por outro lado, fornece um valor numérico que reflete como cada variável original contribui para o gráfico de *scores* e o quanto isso tem em comum com um componente específico. Por meio da avaliação simultânea dos gráficos de *scores* e *loadings* é possível identificar quais variáveis contribuem para diferenciar as amostras. Um exemplo de análise de PCA é representado na Figura 12 (LUBES; GOODARZI, 2017).

Figura 12 – Gráficos de *scores* e *loadings*, respectivamente, inspecionados simultaneamente a fim de entender as relações e tendências entre grupos observados nas amostras e quais variáveis (metabólitos) são responsáveis



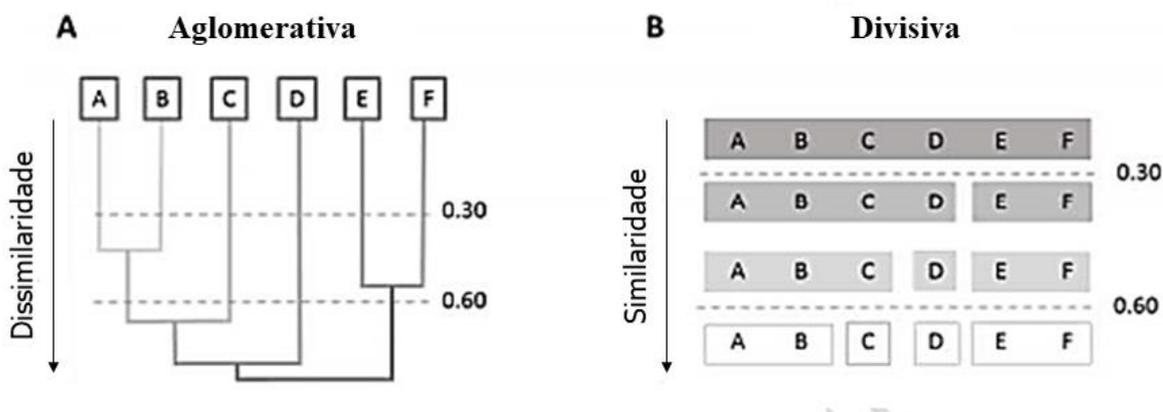
Fonte: (SETHI et al., 2017).

Como já mencionado, a análise hierárquica de cluster (HCA) é uma técnica exploratória, também não-supervisionada, que baseia-se na criação de estruturas ramificadas com uma hierarquia definida, denominada dendrograma, o qual permite uma visualização qualitativa, em um espaço bidimensional, do agrupamento entre amostras de acordo com as medidas de similaridade utilizadas, onde a distância euclidiana é usada como medida de similaridade para explicar o esquema de agrupamento (KUMAR; SHARMA, 2018; LUBES; GOODARZI, 2017). Geralmente, é usado como uma ferramenta de agrupamento para avaliar semelhanças e diferenças intra e intergrupos, sendo semelhante ao PCA. (CORNEJO-BÁEZ et al., 2020; PILON et al., 2020).

As duas principais estratégias para comparar amostras no HCA são a estratégia aglomerativa e a divisiva. A estratégia aglomerativa separa cada caso em seu próprio *cluster* (agrupamento) individual na primeira etapa, para que o número inicial de *clusters* seja igual ao número total de casos. Em etapas sucessivas, casos semelhantes, ou *clusters*, são mesclados até que todos os casos sejam agrupados em um único *cluster* (CORNEJO-BÁEZ et al., 2020; GRANATO et al., 2017; LUBES; GOODARZI, 2017).

Já a estratégia divisiva funciona de maneira inversa, com todos os casos iniciando em um *cluster* grande e gradualmente sendo separados em grupos de *clusters* até que cada caso esteja em um *cluster* individual. Essa última técnica raramente é utilizada devido à sua pesada carga computacional (Figura 13) (CORNEJO-BÁEZ et al., 2020; GRANATO et al., 2017; LUBES; GOODARZI, 2017).

Figura 13 - Dendrograma HCA para algoritmo aglomerativo (A) e fluxo de agrupamento de algoritmo divisivo (B)



Fonte: Adaptado de (GRANATO et al., 2017).

De forma contrária à análise não-supervisionada, nos métodos supervisionados o conjunto de dados é equiparado a um modelo, de forma indispensável, que deve ser robusto e estatístico, produzido por meio de dados biológicos, taxonômicos ou químicos. Durante esse processo, as variáveis são forçadas a se separarem com base no modelo a que foi comparado, sendo considerada uma classificação tendenciosa (PILON et al., 2020).

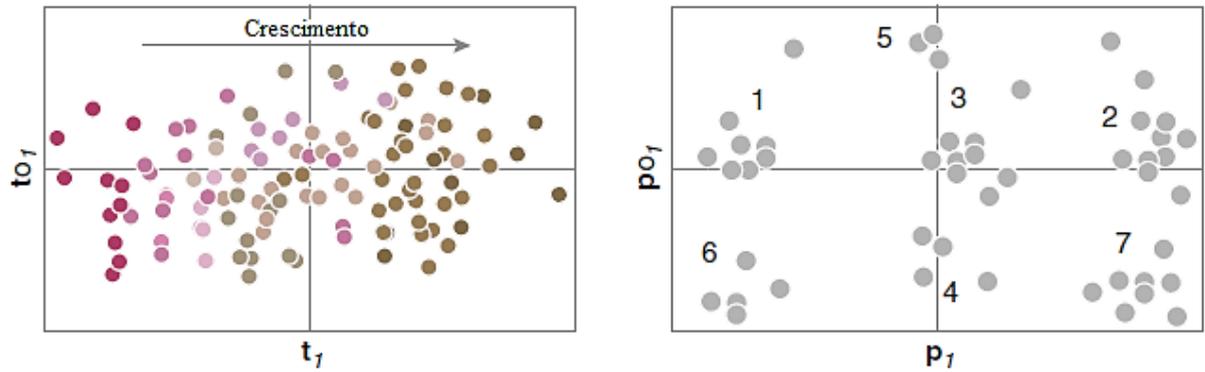
O objetivo é correlacionar as variáveis químicas à um conjunto de dados pré-definido. Em conformidade à finalidade de estudo e do tipo de cálculo empregado, diversos métodos podem ser utilizados para discriminar, classificar ou fazer regressão linear, como por exemplo a PLS-DA, amplamente utilizada na metabolômica na descoberta de metabólitos bioativos (PILON et al., 2020).

Esse último é baseado em mínimos quadrados parciais e utiliza as variáveis independentes de uma matriz "X" e as variáveis categóricas "Y" da amostra predefinida para desenvolver um modelo de treinamento, e a participação no grupo de amostra desconhecida pode ser preditiva usando o valor dos mínimos quadrados parciais de amostras desconhecidas. O modelo obtido tem como característica o R^2 , responsável por julgar o modelo quanto à sua adequação ao conjunto de dados de treinamento (KUMAR; SHARMA, 2018; PILON et al., 2020).

O OPLS, uma extensão da PLS, tem o mesmo poder preditivo, mas fornece uma melhor interpretação das variáveis relevantes. Isso é feito decompondo os dados nas chamadas informações "preditivas" relacionadas à resposta Y (como concentrações, classes), informações estruturadas "ortogonais" não relacionadas à resposta (como variações instrumentais,

biológicas) e variação residual. O OPLS é o método linear multivariado de escolha para, por exemplo, encontrar biomarcadores metabólicos correlacionados à uma variável contínua (SETHI et al., 2017).

Figura 14 - Gráficos de *scores* OPLS (esquerda) e *loadings* (direita) para o componente preditivo 1 *versus* componente ortogonal 1, com pontuações coloridas pela variável y.



Fonte: Adaptado de (SETHI et al., 2017).

3 OBJETIVOS

3.1 Objetivo geral

Estudar a variação dos metabólitos primários e secundários, identificados por RMN, *versus* variação genética de um banco ativo de germoplasma contendo acessos de folhas e cascas de *Spondias mombin*, bem como associar a análise multivariada aos resultados da matriz abordada.

3.2 Objetivos específicos

- Realizar estudo de variabilidade genética entre os acessos do BAG de cajá;
- Caracterizar metabólitos primários e secundários presentes nas folhas e cascas do BAG de cajazeira, a partir da Ressonância Magnética Nuclear;
- Investigar a variabilidade química dos extratos das cascas e folhas de *S. mombin* separadamente, bem como a relação entre os compostos e as amostras, por meio da Análise de Componentes Principais (PCA).

CAPÍTULO 1

Untargeted NMR-based metabolomics to study barks and leaves from *Spondias mombin* L. of an active germplasm bank

**Artigo à ser submetido na revista
Biochemical Systematics and Ecology**

1 **Untargeted NMR-Based metabolomics applied to study of leaves and bark of the**
2 **active germplasm bank of *Spondias mombin* L. using multivariate analysis tools**
3

4 Gladiane S. Silva^a, Elenilson G. Alves Filho^b, Christiane Mendes C. Ramires^c, Yandra
5 Karine L. Bezerra^c, Maria Francilene S. Silva^d, Jhonyson Arruda C. Guedes^a, Gisele S.
6 Silva^e, Lorena Mara A. Silva^e, Paulo Riceli^e, Kirley Canuto^e, Edy S. de Brito^e, Ricardo
7 Elesbão^f, Dávila Zampieri^a, Guilherme J. Zocolo^e
8

9 ^a Departamento de Química Orgânica e Inorgânica, Centro de Ciências, Universidade Federal do Ceará,
10 Campus do Pici, CEP 60451-970, Fortaleza, Ceará, Brasil.

11 ^b Departamento de Engenharia de Alimentos, Centro de Ciências Agrárias, Universidade Federal do Ceará,
12 Campus do Pici, CEP 60356-000, Fortaleza, Ceará, Brasil.

13 ^c Empresa Estadual de Pesquisa Agropecuária da Paraíba, Rodovia BR-230, S/N, Parque Esperança CEP
14 58108502, Cabedelo, Paraíba, Brasil.

15 ^d Núcleo de Pesquisa e Desenvolvimento de Medicamentos, Universidade Federal do Ceará, Rua Coronel
16 Nunes de Melo 1000, Rodolfo Teófilo, CEP 60420-275, Fortaleza, Ceará, Brasil.

17 ^e Embrapa Agroindústria Tropical, Rua Doutora Sara Mesquita 2270, Pici, CEP 60511-110, Fortaleza,
18 Ceará, Brasil.

19 ^f Empresa Brasileira de Pesquisa Agropecuária, Embrapa Alimentos e Territórios, Rua Buarque de Macêdo
20 433-B, Centro, CEP 57020520, Maceió, Alagoas, Brasil.

21

22

23

24

25

26 Corresponding author

27 E-mail: guilherme.zocolo@embrapa.br

28 Phone: +55 (85) 3391-7360

29 ABSTRACT

30 The exploitation of fruit trees in the Northeast region of Brazil occurs in the great majority, in an
31 extractive manner, using inadequate management techniques that lead to the loss of genetic
32 resources and genotypes containing bioactive compounds of pharmacological importance. This
33 practice is repeated with the cajazeira tree, whose cultivations in domestic and commercial
34 orchards are few. As it is a species still in domestication, cytogenetically little known, and with
35 high productive potential in the regions of origin, it is hugely significant to establish strategies for
36 conserving its genetic variability in order to avoid the risk of genetic erosion and maintain it
37 available for use in plant breeding programs. Thus, in the present work, a chemical profile study
38 of the active germplasm bank (AGB) was carried out containing 22 accessions of cajazeira
39 (separated between leaves and barks) divided between clones and progenies. To contribute to
40 generating knowledge and adding value to the AGB, the non-target metabolomics technique was
41 applied through an approach based on Nuclear Magnetic Resonance to elucidate the accessions
42 and genotypes' chemical profiles. It was observed the presence of metabolites of
43 chemical/pharmacological interest with possible technical, medicinal, and nutritional
44 applications, among which quercetin, ellagic acid, and gallic acid can be mentioned. Sugar
45 molecules and amino acids already reported as markers of resistance to stress conditions were
46 also found in abundance in the leaves and bark. Multivariate analysis was used to investigate the
47 chemical variability between accessions and correlate the described metabolites with the place
48 and time of collection. Thus, the results obtained justified the establishment of a chemical base
49 associated with the active germplasm bank, which can be useful to subsidize programs of genetic
50 improvement of the cajazeira tree, essential to increase the agricultural productivity and
51 nutritional quality of the crops in the face of adverse conditions to which they are subjected.

52

53 **Keywords:** Conservation of genetic resources. Structural elucidation. Multivariate
54 analysis. Plant breeding.

55 1. Introduction

56

57 The cajazeira (*Spondias mombin* L.) is a fruit tree belonging to the Anacardiaceae
58 family, native to the most northern humid forests in South America. In Brazil, their centers of
59 diversity are the Atlantic Forest and the Western Amazon [1].

60 This species has an erect stem, with up to 25 m in height, trunk with a diameter between
61 20-40 cm, stem bark gray or light brown, and moderately rough, sparse and extensive branches.
62 Its fruits are of the drupe type with ellipsoid, oblong-ellipsoid, or oblong-ovoid shapes and
63 aromatic, acidic, tasty, edible, and healthy pulp, with excellent market acceptance [1-2].

64 In the North and Northeast regions of Brazil, *Spondias mombin* has been domesticated
65 and exploited in an extractive manner, where it has economic importance due to the
66 commercialization of fruits in open markets and by the industry where they are processed into
67 frozen pulps [3-4]. In Paraíba, it occurs in the mesoregions of Agreste (with temperature varying
68 between 25 and 28 °C, relative humidity between 60 and 80%, rainfall between 700 and 1.600
69 mm/year) and the Coast (with temperature varying between 23 and 24 °C, precipitation between
70 1.100 to 1.800 mm/year and humidity around 80%) [2].

71 In folk medicine, all parts of the cajazeira tree are used to treat various diseases, making
72 up a promising source of bioactive compounds to produce medicines [5]. Based on this
73 information, many scientific studies have been carried out to confirm these biological, as
74 anticancer activities [6] gastroprotective [7-9], antimycobacterial [10-11], antimicrobial [12],
75 anti-inflammatory and antioxidant [13-14], antidiabetic [15-16], cardioprotective [6],
76 hepatoprotective [18], among others. Phytochemical studies have shown the presence of
77 numerous classes of metabolites in the leaf and bark of the tree, such as phenolic acids,
78 flavonoids, saponins, sterols, terpenes, alkaloids, oxalates, phthalates, reducing sugars, cardiac
79 glycosides, cyanogenic glycosides, chlorogenic acid, and tannins [7, 8, 14, 18-22].

80 Even with this potential, anthropic actions and climate change have been causing genetic
81 erosion of this species in its dispersal zones, justifying the introduction of collections and
82 germplasm banks which are repositories where the variability of a species is stored [23-24].

83 Some official research institutions have maintained ex situ collections, such as Embrapa
84 Amazonia Oriental, in Belem, PA, with 30 accesses [25]; Embrapa Agroindustry Tropical-CE,
85 with 17 accesses and a total of 110 plants [26]; the Agronomic Institute of Pernambuco (IPA),
86 in Itambe, PE, with 33 accesses [27] and a germplasm bank from Emepa, PB, in João Pessoa,
87 PB, with 34 accesses and 73 plants [28].

88 Despite the pharmacological and phytotherapeutic potential that this species has, studies
89 identifying the bioactive compounds in the *Spondias mombin* accessions that make up these
90 collections and these germplasm banks are incipient. Chemical substances and active
91 ingredients in appropriate concentrations can be used to separate the accessions in order to know
92 the variability between accessions of the same species [29]. Such chemical characterization can
93 also be used as a strategy to intensify the use of conserved plant genetic resources which are
94 the fuel of genetic improvement in the search for greater productivity and superior genotypes
95 to serve the market [30-31].

96 The use of analytical techniques to investigate metabolic profiles of conserved genetic
97 materials, as well as to study metabolic changes in the accessions of interest, is an extremely
98 rich and innovative approach in the construction of AGB, establishing a more strategic direction
99 from the point of view functional [32-33].

100 Due to the complex chemical profile of plants, the analysis of metabolic profiles has
101 been particularly challenging [34]. Thus, metabolomics has stood out in the search for natural
102 products, with regard to the biotechnological optimization of the discovery and validation of
103 active pharmacological metabolites [35].

104 To simplify and accelerate this process, analytical techniques have been used together
105 with statistical tools [36-38]. Among the techniques applied to the mapping of the metabolic
106 constitution of different species, Nuclear Magnetic Resonance (NMR) stands out for its
107 simplicity, comprehensiveness, and robustness, requiring little or no manipulation of the sample
108 for quantification, in addition to being a specific and reproducible methodology for the
109 structural elucidation of various compounds [38-40].

110 In addition to NMR, multivariate statistical analysis is used to extract relevant
111 information from complex data sets, being responsible for the classification and discrimination
112 of groups of samples by their respective metabolites, as well as for the correlation between the
113 metabolic profile and a given biological activity [38, 40].

114 Given the above, the objective was to discuss the variation of primary and secondary
115 metabolites, identified by NMR, versus the genetic variation of an active germplasm bank
116 containing accessions related to the leaves and barks of *Spondias mombin*, as well as to
117 associate the multivariate analysis of the results of the matrix addressed.

118

119 **2. Material and methods**

120

121 *2.1 Plant material*

122

123 Leaves and bark of *Spondias mombin* accessions, progenies and clones were acquired from
124 the Germplasm Active Bank located at the Experimental Scientist Station José Irineu Cabral,
125 belonging to the Paraíba Company of Research, Rural Extension, and Land Regularization -
126 EMPAER, in João Pessoa, PB.

127 For the construction of the active germplasm bank, 44 accessions were used, 22 of which
128 were references to the leaves, and 22 references to the barks of the plant in question (Table 1).

129 Two collections were made at different times, the first collection was made in 2017, and the
130 second in 2018 (Table 1).

131 The identification and origin of the accesses are shown in Table 1. The progenies are located
132 under the coordinates: 7°11'50,37 "O and 34°48'47,46" W, altitude: 350m, and the clones are
133 located under the coordinates: 7°11'55,29 "O and 34°48'40,26" W, altitude: 341m.

134 The material was dried in the air circulation oven, at 40 °C, for three days and then
135 grinded on the Fortinox STAR FT-80 knife mill (Piracicaba, Brazil). Subsequently, the dry
136 plant material was packed as described by (GUEDES et al., 2020), considering the existing
137 genetic variability.

138

139 Table 1. Coding, procedence, and collection data of clones and progenies of leaves and barks
140 of *Spondias mombin* L.

Propagation	Procedence	First collection			Second collection		
		Collection data	Leaves	Barks	Collection data	Leaves	Barks
	Ingá - PB*	05/17/17	P12L1	P12B1	07/17/18	P12L2	P12B2
	Ingá - PB*	06/19/17	P13L1	P13B1	-	-	-
	João Pessoa – PB**	06/19/17	P15L1	P15B1	07/17/18	P15L2	P15B2
	Ingá - PB*	05/17/17	P23L1	P23B1	-	-	-
	João Pessoa – PB**	06/19/17	P24L1	P24B1	07/17/18	P24L2	P24B2
	Cruz do Espírito Santo – PB**	06/19/17	P30L1	P30B1	-	-	-
	Caruaru – PE*	05/17/17	P34L1	P34B1	07/17/18	P34L2	P34B2
Progenies	João Pessoa – PB**	05/17/17	P36L1	P36B1	07/17/18	P36L2	P36B2
	Ingá – PB*	05/17/17	P53L1	P53B1	-	-	-
	Ingá – PB*	05/17/17	P76L1	P76B1	07/17/18	P76L2	P76B2
	João Pessoa – PB**	05/17/17	P82L1	P82B1	07/17/18	P82L2	P82B2
	João Pessoa – PB**	06/19/17	P91L1	P91B1	07/17/18	P91L2	-
	João Pessoa – PB**	06/19/17	P102L1	P102B1	07/17/18	P102L2	P102B2
	Cruz do Espírito Santo – PB**	05/17/17	P252L1	P252B1	07/17/18	P252L2	P252B2
	Cruz do Espírito Santo – PB**	05/17/17	-	P293B1	07/17/18	P293L2	P293B2

Clones	João Pessoa – PB**	05/24/17	C11L1	C11B1	07/16/18	C11L2	C11B2
	Areia – PB*	05/24/17	C17L1	C17B1	07/16/18	C17L2	C17B2
	Areia – PB*	05/24/17	C18L1	C18B1	07/16/18	C18L2	C18B2
	Areia – PB*	05/24/17	C35L1	C35B1	07/16/18	C35L2	C35B2
	Ingá – PB*	06/19/17	C61L1	C61B1	07/16/18	C61L2	C61B2
	Areia – PB*	05/24/17	C162L1	C162B1	07/16/18	C162L2	C162B2
	Cruz do Espírito Santo – PB**	05/24/17	C252L1	C252B1	07/16/18	C252L2	-

141 * Agreste Mesoregion

142 ** Mata Paraibana Mesoregion

143

144 2.2. NMR analysis

145

146 For NMR analysis, approximately 30 mg of the plant material was solubilized in 700
 147 μ L of deuterated methanol, purchased by Sigma-Aldrich Brazil Ltda (Sao Paulo, SP, BR). The
 148 solution was placed in ultrasonic bath for 2 min for complete solubilization of the plant material
 149 and then, inserted in a micro-centrifuge with rotation for 1 min. The final solution was filtered
 150 and inserted into 5 mm NMR tubes.

151 The NMR experiments were performed on an Agilent 600-MHz spectrometer equipped
 152 with a 5 mm (H-F/¹⁵N-³¹P) inverse detection One Probe™ with actively shielded z-gradient.
 153 The ¹H NMR spectra were acquired using PRESAT pulse sequence for solvent signal
 154 suppression, since this pulse sequence presented less effect in surrounding saturation region
 155 according to the saturation profiling. The spectral data were acquired with 32 scans using
 156 96,000 of time domain points for a spectral window of 16.0 ppm, acquisition time of 5.0 s and
 157 recycling delay of 35.0 s, and controlled temperature to 298 K. The spectral processing was
 158 performed applying exponential Lorentzian broadening of 0.3 Hz and zero filling to 16k points
 159 before Fourier transformation. The phase correction was performed manually, and the
 160 automatic baseline correction was applied over the entire spectral range. The spectra were
 161 referenced by sodium-3-trimethylsilyl propionate signal (TMSP-d₄) – singlet at δ 0.0 for ¹H.

162 The identification of the constituents within the matrices was performed through 2D-
163 NMR evaluation using the correlation spectroscopy (gCOSY), heteronuclear single quantum
164 coherence (gHSQC), heteronuclear multiple bond correlation (gHMBC), assessments using an
165 open access database (www.hmdb.ca), and literature reports (41-43). Two-dimensional (2D)
166 NMR experiments were acquired using the standard spectrometer library pulse sequences. The
167 ^1H - ^1H gCOSY experiments were obtained with spectral width of 7,267.4 Hz in both
168 dimensions; $1\text{k} \times 200$ data matrix; 16 scans per t1 increment and relaxation delay of 1.0 s. One-
169 bond ^1H - ^{13}C gHSQC experiments were acquired with an evolution delay of 3.425 ms for an
170 average $^1J(\text{C,H})$ of 146 Hz; $1\text{k} \times 200$ data matrix; 32 scans per t1 increment; spectral widths of
171 9,615.4 Hz in f2 and 30,165.9 Hz in f1 and relaxation delay of 1.0 s. Long-range ^1H - ^{13}C
172 gHMBC experiments were recorded with an evolution delay of 62.5 ms for $^{\text{LR}}J(\text{C,H})$ of 8 Hz;
173 $1\text{k} \times 200$ data matrix; 64 scans per t1 increment; spectral width 9,615.4 Hz in f2 and 36,199.1
174 Hz in f1 and relaxation delay of 1.0 s.

175

176 2.3. Chemometric analysis

177

178 Chemometric analysis was applied to investigate the composition variability of extracts
179 from barks and leaves of *S. mombin* separately, as well as the relationship between compounds
180 and samples. Therefore, one numerical matrix was constructed for each part of the *S. mombin*
181 (barks and leaves) using the ^1H NMR spectral region between the chemical shifts δ 0.4 and 8.5.
182 For the matrices construction, each spectrum was converted to American Standard Code for
183 Information Interchange (ASCII) file and imported using the OriginTM software 9.4 and then,
184 exported to PLS Toolbox software (8.6.1 – Eigenvector Research Inc., Wenatchee, USA) for
185 unsupervised chemometric method by PCA (Principal Component Analysis) in order to explore
186 both datasets (barks and leaves).

187 Algorithms for baseline correction (with automatic weighted least squared, order 2),
188 variables alignment using COW (Correlation Optimized Warping) with segment length of 50
189 data points and slack of 5 data points, and normalization by area were applied previously to
190 matrices decomposition. After the data pretreatment, the samples were mean-centered and the
191 Singular Value Decomposition (SVD) algorithm was applied to decompose each matrix for
192 PCA modeling. The relevant information was obtained at the first two Principal Components
193 (PC) with confidence level of 95% for both barks and leaves analyses.

194

195 **3. Results and discussion**

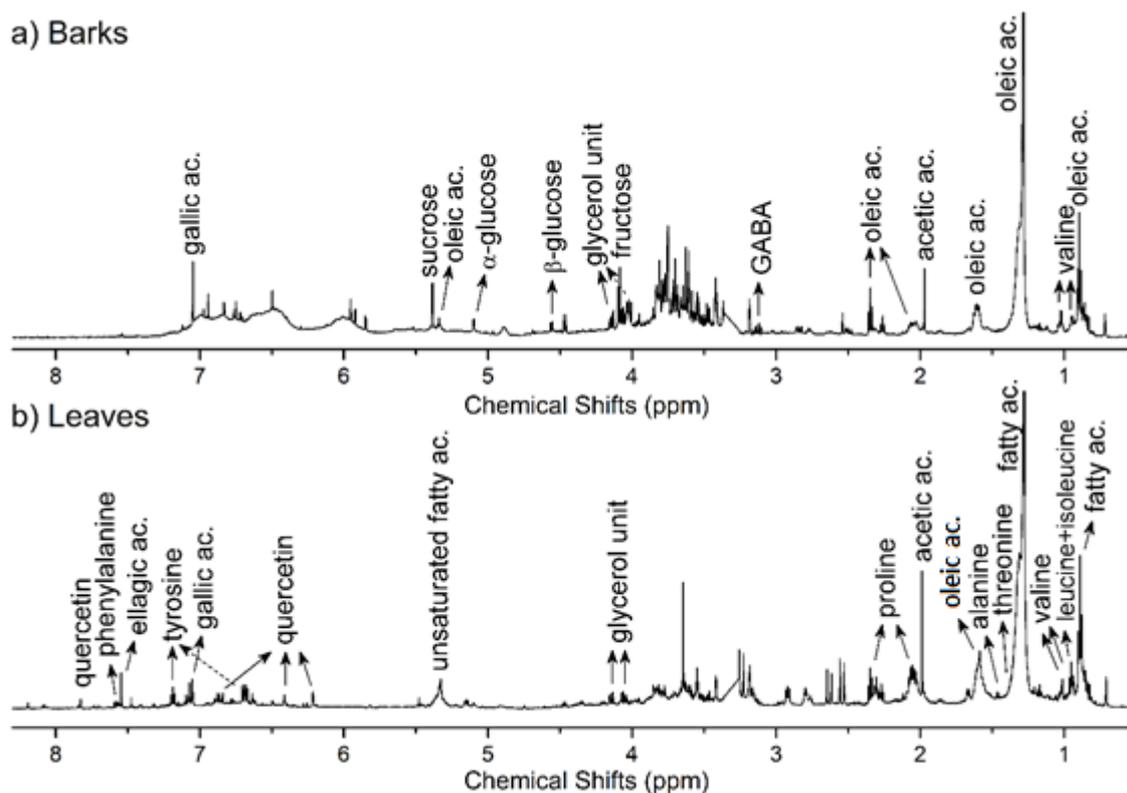
196

197 *3.1. Metabolite profiling of barks and leaves*

198

199 Since the metabolite assignment is an important part of the most complex matrix studies,
200 initially it was developed the metabolomic profiling of barks and leaves from different
201 genotypes of *S. mombin*. Figure 1 presents representative ^1H NMR spectra from extracts of
202 barks (a) and leaves (b) separately, with the identified organic compounds. As expected,
203 different classes of primary metabolites, such as amino acids, sugars and organic acids were
204 identified in both barks and leaves extracts, which are also illustrated in Figure 1. Table 2 and
205 3 describes the NMR identification parameters, such as the ^1H and ^{13}C chemical shifts,
206 multiplicity and constant coupling (made available as Supplementary Material).

207



208

209 Figure 1. Representative ^1H NMR spectra from extracts of the barks (a) and leaves (b) of *S.*
 210 *mombin*.

211

212 According to Figure 1, the leaves of *Spondias mombin* have a chemical profile richer in
 213 quercetin and ellagic acid when compared to the barks, both polyphenols of increasing interest
 214 in technical, medicinal, and nutritional applications, due to their multiple beneficial effects on
 215 the health of human being, being considered potent and promising molecules [44].

216

217 Among the various biological effects present, the ability of these molecules to eliminate
 218 a wide variety of free radicals and prevent cell damage stands out, constituting efficient
 219 multifunctional protectors against oxidative stress and thus preventing the plant from
 220 undergoing reversible or irreversible modifications of biomolecules, such as proteins,
 polynucleic acids, carbohydrates, and lipids [44-57].

221 In addition to the potent oxidative capacity, which implies most of their pharmacological
222 activities, these molecules act as anti-inflammatory, antidiabetic, antidepressant,
223 chemopreventive, hepatic, gastro, neuro, and cardioprotective, immunomodulatory, antiviral,
224 antimicrobial, and anti-obesity agents, being important targets for the development of new
225 cultivars and breeding programs from a nutritional and agronomic point of view. In addition,
226 these have revealed prospective industrial importance for the synthesis of new bioengineering
227 materials [44-57].

228 In more recent studies, the combination of the therapeutic capacity of quercetin and
229 ellagic acid has been associated with coronavirus studies, reported as molecules that are
230 potential candidates for pre-treatment, co-treatment, and post-treatment in inhibiting the SARS-
231 CoV-2 virus [58-61].

232 The amino acids phenylalanine, tyrosine, proline, alanine, threonine, leucine, and
233 isoleucine were also identified in greater concentration in the leaves of *Spondias mombin* when
234 compared to the barks (Figure 1), playing, in general, in addition to the role of protein
235 constituent, the function of signaling molecules of several beneficial physiological processes in
236 the plant, such as growth and development, control of intracellular pH due to the presence of
237 acidic and basic groups, generation of metabolic energy or redox power, hormonal precursors,
238 response to stress and secondary metabolism [62-65].

239 Studies reveal that leucine and isoleucine contribute to plant growth, the response to
240 stress, and the production of flavor components in foods. In addition to being necessary for
241 protein biosynthesis and cell survival, phenylalanine also acts as a precursor to many
242 multifunctional secondary metabolites [66]. Tyrosine acts as a precursor to several secondary
243 metabolites whose physiological roles are involved in electron transport, the attraction of
244 pollinators, defense compounds, structural support, and antioxidants [67].

245 Proline has been the subject of countless studies because it is an amino acid that plays a
246 highly beneficial role in plants in response to stress conditions, among them, salinity, extreme
247 temperatures, the concentration of toxic metal ions, UV radiation, water, and nutrient
248 deficiency, which disrupt cellular redox homeostasis, alter metabolic interactions and impair
249 major plant physiological processes, reducing crop yields worldwide [68-71].

250 In addition to acting as an excellent osmolyte, proline plays the role of a metal chelator,
251 free radical scavenger, cellular redox balancer, energy source, nitrogen and carbon, the
252 stabilizer of subcellular structures and membranes, the activator of ROS detoxification
253 pathways and signaling molecule, being a potential molecule for the development of stress-
254 tolerant crop varieties in breeding programs [68-71].

255 While in the leaves of *Spondias mombin* are present in greater concentration
256 polyphenols and protein amino acids, the presence of sugars (sucrose, α -glucose, β -glucose,
257 fructose), sugar alcohols (glycerol), fatty acids (oleic acid), and GABA is extremely significant
258 in the barks (Figure 1), showing the application of these in subsidies to obtain drought tolerance
259 plants [72-74].

260 Studies indicate that GABA plays a special role in regulating the tolerance of plants to
261 abiotic stresses, especially with regard to increasing drought tolerance, inducing regulation of
262 the stomatal opening, and preventing water loss by the plant [75-77]. In addition, GABA has
263 been described as acting in defense against insect herbivory and regulation of membrane
264 transport [72].

265 Sugars and sugar alcohols, in particular, have been highlighted for their participation in
266 tolerance to abiotic stress (drought, salinity, cold), mainly because they act as osmoprotectants
267 during cell dehydration, stabilizing the cell membrane and preventing its oxidation. Glucose, in
268 particular, induces stomatal closure and increases the adaptability of the plant under water stress

269 conditions [73, 78-79]. In addition, sugars also maintain the leaf's water content, which can be
 270 seen by the low concentration of sugars in the leaf spectrum (Figure 1).

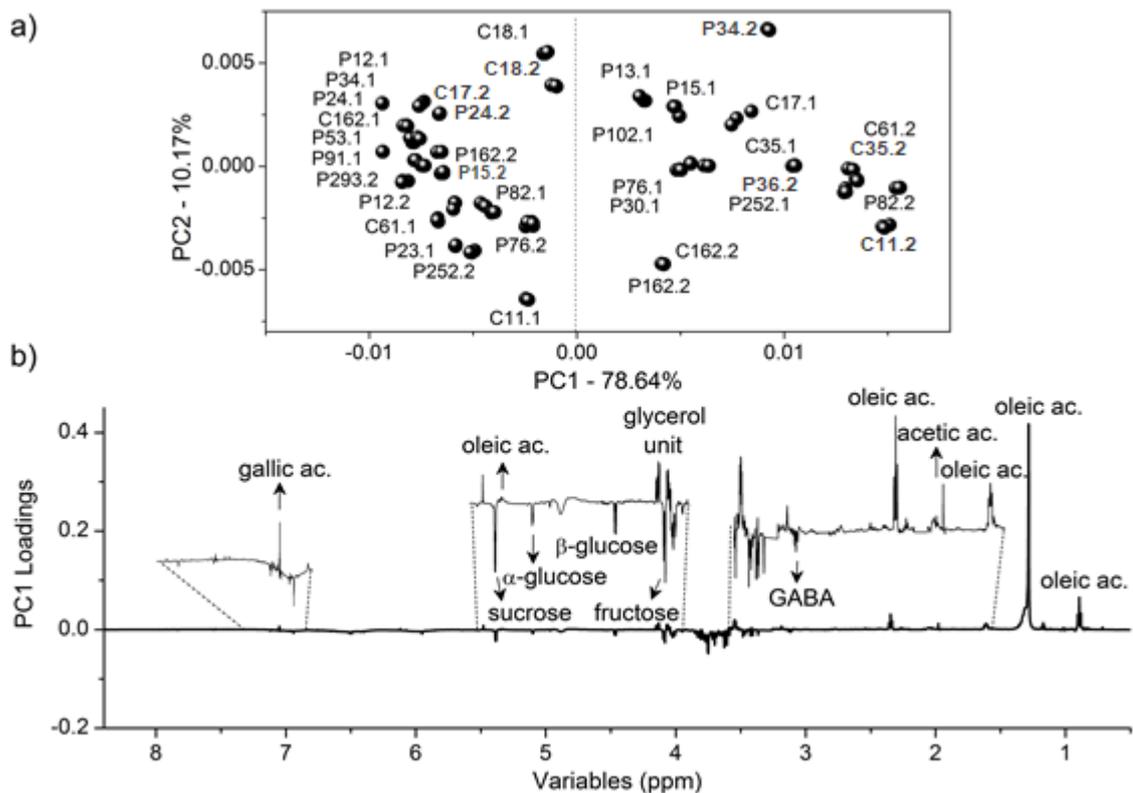
271

272 3.2. Metabolomic variability of barks and leaves according to genotype

273

274 Unsupervised chemometric analysis by PCA was applied to explore the composition
 275 variability of leaves and bark extracts from different accessions of *Spondias mombin* trees in
 276 order to feed the active germplasm bank. Figures 2 and 3 illustrate the PCA results for
 277 evaluation of extracts from barks and leaves respectively.

278



279

280 Figure 2 – PCA modeling using the ^1H NMR dataset of barks from different genotypes of *S.*
 281 *mombin*: a) bidimensional scores coordinate system (PC1 \times PC2), where “C” and “P” mean
 282 clone and progenies, respectively; b) PC1 loadings plotted in lines form with the relevant
 283 identified compounds.

284

285 According to the literature, the clones, which are generated by asexual propagation,
286 continue to be genetically identical, with the advantage of having the same response to
287 environmental factors and contribute to the increase in the economic value of the products
288 obtained and/or productivity. The progenies, however, being propagated by seed, present high
289 genetic variability in terms of size, architecture, crown shape, phenology, and physicochemical
290 characteristics of leaves and fruits [28, 80].

291 According to Figure 2, there is no distinction between the genotypes either from clones
292 or progenies. However, important information may be revealed by the PC1 axis, which retained
293 the main variance on barks extracts according to their composition. Genotypes at positive scores
294 (C61.2, C35.2, P82.2, C11.2, P34.2, C17.1, C35.1, P36.2, P252.1, P13.1, P15.1, P102.1, P76.1,
295 P30.1, C162.2 and P102.2) presented higher amounts of oleic acid, acetic acid, glycerol, and
296 gallic acid.

297 On the other hand, the genotypes at negative scores (P12.1, P34.1, P24.1, C162.1, P53.1,
298 P91.1, P293.2, P12.2, C61.1, P23.1, P252.2, C11.1, P76.2, P82.1, P15.2, P102.2, P24.2, C17.2,
299 C18.2 and C18.1) presented higher amounts of sugar (sucrose, glucose and fructose) and
300 GABA. The intensity of the compounds quantity is directly correlated to the distance of each
301 genotype from the PC1 axis origin (at zero), in other words, the higher the distance the higher
302 the quantity.

303 Thus, the accessions P12B1, P34B1, P24B1, C162B1, P53B1, P91B1, P293B2, and
304 P12B2 registered the highest concentrations of sugar (sucrose, glucose, and fructose) and
305 GABA, possibly due to these being under conditions of hydric stress.

306 According to "Agência Executiva de Gestão das Águas" de Paraíba (AESAPB), in the
307 Agreste Mesoregion, where the accesses P12B1, C162B1, and P53B1 were collected, low
308 rainfall (257.5 mm) was observed between 01/01/17 until the period in which they were

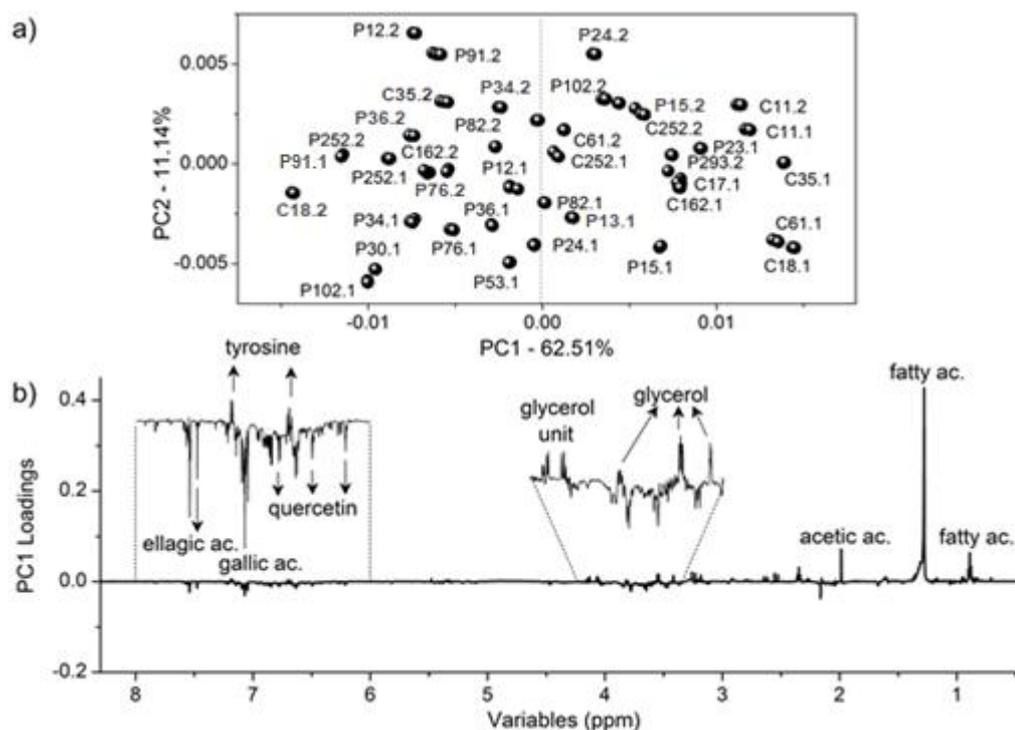
309 collected, 05/17/2017. In contrast, the accessions P24B1, P91B1, and P293B2, possibly
310 presented water stress due to the high rainfall index in the Mata Paraibana Mesoregion in the
311 month in which they were collected, with an average precipitation of 843.6 mm, the highest
312 index of all the Paraiba region in the same period.

313 The other cultivars, corresponding to the accesses close to the zero axis origin in
314 negative PC1, mostly follow the pattern of having been collected in 2018, when the rainfall was
315 not under extreme conditions, which would justify the lower concentration of the compounds
316 mentioned above.

317 Therefore, when the *Spondias mombin* barks are collected in extreme rain periods in
318 Paraiba mesoregions, with low or high rainfall, they have higher sugar and GABA
319 accumulation, compounds associated with resistance to adverse soil and climate conditions. In
320 addition, under the same conditions, a higher concentration of gallic acid (PC1 positive, Figure
321 2) is also observed, proposed as a potent antioxidant, already used by industries as a
322 preservative in food and beverages [81-82].

323 In addition, gallic acid is also described as anticancer, antiulcerogenic, anti-
324 inflammatory, antimicrobial, and antifungal effects [81-82], important characteristics to
325 maintain crop quality and improve productivity, minimizing costs with the possible use of
326 agrochemicals.

327



328

329 Figure 3 – PCA modeling using the ^1H NMR dataset of leaves from different genotypes of *S.*
 330 *mombin*: a) bidimensional scores coordinate system (PC1 \times PC2), where “C” and “P” mean
 331 clone and progenies, respectively; b) PC1 loadings plotted in lines form with the relevant
 332 identified compounds.

333

334 As described for bark extracts, PC1 was the axis retaining the main variance for leaves
 335 extracts discrimination based on genotypes. Leaves extract from the genotypes at positive
 336 scores (C18.1, C61.1, C35.1, C11.1, C11.2, P23.1, P293.2, C17.1, C162.1, C252.2, P15.2,
 337 P24.2, P102.2, C61.2, C252.1, P82.1, P13.1, P24.1 and P15.1) presented higher amounts of
 338 oleic acid, acetic acid, glycerol and tyrosine.

339 On the other hand, the genotypes at negative scores (C18.2, P91.1, P252.2, P36.2,
 340 C35.2, P12.2, P91.2, P34.2, P82.2, C162.2, P252.1, P12.1, P76.2, P36.1, P34.1, P76.1, P30.1,
 341 P102.1 and P53.1) presented higher amount of quercetin, gallic acid and ellagic acid. As
 342 described for barks, the quantity intensity is correlated to the distance from the PC1 origin.

343 As seen in the Principal Component Analysis of *Spondias mombin* barks, the highest
344 concentration of oleic acid, acetic acid, and glycerol was observed in the leaves in the
345 accessions referring to PC1 positive.

346 However, instead of the presence of gallic acid, the amino acid tyrosine has been
347 identified, already described as a precursor to several secondary metabolites whose
348 physiological roles are involved in electron transport, the attraction of pollinators, defense
349 compounds, structural support, and antioxidants.

350 It was noticeable, for the most part, that the compounds mentioned above (PC1 positive)
351 were exhibited in greater concentration by clones, and not by progenies, which was to be
352 expected because they exhibit less genetic variability.

353 In PC1 negative, while in the *Spondias mombin* barks a higher concentration of
354 compounds resistant to hydric stress was observed, the leaves showed higher concentrations of
355 quercetin, gallic acid, and ellagic acid (Figure 3), possibly accumulated by the presence of some
356 pathogen, easily transported in periods of rain and carried by droplets to the leaf surface. The
357 accesses in which such aspects were observed (P252L2, P91L1, C18L2, P102L1, P30L1)
358 belonged, for the most part, to the Mata Paraibana mesoregion, which would justify the
359 occurrence of pathogens in the period in which the leaves were collected.

360 Thus, *Spondias mombin* leaves and barks clones always had the highest concentrations
361 of oleic acid, acetic acid, and glycerol, essential constituents of plant cells, mainly because they
362 provide energy for various metabolic processes. Under these conditions, barks clones are best
363 harvested in July, while leaves clones did not show a standard period. Most progenies, on the
364 other hand, exhibited higher concentrations of sugars and secondary metabolites, both
365 compounds associated with stress and resistance conditions.

366

367 4. Conclusion

368

369 In view of the results obtained, it was possible to elucidate by NMR the structure of 12
370 compounds in the leaves of *Spondias mombin*, among which are included amino acids, organic
371 acids, and flavonoids, both of which have already been described with significant biological
372 activities, such as antioxidant, anti-inflammatory, antifungal effect, antimicrobial, among
373 others. In the shells, the identification of amino acids, organic acids, fatty acids, and
374 carbohydrates allowed us to associate them with the effects of resistance to adverse effects of
375 climate and temperature, which mainly involve water stress.

376 According to the multivariate analysis, is no genotypic difference between *Spondias*
377 *mombin* clones and progenies, however, a higher concentration of glycerol, acetic acid, and
378 oleic acid was observed in the clones, when compared to the progenies. Regarding the peels,
379 given these conditions, it was observed that they showed a pattern of collection, while in the
380 leaves it could not be admitted. Regarding the metabolites found, the progenies showed higher
381 we were of sugars, quercetin, gallic acid, and ellagic acid, which can be better used in future
382 genetic improvement programs, compared to clones.

383 Therefore, the study of the chemical profile of the active germplasm bank of *S. mombin*
384 created, associated with the multivariate analysis, proved to be extremely important, evidencing
385 its importance in the subsidy of plant breeding programs, due to the biologically active
386 compounds found, essential to increase agricultural productivity and nutritional quality of crops
387 in view of the adverse conditions to which they are subjected.

388

389

390

391

392 **Acknowledgments**

393

394 The authors acknowledge the financial support of EMBRAPA and all the support
395 offered (SEG 03.14.01.012.00.00), from the National Council for Scientific and Technological
396 Development (CNPq) for their financial support and scholarships 303791/2016-0, from the
397 Coordination for the Improvement of Higher Education Personnel for the support and
398 scholarship 88887.217187/2018-00 and the National Institute of Science and Technology -
399 INCT BioNat, grant # 465637/2014-0, Brazil.

References

- [1] J.D. Mitchell, D.C. Daly, A revision of *Spondias* L. (Anacardiaceae) in the neotropics, *PhytoKeys*. 55 (2015) 1-92. doi: <https://doi.org/10.3897/phytokeys.55.8849>.
- [2] J. Bosco, K.T. Soares, S.P.A. Filho, R.V. Barros, A cultura da cajazeira, *Emepa Documentos*. 28 (2000) 29.
- [3] E.B. Soares, R.L.F. Gomes, J.G. de M. e Carneiro, F.N. do Nascimento, I.C.V. Silva, J.C.L. da Costa, Caracterização física e química de Frutos de cajazeira, *Revista Brasileira de Fruticultura*. 28 (2006) 518-519. doi: <https://doi.org/10.1590/S0100-29452006000300039>.
- [4] N. Fonseca, C. de F. Machado, J.F. da S. Júnior, R. da S. Carvalho, R. Ritzinger, R.M. Alves, M.C.C. Maia, *Spondias* spp. Umbu, cajá e espécies afins, Instituto Interamericano de Cooperación para la Agricultura. (2018) 29.
- [5] A.O. Ayoka, R.O. Akomolafe, O.S. Akinsomisoye, O. Ukponmwan, Medicinal and economic value of *Spondias mombin*, *African Journal of Biomedical Research*. 11 (2008) 129-136. doi: <https://doi.org/10.4314/ajbr.v11i2.50714>.
- [6] A.C. Akinmoladun, M.F. Khan, J. Sarkar, E.O. Farombi, R. Maurya, Distinct radical scavenging and antiproliferative properties of *Spondias mombin* and antioxidant activity-guided isolation of quercetin-3-O--D-glucopyranoside and undec-1-ene, *African Journal of Pharmacy and Pharmacology*. 9 (2015) 506-513. doi: <https://doi.org/10.5897/AJPP2015.4279>.
- [7] S. Saheed, S.A. Olarewaju, G. Taofeeq, S.T. Olatunde, A.A. Alanamu, Combined administration of *Spondias mombin* and *Ficus exasperata* leaf extracts stall Indomethacin-mediated gastric mucosal onslaught in rats, *African Journal of Traditional, Complementary and Alternative Medicines*. 12 (2015) 45-51. doi: <https://doi.org/10.4314/ajtcam.v12i1.7>.
- [8] S. Sabiu, T. Garuba, T. Sunmonu, E. Ajani, A. Sulyman, I. Nurain, A. Balogun, Indomethacin-induced gastric ulceration in rats: protective roles of *Spondias mombin* and *Ficus exasperata*, *Toxicology Reports*. 2 (2015) 261-267. doi: <https://doi.org/10.1016/j.toxrep.2015.01.002>.

- [9] A.E. Oluwatoyin, O.W. Deborah, Study of aqueous leaf extracts of *Spondias mombin* Linn. (Anacardiaceae) in gastric ulcer models and the possible mechanisms of action, *The Journal of Phytopharmacology*. 8 (2019) 238-247. doi: <https://doi.org/10.31254/phyto.2019.8506>.
- [10] J.A.O. Olugbuyiro, J.O. Moody, M.T. Hamann, Phytosterols from *Spondias mombin* Linn with antimycobacterial activities, *Afr J Biomed Res*. 16 (2013) 19-24. PMID: PMC5096588.
- [11] O.O. Temitope, A.F. Ogunmodede, O.A. Fasusi, A.O. Thonda, A.E. Odunfunwa, Synergistic antibacterial and antifungal activities of *Spondias mombin* extracts and conventional antibiotic and antifungal agents on selected clinical microorganisms, *Scholars Journal of Applied Medical Sciences*. 5 (2017) 307-318. doi: <https://doi.org/10.21276/sjams.2017.5.2.3>.
- [12] O. Aromolaran, O.K. Badejo, Efficacy of fresh leaf extracts of *Spondias mombin* against some clinical bacterial isolates from typhoid patients, *Journal of Tropical Disease*. 4 (2014) 442-446. doi: [https://doi.org/10.1016/S2222-1808\(14\)60603-4](https://doi.org/10.1016/S2222-1808(14)60603-4).
- [13] E.S. Omoregie, E.I. Oikeh, Comparative studies on the phytochemical composition, phenolic content and antioxidant activities of methanol leaf extracts of *Spondias mombin* and *Polyathia longifolia*, *Jordan Journal of Biological Sciences*. 8 (2015) 145-149. doi: <https://doi.org/10.12816/0027561>.
- [14] B. Cabral, E.M.S. Siqueira, M.A.O. Bitencourt, M.C.J.S. Lima, A.K. Lima, C.F. Ortmann, V.C. Chaves, M.F. Fernandes-Pedrosa, H.A.O. Rocha, K.C. Scoretcci, F.H. Reginatto, R.B. Giordani, S.M. Zucolotto, Phytochemical study and anti-inflammatory and antioxidant potential of *Spondias mombin* leaves, *Brazilian Journal of Pharmacognosy*. 26 (2016) 304-311. doi: <https://doi.org/10.1016/j.bjp.2016.02.002>.
- [15] M.E. Goodies, I.E. Emmanuel, O.J. Matthew, E.J.O. Tedwins, A. D. Lotanna, E.O. Earnest, C. Paul, A. Ejiroghene, Antidiabetic activity and toxicity evaluation of aqueous extracts of *Spondias mombin* and *Costus afer* on wistar rats, *British Journal of Pharmaceutical Research*. 6 (2015) 333-342. doi: <https://doi.org/10.9734/BJPR/2015/17519>.
- [16] N. Eluehike, I. Onoagbe, Changes in organ and body weight, serum amylase and antidiabetic effects of tannins from *Spondias mombin* on streptozotocin-induced diabetic rats, *Journal of Insulin Resistance*. 3 (2018) 1-5. doi: <https://doi.org/10.4102/jir.v3i1.40>.

- [17] A.F. Akinmoladun, E.M. Obuotor, M.K. Barthwal, M. Dikshit, E.O. Farombi, Ramipril-like activity of *Spondias mombin* Linn against no-flow ischemia and isoproterenol-induced cardiotoxicity in rat heart, *Cardiovascular Toxicology*. 10 (2010) 295-305. doi: <https://doi.org/10.1007/s12012-010-9089-4>.
- [18] L.L. Nwidu, E. Elmorsy, Y.I. Oboma, W.G. Carter, . Hepatoprotective and antioxidant activities of *Spondias mombin* leaf and stem extracts against carbon tetrachloride-induced hepatotoxicity, *Journal of Taibah University Medical Sciences*. 13 (2018) 262-271. doi: <https://doi.org/10.1016/j.jtumed.2018.03.006>.
- [19] P.C. Njoku, M.I. Akumefula, Phytochemical and nutrient evaluation of *Spondias mombin* leaves, *Pakistan Journal of Nutrition*. 6 (2007) 613-615. doi: <https://doi.org/10.3923/pjn.2007.613.615>.
- [20] C.U. Igwe, G.O.C. Onyeze, V.A. Onwuliri, C.G. Osuagwu, A.O. Ojiako, Evaluation of the chemical compositions of the leaf of *Spondias mombin* L. from Nigeria, *Australian Journal of Basic and Applied Sciences*. 4 (2010) 706-710.
- [21] A.R.A. da Silva, S.M. de Morais, M.M.M. Marques, D.F. de Oliveira, C.C. Barros, R.R. de Almeida, I.G.P. Vieira, M.I.F. Guedes, Chemical composition, antioxidant and antibacterial activities of two *Spondias* species from Northeastern Brazil, *Pharmaceutical Biology*. 50 (2012) 740-746. doi: <https://doi.org/10.3109/13880209.2011.627347>.
- [22] H.C.C. Maduka, A.N. Okpogba, C.E. Ugwu, C.C. Dike, P.N. Ogueche, D.T. Onwuzurike, D.C. Ibe, Phytochemical, antioxidant and microbial inhibitory effects of *Spondias mombin* leaf and stem bark extracts, *IOSR Journal of Pharmacy and Biological Sciences*. 9 (2014) 14-17. doi: <https://doi.org/10.9790/3008-09271417>.
- [23] F.R. Ferreira, Germoplasma de fruteiras, *Revista Brasileira de Fruticultura*. 33 (2011) 1–6. doi: <https://doi.org/10.1590/S0100-29452011000500002>.
- [24] F.X. de Mendes, N.V.B. Mendes, *Spondias mombin* - Cajá. In: CORADIN, L.; CAMILLO, J.; PAREYN, F. G. C. (Ed.). *Espécies nativas da flora brasileira de valor econômico atual ou potencial: plantas para o futuro: região Nordeste*. Brasília, DF: MMA, 2018. (Série Biodiversidade; 51). Disponível em: <http://www.mma.gov.br/publicacoes/biodiversidade/category/142-serie-biodiversidade.html>.

- [25] J.E.U. de Carvalho, R.M. Alves, Recursos genéticos de espécies do táxon *Spondias* na Amazônia Oriental. In: LEDERMAN, I. E.; LIRA JUNIOR, J. S. de; SILVA JUNIOR, F. da. (Ed.). *Spondias* no Brasil: umbú, cajá e espécies afins. Recife: Editora Universitária da UFRPE, 2008, p. 69-74.
- [26] F.X. Souza, Recursos genéticos e melhoramento de *Spondias* no Brasil – cajazeira. In: LEDERMAN, I. E.; LIRA JUNIOR, J. S. de; SILVA JUNIOR, F. da. (Ed.). *Spondias* no Brasil: umbú, cajá e espécies afins. Recife: Editora Universitária da UFRPE, 2008. p. 45-53.
- [27] J.S. de L. Júnior, J.E.F. Bezerra, I.E. Lederman, R.M. de Moura, Recursos genéticos e melhoramento de *Spondias* em Pernambuco: cajazeira, cirigueira e cajá-umbuzeiro. In: LEDERMAN, I. E.; LIRA JUNIOR, J. S. de; SILVA JUNIOR, F. da. (Ed.). *Spondias* no Brasil: umbú, cajá e espécies afins. Recife: Editora Universitária da UFRPE, 2008. p. 80-86.
- [28] C.M. Cassimiro, Recursos genéticos e melhoramento de *Spondias* no estado da Paraíba: cajazeira, cirigueira e cajaraneira. In: LEDERMAN, I. E.; LIRA JUNIOR, J. S. de; SILVA JUNIOR, F. da. (Ed.). *Spondias* no Brasil: umbú, cajá e espécies afins. Recife: Editora Universitária da UFRPE, 2008, p. 63-68.
- [29] J.M.F. Valls, Caracterização de recursos genéticos vegetais. In: NASS, L. L. (Ed.). Recursos genéticos vegetais, Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 281-305).
- [30] R.F. Vieira, T.S. Agostini-Costa, Caracterização química de metabólitos secundários em germoplasma vegetal. In: NASS, L. L. (Ed.), Recursos genéticos vegetais, Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2007. p. 343-372.
- [31] L.L. Nass, M.S. Sigrist, C.S. da C. Ribeiro, F.J.B. Reifschneider, Genetic resources: the basis for sustainable and competitive plant breeding, *Crop Breeding and Applied Biotechnology* S2. 12 (2012) 75-86. doi: <https://doi.org/10.1590/S1984-70332012000500009>.
- [32] R. Almeida, G.G. Cezar, R.R. Catharino, S.G. Villas-Bôas, A. Nunes-Nesi, C.A. Labate, A.H. Iglesias, Anais do Encontro Sobre Metabolômica, Brasília, DF, 14 de dezembro de 2011/ editora técnica Patrícia Verardi Abdelnur – Brasília, DF: Embrapa Agroenergia, 2011. 41 p. – (Documentos / Embrapa Agroenergia, ISSN 2177- 4439 ; 09).
- [33] B.C. Aranha, J.F. Hoffmann, R.L. Barbieri, C.V. Rombaldi, F.C. Chaves, Untargeted metabolomic analysis of *Capsicum* spp. by GC–MS, *Phytochemical Analysis*. 28 (2017) 439–

447. doi: <https://doi.org/10.1002/pca.2692>.

[34] V.C. Trenerry, S. Rochfort, Natural products research and metabolomics, *Comprehensive Natural Products II: Chemistry and Biology*. 9 (2010) 595–628. doi: <https://doi.org/10.1016/B978-008045382-8.00211-2>.

[35] S. Forcisi, F. Moritz, B. Kanawati, D. Tziotis, R. Lehmann, P. Schmitt-Kopplin, Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling, *Journal of Chromatography A*. 1992 (2013) 51-65. doi: <https://doi.org/10.1016/j.chroma.2013.04.017>.

[36] D.A. Dias, S. Urban, U. Roessner, a historical overview of natural products in drug discovery, *Metabolites*. 2 (2012) 303-336. doi: <https://doi.org/10.3390/metabo2020303>.

[37] H.A. Gad, S.H. El-Ahmady, M.I. Abou-Shoer, M.M. Al-Azizi, Application of chemometrics in authentication of herbal medicines: a review, *Phytochemical Analysis*. 24 (2012) 1-24. doi: <https://doi.org/10.1002/pca.2378>.

[38] G.A.B. Canuto, J.L. da Costa, P.L.R. da Cruz, A.R.L. de Souza, A.T. Faccio, A. Klassen, K.T. Rodrigues, M.F.M. Tavares, Metabolômica: definições, estado-da-arte e aplicações representativas, *Química Nova*. 41 (2018) 75-91. doi: <https://doi.org/10.21577/0100-4042.20170134>.

[39] C. Simmler, J.G. Napolitano, J.B. McAlpine, S.N. Chen, G.F. Pauli, Universal quantitative NMR analysis of complex natural samples, *Current Opinion in Biotechnology*. 25 (2014) 51-59. doi: <https://doi.org/10.1016/j.copbio.2013.08.004>.

[40] C.S. Funari, I. Castro-Gamboa, A.J. Cavalheiro, V. da S. Bolzani, Metabolômica, uma abordagem otimizada para exploração da biodiversidade brasileira: estado da arte, perspectivas e desafios, *Química Nova*. 36 (2013) 1605-1609. doi: <https://doi.org/10.1590/S0100-40422013001000019>.

[41] L.C. de Rezende, P. de A. Santos, V.B. Riatto, J.M. David, J.P. David, New alkyl phenols and fatty acid profile from oils of pulped *Spondias mombin* L. seed wastes, *Química Nova*. 41 (2018) 540-543. doi: <https://doi.org/10.21577/0100-4042.20170203>.

[42] E.G.A. Filho, L.N. Braga, L.M.A. Silva, F.R. Miranda, E.O. Silva, K.M. Canuto, M.R. Miranda, E.S. de Brito, G.J. Zocolo, Physiological changes for drought resistance in different species of *Phyllanthus*, *Scientific Reports*. 8 (2018) 1-12. doi: <https://doi.org/10.1038/s41598->

018-33496-7.

[43] E.G.A. Filho, L.M.A. Silva, P.R. Ribeiro, E.S. de Brito, G.J. Zocolo, P.C. Souza-Leão, A.T. Marques, A.L. Quintela, F.H. Larsen, K.M. Canuto, ¹H NMR and LC-MS-based metabolomic approach for evaluation of the seasonality and viticultural practices in wines from São Francisco River Valley, a Brazilian semi-arid region, *Food Chemistry*. 289 (2019) 558-567. doi: <https://doi.org/10.1016/j.foodchem.2019.03.103>.

[44] A. Sharma, D. Kashyap, K. Sak, H.S. Tuli, A.K. Sharma, Therapeutic charm of quercetin and its derivatives: a review of research and patents, *Pharmaceutical Patent Analyst*. 7 (2018) 15-32. doi: <https://doi.org/10.4155/ppa-2017-0030>.

[45] G. D'Andrea, Quercetin: A flavonol with multifaceted therapeutic applications?, *Fitoterapia*. 106 (2015) 256-271. doi: <https://doi.org/10.1016/j.fitote.2015.09.018>.

[46] P. Elumalai, S. Lakshmi, Role of quercetin benefits in neurodegeneration. In: ESSA, M.M.; AKBAR, M.; GUILLEMIN, G. (Ed.). *The benefits of natural products for neurodegenerative diseases*. Springer International Publishing Switzerland, 2016, p. 229-245. doi: https://doi.org/10.1007/978-3-319-28383-8_12.

[47] J.K. Kim, S.U. Park, Quercetin and its role in biological functions: an updated review, *EXCLI Journal*. 17 (2018) 856-863. doi: <https://doi.org/10.17179/excli2018-1538>.

[48] A. Maalik, F.A. Khan, A. Mumtaz, A. Mehmood, S. Azhar, M. Atif, S. Karim, Y. Altaf, I. Tariq, Pharmacological applications of quercetin and its derivatives: a short review, *Tropical Journal of Pharmaceutical Research*. 13 (2014) 1561-1566. doi: <https://doi.org/10.4314/tjpr.v13i9.26>.

[49] J.S. Nam, A.R. Sharma, L.T. Nguyen, C. Chakraborty, G. Sharma, S.S. Lee, Application of bioactive quercetin in oncotherapy: from nutrition to nanomedicine, *Molecules*. 21 (2016) 108. doi: <https://doi.org/10.3390/molecules21010108>.

[50] M. Russo, C. Spagnuolo, I. Tedesco, S. Bilotto, G.L. Russo, The flavonoid quercetin in disease prevention and therapy: facts and fancies, *Biochemical Pharmacology*. 83 (2012) 6-15. doi: <https://doi.org/10.1016/j.bcp.2011.08.010>.

[51] W. Wang, C. Sun, L. Mao, P. Ma, F. Liu, J. Yang, Y. Gao, The biological activities, chemical stability, metabolism and delivery systems of quercetin: a review, *Trends in Food Science and Technology*. 56 (2016) 21-38. doi: <https://doi.org/10.1016/j.tifs.2016.07.00>.

- [52] A. Galano, M.F. Marquez, A. Pérez-González, Ellagic Acid: An Unusually Versatile Protector against Oxidative Stress, *Chemical Research in Toxicology*. 27 (2014) 904-918. doi: <https://doi.org/10.1021/tx500065y>.
- [53] A. Shakeri, M.R. Zirak, A. Sahebkar, Ellagic Acid: A Logical Lead for Drug Development?, *Current Pharmaceutical Design*. 24 (2018) 106-122. doi: <https://doi.org/10.2174/1381612823666171115094557>.
- [54] V. Demidchik, Mechanisms of oxidative stress in plants: From classical chemistry to cell biology, *Environmental and Experimental Botany*. 109 (2015) 212-228. doi: <https://doi.org/10.1016/j.envexpbot.2014.06.021>.
- [55] D.D. Evtugin, S. Magina, D.V. Evtugin, Recent advances in the production and applications of ellagic acid and its derivatives. A review, *Molecules*. 25 (2020) 2745. doi: <https://doi.org/10.3390/molecules25122745>.
- [56] J.L. Ríos, R.M. Giner, M. Marín, M.C. Recio, A pharmacological update of ellagic acid, *Planta Medica*. 84 (2018) 1068-1093. doi: <https://doi.org/10.1055/a-0633-9492>.
- [57] G. Derosa, P. Maffioli, A. Sahebkar, Ellagic acid and its role in chronic diseases. In: GUPTA, S.C.; PRASAD, S.; AGGARWAL, B.B. (Ed.). *Anti-inflammatory nutraceuticals and chronic diseases. Advances in Experimental Medicine and Biology* 928, 2016, p. 473-479. doi: https://doi.org/10.1007/978-3-319-41334-1_20.
- [58] S. Arokiyaraj, A. Stalin, B.S. Kannan, H. Shin, Geranii herba as a potential inhibitor of SARS-CoV-2 main 3CL^{pro}, Spike RBD, and regulation of unfolded protein response: an in silico approach, *Antibiotics*. 9 (2020) 863. doi: <https://doi.org/10.3390/antibiotics9120863>.
- [59] T.H. Fatoki, O. Ibraheem, I.O. Ogunyemi, A.C. Akinmoladun, H.U. Ugboko, C.J. Adeseko, O.A. Awofisayo, S.J. Olusegun, J.M. Enibukun, Network analysis, sequence and structure dynamics of key proteins of coronavirus and human host, and molecular docking of selected phytochemicals of nine medicinal plants, *Journal of Biomolecular Structure and Dynamics*. 2020. doi: <https://doi.org/10.1080/07391102.2020.1794971>.
- [60] O. Abian, D. Ortega-Alarcon, A. Jimenez-Alesanco, L. Ceballos-Laita, S. Vega, H.T. Reyburn, B. Rizzuti, A. Velazquez-Campoy, Structural stability of SARS-CoV-2 3CL^{pro} and identification of quercetin as an inhibitor by experimental screening, *International Journal of Biological Macromolecules*. 164 (2020) 1693-1703. doi: <https://doi.org/10.1016/j.ijbiomac.2020.07.235>.

- [61] M. Aucoin, K. Cooley, P.R. Saunders, V. Cardozo, D. Remy, H. Cramer, C.N. Abad, N. Hannan, The effect of quercetin on the prevention or treatment of COVID-19 and other respiratory tract infections in humans: a rapid review, *Advances in Integrative Medicine*. 7 (2020) 247-251. doi: <https://doi.org/10.1016/j.aimed.2020.07.007>.
- [62] T.M. Hildebrandt, A.N. Nesi, W.L. Araújo, H.P. Braun, Amino acid catabolism in plants, *Molecular Plant*. 8 (2015) 1563-1579. doi: <https://doi.org/10.1016/j.molp.2015.09.005>.
- [63] A.M.R. Abdel-Mawgoud, A.M. El-Bassiouny, A. Ghoname, S.D. Abou-Hussein, Foliar application of amino acids and micronutrients enhance performance of green bean crop under newly reclaimed land conditions, *Australian Journal of Basic and Applied Sciences*. 5 (2011) 51-55.
- [64] A.S. Khan, B. Ahmad, M.J. Jaskani, R. Ahmad, A.U. Malik, Foliar application of mixture of amino acids and seaweed (*Ascophylum nodosum*) extract improve growth and physicochemical properties of grapes, *International Journal of Agriculture and Biology*. 14 (2012) 383-388.
- [65] W.F. Teixeira, E.B. Fagan, L.H. Soares, R.C. Umburanas, K. Reichardt, D.D. Neto, Foliar and seed application of amino acids affects the antioxidant metabolism of the soybean crop, *Frontiers in Plant Science*. 8 (2017) 1-14. doi: <https://doi.org/10.3389/fpls.2017.00327>.
- [66] Q. Yang, D. Zhao, Q. Liu, Connections between amino acid metabolisms in plants: lysine as an example, *Frontiers in Plant Science*. 11 (2020) 1-8. doi: <https://doi.org/10.3389/fpls.2020.00928>.
- [67] C.A. Schenck, H.A. Maeda, Tyrosine biosynthesis, metabolism, and catabolism in plants, *Phytochemistry*. 149 (2018) 82-102. doi: <https://doi.org/10.1016/j.phytochem.2018.02.003>.
- [68] S. Hayat, Q. Hayat, M.N. Alyemeni, A.S. Wani, J. Pichtel, A. Ahmad, Role of proline under changing environments: a review, *Plant Signaling and Behavior*. 7 (2012) 1456-1466. doi: <https://doi.org/10.4161/psb.21949>.
- [69] M.A. Hossain, M.A. Hoque, D.J. Burritt, M. Fujita, Proline protects plants against abiotic oxidative stress: biochemical and molecular mechanisms. In: AHMAD, P. (Ed.). *Oxidative damage to plants*. Academic Press, 2014, p. 477-522. doi: <https://doi.org/10.1016/B978-0-12-799963-0.00016-2>.

- [70] M.I. Dar, M.I. Naikoo, F. Rehman, F. Naushin, F.A. Khan, Proline accumulation in plants: roles in stress tolerance and plant development. In: IQBAL, N.; NAZAR, R.; KHAN, N.A. (Ed.). Osmolytes and plants acclimation to changing environment: emerging omics technologies, 2016, p. 155-166. doi: https://doi.org/10.1007/978-81-322-2616-1_9.
- [71] H. Stein, A. Honig, G. Miller, O. Erster, H. Eilenberg, L.N. Csonka, L. Szabados, C. Koncz, A. Zilberstein, Elevation of free proline and proline-rich protein levels by simultaneous manipulations of proline biosynthesis and degradation in plants, *Plant Science*. 181 (2011) 140-150. doi: <https://doi.org/10.1016/j.plantsci.2011.04.013>.
- [72] A.W. Bown, B.J. Shelp, Plant GABA: not just a metabolite, *Trends in Plant Science*. 21 (2016) 811-813. doi: <https://doi.org/10.1016/j.tplants.2016.08.001>.
- [73] M. Singh, J. Kumar, S. Singh, V.P. Singh, S.M. Prasad, Roles of osmoprotectants in improving salinity and drought tolerance in plants: a review, *Reviews in Environmental Science and Biotechnology*. 14 (2015) 407-426. doi: <https://doi.org/10.1007/s11157-015-9372-8>.
- [74] C. Liu, L. Zhao, G. Yu, The dominant glutamic acid metabolic flux to produce γ -amino butyric acid over proline in *Nicotiana glauca* leaves under water stress relates to its significant role in antioxidant activity, *Journal of Integrative Plant Biology*. 53 (2011) 608-618. doi: <https://doi.org/10.1111/j.1744-7909.2011.01049.x>.
- [75] D.W. Mekonnen, U.I. Flügge, F. Ludewig, Gamma-aminobutyric acid depletion affects stomata closure and drought tolerance of *Arabidopsis thaliana*, *Plant Science*. 245 (2016) 25-34. doi: <https://doi.org/10.1016/j.plantsci.2016.01.005>.
- [76] Z. Li, Y. Peng, B. Huang, Alteration of transcripts of stress-protective genes and transcriptional factors by γ -aminobutyric acid (GABA) associated with improved heat and drought tolerance in creeping bentgrass (*Agrostis stolonifera*), *International Journal of Molecular Sciences*. 19 (2018) 1623. doi: <https://doi.org/10.3390/ijms19061623>.
- [77] B. Yong, H. Xie, Z. Li, Y.P. Li, Y. Zhang, G. Nie, X.Q. Zhang, X. Ma, L.K. Huang, Y.H. Yan, Y. Peng, Exogenous application of GABA improves PEG-induced drought tolerance positively associated with GABA-shunt, polyamines, and proline metabolism in white clover, *Frontiers in Physiology*. 8 (2017) 1107. doi: <https://doi.org/10.3389/fphys.2017.01107>.
- [78] F. Sami, M. Yusuf, M. Faizan, A. Faraz, S. Hayat, Role of sugars under abiotic stress, *Plant Physiology and Biochemistry*. 109 (2016). doi: <https://doi.org/10.1016/j.plaphy.2016.09.005>.

- [79] M.P. Gongola, B.R. Ramadoss, Sugars play a critical role in abiotic stress tolerance in plants. In: WANI, S.H. (Ed.). Biochemical, physiological and molecular avenues for combating abiotic stress tolerance in plants, 2018, p. 17-38. doi: <https://doi.org/10.1016/B978-0-12-813066-7.00002-4>.
- [80] F.F. Santana, Caracterização de genótipos de cajazeiras. Tese (Doutorado em Agronomia) – Faculdade de Ciências Agrárias e veterinárias, Universidade Estadual Paulista. São Paulo. 97 f. 2010. Disponível em: <<http://hdl.handle.net/11449/105213>>.
- [81] F.H.A. Fernandes, H.R.N. Salgado, Gallic acid: review of the methods of determination and quantification, *Critical Reviews in Analytical Chemistry*. 46 (2015) 257-265. doi: <https://doi.org/10.1080/10408347.2015.1095064>.
- [82] L.A. BenSaad, K.H. Kim, C.C. Quah, W.R. Kim, M. Shahimi, Anti-inflammatory potential of ellagic acid, gallic acid and punicalagin A&B isolated from *Punica granatum*, *BMC Complementary and Alternative Medicine*. 17 (2017) 1-10. doi: <https://doi.org/10.1186/s12906-017-1555-0>.
- [83] J.A.C. Guedes, E.G.A. Filho, M.F.S. Silva, T.H.S. Rodrigues, C.M.C. Ramires, M.A.C. Lima, G.S. Silva, C. Ó Pessoa, K.M. Canuto, E.S. Brito, R.E. Alves, R.F. Nascimento, G.J. Zocolo, GC-MS-Based metabolomic profiles combined with chemometric tools and cytotoxic activities of non-polar leaf extracts of *Spondias mombin* L. and *Spondias tuberosa* Arr. Cam., *J. Braz. Chem. Soc.* 31 (2020) 331-340. doi: <https://doi.org/10.21577/0103-5053.20190185>.

5 CONCLUSÃO

Diante dos resultados obtidos, foi possível elucidar por RMN a estrutura de 12 compostos nas folhas de *S. mombin*, dentre os quais estão inclusos aminoácidos, ácidos orgânicos e flavonóides, ambos já descritos por apresentarem atividades biológicas significativas, como efeito antioxidante, antiinflamatório, antifúngico, antimicrobial, dentre outros. Nas cascas, a identificação de aminoácidos, ácidos orgânicos, ácidos graxos e carboidratos nos permitiu associá-las à efeitos de resistência à efeitos adversos de clima e temperatura, os quais envolvem, principalmente, o estresse hídrico.

De acordo com a análise multivariada, não houve diferença genotípica entre os clones e progênies de *S. mombin*, entretanto foi observado uma maior concentração de glicerol, ácido acético e ácido oleico nos clones. Em relação às cascas, dadas essas condições, foi observado que estes exibiram um padrão de coleta, enquanto nas folhas o mesmo não pôde ser admitido. Já com relação aos metabólitos encontrados, as progênies demonstraram concentrações mais elevadas de açúcares, quercetina, ácido gálico e ácido elágico, podendo ser melhor aproveitadas em futuros programas de melhoramento genética, em comparação aos clones.

Sendo assim, o estudo do perfil químico do banco ativo de germoplasma de *S. mombin* criado, associado à análise multivariada se mostrou extremamente importante, evidenciando a importância deste no subsídio de programas de melhoramento de plantas, em virtude dos compostos biologicamente ativos encontrados, essenciais para aumentar a produtividade agrícola e qualidade nutricional das safras em face das condições de estresse a que estão submetidas.

REFERÊNCIAS

A. HUSSEIN, R.; A. EL-ANSSARY, A. Plants Secondary Metabolites: The Key Drivers of the Pharmacological Actions of Medicinal Plants. **Herbal Medicine**, 2019.

ADEDOKUN, M. O. et al. Socio-economic importance and utilization of *Spondias mombin* in Nigeria. **Asian Pacific Journal of Tropical Medicine**, v. 3, n. 3, p. 232–234, 2010.

AHARONI, A.; GALILI, G. Metabolic engineering of the plant primary-secondary metabolism interface. **Current Opinion in Biotechnology**, v. 22, n. 2, p. 239–244, 2011.

ALMEIDA, D. F. L. DOS S. Estudo das Vias Metabólicas das Plantas na Síntese de Pigmentos Naturais. p. 61, 2017.

ALMEIDA, R. et al. **Encontro sobre Metabolômica**, 2012.

ALSEEKH, S.; FERNIE, A. R. Metabolomics 20 years on: what have we learned and what hurdles remain? **Plant Journal**, v. 94, n. 6, p. 933–942, 2018.

ARANHA, B. C. et al. Untargeted Metabolomic Analysis of *Capsicum* spp. by GC–MS. **Phytochemical Analysis**, v. 28, n. 5, p. 439–447, 2017.

ASUQUO, O. R. et al. Effects of ethanolic leaf extract of *Spondias mombin* on the pituitary-gonadal axis of female Wistar rats. **Asian Pacific Journal of Reproduction**, v. 2, n. 3, p. 169–173, 2013.

BADDINI, A. L. Q.; NETTO, A. D. P.; TEIXEIRA, V. L. Metabolomics based on ¹H NMR and partial least squares-discriminant analysis. **Revista Virtual de Química**, v. 6, n. 2, p. 396–415, 2014.

BAILÃO, E. F. L. C. et al. Bioactive compounds found in Brazilian cerrado fruits. **International Journal of Molecular Sciences**, v. 16, n. 10, p. 23760–23783, 2015.

BARATA, A. M. et al. O Banco Português de Germoplasma Vegetal e a conservação dos recursos genéticos em Portugal. **Agrorrural: Contributos científicos**, n. December 2014, p. 964–974, 2011.

BATISTA, F. R. C. et al. **O umbuzeiro e o semiárido brasileiro**. [s.l: s.n.].

BOFF, L. et al. Antiviral Potential of *Spondias mombin* L. Leaves Extract Against Herpes Simplex Virus Type-1 Replication Using In Vitro and In Silico Approaches. 2020.

BRAVERMAN, I. Conservation without nature: The trouble with in situ versus ex situ conservation. **Geoforum**, v. 51, p. 47–57, 2014.

BRENNAN, L. Progress in Nuclear Magnetic Resonance Spectroscopy NMR-based metabolomics : From sample preparation to applications in nutrition research Data NMR analysis - Acquisition - Baseline correction - Peak alignment samples interpretation Statistical Analysis Pat. **PROGRESS IN NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY**, v. 83, p. 42–49, 2014.

BRITO, S. A. et al. Evaluation of gastroprotective and ulcer healing activities of yellow mombin juice from *Spondias mombin* L. **PLoS ONE**, v. 13, n. 11, p. 1–16, 2018a.

BRITO, S. A. et al. Antiulcer activity and potential mechanism of action of the leaves of *Spondias mombin* L. **Oxidative Medicine and Cellular Longevity**, v. 2018, 2018b.

CABRAL, B. et al. Phytochemical study and anti-inflammatory and antioxidant potential of *Spondias mombin* leaves. **Brazilian Journal of Pharmacognosy**, v. 26, n. 3, p. 304–311, 2016.

CALIXTO, J. B. BIODIVERSIDADE COMO FONTE DE MEDICAMENTOS. **Ciência e Cultura**, 2003.

CANUTO, G. A. B. et al. Metabolomics: Definitions, state-of-the-art and representative applications. **Química Nova**, v. 41, n. 1, p. 75–91, 2018a.

CANUTO, G. A. B. et al. METABOLÔMICA: DEFINIÇÕES, ESTADO-DA-ARTE E APLICAÇÕES REPRESENTATIVAS. **Quim. Nova**, v. 41, n. 1, p. 75–91, 2018b.

COHEN, S. D.; KENNEDY, J. A. Plant metabolism and the environment: Implications for managing phenolics. **Critical Reviews in Food Science and Nutrition**, v. 50, n. 7, p. 620–643, 2010.

CORADIN, L.; CAMILLO, J.; PAREYN, F. G. C. **Espécies Nativas da Flora Brasileira de Valor Econômico Atual ou Potencial Planta para o Futuro: Região Nordeste**. [s.l.: s.n.].

CORNEJO-BÁEZ, A. A. et al. Chemometrics: a complementary tool to guide the isolation of pharmacologically active natural products. **Drug Discovery Today**, v. 25, n. 1, p. 27–37, 2020.

COSTA, M. P. et al. Screening of Brazilian plants for anticancer activity: An overview. **Current Topics in Pharmacology**, v. 17, n. 1, p. 45–59, 2013.

CUBERO-LEON, E.; PEÑALVER, R.; MAQUET, A. Review on metabolomics for food authentication. **FRIN**, v. 60, p. 95–107, 2014.

DA SILVA, A. R. A. et al. Chemical composition, antioxidant and antibacterial activities of two *Spondias* species from Northeastern Brazil. **Pharmaceutical Biology**, v. 50, n. 6, p. 740–746, 2012.

DEBORDE, C. et al. Plant metabolism as studied by NMR spectroscopy. **Progress in Nuclear Magnetic Resonance Spectroscopy**, v. 102–103, p. 61–97, 2017.

DIAS, D. A.; URBAN, S.; ROESSNER, U. A Historical Overview of Natural Products in Drug Discovery. p. 303–336, 2012.

DOS SANTOS, D. Y. A. C. **Botânica Aplicada : metabólitos secundários na interação planta-ambiente Déborah Yara Alves Cursino dos Santos**. [s.l: s.n.].

DOS SANTOS, H. M. et al. Evaluation of native and exotic Brazilian plants for anticancer activity. **Journal of Natural Medicines**, v. 64, n. 2, p. 231–238, 2010.

ERNST, M. et al. Mass spectrometry in plant metabolomics strategies: From analytical platforms to data acquisition and processing. **Natural Product Reports**, v. 31, n. 6, p. 784–806, 2014.

EVA, G.; WEBER, B. *Encontrados em Plantas e sua*. 2010.

FANG, C.; FERNIE, A. R.; LUO, J. Exploring the Diversity of Plant Metabolism. **Trends in Plant Science**, v. 24, n. 1, p. 83–98, 2019.

FERREIRA, F. R. Germoplasma de Fruteiras. **Revista Brasileira de Fruticultura**, v. 33, n. SPEC. ISSUE 1, p. 1–6, 2011.

FONSECA, N. et al. *Cajá e espécies afins*. p. 29 p, 2018.

FORCISI, S. et al. Liquid chromatography-mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. **Journal of Chromatography A**, v. 1292, p. 51–65, 2013.

FRANCILENE, M. et al. UPLC-HRMS and NMR applied in the evaluation of solid-phase extraction methods as a rational strategy of dereplication of *Phyllanthus* spp . aiming at the discovery of cytotoxic metabolites. **Journal of Chromatography B**, v. 1120, n. December 2018, p. 51–61, 2019.

FUMAGALI, E. et al. Produção de metabólitos secundários em cultura de células e tecidos de plantas: O exemplo dos gêneros *Tabernaemontana* e *Aspidosperma*. **Brazilian Journal of Pharmacognosy**, v. 18, n. 4, p. 627–641, 2008.

FUNARI, C. S. et al. **METABOLÔMICA, UMA ABORDAGEM OTIMIZADA PARA EXPLORAÇÃO DA BIODIVERSIDADE BRASILEIRA: ESTADO DA ARTE, PERSPECTIVAS E DESAFIOS** *Quim. Nova*. [s.l: s.n.].

GAD, H. A.; EL-AHMADY, S. H.; ABOU-SHOER, M. I. Application of Chemometrics in Authentication of Herbal Medicines : A Review. n. April, 2012.

GALDEANO, D. M.; KLEINGESINDS, C. **Microrganismos x Planta: guerra ou parceria?** [s.l: s.n.].

GIKA, H. G. et al. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. **Journal of Pharmaceutical and Biomedical Analysis**, v. 87, p. 12–25, 2014.

GIULIETTI, A. N. A. M.; HARLEY, R. M. Biodiversidade e conservação das plantas no Brasil. v. 1, 2005.

GRANATO, D. et al. **AC SC. Trends in Food Science & Technology**, n. 2018, 2017.

GRIFFIN, J. L.; MERZABAN, J. NMR-based metabolomics in human disease diagnosis : applications , limitations , and recommendations. p. 1048–1072, 2013.

GUEDES, J. A. C. et al. GC-MS-based metabolomic profiles combined with chemometric tools and cytotoxic activities of non-polar leaf extracts of *Spondias mombin* L. and *Spondias tuberosa* Arr. Cam. **Journal of the Brazilian Chemical Society**, v. 31, n. 2, p. 331–340, 2020.

HOOFT, J. J. J. VAN DER. Updates in metabolomics tools and resources : 2014 – 2015
Keywords : p. 86–110, 2016.

INGRET, T. et al. Leaves of *Spondias mombin* L . a traditional anxiolytic and antidepressant : Pharmacological evaluation on zebra fish (*Danio rerio*). **Journal of Ethnopharmacology**, v. 224, n. February, p. 563–578, 2018.

J. A. O OLUGBUYIRO; J. O MOODY; M. T HAMANN. Phytosterols from *Spondias mombin* Linn with Antimycobacterial Activities. **Afr J Biomed Res.**, v. 16, n. 1, p. 19–24, 2013.

- KABERA, J. Plant Secondary Metabolites: Biosynthesis, Classification, Function and Pharmacological Classification, Function and Pharmacological Properties Ethnobotanical survey of medicinal plants with pesticides properties and their potential use in Rwanda View pro. **Journal of Pharmacy and Pharmacology** **2**, v. 2, p. 377–392, 2014.
- KIM, H. K.; VERPOORTE, R. Sample preparation for plant metabolomics. **Phytochemical Analysis**, v. 21, n. 1, p. 4–13, 2010.
- KRASTANOV, A. Metabolomics - The state of art. **Biotechnology and Biotechnological Equipment**, v. 24, n. 1, p. 1537–1543, 2010.
- KUMAR, R.; SHARMA, V. Chemometrics in forensic science. **TrAC - Trends in Analytical Chemistry**, v. 105, p. 191–201, 2018.
- LACCHIA, A. P. S. et al. Foliar colleters in Anacardiaceae: first report for the family. **Botany**, v. 94, n. 5, p. 337–346, maio 2016.
- LEWINSOHN, T. M. Quantas espécies há no Brasil. n. January 2005, 2017.
- LINS, A. U. et al. Anti-inflammatory and antioxidant activity of hydroethanolic extract of *Spondias mombin* leaf in an oral mucositis experimental model. **Archives of Oral Biology**, v. 9969, n. 19, 2020.
- LIU, X.; LOCASALE, J. W. Metabolomics: A Primer. **Trends in Biochemical Sciences**, v. 42, n. 4, p. 274–284, 2017.
- LUBES, G.; GOODARZI, M. Analysis of Volatile Compounds by Advanced Analytical Techniques and Multivariate Chemometrics. **Chemical Reviews**, v. 117, n. 9, p. 6399–6422, 2017.
- LUZANIN, K.; ABRANTES, M. Ressonância Magnética Nuclear - Ferramenta Versátil em Química Farmacêutica e Imagiologia Médica. **Química, Boletim da Sociedade Portuguesa de Química**, v. 117, n. 2, p. 25–30, 2010.
- MACHADO, C. DE F. et al. Banco Ativo de Germoplasma de *Spondias* da Embrapa. **Embrapa**, p. 2–3, 2015.
- MATTIETTO, R. A.; MATTA, V. M. **Cajá (Spondias mombin L.)**. [s.l.] Woodhead Publishing Limited, 2011.
- MGL, B. et al. **BIODIVERSIDADE, USO TRADICIONAL DE PLANTAS MEDICINAIS E PRODUÇÃO DE FITOTERÁPICOS EM MINAS GERAIS**. [s.l.: s.n.].

- MITCHELL, J. D.; DALY, D. C. A revision of Spondias L. (Anacardiaceae) in the neotropics. **PhytoKeys**, v. 55, n. 1, p. 1–92, 2015.
- MUKHERJEE, P. K. Plant Metabolomics and Quality Evaluation of Herbal Drugs. In: **Quality Control and Evaluation of Herbal Drugs**. [s.l.] Elsevier, 2019. p. 629–653.
- NAGANA GOWDA, G. A.; RAFTERY, D. Recent advances in NMR-based metabolomics. **Analytical Chemistry**, v. 89, n. 1, p. 490–510, 2017.
- NEWMAN, D. J.; CRAGG, G. M. **Natural Products as Sources of New Drugs from 1981 to 2014** *Journal of Natural Products* American Chemical Society, , 25 mar. 2016.
- NWIDU, L. L. et al. Hepatoprotective and antioxidant activities of Spondias mombin leaf and stem extracts against carbon tetrachloride-induced hepatotoxicity. **Journal of Taibah University Medical Sciences**, v. 13, n. 3, p. 262–271, 2018.
- OJO, O. A. et al. Inhibitory effects of solvent-partitioned fractions of two nigerian herbs (Spondias mombin Linn. and mangifera indica L.) on α -amylase and α -glucosidase. **Antioxidants**, v. 7, n. 6, p. 1–11, 2018.
- OLIVOTO, T. et al. Plant secondary metabolites and its dynamical systems of induction in response to environmental factors : A review. v. 12, n. 2, p. 71–84, 2017.
- PELL, S. K. et al. Anacardiaceae. In: **Flowering Plants. Eudicots**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 7–50.
- PESSOA, C. et al. Anticancer potential of Northeast Brazilian plants. **Advances in Phytomedicine**, v. 2, n. C, p. 197–211, 2006.
- PIERCE, K. M. et al. Review of chemometric analysis techniques for comprehensive two dimensional separations data. **Journal of Chromatography A**, v. 1255, p. 3–11, 2012.
- PILON, A. et al. Metabolômica De Plantas: Métodos E Desafios. **Química Nova**, n. March, 2020.
- PIMENTEL, V. et al. Biodiversidade brasileira como fonte da inovação farmacêutica: uma nova esperança? **Revista do BNDES** 43, p. 41–89, 2015.
- POOLS, E. G. et al. Germplasm and Genetic Variability. p. 113–128, 2020.
- POTT, D. M.; OSORIO, S.; VALLARINO, J. G. From central to specialized metabolism: An overview of some secondary compounds derived from the primary metabolism for their role

in conferring nutritional and organoleptic characteristics to fruit. **Frontiers in Plant Science**, v. 10, n. June, 2019.

RAMIRES, C. M. C. Desenvolvimento De Clones De Cajazeiras Sobre Diferentes Porta-Enxertos E Diversidade Genética De Acessos Quanto a. 2016.

RAUF, A. et al. Anticancer potential of quercetin: A comprehensive review. **Phytotherapy Research**, v. 32, n. 11, p. 2109–2130, 2018.

REMILA, S. et al. Antioxidant, cytoprotective, anti-inflammatory and anticancer activities of Pistacia lentiscus (Anacardiaceae) leaf and fruit extracts. **European Journal of Integrative Medicine**, v. 7, n. 3, p. 274–286, 2015.

ROA, C. et al. Plant Genetic Resources: Needs, Rights, and Opportunities. **Trends in Plant Science**, v. 21, n. 8, p. 633–636, 1 ago. 2016.

SABIU, S. et al. Indomethacin-induced gastric ulceration in rats: Ameliorative roles of Spondias mombin and Ficus exasperata. **Pharmaceutical Biology**, v. 54, n. 1, p. 180–186, 2016.

SACRAMENTO, C. K.; SOUZA, F. X. Cajá. **Fruticultura Tropical: espécies regionais e exóticas**, p. 83–105, 2009.

SAMEH, S. et al. Genus Spondias: A Phytochemical and Pharmacological Review. **Evidence-based Complementary and Alternative Medicine : eCAM**, v. 2018, 2018.

SCHRIPSEMA, J. Application of NMR in plant metabolomics: Techniques, problems and prospects. **Phytochemical Analysis**, v. 21, n. 1, p. 14–21, 2010.

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, 24 ago. 2015.

SETHI, S. et al. **Metabolomics: From Fundamentals to Clinical Applications**. [s.l: s.n.]. v. 965

SILVA, A. R. A. et al. Antiviral activities of extracts and phenolic components of two spondias species against dengue virus. **Journal of Venomous Animals and Toxins Including Tropical Diseases**, v. 17, n. 4, p. 406–413, 2011.

SILVA, B. M. et al. Genetic diversity of Cajazeira (*Spondias mombin* L.) in three geographic regions. **Genetics and Molecular Research**, v. 16, n. 1, 2017.

SIMÕES, C. M. O. et al. **Farmacognosia: do produto natural ao medicamento**. Porto Alegre: Artmed, 2017.

SIQUEIRA, E. M. D. S. et al. Antiviral Potential of Spondias mombin LLeaves Extract against Herpes Simplex Virus Type-1 Replication Using in Vitro and in Silico Approaches. **Planta Medica**, v. 86, n. 7, p. 505–515, 2020.

SMOLINSKA, A. et al. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. **Analytica Chimica Acta**, v. 750, p. 82–97, 2012.

SOUZA, A. M. DE; RONEI JESUS POPPI. Experimento didatico de quimiometria parte I. **Química Nova**, v. 35, n. 1, p. 223–229, 2012.

SPEHAR, C. R. **Recursos Genéticos**. [s.l: s.n.].

SPEHAR, C. R.; CLAUDIO, P.; KARIA, T. **Recursos Genéticos**. [s.l: s.n.].

TIBURSKI, J. H. et al. Nutritional properties of yellow mombin (*Spondias mombin* L.) pulp. **Food Research International**, v. 44, n. 7, p. 2326–2331, 2011.

TRENERRY, V. C.; ROCHFORT, S. J. Natural products research and metabolomics. **Comprehensive Natural Products II: Chemistry and Biology**, v. 9, p. 595–628, 2010.

TUGIZIMANA, F.; PIATER, L.; DUBERY, I. Plant metabolomics: A new frontier in phytochemical analysis. **South African Journal of Science**, v. 109, n. 5–6, p. 18–20, 2013.

VU, T. N.; LAUKENS, K. Getting your peaks in line: A review of alignment methods for NMR spectral data. **Metabolites**, v. 3, n. 2, p. 259–276, 2013.

WALKER, J. M. **Metabolomics Tools for Natural Product Discovery IN Series Editor**. [s.l: s.n.].

WANG, D. et al. Ellagic acid inhibits proliferation and induces apoptosis in human glioblastoma cells. **Acta Cirurgica Brasileira**, v. 31, n. 2, p. 143–149, 2016.

WENG, J.-K. The evolutionary paths towards complexity: a metabolic perspective. **New Phytologist**, v. 201, n. 4, p. 1141–1149, 1 mar. 2014.

WINK, M. Modes of Action of Herbal Medicines and Plant Secondary Metabolites. **Medicines**, v. 2, n. 3, p. 251–286, 8 set. 2015.

WOLFENDER, J.-L. et al. Plant Metabolomics: From Holistic Data to Relevant Biomarkers.

Current Medicinal Chemistry, v. 20, n. 8, p. 1056–1090, 2013.

YI, L. et al. Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. **Analytica Chimica Acta**, v. 914, p. 17–34, 2016.

ZENG, Z. et al. Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics. **TrAC - Trends in Analytical Chemistry**, v. 53, p. 150–166, 2014.

APÊNDICE A - SUPPLEMENTARY DATA

Untargeted NMR-Based metabolomics applied to study of leaves and bark of the active germplasm bank of *Spondias mombin* L. using multivariate analysis tools

Gladyane S. Silva^a, Elenilson de Godoy A. Filho^d, Christiane Mendes C. Ramires^e, Yandra Karine L. Bezerra^c, Maria Francilene S. Silva^b, Jhonyson Arruda C. Guedes^a, Gisele S. Silva^c, Lorena Mara A. Silva^c, Paulo Riceli^c, Kirley Canuto^c, Edy S. de Brito^c, Ricardo Elesbão^f, Dávila Zampieri^a, Guilherme J. Zocolo^c

^a Departamento de Química Orgânica e Inorgânica, Centro de Ciências, Universidade Federal do Ceará, Campus do Pici, CEP 60451-970, Fortaleza, Ceará, Brasil.

^b Departamento de Engenharia de Alimentos, Centro de Ciências Agrárias, Universidade Federal do Ceará, Campus do Pici, CEP 60356-000, Fortaleza, Ceará, Brasil.

^c Empresa Estadual de Pesquisa Agropecuária da Paraíba, Rodovia BR-230, S/N, Parque Esperança CEP 58108502, Cabedelo, Paraíba, Brasil.

^d Núcleo de Pesquisa e Desenvolvimento de Medicamentos, Universidade Federal do Ceará, Rua Coronel Nunes de Melo 1000, Rodolfo Teófilo, CEP 60420-275, Fortaleza, Ceará, Brasil.

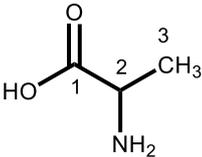
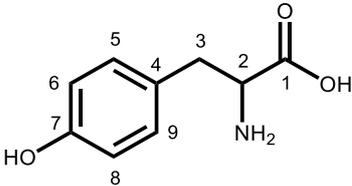
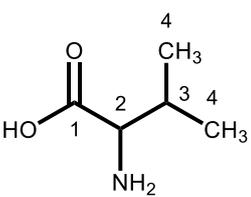
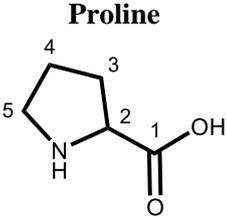
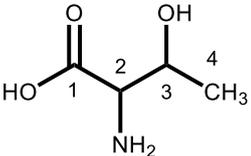
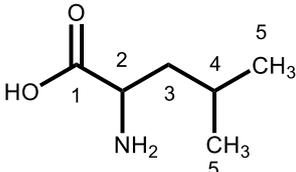
^e Embrapa Agroindústria Tropical, Rua Doutora Sara Mesquita 2270, Pici, CEP 60511-110, Fortaleza, Ceará, Brasil.

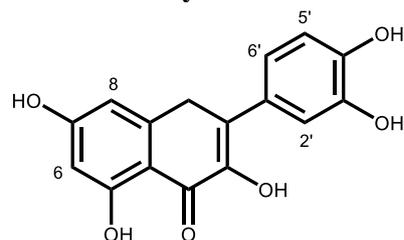
^f Empresa Brasileira de Pesquisa Agropecuária, Embrapa Alimentos e Territórios, Rua Buarque de Macêdo 433-B, Centro, CEP 57020520, Maceió, Alagoas, Brasil.

* Corresponding author:

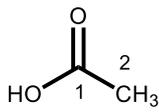
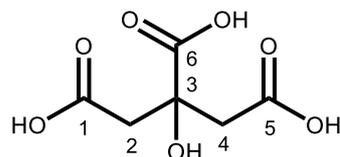
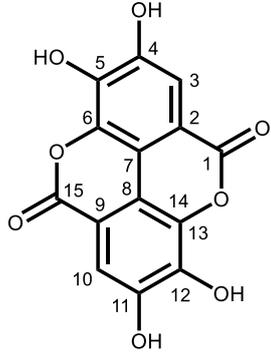
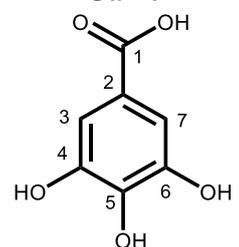
E-mail: guilherme.zocolo@embrapa.br; Phone: +55 (85) 3391-7360

Table 2 – Primary and secondary metabolites identified in *Spondias mombin* L. leaves

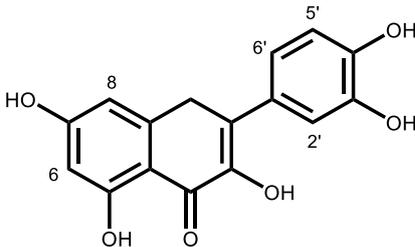
Structures	$\delta^1\text{H}$ (multip.* J in Hz)	$\delta^{13}\text{C}$	$\delta^1\text{H}$ ref.	$\delta^{13}\text{C}$ ref.
AMINO ACIDS				
Alanine				
	2 – 4.00 (o) 3 - 1.51 (d 7.8)	54.1 17.7	3.90 (q 7.3) 1.52 (d 7.3)	53.4 19.1
Tyrosine				
	6,8 - 6.72 (o) 5,9 - 7.22 (d 8.4) 3 - o 2 - o	116.2 133.7 o o	6.84 (d 8.5) 7.17 (d 8.5) 3.17 (n) 3.93 (n)	118.9 133.5 38.3 59.0
Valine				
	2 - 3.62 (o) 3 - 2.36 (no) 4 - 1.06 (d 7.2) 5 - 1.07 (d 7.2)	62.6 33.4 18.4 19.6	3.82 (d 4.4) 2.33 (m) 1.02 (d 7.1) 1.06 (d 7.1)	63.2 32.0 19.1 20.9
Proline				
	6 – 2.90 5 – 1.94 2 – 3.81 3 – 2.44 4 – 2.15	44.0 24.8 60.5 32.7 27.0	3.32 (m) 1.99 (m) 4.12 (dd 8.83; 8.42) 2.34 (m) 2.07 (m)	49.0 26.4 64.0 31.7 31.7
Threonine				
	2 – 3.51 (o) 3 – 4.01 (o) 4 – 1.33 (o)	o 69.0 22.0	3.57 (d 4.87) 4.24 (m) 1.32 (d 6.58)	63.5 68.9 22.3
Leucine				
	2 – no 3 – 2.12 (o) 4 – 1.75 (o) 5 – 0.95 (o)	55.1 46.4 27.3 25.7	3.74 1.70 1.71 0.96	56.2 42.6 26.8 24.8

Phenylalanine				
	5,9 – 7.37 (<i>d</i>)	129.6	7.32 (<i>d</i> 6.98)	132.1
	6.6)	128.7	7.42 (<i>m</i>)	131.8
	6,8 – 7.56 (<i>m</i>)	129.0	7.37 (<i>m</i>)	130.4
	7 – 7.31 (<i>m</i>)			

ORGANIC ACIDS

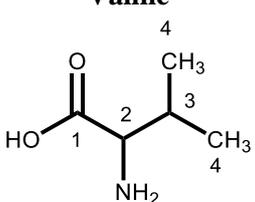
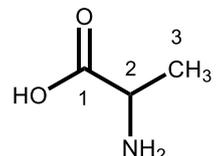
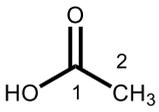
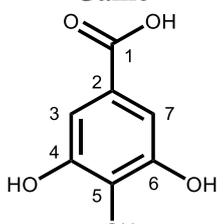
Acetic				
	2 - 1.99 (<i>s</i>)	22.7	1.90 (<i>s</i>)	25.9
Citric				
	2 - 2.43 (<i>o</i>)	45.4	2.52 (<i>d</i> 15.8)	48.6
	4 - 2.70 (<i>o</i>)	45.4	3.66 (<i>d</i> 15.8)	48.6
Ellagic				
	3,10 - 7.47 (<i>s</i>)	112.0	7.58 (<i>s</i>)	113.5
	2 - (-)	116.1	-	115.7
	4 - (-)	143.6	-	146.7
	5 - (-)	157.9	-	159.1
	11 - (-)	151.1	-	148.5
	12 - (-)	142.4	-	140.0
	13 - (-)	138.3	-	136.8
	15 - (-)	163.4	-	160.8
Gallic				
	3,7 - 7.06 (<i>s</i>)	112.0	7.04 (<i>s</i>)	112.2
	2 - (-)	120.7	-	123.9
	4,6 - (-)	142.4	-	138.2
	5 - (-)	148.2	-	147.4
	1 - (-)	168.1	-	167.4

OTHER COMPOUNDS

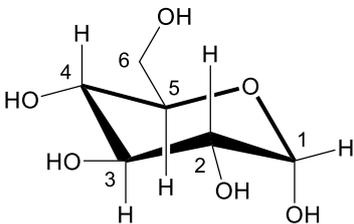
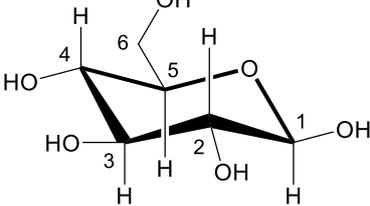
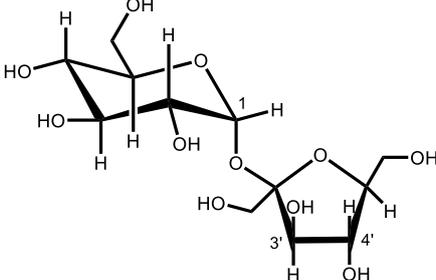
Quercetin				
	6 – 6.20	99.6	6.19	97.9
	8 – 6.39	95.5	6.41	93.1
	2' - 7.46	111.5	7.68	114.9
	5' - 6.90	115.7	6.89	115.3
	6' - 7.83	117.4	7.54	119.8

s – simplet; *d* – duplet; *t* – triplet; *q* – quadruplet; *quin* – quintet; *dd* – double duplet; *o* – overlapping signal; *n* – no information; *no* – not observed.

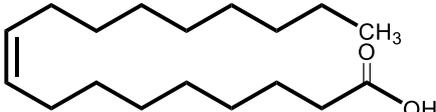
Table 3 – Primary and secondary metabolites identified in *Spondias mombin L.* barks

Structures	$\delta^1\text{H}$ (multip.* J in Hz)	$\delta^{13}\text{C}$	$\delta^1\text{H}$ ref.	$\delta^{13}\text{C}$ ref.
AMINO ACIDS				
Valine				
	2 - no	57.5	3.82 (<i>d</i> 4.4)	63.2
	3 - 2.36 (<i>no</i>)	32.7	2.33 (<i>m</i>)	32.0
	4 – 0.87 (<i>o</i>)	20.2	1.02 (<i>d</i> 7.1)	19.1
	5 – 0.95 (<i>d</i> 7.2)	19.2	1.06 (<i>d</i> 7.1)	20.9
Alanine				
	2 - 3.80 (<i>o</i>)	no	3.90 (<i>q</i> 7.3)	52.1
	3 - 1.49 (<i>d</i> 7.2)	19.2	1.52 (<i>d</i> 7.3)	18.9
ORGANIC ACIDS				
Acetic				
	2 - 1.99 (<i>s</i>)	no	1.90 (<i>s</i>)	25.9
Gallic				
	3,7 - 7.05 (<i>s</i>)	109.6	7.04 (<i>s</i>)	112.2
	2 – (-)	no	-	123.9
	4,6 – (-)	140.0	-	138.2
	5 – (-)	no	-	147.4
	1 – (-)	167.7	-	167.4

CARBOHYDRATES

	α-glucose	1 - 5.11 (<i>d</i> 4.0)	92.5	5.25 (<i>d</i> 3.8)	95.4
		2 - 3.48 (<i>o</i>)	72.4	3.89-3.36 (<i>o</i>)	72.2
		3 - 3.79 (<i>o</i>)	75.0	n	76.0
		4 - 3.52 (<i>o</i>)	74.1	n	72.8
		5 - 3.70 (<i>o</i>)	64.2	n	64.2
		6 - 3.88 (<i>o</i>)	74.2	n	74.5
	β-glucose	1 - 4.47 (<i>d</i> 7.80)	96.9	4.66 (<i>d</i> 8.1)	99.2
		2 - 3.27 (<i>m</i>)	77.2	3.25 (<i>t</i> 8.4)	77.6
		3 - 3.71 (<i>m</i>)	63.7	n	56.1
		4 - 3.50 (<i>m</i>)	79.3	n	79.0
		5 - 3.43 (<i>m</i>)	72.0	n	72.8
		6 - 3.87 (<i>m</i>)	63.1	n	63.1
	Sucrose	1 - 5.39 (<i>d</i> 3.8)	92.2	5.44 (<i>d</i> 3.8)	94.7
		2 - 3.56 (<i>o</i>)	74.0	3.89-3.57 (<i>m</i>)	73.5
		3 - 3.76 (<i>o</i>)	75.8	3.74 (<i>m</i>)	75.3
		4 - 3.48 (<i>o</i>)	72.3	3.47 (<i>m</i>)	71.8
		5 - 3.85 (<i>o</i>)	75.2	3.83 (<i>m</i>)	74.9
		6 - 3.82 (<i>o</i>)	62.5	3.81 (<i>m</i>)	62.8
		1' - 3.82 (<i>o</i>)	65.1	3.66 (<i>m</i>)	64.0
		2' - no	105.2	no	106.3
		3' - 4.22 (<i>m</i>)	80.0	4.20 (<i>t</i> 8.4)	79.0
		4' - 4.05 (<i>m</i>)	77.4	4.08 (<i>d</i> 9.0)	76.6
		5' - 3.89 (<i>o</i>)	84.5	3.87 (<i>m</i>)	84.0
	6' - 3.68 (<i>o</i>)	64.1	3.81 (<i>m</i>)	65.0	

OTHER COMPOUNDS

	Oleic acid	1 -	173.5	-	no
		2 - 2.35 (<i>t</i> 7.8)	34.3	2.35 (<i>t</i> 7.3)	34.0
		3 - 1.63 (<i>m</i>)	24.8	1.64 (<i>m</i>)	24.6
		4-7,12-17 - 1.24-1.38 (<i>m</i>)	22.5-	1.25-1.33 (<i>m</i>)	22.5-31.9
		8,11 - 2.02 (<i>m</i>)	29.1	2.02 (<i>m</i>)	27.1
		9,10 - 5.34 (<i>m</i>)	26.7	5.34 (<i>m</i>)	129.9
		18 - 0.90 (<i>t</i> 7.2)	130.0	0.89 (<i>t</i>)	14.1
			13.0		

s – singlet; *d* – doublet; *t* – triplet; *q* – quadruplet; *quin* – quintet; *dd* – double doublet; *o* – overlapping signal; *n* – no information; *no* – not observed.

Ellagic: Sono-transformation of tannic acid into biofunctional ellagic acid micro/nanocrystals with distinct morphologies.

References

- [41] L.C. de Rezende, P. de A. Santos, V.B. Riatto, J.M. David, J.P. David, New alkyl phenols and fatty acid profile from oils of pulped *Spondias mombin* L. seed wastes, *Química Nova*. 41 (2018) 540-543. doi: <https://doi.org/10.21577/0100-4042.20170203>.
- [42] E.G.A. Filho, L.N. Braga, L.M.A. Silva, F.R. Miranda, E.O. Silva, K.M. Canuto, M.R. Miranda, E.S. de Brito, G.J. Zocolo, Physiological changes for drought resistance in different species of *Phyllanthus*, *Scientific Reports*. 8 (2018) 1-12. doi: <https://doi.org/10.1038/s41598-018-33496-7>.
- [43] E.G.A. Filho, L.M.A. Silva, P.R. Ribeiro, E.S. de Brito, G.J. Zocolo, P.C. Souza-Leão, A.T. Marques, A.L. Quintela, F.H. Larsen, K.M. Canuto, ¹H NMR and LC-MS-based metabolomic approach for evaluation of the seasonality and viticultural practices in wines from São Francisco River Valley, a Brazilian semi-arid region, *Food Chemistry*. 289 (2019) 558-567. doi: <https://doi.org/10.1016/j.foodchem.2019.03.103>.