UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

JONATHAN ANDRE MORALES MARROQUÍN

VARIAÇÕES GENÔMICAS ASSOCIADAS À ADAPTAÇÃO EM PALMEIRAS DO GÊNERO ACROCOMIA NO NEOTRÓPICO

GENETIC VARIATIONS ASSOCIATED WITH ADAPTATION IN ACROCOMIA PALMS ACROSS THE NEOTROPICS

CAMPINAS
2024

JONATHAN ANDRE MORALES MARROQUÍN

VARIAÇÕES GENÔMICAS ASSOCIADAS À ADAPTAÇÃO EM PALMEIRAS DO GÊNERO ACROCOMIA NO NEOTRÓPICO

GENETIC VARIATIONS ASSOCIATED WITH ADAPTATION IN ACROCOMIA PALMS ACROSS THE NEOTROPICS

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento

Thesis presented to the Institute of Biology of the University of Campinas for fulfillment of the requirements for the degree of Doctor in Genetics and Genetic Breeding.

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO JONATHAN ANDRE MORALES MARROQUÍN, E ORIENTADO PELA PROFA. MARIA IMACULADA ZUCCHI

Orientadora: Profa. Dra. Maria Imaculada Zucchi
Co-orientador: Prof. Dr. Carlos Augusto Colombo

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Biologia
Gilberto Martins Vicentin - CRB 8/8620

M792v

Morales-Marroquín, Jonathan Andre, 1989-
Variações genômicas associadas à adaptação em palmeiras do gênero *Acrocomia* no Neotrópico / Jonathan Andre Morales Marroquín. – Campinas, SP : [s.n.], 2024.

Orientador(es): Maria Imaculada Zucchi.
Coorientador(es): Carlos Augusto Colombo.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Biologia.

1. Macaúba. 2. Palmeira. 3. Genotipagem por sequenciamento. 4. Genética de populações. 5. Genômica. I. Zucchi, Maria Imaculada, 1971-. II. Colombo, Carlos Augusto. III. Universidade Estadual de Campinas (UNICAMP). Instituto de Biologia. IV. Título.

Informações complementares

Título em outro idioma: Genetic variations associated with adaptation in *Acrocomia* palms across the Neotropics
Palavras-chave em inglês:
Acrocomia aculeata
Palms
Genotyping by sequencing
Population genetics
Genomics
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
Maria Imaculada Zucchi [Orientador]
Anete Pereira de Souza
Fernanda Amato Gaiotto
Evellyn Giselly de Oliveira Couto
Suelen Alves Vianna
Data de defesa: 20-06-2024
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-7946-577X
- Currículo Lattes do autor: http://lattes.cnpq.br/1821805203521800

Campinas, 20 de junho 2024

Comissão examinadora

Profa. Dra. Maria Imaculada Zucchi

Dra. Evellyn Giselly de Oliveira Couto

Profa. Dra. Anete Pereira de Souza

Profa. Dra. Fernanda Amato Gaiotto

Dra. Suelen Alves Vianna

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de Genética e Biologia Molecular do Instituto de Biologia da UNICAMP.

A mis padres,

Oscar *y* ***Amarilis***

A mis hermanos,

Hoskar*,* ***Oscar Leonardo*** *y* ***Johan***

A mi tribu,

Erick*,* ***Daniela****,* ***Regina****,* ***Mafer****,*
Adriana*,* ***Rocío*** *y* ***Ramón***

A mi tierra,

Guatemala *y* ***Centroamérica***

A mi hogar en el Sur,

Brasil

y a la fuerza que
mueve el ***Universo***

Con todo mi amor se los dedico.

Agradecimentos

> “*Respeito muito minhas lágrimas, ainda mais minha risada.*
> *Escrevo assim minhas palavras, na voz de uma mulher sagrada*”.
> – **Gal Costa** (Compositor: Caetano Veloso, 1984)

Gostaria de agradecer a todos que contribuíram para a realização desse trabalho, em especial:

À minha orientadora, Profa. Maria Imaculada Zucchi, pela oportunidade de desenvolver e escutar cada uma das ideias que propus com tanto carinho e profissionalismo, sem dúvidas cresci como profissional e muito mais como humano. Obrigado Professora por me ensinar a ser um pesquisador. Levarei todos seus ensinamentos sempre no coração.

À meu co-orientador, Prof. Carlos Colombo, pelo apoio em cada momento. Sempre com um sorriso, um ensinamento e uma mão para dar. Obrigado por fazer possível a nossa viagem de prospecção pela América Central.

Ao Prof. José Baldin Pinheiro, pelo apoio incondicional e por todas as oportunidades e colaborações. Obrigado pela confiança e por ser um exemplo.

Ás minhas colaboradoras, as pesquisadoras Suelen Vianna, Brenda Díaz-Hernández, e Evellyn Couto pela paciência e todos os ensinamentos na cultura da Macaúba.

Aos pesquisadores da rede de colaboradores formada pelo “Proyecto Coyol” em cada um dos países da América Central, especialmente a Erick López, Rocío Silva, Roger Orellana, Emmanuel Araya Valverde, Elizabeth Arnaez e José Palácios.

Ao meio jornalístico El Periódico de Guatemala e Nestlé Guatemala por acreditar no “Proyecto Coyol” e ser patrocinadores da nossa jornada pela América Central.

Aos meus camaradas do coração: Marcela, Flaviane, Carlão, Igor, Marcones, Caroline, João, Jota, Daya, Moro e Aninha. Obrigado por tornarem meus dias mais leves. Pelas palavras, pelas

aventuras e, sobretudo, por me ensinarem que um bom pesquisador sorri diante do erro e persevera com disciplina até alcançar um bom resultado. Obrigado, queridos.

Aos meus queridos amigos do Grupo de Genética e Genômica da Conservação: Marcão, Gabriel, Kauanne, Scaketti, Gabi, Man-Gabis, André, Du, Zirlane, Marcos, Roberto, Ana Paula, Luciano, Patrícia e Cesar. Obrigado pela amizade e os ensinamentos. Contem comigo sempre!

À Universidade Estadual de Campinas e ao Programa de Pós-graduação em Genética e Biologia Molecular, pelo ensino público de qualidade.

Á Escola Superior de Agricultura Luiz de Queiroz da Universidade Estadual de São Paulo (Esalq – USP), por me acolher e fazer possível desenvolver esta pesquisa no Laboratório de Diversidade Genética e Melhoramento dirigido pelo Prof. José Baldin Pinheiro.

A minha família piracicabana, Mari, Jaime, Ramón, Dona Débora, Dona Neusa, Lalá e toda a família Pedrozo, por alimentarem meu coração e me darem todo o apoio que preciso.

E, finalmente, aos meus pais, Amarilis e Oscar, por me amarem incondicionalmente. Este trabalho é por e para vocês. Obrigado por me ensinarem a falar com Deus. Juntos até o fim.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, e também contou com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil, Processos nº: 2021/10319-0, 2019/20307-0, 2014/23591-7, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) , Processo nº 431226/2018-0.

Sobre os ombros de gigantes, a todos, muito obrigado!

Resumo

Nossa pesquisa se concentra na adaptação genética e na estrutura populacional de três espécies de palmeiras neotropicais do gênero *Acrocomia*, conhecidas por seu alto potencial para extração de óleo. Utilizando a técnica Genotyping-by-Sequencing (GBS), estudamos os processos genéticos adaptativos em *A. aculeata*, *A. totai* e *A. intumescens* na região Neotropical. Para *A. aculeata*, identificamos dois grandes grupos genômicos, um Norte-Americano e outro Sul-Americano, influenciados pela dispersão e pelas barreiras biogeográficas, com assinaturas seletivas ligadas à biossíntese de ácidos graxos e carotenoides, resistência a patógenos e adaptação ao estresse ambiental. *A. totai* apresentou uma estrutura genética que sugere uma diversificação recente impulsionada por fatores climáticos e geológicos especialmente no bioma do Pantanal, enquanto a estruturação genética de *A. intumescens* foi moldada por barreiras biogeográficas no bioma da Caatinga. Concentrando-se no grupo da América do Norte, na América Central, nossa pesquisa revelou três grupos genômicos primários em *A. aculeata*: Mesoamericano, Costa-riquenho e Panamenho, cada um com subpopulações distintas. O grupo Mesoamericano apresentou a maior diversidade genética, influenciada por rotas de migração passadas e barreiras biogeográficas, como a depressão nicaraguense e a cadeia montanhosa de Talamanca. A modelagem do nicho ecológico mostrou alta adequação das terras baixas da América Central para *A. aculeata*. Nossos resultados discutem a história evolutiva e a divergência genética moldada por barreiras biogeográficas, essenciais para a conservação e estratégias de reprodução para compreender a diversidade genética nas populações naturais de *A. aculeata*. O estudo destaca o potencial de *Acrocomia* como uma nova cultura para a produção de óleo, biocombustíveis e agricultura sustentável. Pesquisas futuras devem priorizar o sequenciamento do genoma completo e estudos de interação genótipo-ambiente para otimizar o papel de *Acrocomia* no desenvolvimento sustentável e na energia renovável.

Abstract

Our research focus on the genetic adaptation and population structure of three neotropical palm species of the *Acrocomia* genus, known for their high oil extraction potential. Using Genotyping-by-Sequencing (GBS), we study the adaptive genetic processes in *A. aculeata, A. totai*, and A. *intumescens* across the Neotropics. For *A. aculeata*, we identified two major gene pools, a North American and a South American, influenced by dispersal and biogeographic barriers, with selective signatures linked to fatty acid and carotenoid biosynthesis, pathogen resistance, and environmental stress adaptation. *A. totai* showed a genetic structure suggesting recent diversification driven by climatic and geological factors in the Pantanal biome, while *A. intumescens*' genetic structuring was shaped by biogeographic barriers in the Caatinga biome. Focusing in the North American group, in Central America, our research revealed three primary genomic groups within *A. aculeata*: Mesoamerican, Costa Rican, and Panamanian, each with distinct subpopulations. The Mesoamerican group exhibited the highest genetic diversity, influenced by past migration routes and biogeographic barriers such as the Nicaraguan depression and the Talamanca mountain chain. Ecological niche modeling indicated high suitability of Central American lowlands for *A. aculeata*. Our results discuss the evolutionary history and genetic divergence shaped by biogeographic barriers, key for conservation and breeding strategies to understand genetic diversity in *A. aculeata* natural populations. The study highlights *Acrocomia*'s potential as a novel crop for oil production, biofuels, and sustainable agriculture. Future research should prioritize whole-genome sequencing and genotype-environment interaction studies to optimize *Acrocomia*'s role in sustainable development and renewable energy.

Sumario

1. Organização da Tese

Para começar, fornece-se um referencial bibliográfico que descreve os principais conceitos discutidos nos capítulos que compõem o documento, sobre os processos adaptativos, pré-melhoramento e conservação das palmeiras do gênero *Acrocomia* especialmente para a espécie com maior importância económica, *A. aculeata* (Jacq.) Lodd. ex Mart., como uma cultura emergente no Neotrópico.

O **Capítulo I** descreve as variações genômicas associadas aos processos de adaptação do gênero *Acrocomia* através da América, para o aprimoramento da cultura visando a seleção assistida por marcadores. Neste trabalho, explora-se a estrutura genômica populacional das três espécies de maior importância econômica do gênero, *A. aculeata*, *A. totai* Mart. e *A. intumescens* Drude em toda sua distribuição no continente Americano. Descrevem-se também genes candidatos que possam estar sujeitos a processos de seleção e adaptação. Foram identificados genes associados ao metabolismo de lipídios e triglicerídeos, o que é de suma importância na domesticação da espécie. Este é o primeiro trabalho abrangendo os processos adaptativos em *Acrocomia* a nível genômico em todo o Neotrópico. O Capítulo I já foi submetido a uma revista internacional renomada e foi escolhido para compor a tese por ser o tema principal deste projeto de doutorado.

O **Capítulo II** foca no grupo genético da América do Norte de *Acrocomia aculeata*, especificamente nas populações naturais dos países da América Central. Pouco se sabe sobre a espécie na região, já que historicamente estes países tem sido negligenciados, tanto na produção científica quanto no desenvolvimento humano. Além disso, nas últimas décadas o melhoramento da espécie tem-se concentrado no grupo genético da América do Sul. Este é o primeiro trabalho de genotipagem de uma planta terrestre em toda a sua distribuição na América Central (Guatemala, Honduras, Nicarágua, Costa Rica e Panamá) utilizando sequenciamento de nova geração (NGS) abrangendo todo o genoma. Neste capítulo, descreve-se o hábito da espécie na região, a estrutura genômica e os possíveis processos biogeográficos que explicam a grande diversidade e a estrutura populacional da Macaúba na América Central. Este capítulo está em aprimoramento para sua futura publicação, por isso foi escolhido para compor a tese.

Como parte da pergunta inicial do trabalho, o aluno trabalhou na delimitação das 3 espécies de maior relevância económica do gênero *Acrocomia* fazendo o sequenciamento, anotação e descrição dos genomas dos cloroplastos (plastomas) de *Acrocomia totai* e *Acrocomia intumescens* para ser comparados com *A. aculeata*. O trabalho foi publicado em uma revista de alto impacto e pode ser acessado na seção de **Anexos**. Como o trabalho já foi publicado, foi decidido não incluí-lo no documento de tese.

Além disso, as implicações das políticas públicas na conservação de recursos genéticos, biodiversidade e governança de recursos naturais da América Central são discutidas em dois trabalhos anexados concebidos e publicados pelo aluno. O primeiro é um artigo científico publicado em uma revista de alto impacto onde se discute a produção científica em temas de biodiversidade na América Central. O segundo trabalho foi parte da chamada da IATT (Inter-Agency Task Team) das Nações Unidas para iniciativas de políticas científicas para o Fórum STI 2023 (Fórum de Múltiplas Partes Interessadas em Ciência, Tecnologia e Inovação para os Objetivos de Desenvolvimento Sustentável). Ambos trabalhos podem ser encontrados na seção de **Anexos**. O aluno também contribuiu significativamente com outros trabalhos dentro do grupo de pesquisa, colaborando com outros alunos e pesquisadores. Os trabalhos em colaboração também estão anexados ao final da tese.

2. Referencial Bibliográfico

2.1 Introdução e Justificativa para estudos moleculares em *Acrocomia*

A atual crise energética requer a busca por novas fontes de energia de base biológica, renováveis e recicláveis e de tecnologias que priorizem a qualidade de vida e do meio ambiente. Isso permitirá a diminuição das emissões de gás carbônico, a correta utilização da biodiversidade e a segurança alimentar - o qual abrirá amplas perspectivas para utilização de óleos de origem vegetal como substitutos aos combustíveis fósseis. Como recursos vegetais amplamente utilizados, citam-se as espécies oleaginosas, que tem seu óleo empregado nas indústrias de alimentação humana, cosmética, farmacêutica, automação, biocombustíveis entre outras. Segundo (Vargas-Carpintero *et al.*, 2022; Shahbandeh, 2024), o consumo mundial de óleos vegetais em 2023/24 foi de 217,99 milhões de toneladas, com estimativa para 2024/25 de mais de 223 milhões de toneladas. Do total de óleo consumido mundialmente, cerca de 30% é extraído do dendezeiro (*Elaeis guineensis* Jacq.) (77,99 milhões de toneladas) sendo 90% aplicado no setor de alimentos e os outros 10% no setor industrial e de biocombustíveis. No Brasil, mais de 84% do óleo produzido é utilizado para fins alimentícios e aproximadamente 16% para fins industriais, sendo o óleo de soja o mais consumido nos dois tipos de uso (86%), seguido pelo óleo de algodão e de dendê (Colombo *et al.*, 2018). Embora o setor de alimentos tenha a maior participação no mercado de consumo, espera-se que o segmento de biocombustíveis, em virtude do apelo ambiental e da diversificação da matriz energética cresça rapidamente sabendo que o Brasil é o segundo maior consumidor de biocombustíveis (Vargas-Carpintero *et al.*, 2021).

Pensando em novas fontes de energia renovável, as palmeiras representam uma excelente alternativa para produção de óleo, além de serem utilizadas como recurso de subsistência para muitos povos e apresentarem um grande potencial econômico devido à sua multiplicidade de usos. No Brasil, existem extensões de palmeiras nativas, como por exemplo, as palmeiras do gênero *Acrocomia*, que ocorrem em quase todo o território nacional (Lorenzi *et al.*, 2010). No entanto, sua exploração é realizada, em grande parte, de forma extrativista por comunidades tradicionais e rurais, e com baixa produtividade devido à falta de conhecimento sobre sua morfologia, fisiologia, taxonomia, aspectos produtivos, ecologia e diversidade genética e biologia da reprodução (Suelen Alves Vianna *et al.*, 2017; Colombo *et al.*, 2018; de Lima *et al.*, 2018).

As palmeiras do gênero *Acrocomia* podem ter todas as suas partes utilizadas: estipe, folhas, espinhos e palmito, destacando-se o grande potencial para geração de energia a partir da utilização do óleo de seus frutos como matéria-prima para produção de biocombustíveis. A produção de frutos de acordo com estimativas, pode chegar a 140 t.ha.ano-1 na espécie *Acrocomia aculeata* (Jacq.) Lodd. ex Mart., conhecida popularmente no Brasil como macaúba. Ela apresenta até 78% de óleo na polpa seca. Estudos tem demostrado grande diversidade genética, sendo o Brasil provavelmente o país que apresenta a maior variabilidade. Populações promissoras (fruto com maior teor oleaginoso) foram encontradas principalmente nos estados de Minas Gerais e São Paulo as quais apresentam genótipos interessantes para os programas de melhoramento (Ciconini *et al.*, 2013; Colombo *et al.*, 2018; Vargas-Carpintero *et al.*, 2021). Porém, a pesquisa é ainda incipiente e há uma grande necessidade de ter acesso a recursos genômicos como o genoma de referencia da espécie. Estes recursos ampliariam o conhecimento da estrutura genética das populações para poder associar fenótipo/genótipos e entender os processos de diversificação do gênero na América, tanto para sua conservação como para seu melhoramento.

A produção de frutos é o interesse principal destes programas com foco no desenvolvimento socio-econômico. Compreender a estrutura genética das populações naturais em toda sua distribuição e as marcas genômicas associadas a seleção é de suma importância para melhorar a produtividade da Macaúba. O gênero *Acrocomia* possui nove espécies que apresentam diferentes distribuições ao longo da América tropical e subtropical, sendo *A. aculeata* a que apresenta a maior distribuição e importância econômica no Brasil. É importante salientar que, embora *A. aculeta* seja a mais estudada no gênero, as outras espécies do complexo também podem ser utilizadas comercialmente, dentre elas as espécies *A. totai* Mart. e *A. intumescens* Drude.

A proposta deste trabalho foi a coleta e análises genômicas de populações nativas das espécies mais produtivas é importantes economicamente do gênero: *A. aculeata*, *A. intumescens* e *A. totai* representativas das diferentes ecorregiões (regiões biogeográficas) propostas para as palmeiras na América. O estudo foi conduzido utilizando-se três abordagens: (1) genômica de populações e estruturação da diversidade das três espécies principais, (2) detecção de genes por sinais putativos de seleção e adaptação utilizando SNPs para o melhoramento assistido com marcadores, (3) a descrição do grupo genético de *Acrocomia aculeata* para América Central e (4) comparação estrutural e anotação do genoma do cloroplasto (plastoma) de *A. intumescens* e *A. totai*. A genômica de populações e os recursos genéticos serão fundamentais para ampliar a visão sobre a diversidade, elucidar as dinâmicas do fluxo gênico e estrutura das populações e assim compreender melhor a história natural das diferenças dentro do gênero *Acrocomia*. Este trabalho permitirá entender melhor as diferenças entre diversidade do pool gênico norte-americano e sul-americano como também gerará informações de extrema importância para o manejo da espécie, tanto na manutenção de plantios comerciais em programas de pre-melhoramento genético, assim como na conservação da diversidade.

2.2 *Acrocomia* – taxonomia, distribuição e aspectos socioeconômicos

O gênero *Acrocomia* pertence à tribo Cocoseae (Arecaceae: Arecoideae), subtribo Bactridinae nativa do Neotrópico. O gênero é representado por espécies neotropicais sendo reconhecidas atualmente nove espécies (Lorenzi *et al.*, 2010; 'The Plant List', 2013; de Lima *et al.*, 2018): *A. aculeata* (Jacq.) Lodd ex Mart., *A. corumbaensis* S. A. Vianna, *A. crispa* (Kunth) C. F. Baker ex Becc., *A. intumescens* Drude, *A. media* O. F. Cook e *A. totai* Mart. de porte arbóreo, *A. glaucescens* Lorenzi de baixo porte e, *A. emensis* (Toledo) Lorenzi e *A. hassleri* (Barb. Rodr.) W. J. Hahn rizomatosas. Destas, três apresentam considerável importância socioeconômica em suas respectivas áreas de distribuição: *A. aculeata*, *A. intumescens* e *A. totai* (Fig 1).

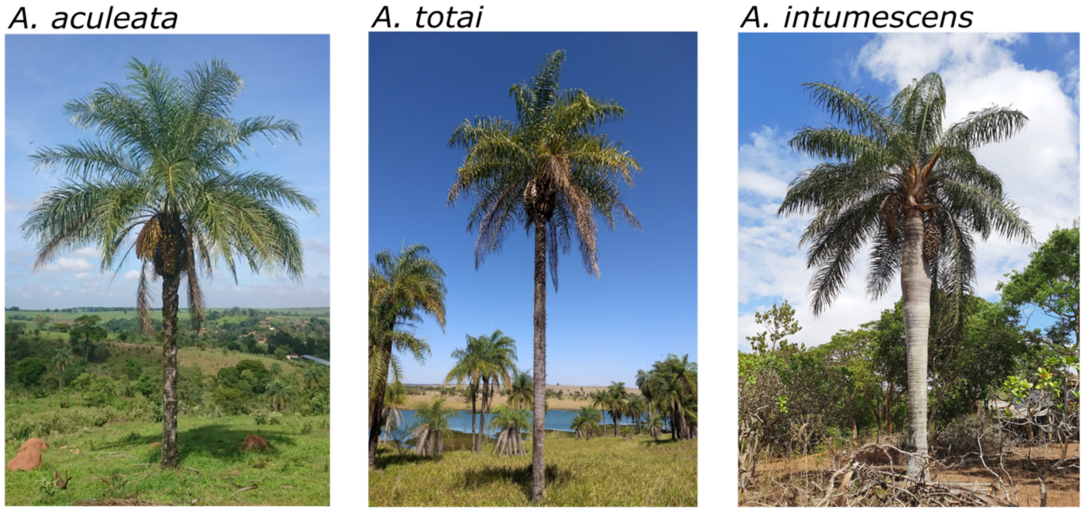


Fig 1. Plantas adultas de *Acrocomia aculeata*, *Acrocomia totai*, *Acrocomia intumescens*. Localização: *A. aculeata* - Itapira, São Paulo, Brasil. *A. totai* - Presidente Epitácio, São Paulo, Brasil. *A. intumescens* - Vale da Neblina, Paraíba, Brasil. Créditos das fotos: Brenda Díaz (*A. aculeata* e *A. totai*) e Eulampio Duarte (*A. intumescens*).

Acrocomia aculeata é a espécie com maior distribuição geográfica do gênero, ocorrendo desde a região central do México, passando por toda a América Central até o norte da Argentina (exceto no Equador, Peru e Chile) e maior parte do Brasil (de Lima *et al.*, 2018) preferencialmente em áreas abertas associadas ao Cerrado (Fig 2B). A espécie pode ser facilmente reconhecida pela grande quantidade de espinhos no estipe, frequente presença dos resquícios da bainha das folhas já caídas, presença de folhas senis e frutos grandes (Lorenzi *et al.*, 2010). Devido à sua ampla distribuição geográfica, sua utilização é feita de diferentes formas. No México e alguns países da América Central a seiva do estipe é utilizada para fabricação de uma bebida fermentada (Díaz *et al.*, 2021). No Brasil os frutos são explorados comercialmente para extração de óleo da polpa para biocombustíveis e da amêndoa para produtos farmacêuticos e indústria de cosméticos (Motoike *et al.*, 2013; Aires and de Carvalho Junior, 2023). Destaca-se que dentre as espécies de *Acrocomia* utilizadas comercialmente, *A. aculeata* apresenta os maiores teores de óleo na polpa (33 a 78% de óleo na polpa seca) (Colombo *et al.*, 2018). Na produção de óleo, é muito superior a outras culturas destinadas para tal fim, estando abaixo somente do dendê (*Elaeis guineensis* Jacq.), mas com potencial para ultrapassar em produção, já que os dados atuais são somente de populações nativas e não genótipos melhorados. Ela pode se hibridizar com *Acrocomia totai*. Embora *A. aculeata* seja uma espécie aceita, alguns autores consideram os indivíduos das regiões norte-americanas (América Central, Antilhas e México) como espécies diferentes. Em muitas citações os indivíduos pertencentes ao grupo norte-americano são tratados como *Acrocomia mexicana* Karw. ex Mart., *Acrocomia quisqueyana* L.H. Bailey, *Acrocomia vinífera* Oerst. Isto pode ser devido a diferenças vegetativas e por ambas apresentarem grande plasticidade dos atributos morfológicos, porém estudos recentes sugerem que existe uma grande diferenciação entre o pool genético da América do Sul e da América do Norte, mas não possuem amostragem populacional suficiente (Díaz *et al.*, 2021).

Acrocomia totai é a segunda espécie do gênero com maior distribuição, ocorrendo nos estados de Mato Grosso do Sul e Tocantins, região norte do Paraná, região sudoeste de São Paulo, Bolívia, Paraguai e norte da Argentina (Lorenzi *et al.*, 2010) preferencialmente em terrenos temporariamente alagados de várzeas úmidas e às vezes em áreas mais secas (Fig 2B). Comparativamente à *A. aculeata*, apresenta menor quantidade de espinhos no estipe, abscisão nítida das folhas, cachos e frutos menores, menor teor de óleo na polpa e maior variabilidade na coloração da casca e da polpa (Fig 2) (Vianna *et al.*, 2021). Tem sido descritos híbridos entre *A. aculeata* e *A. totai*. No Paraguai, os frutos da espécie são utilizados para extração de óleo da amêndoa para produção de sabão. Nos demais locais, sobretudo na região do Pantanal sul-mato-grossense, a polpa é utilizada para consumo *in natura* ou processada em diferentes alimentos apresentando excelente qualidade nutricional (Vianna *et al.*, 2021).

Acrocomia intumescens é uma das espécies endêmicas do gênero, ocorrendo somente em algumas áreas do nordeste brasileiro associadas a áreas de Mata Atlântica principalmente nos estados de Pernambuco e Ceará (Fig 2B) (Lorenzi *et al.*, 2010). Morfologicamente, pode ser facilmente identificada pelo intumescimento no terço médio do estipe, ausência de espinhos no estipe na fase adulta, abscisão completa de folhas senis (Fig 1) cachos e frutos com tamanhos intermediários à *A. aculeata* e *A. totai,* e menor variação das cores da casca e da polpa (Suelen Alves Vianna *et al.*, 2017). Nesses Estados a exploração dos frutos é principalmente para consumo da polpa *in natura* e utilização da mesma para processamento de diferentes alimentos e medicamentos naturais. Dentre as espécies do gênero, *A.* ***intumescens*** pode ser considerada a segunda mais produtiva em óleo, apresentando de 34 a 41% de óleo na polpa em base seca (Bora and Rocha, 2004). A espécie apresenta potencial ornamental e em alguns locais já é utilizada para tal fim, a exemplo do Estado de Pernambuco, onde é muito utilizada em praças, parques municipais e jardins particulares.

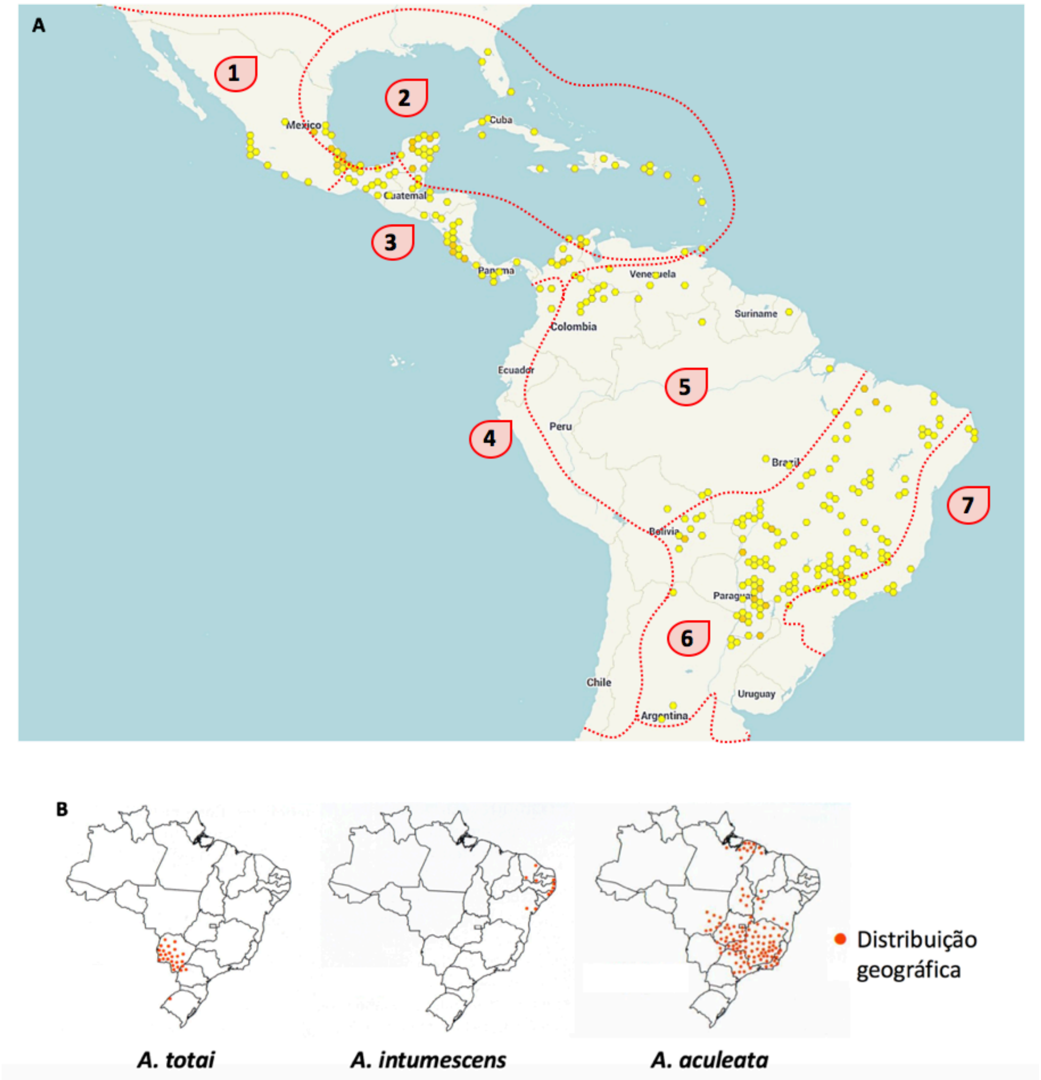


Fig 2. Mapas de distribuição de *Acrocomia aculeata*, *A. intumescens* e *A. totai*. (A) Mapa das Regiões ecogeográficas (biogeográficas) propostas para a família Arecaceae, sendo elas: (1) Região Mexicana, (2) Região do Caribe, (3) Região da América Central, (4) Região Andina, (5) Região Amazônica, (6) Região do planalto central brasileiro, (7) Região da Mata Atlântica (Henderson, Galeano-Garces and Bernal, 1995; Eiserhardt *et al.*, 2013; Roncal *et al.*, 2013). Os hexágonos representam a localização dos registros do gênero Acrocomia na base de dados GBIF (acesso 27/06/2023) na América. Os hexágonos mais escuros reúnem maior quantidade de pontos de ocorrência presentes na base de dados. (B) Distribuição no território brasileiro das espécies *A. totai*, *A. intumescens* e *A. aculeata* (Lorenzi *et al.*, 2010)

2.3 Historia natural do gênero *Acrocomia*

A família Arecaceae está distribuída pantropicalmente, sendo organismos modelo para estudos de ecologia evolutiva dos biomas que compõem as florestas tropicais no mundo

(Eiserhardt *et al.*, 2011; Kissling *et al.*, 2012; William J Baker and Couvreur, 2013; William J. Baker and Couvreur, 2013; Eiserhardt, Couvreur and Baker, 2017). A história evolutiva do gênero *Acrocomia* ainda não é clara, e estudos com relógio molecular sugerem que os primeiros representantes diversificaram há 22,3 Ma no Mioceno na América do Sul (provável centro de diversificação ou origem) (Cano *et al.*, 2022). Após isso, expande-se para América do Norte. As interações bióticas dos biomas como os processos de polinização e dispersão de sementes feita por aves e megamastofauna no Grande Intercâmbio da Biota Americana (GABI) ocorrida no Plioceno (2,8 Ma) também modularam a diversidade (Janzen and Martin, 1982; O'Dea *et al.*, 2016; Eiserhardt, Couvreur and Baker, 2017).

No presente, *Acrocomia* encontra-se em todas as ecoregiões (regiões biogeográficas) propostas para a família Arecaceae do Neotrópico (Fig 2). Apresenta ampla distribuição geográfica, ocorrendo desde o México e Antilhas até Argentina (Henderson, Galeano-Garces and Bernal, 1995; Eiserhardt *et al.*, 2013; Roncal *et al.*, 2013). Sendo *A. aculeata* a única espécie que ocorre em todas elas. *Acrocomia* responde aos processos de radiação adaptativa, sendo um grupo monofilético. O registro fóssil de *Acrocomia* também está associado a eventos mais modernos (≈300 mil anos) como na etnobotânica e usos extrativistas das diferentes expressões das sociedades e civilizações humanas. *Acrocomia aculeata* considera-se uma espécie incipientemente domesticada, ou seja as populações tem tanto a média quanto a variação de uma característica selecionada dentro da variação da espécie. (Clement, 1999a, 1999b; Levis *et al.*, 2018; Clement *et al.*, 2021). Cabe falar também que elas foram manejadas por diferentes civilizações pré-colombianas, sendo na América do Norte consumido pelos povos ancentrais Olmecas, Maias e Aztecas e seus descendentes (Lentz, 1990; Ramírez-Núñez *et al.*, 2019). Já os indivíduos Sulamericanos pelos povos Amazônicos, Tupi e Guaranis (Clement, 1999a; Morcote-Ríos and Bernal, 2001; Cámara-Leret, Fortuna and Bascompte, 2019). Este tipo de separação pode ser observado em outras plantas domesticadas como o feijão e milho (Bitocchi *et al.*, 2012; Bedoya *et al.*, 2017).

2.4 Diversidade genética do gênero *Acrocomia*

Fatores ambientais, ecológicos e populacionais afetam a taxa de fixação das frequências alélicas que dirigem a adaptação das populações naturais ao longo de muitas gerações. Estas linhagens ou metapopulações podem experimentar um isolamento reprodutivo seja simpátrico ou alopátrico e, consequentemente, algum evento de especiação ou extinção que possa ser perceptível a nível populacional (Yardeni *et al.*, 2016). Estudar a genética de populações e filogenia de grupos nativos, como as palmeiras, pode ser uma janela para entender a microevolução da hiperdiversidade das florestas tropicais (Gutiérrez-García and Vázquez-Domínguez, 2013; Eiserhardt, Couvreur and Baker, 2017).

Nos últimos anos, vários estudos têm procurado avaliar a variabilidade genética de *Acrocomia* spp. utilizando dados fenotípicos e genotípicos. Por serem plantas pioneiras, de polinização aberta e com ampla ocorrência no neotrópico, apresentam grande plasticidade fenotípica (Suelen A Vianna *et al.*, 2017; Suelen Alves Vianna *et al.*, 2017; Colombo *et al.*, 2018; Vianna *et al.*, 2021) e variabilidade dos constituintes químicos dos frutos (Bora and Rocha, 2004; Colombo *et al.*, 2018; Alfaro-Solís *et al.*, 2020; Vianna *et al.*, 2021; Aires and de Carvalho Junior, 2023) e diversidade genética (Nucci *et al.*, 2008; Abreu *et al.*, 2012; Mengistu, Motoike and Cruz, 2016; Díaz *et al.*, 2021; Couto *et al.*, 2024). No entanto, destaca-se que a maioria dos estudos com marcadores moleculares em *Acrocomia* são avaliações sobre diversidade genética da espécie *A. aculeata* provenientes dos estados brasileiros, de São Paulo e Minas Gerais com o uso de marcadores microssatélites (Nucci *et al.*, 2008; Abreu *et al.*, 2012). Apenas dois trabalhos avaliam a diversidade da espécie utilizando sequenciamento de nova geração (NGS), focando-se principalmente na América do Sul (Díaz *et al.*, 2021; Couto *et al.*, 2024). Em todos os casos os autores reportaram altos níveis de diversidade, principalmente dentro de populações de *A. aculeata* (Mengistu, Motoike and Cruz, 2016; Neiva *et al.*, 2016; Araújo *et al.*, 2017; Cláudio *et al.*, 2017; Coelho *et al.*, 2018).

No entanto, inferir as relações filogenéticas e populacionais entre espécies ou populações recentemente divergentes continua a ser um dos principais desafios, sobretudo no tocante à escolha de marcadores moleculares com sinal filogenético suficiente para visualizar as relações em níveis taxonômicos mais baixos (Escudero *et al.*, 2014). Neste sentido as tecnologias de sequenciamento de alto rendimento revolucionaram as questões que podem ser tratadas e os táxons que podem ser estudados usando abordagens genômicas de nova geração (NGS). Grandes conjuntos de dados em escala genômica podem ser usados tanto para abordar as relações filogenéticas entre espécies muito próximas utilizando marcas de polimorfismos de nucleotídeo único (SNPs). Devido à dificuldade do uso de sequências do genoma integral, métodos de representação reduzida tal como a genotipagem por sequenciamento (GBS) têm sido utilizados com sucesso em estudos populacionais e filogenômicos, apresentando a vantagem de permitir seu uso em organismos sem dados de sequência disponíveis e por apresentar grande potencial para detecção de um maior número e com melhor qualidade de marcadores em todo o genoma (Elshire *et al.*, 2011).

As bibliotecas GBS, possuem três passos fundamentais: (1) fragmentação com enzimas de restrição, (2) ligação de adaptadores com barcodes específicos e (3) formação de um pool de amostras para ser enriquecidas por PCR (reação em cadeia da polimerase). Após a construção, a biblioteca genômica é sequenciada em plataformas NGS (Elshire *et al.*, 2011). Após a etapa de enriquecimento, áreas com baixo número de cópias são amostradas com sucesso; por outro lado, regiões repetitivas são evitadas por enzimas sensíveis à metilação. Com isso, a genotipagem se torna mais econômica e os dados são mais fáceis de analisar em espécies que não possuem genoma de referencia como *Acrocomia*.

Os métodos de análise de dados de GBS incluem identificação de barcodes, demultiplexamento de indivíduos e filtragem para eliminar sequencias de baixa qualidade. Depois disso, existem duas abordagens possíveis para a identificação dos SNPs: *de novo* ou com o genoma de referência. Pipelines baseados em uma referência, como TASSEL-GBS (Glaubitz *et al.*, 2014) e Stacks (Catchen *et al.*, 2013), mapeiam os *loci* e identificam SNPs usando a sequência do genoma de referência da espécie ou de outra espécie filogeneticamente próxima. Para as espécies que não possuem genoma referência como *Acrocomia*, os SNPs são identificados por meio do alinhamento das sequências contra elas mesmas. O pipeline de Stacks é exemplo para esta estratégia.

2.5 Estudos genômicos em *Acrocomia aculeata*

Acrocomia aculeata apresenta um genoma de aprox. 5.6 Gbp (2C = 5.81 pg e AT = 58.3%), pouco se sabe da organização estrutural e composição do genoma total (Abreu *et al.*, 2011). O seu plastoma é uma molécula circular de 155 kbp com uma estrutura quadripartida que inclui duas regiões de repetições invertidas (IRs) entre duas regiões de cópia simples, uma região de cópia única maior (LSC – "Large Single Copy") e uma região de cópia única menor (SSC – "Small Single Copy") (de Santana Lopes *et al.*, 2018). Como parte deste trabalho, foi publicado o plastoma de *Acrocomia totai* e *Acrocomia intumescens* (Francisconi and Morales-Marroquín *et al.*, 2023).

Nas Angiospermas o genoma do cloroplasto pode ser circular ou linear e vai dos 120 a 180 kbp com a estruturação quadripartida. Ele possui aproximadamente 100 genes codificadores de proteínas, genes tRNA e RNA ribossômico (Cosner, Raubeson and Jansen, 2004). O plastoma é altamente conservado, entretanto também tem sido observados rearranjos como inversões que tem ajudado a definir a diversificação e padrões evolutivos como em algumas espécies de *Passiflora* e *Hevea brasiliensis* (Euphorbiaceae) (Tangphatsornruang *et al.*, 2011; Cauz-Santos *et al.*, 2017). Regiões com duplicações, deleções e inclusive transferência de genes do genoma nuclear tem sido descritos. Essas regiões também podem ser usadas como marcadores moleculares (Martínez-Alberola *et al.*, 2013; Kim *et al.*, 2015). Por exemplo, nos plastomas das palmeiras *Astrocaryum aculeatum* e *A. murumuru* existe uma inversão característica que engloba um conjunto de genes envolvidos na clororrespiração e tradução dos plastídios (de Santana Lopes *et al.*, 2019). Neste sentido a análise total das sequências do cloroplasto tem permitido inferir relações filogenéticas entre espécies próximas e também utilizando polimorfismos de nucleotídeo único (SNPs). As diferencias entre as arquiteturas do cloroplasto de *Acrocomia aculeata*, *Acrocomia intumescens* e *Acrocomia totai* ajudaram a separar as espécies e demostrar suas diferencias filogenéticas (Francisconi and Morales-Marroquín *et al.*, 2023).

Em virtude do exposto e da importância econômica do gênero, estudos com *A. aculeata* têm se intensificado nos últimos anos, sendo prioridade para definir estratégias no melhoramento genético e conservação própria para cada região. Além disso, estudos genômicos e a comprovação de estruturação e alta diversidade intraespecífica pode ser uma alternativa para aumentar a diversidade genética, assim como identificar genótipos com características complementares e de interesse agronômico. Em *A. aculeata*, tem sido estudada a diversidade genética utilizando SNP, no entanto os estudos tem negligenciado a diversidade e estruturação entre os indivíduos das populações da América do Norte, principalmente as populações da América Central. Neste sentido, o trabalho foi conduzido utilizando-se quatro abordagens: (1) genômica de populações e estruturação da diversidade das três espécies principais sobre processos de seleção e adaptação, (2) detecção de genes por sinais putativos de seleção e adaptação utilizando SNPs para o melhoramento assistido com marcadores, (3) a descrição do grupo genético de *Acrocomia aculeata* para América Central, e (4) comparação estrutural e anotação do genoma do cloroplasto (plastoma) de *A. aculeata*, *A. intumescens* e *A. totai*.

Este trabalho permitiu entender melhor as diferencias entre a estrutura populacional como a nível interespecífico das três espécies mais importantes economicamente do gênero *Acrocomia*. É de suma importância estudar a diversidade genética de *Acrocomia* para criar incentivos e recursos genômico para apressurar o melhoramento assistido com marcadores. Este trabalho pretende ajudar a elucidar um pouco mais a história natural e filogenética do gênero e contribuir na compreensão das dinâmicas naturais que fizeram das florestas tropicais o bioma mais diverso do planeta.

3. Objetivos

3.1 Objetivo Geral

Desenvolver ferramentas moleculares que permitam ajudar a compreender a história natural e a diversidade genética do germoplasma da Macaúba (*Acrocomia* sp.) no Neotrópico para seu uso nos programas de melhoramento.

3.2 Objetivo Específico

Capítulo I

- Identificar marcadores SNPs, focados em *loci* sobre seleção descobertos por GBS em populações naturais em toda a distribuição de *Acrocomia aculeata*, *Acrocomia totai* e *Acrocomia intumescens.*
- Identificar a estrutura genômica utilizando marcadores sobre seleção das três espécies de *Acrocomia* em todo o continente Americano.
- Entender as implicações sobre o processo de adaptação e historia natural das três espécies de *Acrocomia* no Neotrópico.
- Identificar genes candidatos que estejam influenciados pelo processo de adaptação das três espécies de *Acrocomia* que possam ser importantes para o pre-melhoramento e domesticação das culturas.

Capitulo II

- Identificar a estrutura genômica do gene pool Norte Americano de *Acrocomia aculeata*, especificamente na América Central.
- Entender as possíveis influencias biogeográficas que modularam a diversidade e historia natural de *Acrocomia aculeata* na América Central.
- Modelar a distribuição potencial do nicho ecológico de *Acrocomia aculeata* na América Central.

4. Capítulo I

Genetic variations associated with adaptation processes in Acrocomia palms: A comparative study across the Neotropic for future crop improvement

Jonathan A. Morales-Marroquín[1]*, Brenda Gabriela Díaz-Hernández[2], Suelen Alves Vianna[5], Alessandro Alves-Pereira[3], Carlos Eduardo de Araújo Batista[2], Carlos A. Colombo[5], José Baldin Pinheiro[2], Maria Imaculada Zucchi[1,2,4]*

[1] Genetics and Molecular Biology Department, Biology Institute, University of Campinas (UNICAMP), Brazil

[2] Department of Genetics, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Brazil

[3] Plant Biology Department, Biology Institute, University of Campinas (UNICAMP), Brazil

[4] Unidade Regional de Pesquisa e Desenvolvimento (APTA), Piracicaba, São Paulo, Brazil

[5] Institute of Agronomy-IAC, Research Center of Plant Genetic Resources, Campinas, São Paulo, Brazil

*** Correspondence:**

Jonathan A. Morales-Marroquín
jmorales1089@gmail.com

Maria I. Zucchi
mizucchi@unicamp.br

Submitted to PLOS ONE, April 4th 2024.

Keywords: *Acrocomia aculeata*, *Acrocomia intumescens*, *Acrocomia totai*, Arecaceae, GBS, genome scans, population genomics, SNP

Abstract

Population genetic research has evolved, focusing on selection processes using single nucleotide polymorphisms (SNP) genotyping techniques to study crop traits and domestication. This study explores the adaptation process of three neotropical palms of *Acrocomia*, a genus that has high potential for oil extraction. Our research focuses on their genetic structure, evolutionary significance, and implications of the selection signatures for breeding efforts. We employed Genotyping-by-Sequencing (GBS) with a focus on outlier SNP markers for identifying adaptive genetic processes in *A. aculeata*, *A. totai*, and *A. intumescens* across the Neotropic. Our results reveal two major gene pools in *A. aculeata*, a North American and a South American group mainly influenced by dispersal and biogeographic barriers, with putative selective signatures linked to fatty acid and carotenoid biosynthesis, pathogen resistance, and environmental stress adaptation. *A. totai* presented a pronounced genetic structure, with SNPs under selection indicating a recent diversification driven by climatic and geological factors, especially in the Pantanal biome. *A. intumescens* displays genetic structuring shaped by the endemic process of biogeographic barriers within the Caatinga biome, with potential shared ancestry with *A. aculeata.* Correlations between allele frequencies and climatic variables highlight adaptation to diverse environments, principally semi-arid, with the annual mean temperature being one of the most influential. Candidate genes associated with fatty acid and carotenoid biosynthesis, as well as pathogen resistance and drought tolerance, indicate targets for domestication. Despite challenges in reduced representation sequencing, this study underlines the potential of *Acrocomia* as a novel crop, offering prospects in oil production, biofuels, and sustainable agriculture. Future efforts should prioritize whole-genome sequencing and genotype-environment interaction studies to realize the full potential of *Acrocomia* in sustainable development and renewable energy production.

Introduction

The Neotropical palm genus *Acrocomia* (Arecaceae: Arecoideae: Cocoseae: Bactridinae subtribe) holds significant agricultural potential. This genus comprises nine species with diverse distributions across tropical and subtropical America, and is characterized by its solitary, spiny trunk (stipe), persistent remnants of fallen leaf sheaths, and a yellow, coccoid drupe rich in oil (de Lima *et al.*, 2018). Within the group, *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. presents the widest distribution and greatest economic importance, also *A. totai* Mart. and *A. intumescens* Drude. show promising potential for oil production and bioeconomy (Vargas-

Carpintero *et al.*, 2021). The species are popularly known as Macauba, Coyol, Corozo, Macaw palm (*A. aculeata*), Bocaiuva (*A. totai*), and, Macaiba (*A. intumescens*) (Díaz *et al.*, 2021). It is worth mentioning that *A. aculeata* and the African oil palm (*Elaeis guineensis* Jacq.) exhibit analogous oil yields and fatty acid compositions (Colombo *et al.*, 2018).

Acrocomia aculeata, *A. totai*, and *A. intumescens* can be distinguishable by their distribution and distinct trunk (stipe) traits (Fig. 1 and 2). Wild populations of these palms have exhibited genetic and phenotypic variations linked to environmental adaptability (Ciconini *et al.*, 2013; Suelen A Vianna *et al.*, 2017; Resende *et al.*, 2020). *A. aculeata* is characterized by its spiky trunk and persistent leaf bases from fallen leaves. Its yellow fruits have a diameter ranging from 3.0 to 5.0 cm (Colombo *et al.*, 2018). The total yield obtained (dry basis) in the extraction of *A. aculeata* fruit pulp oil was 65% w/w from populations of the Brazilian state of Minas Gerais (Colombo *et al.*, 2018), and 53.6% from populations of Costa Rica (Alfaro-Solís *et al.*, 2020). Its distribution ranges from the subtropical to tropical regions of northern Mexico through Central America (Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, and Panama), Caribbean Islands, into Colombia, Venezuela, Guyana, Brazil, and northern Argentina (de Lima *et al.*, 2018). It is present in six of the seven proposed Neotropical palm biogeographic areas, predominantly in transitions between macro ecoregions like: tropical rainforests (moist broadleaf forest) to drier ecosystems like savannas, grasslands, dry broadleaf forests, and xeric shrublands (Fig. 2) (Eiserhardt *et al.*, 2013; Dinerstein *et al.*, 2017). It occurs in the following local biomes: Brazilian Cerrado, Atlantic coastal forest, Amazon, Central America moist forest, Caribbean dry forests, and Llanos (Lorenzi *et al.*, 2010; Freitas *et al.*, 2019).

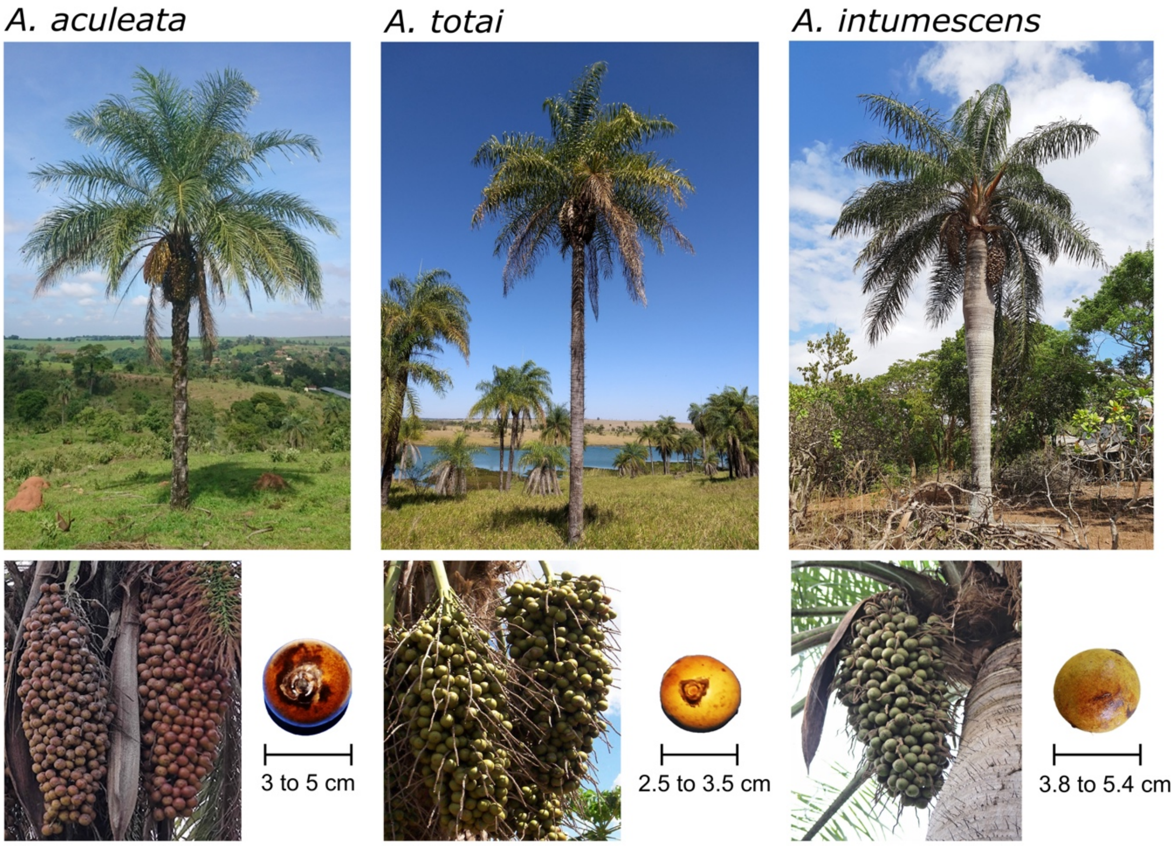


Fig 1. Adult plants of *Acrocomia aculeata*, *Acrocomia totai*, *Acrocomia intumescens*, and their fruits. Location: *A. aculeata* – Palm from Itapira, São Paulo, Brazil. Fruits from Minas Gerais, Brazil. *A. totai* – Palm from Presidente Epitácio, São Paulo, Brazil. Fruits from Corumbá, Mato Grosso do Sul, Brazil. *A. intumescens* – Palm and fruits from Vale da Neblina, Paraíba, Brazil. Photo credits: Brenda Díaz-Hernández (Palms: *A. aculeata* and *A. totai*), Suelen Alves Vianna (Fruits: *A. aculeata* and *A. totai*), and Eulampio Duarte (*A. intumescens* Palm and Fruits).

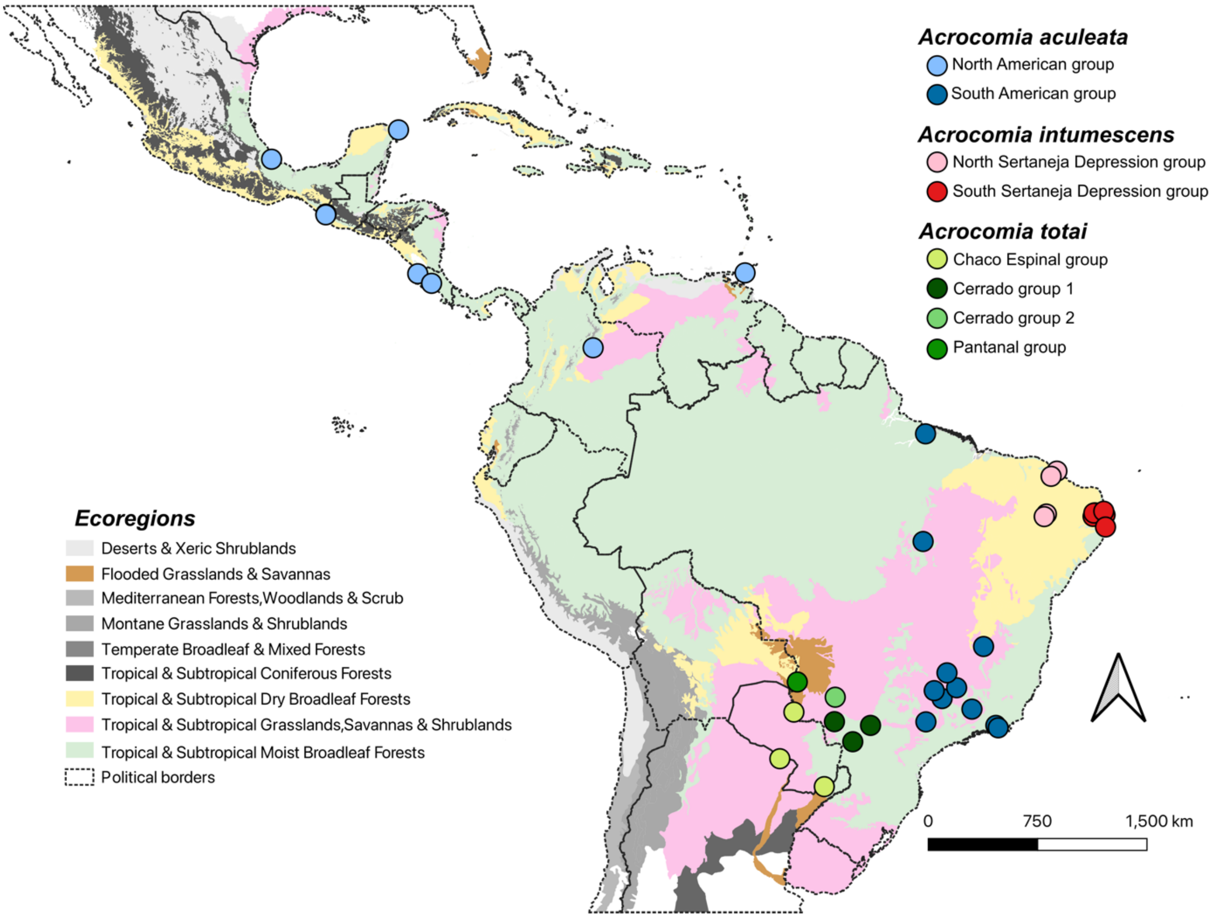


Fig 2. Map showing sampling locations of *Acrocomia* natural populations in the various macro ecoregions of the Neotropic. The map displays populations and major genetic groups within them of *Acrocomia aculeata* (blue dots), *A. totai* (green dots), and *A. intumescens* (red dots), as described in Table 2. The different macro ecoregions of the Neotropic are defined according to Dinerstein et al. (Dinerstein *et al.*, 2017).

A. totai generally features smooth trunk, occasionally with spikes, and rarely covered in persistent leaf bases. The diameter of its yellowish orange fruits ranges from 2.5 to 3.5 cm (Fig. 1) (Suelen Alves Vianna *et al.*, 2017). The total yield obtained (dry basis) in the extraction of *A. totai* pulp oil was 34.70% w/w from populations of the Brazilian state of Mato Grosso do Sul (Vianna *et al.*, 2021). It is distributed across a preeminent subtropical region in southern Brazil, northern Argentina, Bolivia, and Paraguay, occurring in the local biomes of Brazilian Pantanal, Chaco, and Espinal, as well as the Brazilian Cerrado transition biome. In macro-ecoregion classifications, it occurs in moist broadleaf forests, flooded grasslands, and savannas, as well as in drier ecosystems such as savannas, grasslands, and shrublands (Fig. 2) (Dinerstein *et al.*, 2017; Freitas *et al.*, 2019). *A. intumescens* produces yellow fruits with a diameter ranging from 3.8 to 5.4 cm. In its early life, the palm has a swollen trunk with spikes when young, which later becomes smooth in adulthood (Fig. 1) (de Lima *et al.*, 2018). The total yield obtained (dry basis) in the extraction of *A. intumescens* pulp oil was 26.90% w/w from populations of the Brazilian state of Pernambuco (Nascimento *et al.*, 2016). *A. intumescens* has a more restricted distribution, endemic to the arid and semi-arid tropical regions of the transition zones between the local biomes of Brazilian Caatinga and Cerrado, also in Atlantic coastal forest in the northeastern states of Pernambuco, Ceará, Paraíba, and Maranhão (Lorenzi *et al.*, 2010). In macro-ecoregion classifications, it occurs in dry broadleaf forest (Fig. 2) (Dinerstein *et al.*, 2017).

Acrocomia aculeata is increasingly recognized as a promising minor crop, offering a diverse array of emerging high-value commodities. These include products derived from its fruits, kernels, nut and leaves, as well as high-content oils suitable for various applications including biofuels, human and animal consumption (vegetable oil and flours), cosmetics, and pharmaceuticals. Furthermore, *A. aculeata* holds significant promise for sustainable biomass production, raising important considerations regarding land use, biodiversity conservation, carbon sequestration and storage, and its potential contributions to climate and environmental protection (Colombo *et al.*, 2018; Vargas-Carpintero *et al.*, 2021). Biofuels derived from *Acrocomia* could facilitate the transition from fossil fuels to more sustainable energy options in both developed and developing countries. The *Acrocomia* ability to thrive in marginal areas with poor, arid soils, such as rangeland, bare ground, degraded cropland, or grassland, aligns with multiple United Nations Sustainable Development Goals (SDGs): (1) SDG 7 (Affordable and Clean Energy) by providing a renewable fuel source. (2) SDG 13 (Climate Action) by potentially mitigating carbon emissions and promoting climate-resilient agriculture. (3) SDG 15 (Life on Land) by restoring degraded land and potentially mitigating biodiversity loss.

These three species easily occur in anthropized areas, and their dispersion is favored by human interaction (Colombo *et al.*, 2018). There is evidence that *Acrocomia* was used by Pre-Columbian civilizations as a ritual and edible plant (Clement, 1999a, 1999b; Morcote-Ríos and Bernal, 2001). The oldest archaeological record of its utilization in North America is dated from the Pre-Olmec and Olmec civilizations in the southern Mexican state of Veracruz (4000 BP) (Ramírez-Núñez *et al.*, 2019). The oldest human interaction in South America is with the Paleo-Indians and other proto-Tupi-Guarani civilization in the Brazilian lowlands dated 10,030 BP. The oldest known dated fossil was described in the northern Brazilian state of Pará (Santarén - 11,200 BP), followed by Argentina (8500 BP), Panama (8040 BP), and Mexico

(6750 BP) (Rodríguez and Aschero, 2005; de Lima *et al.*, 2018). *Acrocomia aculeata* is considered an incipiently domesticated crop and probably share the same evolutionary pressures as other Neotropical domesticated crops for new ecological niche of cultivation and adaptation (Clement, 1999b). Despite the growing economic interest in some *Acrocomia* species, understanding of their genomic information remains limited, especially adaptation strategies and gene flow relationships.

The development of new crop varieties with improved traits such as disease resistance, drought tolerance, and high oil yield can be achieved through the identification of local varieties (ecotypes), genetic markers, and quantitative trait loci (QTL) mapping (Schreiber, Stein and Mascher, 2018; Huang *et al.*, 2022). Understanding the genetic diversity and QTLs of *Acrocomia* palms is essential for identifying desirable characteristics and selecting suitable breeding materials. The ability of crop species to adapt and maintain resilience in response to changing environmental conditions is significantly influenced by genetic diversity (Petereit *et al.*, 2022). Using loci under selection, breeders can identify and characterize candidate genes responsible for key adaptive traits (Alves-Pereira *et al.*, 2022). Furthermore, research can provide insights on the mechanisms underlying palm diversification in the Neotropic by identifying important evolutionary processes such speciation, adaptation, and dispersal.

This study assessed selective signatures and genome-wide diversity in natural populations of *Acrocomia aculeata*, *Acrocomia totai*, and *Acrocomia intumescens* from different Neotropic local biomes. We used double-digest genotyping by sequencing (GBS) to identify single nucleotide polymorphisms (SNPs) for putative signatures of selection. Our analysis explores the potential role of these SNPs in adaptation, diversification processes, and genetic population structures, particularly in relation to cultivation in diverse environmental contexts, with a focus on important crop traits. This research aims to generate novel insights into the evolution of *Acrocomia* across the Americas and contribute to the domestication and breeding of this promising tropical palm for sustainable vegetable oil production, bioeconomy, and biofuels.

Results

SNP discovery in *Acrocomia* sequencing

Sequencing of the two ddGBS libraries for *A. aculeata* and *A. totai* generated a total of 219,264,253 reads, while the two *A. intumescens* ddGBS libraries generated a total of

554,630,990 reads. After quality-control filtering, the number of retained reads were: 60,786,924 (mean = 779,319.5 reads per sample, SD ± 449,982) for *A. aculeata*; 29,283,574 (mean = 697,227.9 reads per sample, SD ± 355,813) for *A. totai*; and 276,741,447 (mean = 1,990,945.6 reads per sample, SD ± 1,333,058) for *A. intumescens*. After the analysis in Stacks, the final data set had 1997 SNPs for the 78 *A. aculeata* samples (mean depth per locus = 18.7X, SD ± 6.2, 8.8% of missing data). For the 40 *A. totai* samples, 1629 SNPs were identified (mean depth per locus = 17.2X, SD ± 6.4, 2.9% of missing data). For the 131 *A. intumescens* samples a total of 3466 SNPs were identified (mean depth per locus = 11.5X, SD ± 4.4, 13.4% of missing data).

Putative signatures of selection and population structuration in *Acrocomia*

Because pcadapt and LFMM are methods that account for the global genetic structure of the data, some interesting patterns may be considered for the analyses of putative signatures of selection. Both analyses suggested that the genetic structure within *A. aculeata* and *A. intumescens* is not explain by the local biome groups assumed according to the classification proposed by Freitas *et al.*, (Freitas *et al.*, 2019) but rather the outcome of genomic isolation of biogeographic barriers and historical events. In the case of *A. totai*, a genetic structuring was observed in accordance with local biomes and macro ecoregions (Fig. 3, S1 and S3) suggesting local ecosystem adaptation. Similar DAPC results were obtained, indicating a significant degree of divergence between natural populations from specific biome. The major genetic structure in *A. aculeata* is observed among sampling locations (or biogeographic groups) within and outside Brazilian lowlands (<1500 masl), resulting in the separation of the samples into two clusters along the first principal component (43.88% of the variance), with a distinct divergence between samples from Mexico and Central America and those from Brazil (Fig. 3 and S1). In *A. totai*, samples from Pantanal (PAN) were the most divergent, the Argentinean-Paraguayan Chaco and Espinal (CHE) samples tend to cluster in the same genetic group, while samples from the Brazilian Cerrado (CER) were more admixed than those from the other groups (Fig. 3 and S1). The major genetic divergence in *A. intumescens* was observed between the Brazilian state of Ceará, indicating interrupted gene flow likely due to genetic isolation or a biogeographic barrier, rather than between biomes (Fig. 3 and S1).

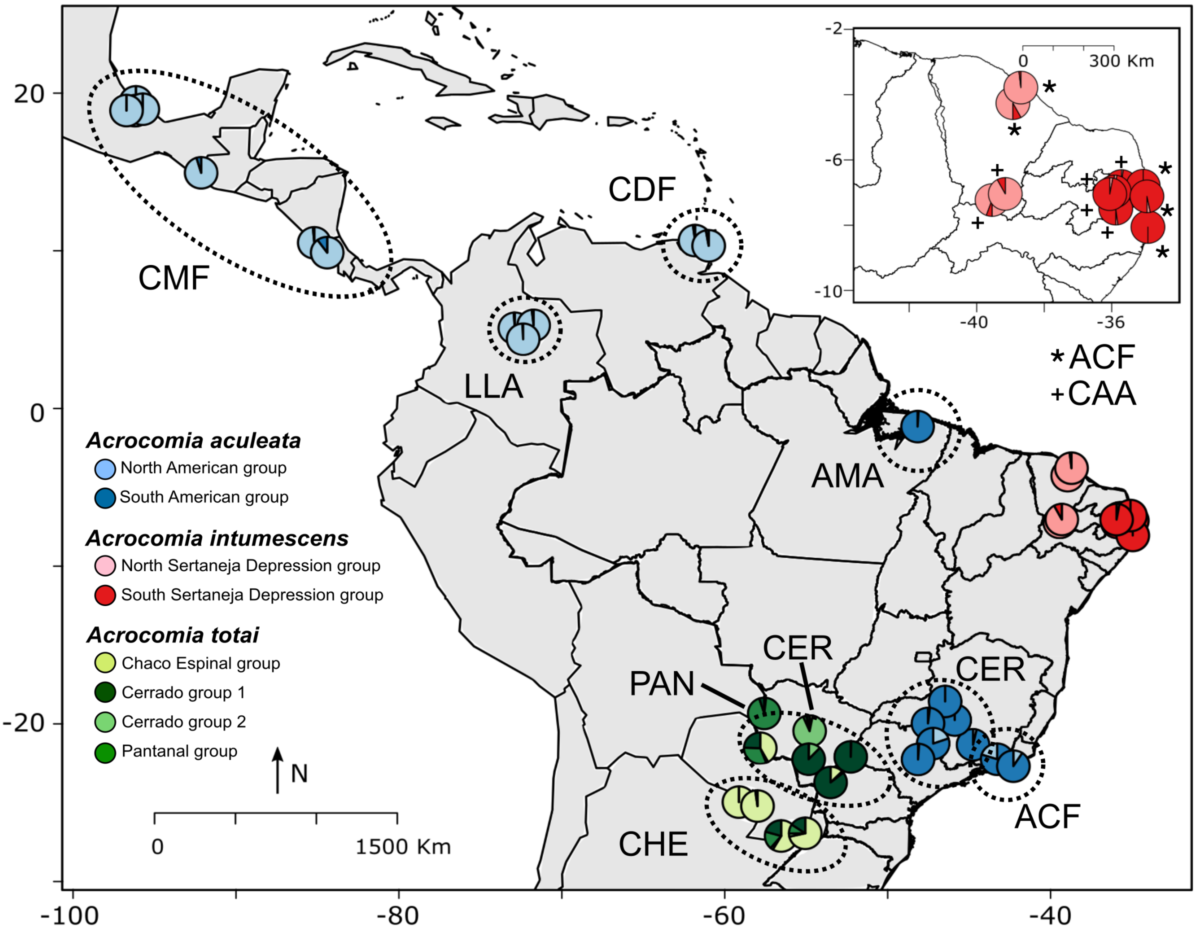


Fig 3. Map of sampling locations for *Acrocomia aculeata* (blue), *A. totai* (green), and *A. intumescens* (red), along with major genetic divergences among local biome groups as suggested by sparse non-negative matrix factorization (sNMF) analyses. Pie charts show the average sNMF ancestry coefficients across the genetic clusters represented by different color shades. Some pie charts were slightly moved to ease visualization. Acronyms follow Table 2 local biomes: CMF Central American Moist Forest; LLA Llanos; CDF Caribbean Dry Forest; ACF Atlantic Coastal Forest; CAA Caatinga; AMA Amazon; CER Cerrado; PAN Pantanal; CHE Chaco and Espinal.

The numbers of outlier SNPs detected for *A. aculeata* were 500 based on pcadapt and 129 based on FstHet, while LFMM indicated 525 loci associated with at least one bioclimatic variable. Of these, 326 markers were considered as putative under selection because they were indicated by at least two of these methods (Fig. 4A). A total of 67 *A. aculeata* sequences with an outlier marker had at least one blastx hit, and 54 had at least one GO annotation. These GO terms could be summarized in 33 different classes (Fig. 5A). For *A. totai*, pcadapt indicated 218 outlier SNPs, FstHet 306, and 446 SNPs were associated to a bioclimatic variable in LFMM. Of these, 192 SNPs were considered as putative under selection (Fig. 4B). A total of 38 *A. totai* sequences with an outlier marker had at least one blastx hit, and 33 had at least one GO annotation. These GO terms could be summarized in 28 different classes (Fig. 5B). For *A. intumescens*, pcadapt indicated 688 outlier SNPs, FstHet 157, and 1024 markers were associated with a bioclimatic variable in LFMM. Of these, 104 SNPs were considered as markers putatively under selection (Fig. 4C). Only 8 *A. intumescens* sequences with an outlier marker had at least one blastx hit, and 2 had GO annotations. These GO terms could be summarized in 7 different classes (Fig. 5C). For the three species, the most common GO annotations were related with the biological processes of "metabolic process" and "cellular process", and the molecular functions of "binding" and "catalytic activity". GO annotations related to "response to stimulus" and "regulation" were also frequent for the proteins with similarities to the sequences with outlier SNPs of *A. aculeata* and *A. totai*.

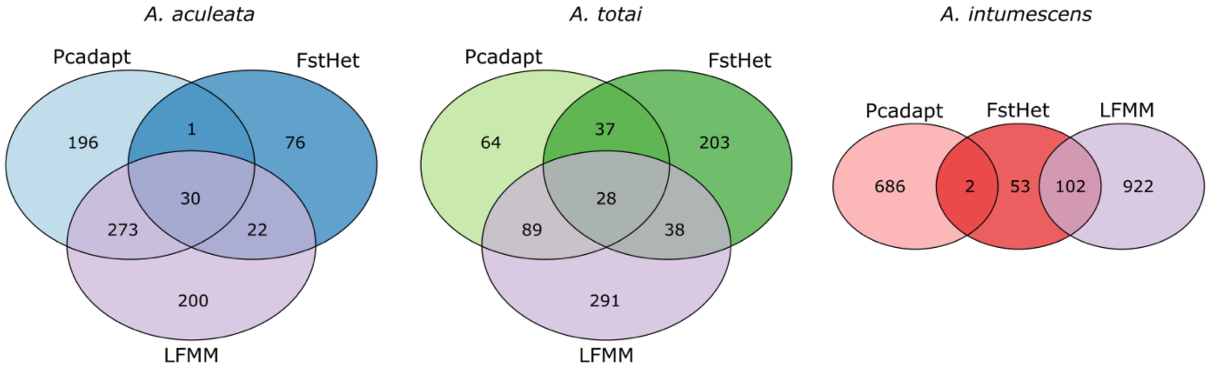


Fig 4. Venn diagrams showing the number of loci identified as outliers with different methods. The outlier SNPs detected in Pcadapt are based on the genetic structure of the PCA, while in FstHet, they are based on the betahat statistic. In LFMM, the outliers are associated with bioclimatic variables. A) *Acrocomia aculeata*, B) *A. totai*, and C) *A.intumescens*.

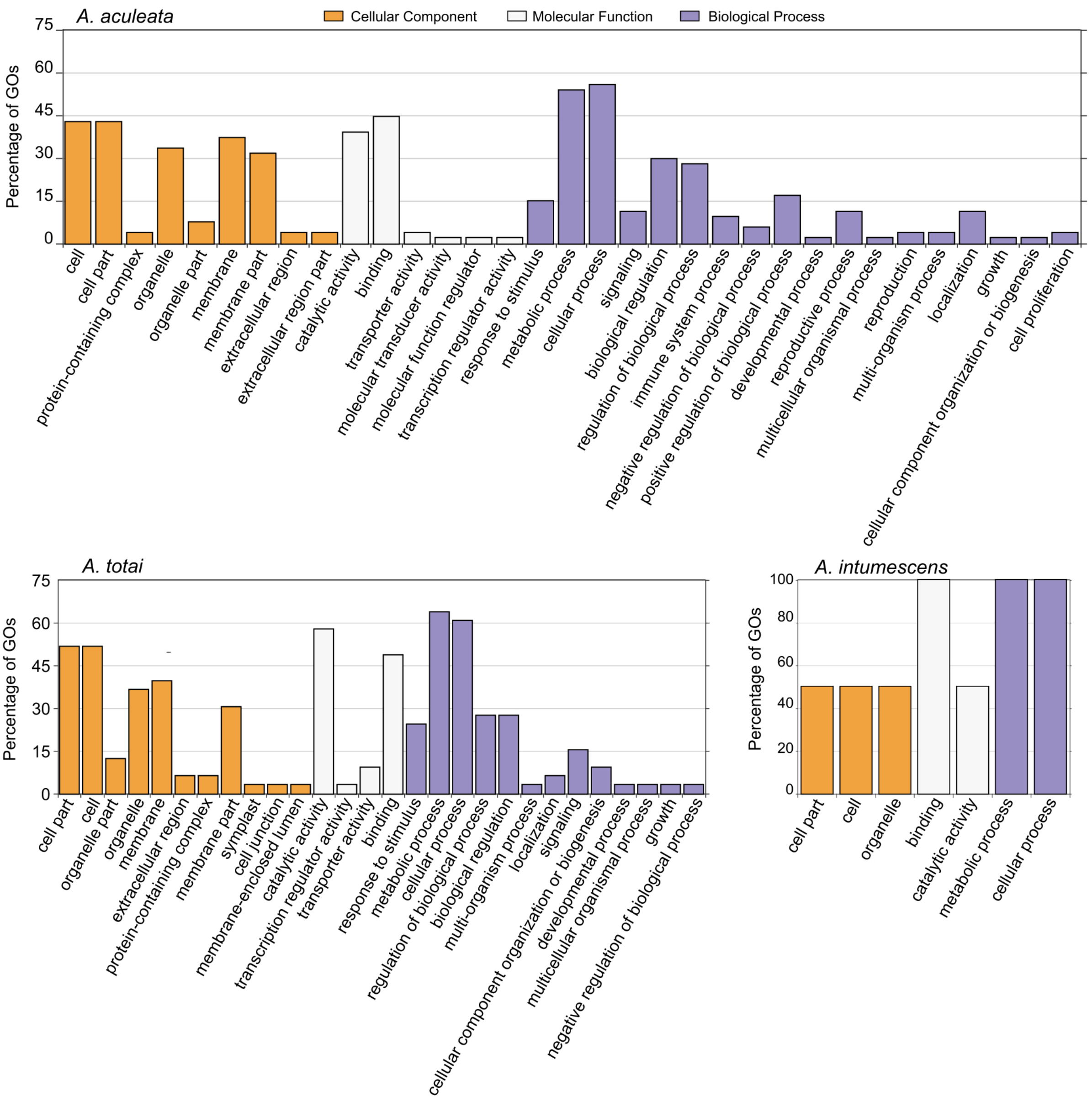


Fig 5. Summary of Gene Ontology (GO) annotations associated with proteins. The outlier SNPs in gene sequences associated with proteins are deposited in GenBank. GO terms summarize annotations according to cellular components, molecular functions, and biological processes. A) *Acrocomia aculeata*, B) *A. totai*, and C) *A.intumescens*.

These genes exhibited high similarity to genes annotated and described in palm species such as *Elaeis guineensis*, *Phoenix dactylifera* L., and *Cocos nucifera* L., all of which have fully sequenced genomes (Table 1, S2). The genes observed were related to many functions, including oil production metabolism/catabolism, carotenoid biosynthesis, plant growth and development, organ sizes, root, flowering, pathogen resistance, and biotic and abiotic stresses. The most relevant loci identified as outliers in different methods and for each species are listed in Table 1 (the whole list is in Table S1), while a summary of blastx results and their associated GO annotations are in Table S2.

Table 1. Most relevant gene loci with putative selective signatures in *Acrocomia aculeata*, *A. totai,* and *A. intumescens*. The complete list of gene loci and their description can be found in Table S2.

Acrocomia aculeata						
Sequence name	**Sequence desc.**	**Species**	**Function**	**E-Value**	**Similarity**	**Reference**
CLocus_16512	Carotenoid cleavage dioxygenase 7 (chloroplastic)	*Elaeis guineensis*	Carotene catabolic process, Strigolactone biosynthesis that modulates plant growth, reproduction, senescence	4.87E-09	89.7	(Booker *et al.*, 2004; Mishra, Upadhyay and Shukla, 2017; Saeed, Naseem and Ali, 2017; Yao *et al.*, 2022)
CLocus_27100	Putative WRKY transcription factor 19	*Elaeis guineensis*	Modifies the content of fatty acids and influences the accumulation of seed oil; Abiotic stress responses, and plant pathogen interactions	1.96E-10	100.0	(Cheng *et al.*, 2016; Khoso *et al.*, 2022)
CLocus_25725	Lipid binding protein	*Elaeis guineensis*	Lipid biosynthesis Lipid transport; Lipid binding	2.37E-09	100.0	(Guo *et al.*, 2022)
CLocus_317	Polyprenol reductase 1 isoform X1	*Elaeis guineensis*	Abiotic stress responses, and plant pathogen interactions, lipase activity	9.80E-08	96.6	(Habsyah *et al.*, 2021; Van Gelder *et al.*, 2021)
CLocus_11291	Patatin-like protein 2	*Cocos nucifera*	Lipid Metabolism (phospholipase activity; lipid catabolic process; acylglycerol lipase activity); Resistance to pathogens	3.83E-11	96.7	(La Camera *et al.*, 2009; Ting *et al.*, 2016)
CLocus_11820	carotenoid cleavage dioxygenase 7, chloroplastic	Phoenix dactylifera	Carotene catabolic process, modulates plant growth, reproduction, senescence	4.37E-09	90.0	(Booker *et al.*, 2004; Liu *et al.*, 2013; Yao *et al.*, 2022)
CLocus_24655	pentatricopeptide repeat-containing protein At5g15340, mitochondrial	Elaeis guineensis	Zinc ion binding. Influence organellar expression function and, consequently, on photosynthesis, respiration, plant development, and environmental responses	9.48E-10	100.0	(Barkan and Small, 2014)
CLocus_12388	pentatricopeptide repeat-containing protein, mitochondrial-like	Cocos nucifera	Zinc ion binding. Influence organellar expression function and, environmental responses	1.20E-06	86.7	(Barkan and Small, 2014)
CLocus_16644	Polyprenol reductase 2	*Phoenix dactylifera*	Essential in dolichol biosynthesis	7.80E-10	100.0	(Habsyah *et al.*, 2021; Van Gelder *et al.*, 2021)

CLocus_17318	Carotenoid cleavage dioxygenase 8 homolog B (chloroplastic)	*Morus notabilis*	Apocarotenoids, carotenoid cleavage dioxygenases (CCDs), inhibit shoot branching. Also active on other carotenoid substrates like lycopene or zeaxanthin	1.92E-11	100.0	(Sorefan *et al.*, 2003; Schwartz, Qin and Loewen, 2004)
CLocus_18527	Protein OBERON 3	*Phoenix dactylifera*	Embryonic root meristem initiation, rooting	2.16E-10	100.0	(Saiga *et al.*, 2012; Lin *et al.*, 2016)
CLocus_21286	GRF-interacting factor 1	*Elaeis guineensis*	Leaf development and growth, reproductive system development and growth	3.67E-09	96.6	(Lee *et al.*, 2014; Lu *et al.*, 2020)
CLocus_22454	Momilactone A synthase	*Cocos nucifera*	Momilactone A biosynthesis, diterpenoid secondary metabolites, involved in the defense mechanism of the plant produced in response to attacks	9.13E-12	100.0	(Shimura *et al.*, 2007; Zhang, Zhang and Peters, 2020)
CLocus_23382	Phosphoinositide phospholipase C 2	*Elaeis guineensis*	Signal transduction in events such as guard cell signaling, salt stress, osmotic stress, acquired resistance, Nod factor signaling, drought, systemic acquired resistance, carbon fixation in C4 plants, and response to the pathogen	4.27E-12	100.0	(Rupwate and Rajasekharan, 2012)
CLocus_29621	BTB/POZ domain and ankyrin repeat-containing protein NPR2	*Cocos nucifera*	Plant defense and plant disease resistance	4.59E-10	100.0	(Backer, Naidoo and van den Berg, 2019)
CLocus_30624	Subtilisin-like protease SBT1.5	*Elaeis guineensis*	Plant–pathogen recognition and immune priming	5.86E-09	96.6	(Figueiredo, Monteiro and Sebastiana, 2014)
CLocus_35203	BTB/POZ domain and ankyrin repeat-containing protein NPR2	*Elaeis guineensis*	Plant defense and plant disease resistance	3.68E-10	100.0	(Backer, Naidoo and van den Berg, 2019)
CLocus_21703	FRIGIDA-like protein 4a	*Elaeis guineensis*	Regulates flowering induction, flowering time	2.22E-04	100.0	(Zhang and Jim enez-G omez, 2020)
CLocus_25856	Putative pentatricopeptide repeat-containing protein At5g37570	*Elaeis guineensis*	Organelle biogenesis and function, photosynthesis, respiration, plant development, and environmental responses	1.15E-10	96.6	(Barkan and Small, 2014)

CLocus_32860	ABC transporter D family member 1	*Capsicum annuum*	Lipid metabolism (long-chain fatty acid transporter activity, Long-chain fatty acid import into peroxisome; fatty-acyl-CoA transport), Organ growth, plant nutrition, plant development, response to abiotic stress, and the interaction of the plant with its environment	2.37E-04	100.0	(Kang *et al.*, 2011; Ting *et al.*, 2016)
CLocus_82942	Endoglucanase 8-like	*Elaeis guineensis*	Synthesis, remodeling and turnover of cell wall components in symbiotic and hostile plant–microbe interactions	1.09E-06	100.0	(Perrot, Pauly and Ramírez, 2022)

Acrocomia totai

Sequence name	**Sequence desc.**	**Species**	**Function**	**E-Value**	**Similarity**	**Reference**
CLocus_17145	GDSL esterase/lipase	*Elaeis guineensis*	Lipid metabolism: produce and storage oil in seeds, cuticular lipids to cover and decorate organ surfaces, oxylipins, and other signaling molecules	7.53E-11	100.0	(Shen *et al.*, 2022; Ding *et al.*, 2023)
CLocus_37219	Subtilisin-like protease SBT1.2	*Elaeis guineensis*	Plant–pathogen recognition and immune priming	3.18E-11	100.0	(Figueiredo, Monteiro and Sebastiana, 2014)
CLocus_15942	Low affinity sulfate transporter 3	*Cocos nucifera*	Plant response to drought and salinity stress	7.41E-10	100.0	(Gallardo *et al.*, 2014)
CLocus_16267	Mitogen-activated protein kinase kinase kinase NPK1	*Elaeis guineensis*	Regulates innate immunity and development in plants	1.33E-08	93.1	(Jin *et al.*, 2002)
CLocus_20717	Fatty acid desaturase 4, chloroplastic	*Cocos nucifera*	Influence levels of susceptibility to multiple stresses, including insect infestations	2.21E-09	100.0	(Li *et al.*, 2021)
CLocus_21866	Phosphatidylinositol 4-kinase alpha 1 isoform X2	*Elaeis guineensis*	Essential for pollen, embryonic, and post-embryonic development also cell signaling	5.17E-10	100.0	(Noack *et al.*, 2022)
CLocus_24114	9-cis-epoxycarotenoid dioxygenase NCED6, chloroplastic	*Cocos nucifera*	Increases drought tolerance, multi-abiotic stress tolerance, regulates plant growth, also enhances seed dormancy	3.42E-13	96.6	(Martínez-Andújar *et al.*, 2011; He *et al.*, 2018; Huang *et al.*, 2018)
CLocus_26989	Momilactone A synthase	*Cocos nucifera*	Drought, salinity, and oxidative stress conditions tolerance	1.58E-11	100.0	(Xuan *et al.*, 2016; Kato-Noguchi, 2023)
CLocus_29376	Subtilisin-like protease SBT1.2	*Elaeis guineensis*	Plant–pathogen recognition and immune priming	2.36E-10	93.3	(Figueiredo, Monteiro and Sebastiana, 2014)

CLocus_37352	Protein OBERON 3	*Phoenix dactylifera*	Embryonic root meristem initiation, rooting	2.21E-10	100.0	(Saiga *et al.*, 2012; Lin *et al.*, 2016)
CLocus_54281	Phospholipase A2-alpha	*Elaeis guineensis*	Signaling roles during plant abiotic and biotic stress responses	2.52E-09	96.6	(Takáč, Novák and Šamaj, 2019)
CLocus_12178	Protein ENHANCED DISEASE RESISTANCE 4	*Cocos nucifera*	Regulation of defense response to fungus	1.61E-05	93.3	(Wu *et al.*, 2015)

Acrocomia intumescens						
Sequence name	**Sequence desc.**	**Species**	**Function**	**E-Value**	**Similarity**	**Reference**
CLocus_111127	Phosphatase 2C BIPP2C1	*Elaeis guineensis*	Plant signal transduction processes and stress signaling	3.23E-09	100.0	(Fuchs *et al.*, 2013)

Discussion

The focus of population genetic research has rapidly shifted from spatially neutral genetic processes to adaptive genetic processes, driven by advancements in SNP genotyping techniques for studying crop traits and domestication processes. Considerable deviation of outlier markers from the distribution of a specific statistical evaluation under a particular model can serve as the basis for methods aimed at detecting selective signatures for understanding adaptation events (Lotterhos and Whitlock, 2015; Alves-Pereira *et al.*, 2022). A significant future development in sustainable agriculture will involve the domestication of novel native plant species, such as *Acrocomia*, to facilitate the transition towards cleaner energy sources. Below, we discuss the most interesting findings regarding outlier SNPs in the adaptation process of *Acrocomia*, their biological and evolutionary significance, and their implications for breeding efforts.

The adaptation and evolutionary process in *Acrocomia*: Dispersal and Biogeographic Barriers

Acrocomia phenotype has been shaped by ecological interactions that have influenced the evolution of specific traits for its survival (Nascimento *et al.*, 2020). *Acrocomia* species are distributed across tropical America, extending from North America (Northern Mexico) to South America, inhabiting areas transitioning from tropical and subtropical rainforests (moist broadleaf forests) to savannah and xeric shrubland regions, including dry broadleaf forests (semi-arid and arid ecosystems) (Fig. 2). Their wide distribution suggests their ability to adapt to various soil and climate conditions. sNMF analyses can help identify genetic signatures associated with adaptation to specific environmental conditions or phenotypic traits. Our findings provide insights into the genetic structure and the association of specific SNP loci under selection forces in the adaptation process of *Acrocomia aculeata*, *A. totai*, and *A. intumescens* (Fig. 4, Table 1).

Our results support the division of *A. aculeata* into two major gene pools: a North American group and a South American group, consistent with previous observations based only on neutral loci (Díaz *et al.*, 2021) (Fig. 2 and 3). Notably, we found no significant differences in genetic structure within minor local biomes within each gene pool, indicating that the Amazon basin likely served as a source region for diversification across the species distribution. This pattern in genetic structure is also present in other crop species such as common beans and maize (Bitocchi *et al.*, 2012; Bedoya *et al.*, 2017). Both dispersal and biogeographic barriers likely influenced the genomic structure of *Acrocomia aculeata*. Also, *A. aculeata* exhibited the greatest number of putative selective signatures (Table 1; S1 and S2) in genes associated with fatty acid and triacylglycerol biosynthesis, carotenoid biosynthesis, pathogen resistance and defense, as well as genes specialized in adapting to abiotic and environmental stress. From the nine species of the genus, *A. aculeata* exhibits the highest fruit pulp oil content (65% w/w dry basis) (Colombo *et al.*, 2018).

Individual plant fitness is influenced by animal–plant interactions, encompassing mutualisms such as seed dispersal and antagonistic relationships like herbivory and seed predation (Nascimento *et al.*, 2020). For example, the morphology of *Acrocomia aculeata* fruit, including its size, color, and oil content, appears to have undergone correlated evolution as predicted by the seed-dispersal syndrome hypothesis before the beginning of its domestication process through very recently human interaction (~10,000 BP) (Rodríguez and Aschero, 2005; de Lima *et al.*, 2018). Additionally, spines and other mechanical defense features are common mechanisms employed by plants to avoid herbivory. Numerous studies on palms indicate that spinescence reduces herbivory by large mammals (Göldel *et al.*, 2016). The same herbivory and seed dispersal pattern is present in A. *totai* and *A. intumescens*.

Frugivorous mammals have a preference for consuming dull-colored fruits, such as those that are green, yellow, brown, or pale orange (Janzen and Martin, 1982; Nascimento *et al.*, 2020) (Fig. 1). It is hypothesized that the coccoid type fruit (drupe) of *Acrocomia* ancestor species may have been consumed by large herbivores in the past, with migration events like the Great American Biotic Interchange (GABI; 2.5 million BP) likely influencing its distribution through seed dispersal (Janzen and Martin, 1982). Consequently, certain observable fruit traits and defenses against herbivory, such as spiny trunks and hard epicarp, could be considered anachronic, representing interactions that occurred in the past (Janzen and Martin, 1982). The term "anachronic" denotes traits or features that are outdated or no longer relevant in the current context but may have been adaptive or functional in the past (Janzen and Martin, 1982; Nascimento *et al.*, 2020). While most large mammals are now extinct in the Neotropics following the Late Quaternary Extinction episode, Neotropical palms like *Acrocomia* diversified in an ecological environment where mammalian assemblages were abundant in large-bodied species, approximately from 15 million to 23,000 BP (William J. Baker and Couvreur, 2013; Villavicencio *et al.*, 2016; Nascimento *et al.*, 2020). However, our findings suggest only a certain level of connectivity between populations within the two major gene pools. Currently, cattle and humans are regarded as the primary contributors to seed dispersal for this species, highlighting its recent close association with human activity (Lentz, 1990; Göldel *et al.*, 2016).

Archaeological evidence encompassing Pre-Columbian civilizations across the neotropics, from Mesoamerica (northern Mexico, Central America) to South America, suggests the cultivation and ritualistic utilization of *Acrocomia aculeata* (Rodríguez and Aschero, 2005; Ramírez-Núñez *et al.*, 2019). The presence of *Acrocomia* in these regions indicates a potential influence of Amerindian migration on species diversification and offers insights into its domestication process. Such historical records reinforce the relationship between indigenous populations and *Acrocomia*, potentially shaping its genetic diversity and evolutionary trajectory through intentional cultivation and cultural practices. Further investigation is necessary to elucidate the specific historical events and ecological factors shaping the observed genetic structure and center of origin.

Acrocomia totai exhibits a pronounced genetic structuring, showing a clear differentiation between populations from the Brazilian Cerrado, Pantanal, and Chaco and Espinal (Fig. 2 and 3). This contrasts with previous studies focused only on neutral loci using SNPs and microsatellites where they observed less genetic structuring among populations (de Lima, Meerow and Manfrin, 2020; Díaz *et al.*, 2021). This indicates that the populations of *A. totai* are experiencing a recent diversification and ongoing speciation processes. *A. totai* has a subtropical distribution and is adapted to temperate conditions, with selective signatures found in genes associated with fatty acid and triacylglycerol biosynthesis, pathogen resistance and defense, as well as genes specialized in adapting to abiotic and environmental stress (Table 1;

S1 and S2). Pantanal populations exhibit lower levels of admixture compared to other genomic groups within *A. totai*, particularly those distributed in the Brazilian Cerrado. The Cerrado lowlands and ecotones with other local biomes serve as a bridge, including for the South American gene pool of *A. aculeata*. Hybridization between these two species has been reported (Díaz *et al.*, 2021). Phenotypic differences were also observed in other studies, compared to fruits from the Pantanal, fruits from Cerrado had a higher proportion of epicarp and a lower proportion of kernels (Ciconini *et al.*, 2013). The Pantanal biome, an active sedimentary basin characterized by faults and earthquakes, experiences subsidence and depressions that are highly susceptible to flooding. During the late Pleistocene (14,000–10,000 BP), arid conditions predominated, possibly allowing *A. totai* to recently colonize the biome (Assine *et al.*, 2016). The Pantanal lacks endemic tree species, with the majority of terrestrial species being immigrants from the Cerrado (Junk *et al.*, 2006). Similar biogeographic patterns are observed in the palm *Copernicia alba* Morong. (Ferreira Costa *et al.*, 2023). The domestication process of *A. totai* remains poorly understood, although archaeological records indicate its use for fiber production in Argentina (Rodríguez and Aschero, 2005). Human utilization varies among *Acrocomia* species, with *A. totai* being utilized for the yellowish pulp of the fruit for flour production, and the oil derived from the seed, which are the two products of greatest interest (Sanjinez-Argandoña and Chuba, 2011; Ciconini *et al.*, 2013; Vianna *et al.*, 2021).

A. intumescens presented a genetic structure influenced by an allopatric isolation across various phytophysiognomies within Caatinga, rather than by variations in biomes (Fig. 2 and 3) (da Silva *et al.*, 2018; Uchôa *et al.*, 2022). Specifically, notable genetic structuring was observed between populations in the North Sertaneja depression and those transitioning from the Borborema highlands to the South Sertaneja depression within the Caatinga biome. The Caatinga exhibits high levels of endemism, with species adapted to survive in its arid climate characterized by drought conditions. The long-term stability of the Caatinga, along with the assembly of ancient plants, has been significantly influenced by aridification processes, while recent vegetation shifts and climate change have driven in situ diversifications. The increased environmental variability has led to the appearance of modern species through Pleistocene/Pliocene (2.6 million BP) ecological specialization (Fernandes *et al.*, 2022). This could elucidate the phylogenomic proximity observed between *A. aculeata* and *A. intumescens*, suggesting a shared common ancestor in the past and probably, *Acrocomia* recently colonized and diversified in the Caatinga (Freitas *et al.*, 2021). This shared ancestry may be attributed to the dominance of tropical rainforests in the region, which connected the area to the Amazon during the Early Cenozoic period (Díaz *et al.*, 2021; Francisconi *et al.*, 2023). This pattern of endemism and genetic differentiation is shared with other plant species in the Caatinga, including other palms like *Copernicia prunifera* (Miller) H.E. Moore, legumes (*Coursetia*, *Vatairea*, and *Luetzelburgia* genera), as well as other clades such as *Conopophaga cearae* (Aves. common name: Caatinga gnateater) and lizard species (Costa *et al.*, 2022; Fernandes *et al.*, 2022; Uchôa *et al.*, 2022; Batalha-Filho *et al.*, 2023). In *A. intumescens*, the only selective signature found in a gene was associated with stress signaling.

The close relationships observed among *Acrocomia* species may be attributed to the radiation of the Bactridinae subtribe towards the middle of the Miocene (23 million BP), when the ancestor of *Acrocomia* diversified (William J Baker and Couvreur, 2013). This period of divergence coincided with the terminal Eocene cooling event, characterized by substantial climatic changes that led to accelerated turnover in flora and fauna (Hoorn *et al.*, 2010). This pattern is also evident in the conserved genomic structures and similarity of *Acrocomia aculeata*, *A. totai* and, *A. intumescens* in their plastomes architecture and nuclear gene phylogeny, low pairwise Fst indicating a close relationships (Díaz *et al.*, 2021; Francisconi *et al.*, 2023). The high admixture between populations and the short branch lengths observed in phylogenomic studies, which separate the three species, presents challenges in establishing unambiguous intergeneric relationships within the genus *Acrocomia*. This may suggest a large-scale species-level extinction followed by rapid diversification of surviving lineages.

Correlations between allele frequencies and climatic variables in the LFMM analysis revealed five climatic variables influencing adaptation in the three species: annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), Annual Precipitation (BIO12), and precipitation of the driest month (BIO14), indicating selection as a driving force of evolutionary adaptation (Fig. S2). Similar patterns were observed in photosynthesis rates and fruit pulp mass development in other *Acrocomia* studies and in the neotropical palm *Mauritia flexuosa* Mart., which is influenced by BIO1, BIO2, and BIO14 (Melo *et al.*, 2020; Resende *et al.*, 2020). Although *Acrocomia* can thrive in diverse climatic conditions occurring from humid to semi-arid ecosystems, they are exclusively found in open areas, often influenced by human activity. Our results are supported by previous studies that suggest temperature played a significant role in the evolution of the climate niche in Neotropical palm species (Walther *et al.*, 2007; Eiserhardt *et al.*, 2013).

The sNMF analysis (Fig. S2) showed low differentiation among populations from different biomes in putative SNPs under selection. However, there was evidence of admixture in *Acrocomia totai*, particularly between natural populations from Cerrado in Paraguay, Mato Grosso do Sul (Brazil), Parana (Brazil), and Sao Paulo (Brazil). Biogeographic isolation by barriers is an important driver of genetic differentiation in palm as described by Eiserhardt *et al.:* “The significant effect of spatial distance on clade turnover in American palms can thus be interpreted as evidence for dispersal limitation on evolutionary timescales caused either by barriers or by time” (Eiserhardt *et al.*, 2013). Even though Eiserharth's results focus on macroevolutionary diversification and phylogenetic turnover by dispersal and niche evolution, they indicate that barrier distance explains more variation in diversity than spatial distance. We observe a similar pattern in our results, where biogeographic barriers are more relevant in shaping genetic structure than biomes proximity and microclimatic variables in *Acrocomia* niches (Fig. 3).

Palms are not naturally adapted to withstand extremely high or low temperatures. Interestingly, the evolution of these geographic range restrictions does not appear to be a limiting factor for *Acrocomia*, despite the fact that water supply most likely limits the regional ranges of many Neotropical palms (Blach-Overgaard *et al.*, 2010). *Acrocomia* appears to be highly adapted to semi-arid ecosystems. The distributions of Neotropical palm species are assumed to be impacted by edaphic environments, despite the absence of empirical data. It is known that the natural occurrence in Brazilian populations of *Acrocomia aculeata* is associated with eutrophic (high fertility) soils with medium to clayey textures and an average pH of 5.5 (Motoike *et al.*, 2013). However, further investigation is needed for the North American gene pool characteristics.

Domestication syndrome in palms is poorly understood (Clement, 1999a; Levis *et al.*, 2018). Detecting selective sweep regions associated with domestication in *Acrocomia* is crucial. These genes probably have undergone strong positive selection during this process. Selective sweeps are characterized by a reduction in genetic diversity around the selected allele due to the rapid increase in frequency of advantageous alleles favored by human-driven selection (Pankin *et al.*, 2018). Maybe there is a heterogeneous origin in the domestication of *Acrocomia*. However, the existence of several founder lineages and the link between a mosaic ancestry patterns only will be clear with a sequenced reference genome for the three species. The identification of loci under selection can be greatly enhanced by having access to a reference genome (Huang *et al.*, 2022). *Acrocomia* may respond polygenically to environmental variation like other palms (Ting *et al.*, 2016; Melo *et al.*, 2020; Salgado *et al.*, 2022), meaning that selection may result in minimal changes in allele frequencies, leading to a poor adaptation signal at a single locus.

Our results demonstrate that genetic structure in populations from different biomes is homogenized within the same biogeographic regions due to gene flow. This also suggests the potential for migration to disperse advantageous alleles among populations sharing the same biogeographic region, leading to allele fixation in genes involved in adaptation to environmental change. This potentially explains *Acrocomia*'s tolerance to seasonality and sheds light on the species' interaction with humans and its domestication process. We suggest investigating the functionality of distinct biological pathways under diverse conditions using a transcriptomic approach for understanding better its domestication syndrome and its potential resilience to climate change (Walther *et al.*, 2007; Resende *et al.*, 2020). Also, conservation strategies can be formulated by identifying genetically distinct populations that may require special attention for conservation.

Genes putatively involved in adaptation and implications for domestication in *Acrocomia*

A high-quality reference genome sequence and annotation at chromosomal level of plant species is essential for genetic research on crop breeding and domestication (Schreiber, Stein and Mascher, 2018). *Acrocomia* is considered an incipiently domesticated plant and does not have a reference genome yet, so a population genomics approach using GBS or RAD-Seq can help investigate the evolutionary mechanisms underlying its diversification and variation. Significant genomic alterations have occurred during plant domestication as a result of evolutionary processes such as genetic drift and artificial selection. The study of artificial selection has used genomic techniques for detecting signatures putatively under selection, which focus on identifying polymorphisms in trait loci that have been disguised by neutral variation in species under a domestication process (Huang *et al.*, 2022).

Despite lacking a complete reference genome for *Acrocomia*, we were able to analyze genes potentially under selection. To do this, Blast2GO and Blastx use different reference genomes of related palm species to compare the loci where signatures of potential adaptation were located. The most common similarity matches between our GBS tags and annotated sequences were found mostly in homologous genes of *Elaeis guineensis* (oil palm), *Cocos nucifera* (coconut), and *Phoenix dactylifera* (date palm) (Table 1). For example, phylogenomic studies utilizing chloroplast genome sequences of *Acrocomia aculeata* and

Elaeis guineensis demonstrated that all nodes had a posterior probability of 1.0 (PP = 1.0), indicating a close relationship between these species (Francisconi *et al.*, 2023). Here, we present some relevant candidate genes that exhibit selection signatures and could be involved in the adaptation processes (and maybe the domestication) of *Acrocomia*.

Some significant selection signatures were identified in genes associated with fatty acid and triacylglycerol biosynthesis pathways in *Acrocomia aculeata* (Table 1). For example, loci such as CLocus_25725 correspond to the gene encoding a lipid-binding protein that facilitates lipid transport. CLocus_11291 corresponds to the gene encoding a patatin-like protein 2, which exhibits phospholipase activity in lipid catabolic processes and acylglycerol lipase activity. Additionally, CLocus_32860 was found in the gene encoding an ABC transporter D family member 1, involved in lipid metabolism functions such as the transport of long-chain fatty acids, import of long-chain fatty acids into peroxisomes, and fatty-acyl-CoA transport. Furthermore, CLocus_27100 corresponds to the gene encoding a putative WRKY transcription factor that modulates the content of fatty acids and influences the accumulation of seed oil. Additionally, we found signatures like CLocus_24655 and CLocus_12388 related to genes involved in ion binding processes. In the research led by Couto *et al.*, candidate genes related to oil production traits in *A. aculeata* were identified (Couto *et al.*, 2024). The genome-wide association study (GWAS) suggested that candidate genes controlling oil production were associated with metal ion binding and correlated with traits such as fruit oil content, fruit pulp fresh mass, leaf number, and leaf length. In *Acrocomia totai*, the locus CLocus_17145 was present in the gene encoding GDSL esterase/lipase, which mobilizes the lipids stored in seeds and plays a major role in seed germination and early seedling establishment (Ding *et al.*, 2023). These SNPs may be of agronomic importance because one of the primary objectives of the *Acrocomia* breeding program is to increase oil content for industrial purposes such as biofuels.

These fatty acids and triacylglycerols biosynthesis pathways were also reported in the fine mapping and cross-validation of QTLs linked to fatty acid composition in different varieties of *Elaeis guineensis* made by Ding *et al.* (Ding *et al.*, 2023). The Blast results for their QTL included genes and transcription factors linked to diacylglycerol acyltransferase (DGAT1) and long chain acyl-CoA synthetase. The synthesis of triacylglycerol and fatty acids occurs in different compartments within plants. Specifically, new fatty acid synthesis takes place in the plastid (leucoplast), where the acyl carrier protein (ACP) holds the fatty acid chain as it elongates. Then, acyl-ACP thioesterases hydrolyze the acyl-ACPs, releasing non-esterified fatty acids. These fatty acids are then exported to the endoplasmic reticulum, where they assemble to form triacylglycerol (Chapman and Ohlrogge, 2012; Ting *et al.*, 2016).

Another significant finding were signatures associated in carotenoid biosynthesis pathways (Table 1). In *Acrocomia aculeata*, loci such as CLocus_16512 and CLocus_11820 were associated with genes encoding carotenoid cleavage dioxygenase 7 and carotenoid cleavage dioxygenase 8 homolog B, which play roles in carotene catabolic processes and the production of apocarotenoids. These enzymes, known as carotenoid cleavage dioxygenases (CCDs), also influence shoot and branching inhibition and act on various carotenoid substrates like lycopene or zeaxanthin. Additionally, in *A. totai*, the locus CLocus_24114 was found in the gene encoding 9-cis-epoxycarotenoid dioxygenase NCED6, which is involved in regulating plant growth and seed dormancy. Carotenoids are vital for plant growth and development as they serve as precursors for the synthesis of plant hormones such as strigolactones and abscisic acid (ABA) (Mishra, Upadhyay and Shukla, 2017; Saeed, Naseem and Ali, 2017). From a nutritional perspective, the high content of carotenoids in crops is an attractive trait, offering benefits such as mitigating vitamin A deficiency through dietary intake of plant-derived carotenoids. For example, the increasing popularity of orange carrots may be attributed, at least in part, to the numerous studies highlighting the health advantages associated with carotenoids (Coe *et al.*, 2023).

Additionally, we observe the presence of outlier SNPs in putative resistance genes from different classes, indicating that natural populations under selection possess important genetic resources for crop defense against pathogens (Table 1). For instance, outlier SNPs such as CLocus_27100, CLocus_317, CLocus_22454, CLocus_29621, CLocus_35203, and CLocus_11291 were identified in genes involved in plant defense and disease resistance responses in *Acrocomia aculeata*. In *A. totai*, SNPs such as CLocus_37219, CLocus_29376, and CLocus_12178 were associated with genes involved in plant-pathogen recognition, immune priming, and regulation of defense response to fungus, respectively. Currently, there is no evidence of pathogens, pests, or diseases in either natural populations or commercial plantations of *A. aculeata* that are currently in the production phase located in Brazil's central region (Colombo *et al.*, 2018). Phytosanitary issues with this novel crop would likely arise during the domestication bottleneck in response to genetic diversity and loss of heterozygosity.

We have identified signatures in genes specialized in adapting to abiotic and environmental stress in *Acrocomia aculeata*, such as CLocus_317, CLocus_12388, CLocus_23382, CLocus_25856, and CLocus_32860 (Table 1). In *Acrocomia totai*, a signature like CLocus_15942 and CLocus_24114 were found in the gene encoding low-affinity sulfate transporter 3 and 9-cis-epoxycarotenoid dioxygenase NCED6 respectively, responsible for the plant's response to drought and salinity stress (Gallardo *et al.*, 2014). Similarly, in *Acrocomia intumescens*, the only signature (CLocus_111127) present on a gene was also associated with stress responses. This may help elucidate how *Acrocomia* species can withstand extreme temperatures and droughts. Drought tolerance is a polygenic, complex quantitative trait (Wang *et al.*, 2020). Developments in agricultural physiology and genetics have contributed with important insights into drought tolerance in palm species (Wang *et al.*, 2020; Salgado *et al.*, 2022; Suraninpong *et al.*, 2023). Therefore, improving yield under water-deficient conditions is a major objective in plant breeding. Increasing drought tolerance through traditional breeding is a slow process (Golldack, Lüking and Yang, 2011). *Acrocomia*'s inherited characteristics give it the advantage of being adapted to semi-arid, even arid regions like the Brazilian Caatinga. For the majority of non-model species, the mechanisms driving drought tolerance remain poorly understood and vary among species. In comparison to domesticated species, incipiently or less domesticated populations may have a larger range of ecological adaptations (Petereit *et al.*, 2022). Crop adaptability to perturbed areas depends on its resistance and resilience to environmental stress, which could be crucial for *Acrocomia*'s adaptation in the face of climate change.

Challenges of reduced representation sequencing (GBS/RAD-seq) in adaptation studies

Reduced representation sequencing methods in adaptation studies such as genotyping-by-sequencing (GBS) or restriction-site associated DNA sequencing (RAD-seq) employ high-throughput sequencing to generate high genome-wide marker data. The advantages of these techniques is their capacity to develop genotyping assays without requiring prior genomic knowledge or substantial costs without needing a reference genome like *Acrocomia* (Schreiber, Stein and Mascher, 2018). GBS serves as a great alternative for investigating unconventional food plants, natural populations, and wild relatives. Additionally, it has been employed to assess diversity and gene flow between crops and their wild types, as demonstrated in Brazilian manioc varieties and other crops (Alves-Pereira *et al.*, 2022). Some challenges in adaptation analysis using GBS data include allelic dropout (Gautier *et al.*, 2013), missing targets in resequencing, false positives selection signatures (Ahrens *et al.*, 2018) and the management of missing data (Arnold *et al.*, 2013). However, the identification of false positives may arise from departures from the model's assumptions and covariance with sampling techniques, population structure, and demographic variables (Ahrens *et al.*, 2018). Generally, combining the SNPs identified by various outlier tests is one approach to address these limitations (Pankin *et al.*, 2018; Alves-Pereira *et al.*, 2022).

Meanwhile, it is becoming increasingly evident that many crops have complex ancestries, necessitating additional archaeogenomics data (Huang *et al.*, 2022). Because of its wide distribution, to gain a better understanding of the domestication process of *Acrocomia* and possible genes targeted by human selection, a more comprehensive examination of variables including the diversity of wild population substructure and the proportional genetic gains of various natural subpopulations is recommended, especially in the North American genepool. A pangenomic approach is crucial to consider because plant genomes are dynamic, containing a reservoir of genetic diversity that enables adaptation to different biomes (Petereit *et al.*, 2022). Future crop varieties will likely require a broader range of genes than those provided by single reference genomes. Developing a pangenome of *Acrocomia* could capture the genomic diversity present in the various gene pools observed in our research.

The promising aspects of *Acrocomia* and future perspectives

After exploring various aspects of population genetics, adaptation processes, and potential domestication in *Acrocomia* species, our study presents a comprehensive understanding of the genetic structure, evolutionary history, and adaptive mechanisms within this genus. Through advanced genomic techniques such as SNP genotyping and adaptive signatures analyses, we explored the relationships between different *Acrocomia* species and their adaptation to diverse environmental conditions across neotropical biomes. A high degree of diversity was observed in *Acrocomia* species from various neotropical biomes. This variation may have arisen from the crop's natural history, domestication process and its cultivation in diverse ecosystems with varying human preferences. It is plausible that some of the selection signals are associated with desirable agronomic traits, which could be crucial for the breeding of these palms.

Acrocomia emerges as a promising crop due to its multipurpose potential, offering prospects in oil production, biofuels, food, and pharmaceutical industry (Aires and de Carvalho Junior, 2023). Notably, the oil extracted from *A. aculeata* exhibits quality comparable to palm oil (*Elaeis guineensis*) while requiring lower water resources and demonstrating adaptability to semi-arid to arid environments and degraded ecosystems. Given that approximately 80% of Brazilian grasslands are classified as degraded, intercropping with perennial plants like

Acrocomia emerges as a promising solution for combating soil degradation (Colombo *et al.*, 2018; Resende *et al.*, 2020). Currently, Brazil is leading its breeding programs through public-private initiatives focusing on *Acrocomia aculeata* and *A. totai* (Vargas-Carpintero *et al.*, 2022). It is projected that experimental and commercial plantations of *A. aculeata* will reach an area of 200,000 hectares in the next decade (Bertão, 2023), with a guaranteed market for the oil and biodiesel production. Significant research progress has been made, particularly in seed germination, seedling production, and integrated agricultural systems, although molecular breeding remains necessary (Colombo *et al.*, 2018; Vargas-Carpintero *et al.*, 2021). Prioritizing the sequencing and annotation of the entire genomes of the three most economically important species in the genus is crucial for future genomic efforts in their domestication process. Addressing challenges such as genotype-environment interaction, biodiversity prospecting for phenotypic breeding ideotypes (plant models), and crop management are necessary. Most research efforts have focused only on the South American gene pool.

Our findings reveal significant genetic structuring within and between species, highlighting the influence of dispersal, biogeographic barriers, and historical ecological interactions on the evolution of *Acrocomia*. We observed distinct genetic signatures associated with adaptation to specific environmental factors, including climatic variables, pathogen resistance, and stress responses. Notably, *Acrocomia* species exhibit potential resilience to climate change, particularly in semi-arid ecosystems, making them promising candidates for sustainable agriculture and biofuel production. Furthermore, our study sheds light on the potential domestication of *Acrocomia* species, emphasizing the importance of genetic resources and candidate genes associated with desirable agronomic traits such as oil content, carotenoid biosynthesis, and stress tolerance. While facing challenges inherent to reduced representation sequencing methods, such as GBS, our research underscores the necessity of comprehensive genomic resources, including reference genomes and pangenomic approaches, to fully elucidate the domestication syndrome and facilitate breeding efforts.

Acrocomia is a valuable crop offering opportunities for future industries while contributing to sustainable development and biodiversity conservation. By prioritizing collaborative research efforts in private-public programs, genetic conservation, and policy support, different stakeholders can unlocked the potential of genetic diversity and adaptive capacity of *Acrocomia* species to address global challenges, including food security, renewable energy production, and climate resilience in agroecosystems.

Materials and methods

Plant material and DNA isolation

We sampled leaves from natural populations, 78 for *Acrocomia aculeata*, 40 for *Acrocomia totai*, and 131 for *Acrocomia intumescens*, for a total of 249 samples (Table 2). The collections of these samples were registered in the Brazilian National Council for Genetic Patrimony CGEN (numbers A69E071 and A5D139D). The leaves were dehydrated using silica gel and stored in paper bags at -20 °C. Following Doyle and Doyle's protocol, we extracted whole genomic DNA from 50 mg leaf samples. Agarose gel electrophoresis (1% w/v) with GelRed stain (Sigma-Aldrich) was used to assess DNA quality and integrity. We quantified and normalized DNA concentrations to 30 ng/µL using the dsDNA BR Assay quantification kit for the Qubit3 fluorometer (Invitrogen).

Table 2. *Acrocomia* species samples. Geographical location and biome of the Acrocomia species samples.

Acrocomia aculeata						
Country	**State**	**Location**	**Local Biome**	**Macro Ecoregion**	**Lat**	**Lon**
Brazil	Minas Gerais	Cassia	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-20.566412	-46.933644
Brazil	Minas Gerais	Ibituruna	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-21.343044	-44.739736
Brazil	Minas Gerais	Luz	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-19.773972	-45.864639
Brazil	Minas Gerais	Montes Claros	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-16.747219	-43.886300
Brazil	Minas Gerais	Patos de Minas	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-18.682009	-46.571756
Brazil	Para	Belem	Amazon	Tropical Moist Broadleaf Forest	-1.1446780	-48.146053
Brazil	Rio de Janeiro	Guapimirim	Atlantic coastal forest	Subtropical Moist Broadleaf Forest	-22.537222	-42.981944
Brazil	Rio de Janeiro	Itaborai	Atlantic coastal forest	Subtropical Moist Broadleaf Forest	-22.714332	-42.810942
Brazil	Sao Paulo	Brotas	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-22.276083	-48.118500
Brazil	Sao Paulo	Rifania	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-19.986278	-47.508472
Brazil	Tocantins	Palmas	Amazon	Tropical Moist Broadleaf Forest	-9.043889	-48.324250
Colombia	Casanare	Aguazul	Llanos	Tropical Dry Broadleaf Forest	5.169456	-72.552262
Costa Rica	Guanacaste	Liberia	Central American moist forest	Tropical Moist Broadleaf Forest	10.603496	-85.429016
Costa Rica	San Jose	San Jose	Central American moist forest	Tropical Moist Broadleaf Forest	9.897481	-84.413358
Mexico	Chiapas	Cacahuatan	Central American moist forest	Tropical Moist Broadleaf Forest	14.980654	-92.172965
Mexico	Chiapas	Tuxtla chico	Central American moist forest	Tropical Moist Broadleaf Forest	14.922992	-92.179134
Mexico	Quintana Roo	Cancún	Central American moist forest	Tropical Moist Broadleaf Forest	21.153694	-86.842000
Mexico	Veracruz	Mangal	Central American moist forest	Tropical Moist Broadleaf Forest	19.002444	-96.159917
Trinidad and Tobago	Tunapuna piarco	Saint George	Caribbean dry forest	Tropical Moist Broadleaf Forest	10.664417	-61.399028

Acrocomia totai

Country	State	Location	Local Biome	Macro Ecoregion	Lat	Lon
Argentina	Formosa	Misión Tacaaglé	Chaco and Espinal	Subtropical Grasslands, Savannas and Shrublands	-24.981556	-58.843472
Paraguay	Itapúa	Bella Vista	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-27.029879	-55.579355
Brazil	Mato Grosso do Sul	Campo Grande	Cerrado	Subtropical Grasslands, Savannas and Shrublands	-20.469056	-54.777361
Brazil	Mato Grosso do Sul	Corumba	Pantanal	Flooded Grasslands and Savannas	-19.351261	-57.563631
Brazil	Mato Grosso do Sul	Dourados	Cerrado	Subtropical Moist Broadleaf Forest	-22.262750	-54.837889
Brazil	Mato Grosso do Sul	Porto Murtinho	Chaco and Espinal	Subtropical Grasslands, Savannas and Shrublands	-21.561294	-57.811203
Brazil	Parana	Xambrê	Cerrado	Subtropical Moist Broadleaf Forest	-23.736111	-53.490000
Brazil	Sao Paulo	Teodoro Sampaio	Cerrado	Subtropical Moist Broadleaf Forest	-22.536222	-52.183842

Acrocomia intumescens

Country	State	Location	Local Biome	Macro Ecoregion	Lat	Lon
Brazil	Ceara	Caririaçu	Caatinga	Tropical Dry Broadleaf Forest	-7.030754	-39.273128
Brazil	Ceara	Crato	Caatinga	Tropical Dry Broadleaf Forest	-7.218414	-39.432748
Brazil	Ceara	Fortaleza	Atlantic coastal forest	Tropical Dry Broadleaf Forest	-3.881995	-38.502484
Brazil	Ceara	Guaramiranga	Atlantic coastal forest	Tropical Dry Broadleaf Forest	-4.261151	-38.928265
Brazil	Paraiba	Areia	Caatinga	Tropical Dry Broadleaf Forest	-6.962329	-35.688065
Brazil	Paraiba	Campina Grande	Caatinga	Tropical Dry Broadleaf Forest	-7.203068	-35.845533
Brazil	Paraiba	Joao Pessoa	Atlantic coastal forest	Tropical Dry Broadleaf Forest	-7.098467	-34.966734
Brazil	Paraiba	Mata Limpa	Caatinga	Tropical Dry Broadleaf Forest	-6.938426	-35.712936
Brazil	Paraiba	Remigio	Caatinga	Tropical Dry Broadleaf Forest	-6.959532	-35.789893
Brazil	Paraiba	Rio Tinto	Atlantic coastal forest	Tropical Dry Broadleaf Forest	-6.823933	-35.060280
Brazil	Pernambuco	Recife	Atlantic coastal forest	Tropical Dry Broadleaf Forest	-7.980871	-34.931509

GBS libraries and SNP discovery

Genomic libraries were prepared following the protocol of genotyping-by-sequencing using two restriction enzymes (ddGBS) as described by Poland *et al*. (Poland *et al.*, 2012). The combination of *NsiI* and *MspI* (New England Biolabs) was used for the libraries of *A. aculeata* and *A. totai*, while the combination of *NsiI* and *MseI* (New England Biolabs) was used for *A. intumescens*. Two libraries were prepared mixing *A. aculeata* and *A. totai* samples, each using a 96-plex set of *NsiI* barcode adapters and a common *MspI* adapter. Another two libraries were prepared for *A. intumescens*, each using a 96-plex set of *NsiI* barcode adapters and a common *MseI* adapter. The ddGBS libraries were quantified through RT-PCR on the CFX 384 Touch Real Time PCR (BioRad) equipment using a KAPA Library Quantification kit (KAPA Biosystems, cat. KK4824), and the fragments' profiles were inspected using the Agilent DNA 1000 Kit on a 2100 Bioanalyzer (Agilent Technologies). The libraries of *A. aculeata* and *A. totai* were sequenced on two separate runs in an Illumina NextSeq500, with single-end and 150 bp configurations. The two libraries of *A. intumescens* were sequenced on a single run, but in separate lanes, in an Illumina HiSeq2500, with single-end and 101 bp configurations.

The general sequencing quality and the presence of adapters were assessed with FASTQC (Andrews, 2010). The SNP discovery was performed separately for each species following the *de novo* pipeline of the program Fv.1.42 (Catchen *et al.*, 2013) with similar filtering criteria. Due to the presence of adapters, sequences were trimmed to 90 bp for *A. aculeata* and *A. totai*, and 80 pb *A. intumescens*. Trimming, quality control (removal of sequences with uncalled bases and with Phred scores <10), and demultiplexing were performed with the module *process_radtags*. For each sample, loci were assembled using the module *ustacks* considering minimum sequencing depth (-m) of 3, and distance between reads from the same loci (-N) of 2. A catalog of loci across samples was built with the module *cstacks*, considering the distance between locus (-n) of 2, and loci of the samples were compared to the catalog using *sstacks*. The loci with lower probabilities (-lnl_lim 10) were discarded using the module *rxstacks*. For the three species, candidate SNPs were identified using the module *populations* considering a minimum depth of 3, minor allele frequency of 0.01, the presence of SNP in at least 75% of the samples in each of 14 (*A. aculeata*), 8 (*A. totai*), or 10 (*A. intumescens*) sampling locations. To avoid explicit linkage only a single SNP was retained per GBS locus. Quality metrics (mean sequencing depth per locus and per sample, percentage of missing data per sample) were assessed with VCFTools (Danecek *et al.*, 2011). Additionally, samples with more than 50% of missing genotypes were removed from the final data set, resulting in 69 samples for *A.aculeata*, 40 for *A. totai*, and 131 for *A. intumescens*.

Detection of putative signatures of selection

Putative signatures of selection were investigated using complimentary approaches (Lotterhos and Whitlock, 2015; Manel *et al.*, 2018) either based on significant deviations of F_{ST} estimates among populations, based on principal component analyses (PCA), or based on environmental association analyses, which identify association between environmental variables and individual markers.

A classic approach to identify outlier SNPs is the identification of loci with extremely high or low values of F_{ST} (or related statistics) estimated among populations. This method was performed using the FstHet (Flanagan and Jones, 2017), package for R (R Core Team, 2016) which implements a model similar to FDist2 (Beaumont and Nichols, 1996) to construct a neutral envelope for the distribution of F_{ST}s estimated based on the given data set. The neutral envelope was constructed based on 1,000 bootstraps of the F_{ST}-analogous betahat statistic (Cockerham and Weir, 1993), which considers variations in sampling sizes across populations. In this analysis, the markers below or above the 95% envelope of betahat estimates were considered as outlier SNPs.

Pcadapt (Luu, Bazin and Blum, 2017) was used to identify SNP markers significantly associated with the genetic structure of the data based on a PCA, without assuming any genetic model. This analysis was performed in the pcadapt package for R (R Core Team, 2016), retaining the first 3, 4, and 1 K principal components for *A. aculeata*, *A. totai* and *A. intumescens*, respectively (Fig. S1). In this analysis, the SNPs with q-values ≤ 0.1 (corrected p-values for their association with the first K principal components) were considered as outliers.

The Latent Factor Mixed Models (LFMM) analysis (Frichot *et al.*, 2013) was used to assess the correlations of SNP markers with environmental variables obtained in the *WorldClim2* data base (Fick and Hijmans, 2017). The information recovered from *WorldClim2* refer to 19 bioclimatic variables, of which 11 are primarily related to annual trends of temperature (BIO1 to BIO11) and eight are related to annual trends of precipitation (BIO12 to BIO19). For each species, the variables were extracted for the sampling points, and a Pearson's correlation test was employed to minimize interdependence among them, and only the variables with a correlation coefficient < 0.8 were retained (Fig. S2). A sparse non-negative matrix factorization (sNMF) analysis (Frichot *et al.*, 2014) was performed to estimate the most likely number of genetic clusters for each species to model the covariation of the subjacent genetic structure in LFMM. Both, sNMF and LFMM were performed with the package LEA (Frichot and François, 2015) for R (R Core Team, 2016). For each species, 10 independent repetitions of sNMF simulating from 1 to 10 K ancestral groups were performed with 200,000 iterations. The most probable number of $K = 2$ for *A. aculeata* and *A. intumescens*, and $K = 4$ for *A. totai* were estimated according to the cross-entropy estimates of the algorithm (Fig. S3). These numbers were used in subsequent LFMM analyses, which were performed based on 10 repetitions of 50,000 burn-in followed by 100,000 iterations of the algorithm for each species. The generated p-values were corrected considering a false discovery rate (FDR) of 0.1 as threshold for the identification of SNPs significantly associated with the environmental variables.

SNPs were declared as putatively under selection if identified as outliers (or significantly associated to environmental variables) by at least two of the three methods described above. Blast2GO (Götz *et al.*, 2008) was used to assess similarities between the GBS tags with outlier SNPs and proteins with described putative functional annotations. The similarity between GBS tags in which outlier SNPs were identified and proteins deposited in GenBank was evaluated using blastx with default configurations but restricting the search to Viridiplantae data. Sequences with blastx hits were then screened against the Pfam database to search for protein domains and their associated Gene Ontology (GO) annotations. The putative annotations were summarized based on the GO terms and visually represented with bar plots using the on-line tool WEGO (http://wego.genomics.cn/).

Acknowledgements

We acknowledge the financial support from the São Paulo Research Foundation (FAPESP—grants: 2021/10319-0, 2019/20307-0, 2014/23591-7) and the National Council for Scientific and Technological Development (CNPq—grant #431226/2018-0). We also thank the Coordination for the Improvement of Higher Education Personnel of Brazil (CAPES) for the granted scholarship Ph.D. to JM-M. Also thanks the São Paulo Research Foundation for a post-doctoral scholarship to AA-P (FAPESP 2018/00036-9).

Supporting information

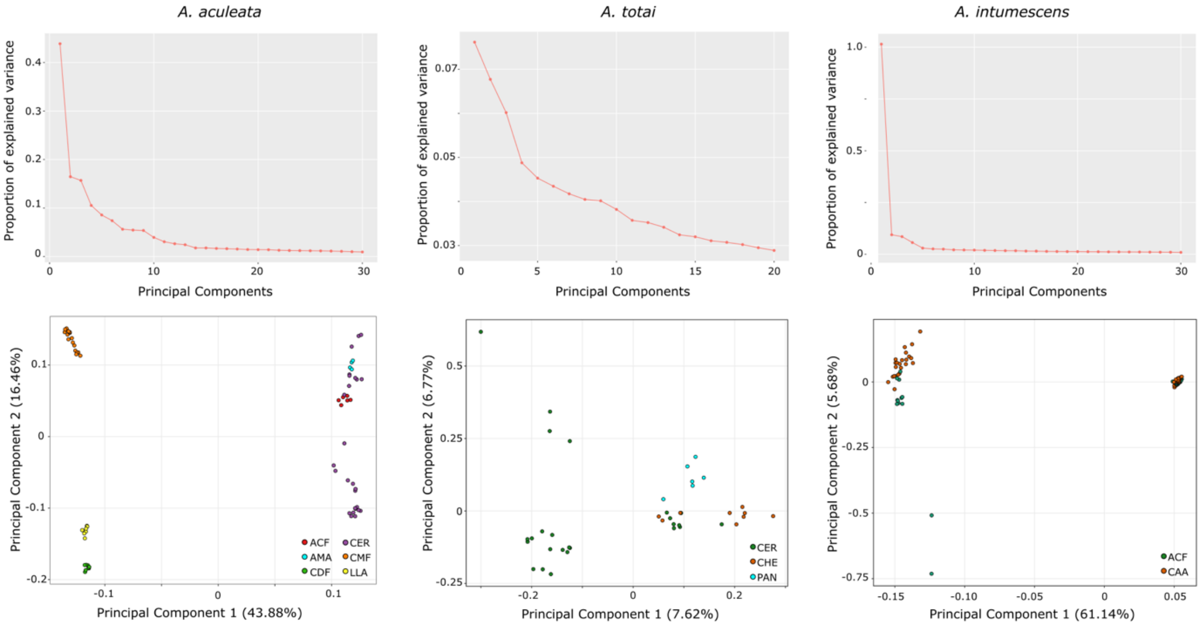
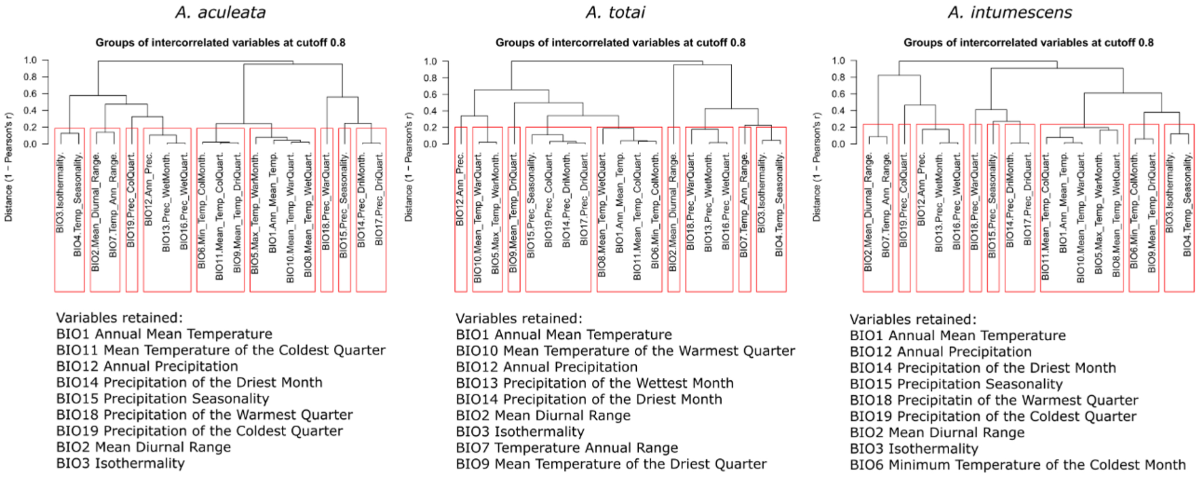


Fig. S1 Principal component analysis (PCA) performed for outlier detection using pcadapt for *Acrocomia aculeata*, *A. totai* and *A. intumescens*. Top row: scree plots of the explained variance (y-axis) retained in each principal component (PC) (x-axis). The number of retained PCs was chosen as the first inflection point (Cattell's rule), where the amount of genetic variation added by successive PCs reaches a plateau. Bottom row: the associated scatter plots of the first two PCs.



Variables retained:
BIO1 Annual Mean Temperature
BIO11 Mean Temperature of the Coldest Quarter
BIO12 Annual Precipitation
BIO14 Precipitation of the Driest Month
BIO15 Precipitation Seasonality
BIO18 Precipitation of the Warmest Quarter
BIO19 Precipitation of the Coldest Quarter
BIO2 Mean Diurnal Range
BIO3 Isothermality

Variables retained:
BIO1 Annual Mean Temperature
BIO10 Mean Temperature of the Warmest Quarter
BIO12 Annual Precipitation
BIO13 Precipitation of the Wettest Month
BIO14 Precipitation of the Driest Month
BIO2 Mean Diurnal Range
BIO3 Isothermality
BIO7 Temperature Annual Range
BIO9 Mean Temperature of the Driest Quarter

Variables retained:
BIO1 Annual Mean Temperature
BIO12 Annual Precipitation
BIO14 Precipitation of the Driest Month
BIO15 Precipitation Seasonality
BIO18 Precipitatin of the Warmest Quarter
BIO19 Precipitation of the Coldest Quarter
BIO2 Mean Diurnal Range
BIO3 Isothermality
BIO6 Minimum Temperature of the Coldest Month

Fig. S2 bioclimatic variables recovered from WorldClim2 and extracted for the sampling points of *Acrocomia aculeata*, *A. totai* and *A. intumescens*. Groups of intercorrelated variables highlighted within red boxes (Pearson's correlation coefficients > 0.8).

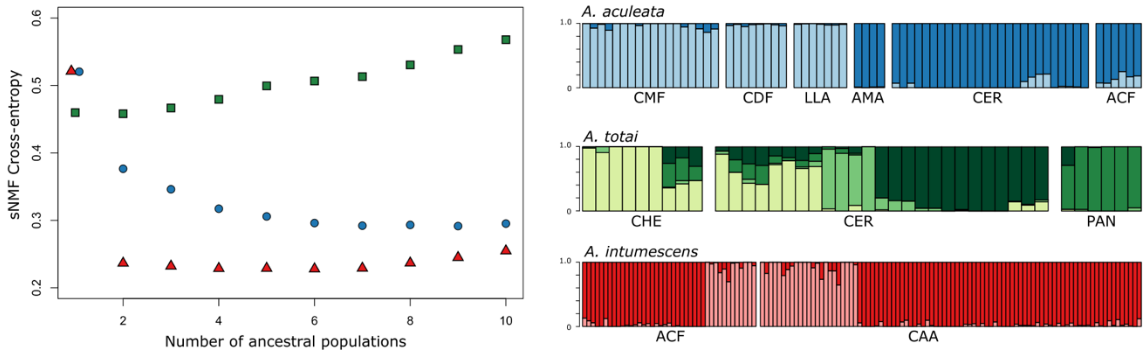


Fig. S3 Sparse non-negative matrix factorization (sNMF) performed for outlier detection using Latent Factor Mixed Models (LFMM) for *Acrocomia aculeata*, *A. totai* and *A. intumescens*. Left: plot of the cross-entropy estimates for each number of simulated ancestral populations. Right: Bar plots representing the sNMF ancestry coefficients across samples from different biogeographic group in each species. Each sample is represented by a bar and different shade colors represent their associated ancestry proportion from distinct genetic groups. Acronyms follow Table 2: CMF Central American Moist Forest; LLA Llanos; CDF Caribbean Dry Forest; ACF Atlantic Coastal Forest; CAA Caatinga; AMA Amazon; CER Cerrado; PAN Pantanal; CHE Chaco and Espinal.

Table. S1 List of loci identified as outliers for each species, based on Pcadapt, FstHet and LFMM. For Pcadapt, “PC” refers to the principal component associated to the outlier locus. For FstHet, estimates of total genetic diversity (Ht) and genetic divergence (Fst) are presented. For LFMM, values within cells are the p-values estimated for the association between loci and the given bioclimatic variable (Excel spreadsheet). https://drive.google.com/drive/folders/1x8oQInx-hu77PPLNU4mfhAAFn7k6fElR?usp=drive_link

Table. S2 Summary of Blast2GO analysis. Information of SNP loci putatively under selection and their associated GBS tags are followed by the first significant blastx hit (including protein name, the species for which it was described, and GenBank accession number), and putative annotations recovered as Gene Ontology (GO) terms. E-value = expected probability of spurious hits (Excel spreadsheet). https://drive.google.com/drive/folders/1x8oQInx-hu77PPLNU4mfhAAFn7k6fElR?usp=drive_link

References

Ahrens, C. W. et al. (2018) ‘The search for loci under selection: trends, biases and progress’, Molecular Ecology. John Wiley & Sons, Ltd, 27(6), pp. 1342–1356. doi: 10.1111/MEC.14549.

Aires, G. C. M. and de Carvalho Junior, R. N. (2023) ‘Potential of Supercritical Acrocomia aculeata Oil and Its Technology Trends’, Applied Sciences 2023, Vol. 13, Page 8594. Multidisciplinary Digital Publishing Institute, 13(15), p. 8594. doi: 10.3390/APP13158594.

Alfaro-Solís, J. D. et al. (2020) ‘Acrocomia aculeata fruits from three regions in Costa Rica: an assessment of biometric parameters, oil content and oil fatty acid composition to evaluate industrial potential’, Agroforestry Systems. Springer Science+Business Media B.V., 94(5), pp. 1913–1927. doi: 10.1007/S10457-020-00511-8/TABLES/8.

Alves-Pereira, A. et al. (2022) ‘Selective signatures and high genome-wide diversity in traditional Brazilian manioc (Manihot esculenta Crantz) varieties’, Scientific Reports 2022 12:1. Nature Publishing Group, 12(1), pp. 1–14. doi: 10.1038/s41598-022-05160-8.

Andrews, S. (2010) ‘FastQC: A Quality Control Tool for High Throughput Sequence Data’.

Arnold, B. et al. (2013) ‘RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling’, Molecular Ecology. John Wiley & Sons, Ltd, 22(11), pp. 3179–3190. doi: 10.1111/MEC.12276.

Assine, M. L. et al. (2016) ‘Geology and geomorphology of the pantanal basin’, Handbook of Environmental Chemistry. Springer Verlag, 37, pp. 23–50. doi: 10.1007/698_2015_349/COVER.

Backer, R., Naidoo, S. and van den Berg, N. (2019) ‘The NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1) and related family: Mechanistic insights in plant disease resistance’, Frontiers in Plant Science. Frontiers Media S.A., 10, p. 430215. doi: 10.3389/FPLS.2019.00102/BIBTEX.

Baker, William J. and Couvreur, T. L. P. (2013) ‘Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography’, Journal of Biogeography. Edited by W. D. Kissling, 40(2), pp. 274–285. doi:

10.1111/j.1365-2699.2012.02795.x.

Baker, William J and Couvreur, T. L. P. (2013) ‘Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages’, Journal of Biogeography. Edited by W. D. Kissling, 40(2), pp. 286–298. doi: 10.1111/j.1365-2699.2012.02794.x.

Barkan, A. and Small, I. (2014) ‘Pentatricopeptide repeat proteins in plants’, Annual review of plant biology. Annu Rev Plant Biol, 65, pp. 415–442. doi: 10.1146/ANNUREV-ARPLANT-050213-040159.

Batalha-Filho, H. et al. (2023) ‘Disentangling the contemporary and historical effects of landscape on the population genomic variation of two bird species restricted to the highland forest enclaves of northeastern Brazil’, Heredity 2023 132:2. Nature Publishing Group, 132(2), pp. 77–88. doi: 10.1038/s41437-023-00662-1.

Beaumont, M. A. and Nichols, R. A. (1996) ‘Evaluating loci for use in the genetic analysis of population structure’, Proceedings of the Royal Society of London. Series B: Biological Sciences. The Royal SocietyLondon, 263(1377), pp. 1619–1626. doi: 10.1098/RSPB.1996.0237.

Bedoya, C. A. et al. (2017) ‘Genetic diversity and population structure of native maize populations in Latin America and the Caribbean’, PLOS ONE. Public Library of Science, 12(4), p. e0173488. doi: 10.1371/JOURNAL.PONE.0173488.

Bertão, N. (2023) Acelen dá início a projeto para uso da macaúba como fonte de energia, Globo Rural: Negócios.

Bitocchi, E. et al. (2012) ‘Mesoamerican origin of the common bean (Phaseolus vulgaris L.) is revealed by sequence data’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 109(14), pp. E788–E796. doi: 10.1073/PNAS.1108973109/SUPPL_FILE/ST02.DOC.

Blach-Overgaard, Anne et al. (2010) ‘Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints’, Ecography. John Wiley & Sons, Ltd, 33(2), pp. 380–391. doi: 10.1111/J.1600-0587.2010.06273.X.

Booker, J. et al. (2004) ‘MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule’, Current Biology. Elsevier, 14(14), pp. 1232–1238. doi: 10.1016/j.cub.2004.06.061.

La Camera, S. et al. (2009) 'The Arabidopsis patatin-like protein 2 (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens', Molecular Plant-Microbe Interactions, 22(4), pp. 469–481. doi: 10.1094/MPMI-22-4-0469.

Catchen, J. et al. (2013) 'Stacks: an analysis tool set for population genomics', Molecular Ecology, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.

Chapman, K. D. and Ohlrogge, J. B. (2012) 'Compartmentation of triacylglycerol accumulation in plants', Journal of Biological Chemistry. Elsevier, 287(4), pp. 2288–2294. doi: 10.1074/jbc.R111.290072.

Cheng, Y. et al. (2016) 'Putative WRKYs associated with regulation of fruit ripening revealed by detailed expression analysis of the WRKY gene family in pepper', Scientific Reports 2016 6:1. Nature Publishing Group, 6(1), pp. 1–11. doi: 10.1038/srep39000.

Ciconini, G. et al. (2013) 'Biometry and oil contents of Acrocomia aculeata fruits from the Cerrados and Pantanal biomes in Mato Grosso do Sul, Brazil', Industrial Crops and Products. Elsevier, 45, pp. 208–214. doi: 10.1016/J.INDCROP.2012.12.008.

Clement, C. R. (1999a) '1492 and the loss of amazonian crop genetic resources. I. The relation between domestication and human population decline', Economic Botany, 53(2), pp. 188–202. doi: 10.1007/BF02866498.

Clement, C. R. (1999b) '1492 and the loss of amazonian crop genetic resources. II. crop Biogeography at contact', Economic Botany, 53(2), pp. 203–216. doi: 10.1007/BF02866499.

Cockerham, C. C. and Weir, B. S. (1993) 'Estimation of Gene Flow from F-Statistics', Evolution. Oxford University Press (OUP), 47(3), p. 855. doi: 10.2307/2410189.

Coe, K. et al. (2023) 'Population genomics identifies genetic signatures of carrot domestication and improvement and uncovers the origin of high-carotenoid orange carrots', Nature Plants 2023 9:10. Nature Publishing Group, 9(10), pp. 1643–1658. doi: 10.1038/s41477-023-01526-6.

Colombo, C. A. et al. (2018) 'Macauba: a promising tropical palm for the production of vegetable oil', OCL. doi: 10.1051/ocl/2017038.

Costa, M. F. et al. (2022) 'Population genomics of the neotropical palm Copernicia prunifera (Miller) H. E. Moore: Implications for conservation', PLOS ONE. Public Library of Science, 17(11), p. e0276408. doi: 10.1371/JOURNAL.PONE.0276408.

Couto, E. G. de O. et al. (2024) 'Genome-Wide Association Insights into the Genomic Regions Controlling Oil Production Traits in Acrocomia aculeata (neotropical native palm)', bioRxiv. Cold Spring Harbor Laboratory, p. 2024.01.17.576138. doi: 10.1101/2024.01.17.576138.

Danecek, P. et al. (2011) 'The variant call format and VCFtools', Bioinformatics. Oxford Academic, 27(15), pp. 2156–2158. doi: 10.1093/BIOINFORMATICS/BTR330.

Díaz, B. G. et al. (2021) 'Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species', PLoS ONE, 16(7 July), pp. 1–24. doi: 10.1371/journal.pone.0241025.

Dinerstein, E. et al. (2017) 'An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm', BioScience. Oxford Academic, 67(6), pp. 534–545. doi: 10.1093/BIOSCI/BIX014.

Ding, Y. et al. (2023) 'Genome-wide exploration of the GDSL-type esterase/lipase gene family in rapeseed reveals several BnGELP proteins active during early seedling development', Frontiers in Plant Science. Frontiers Media S.A., 14, p. 1139972. doi: 10.3389/FPLS.2023.1139972/BIBTEX.

Eiserhardt, W. L. et al. (2013) 'Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents', Scientific Reports, 3(1), p. 1164. doi: 10.1038/srep01164.

Fernandes, M. F. et al. (2022) 'The Origins and Historical Assembly of the Brazilian Caatinga Seasonally Dry Tropical Forests', Frontiers in Ecology and Evolution. Frontiers Media S.A., 10, p. 723286. doi: 10.3389/FEVO.2022.723286/BIBTEX.

Ferreira Costa, M. et al. (2023) 'Population Genomics and Conservation of Copernicia alba: a Non-Timber Palm of Economic Importance in the Brazilian Chaco', Plant Molecular Biology Reporter 2023. Springer, pp. 1–11. doi: 10.1007/S11105-023-01419-0.

Fick, S. E. and Hijmans, R. J. (2017) 'WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas', International Journal of Climatology. John Wiley &

Sons, Ltd, 37(12), pp. 4302–4315. doi: 10.1002/JOC.5086.

Figueiredo, A., Monteiro, F. and Sebastiana, M. (2014) ‘Subtilisin-like proteases in plant–pathogen recognition and immune priming: A perspective’, Frontiers in Plant Science. Frontiers Media S.A., 5(DEC), p. 123235. doi: 10.3389/FPLS.2014.00739/BIBTEX.

Flanagan, S. P. and Jones, A. G. (2017) ‘Constraints on the FST–Heterozygosity Outlier Approach’, Journal of Heredity. Oxford Academic, 108(5), pp. 561–573. doi: 10.1093/JHERED/ESX048.

Francisconi, A. F. et al. (2023) ‘Complete chloroplast genomes of six neotropical palm species, structural comparison, and evolutionary dynamic patterns’, Scientific Reports 2023 13:1. Nature Publishing Group, 13(1), pp. 1–15. doi: 10.1038/s41598-023-44631-4.

Freitas, C. et al. (2021) ‘Incongruent Spatial Distribution of Taxonomic, Phylogenetic, and Functional Diversity in Neotropical Cocosoid Palms’, Frontiers in Forests and Global Change. Frontiers Media S.A., 4, p. 739468. doi: 10.3389/FFGC.2021.739468/BIBTEX.

Freitas, C. G. et al. (2019) ‘Adjacency and Area Explain Species Bioregional Shifts in Neotropical Palms’, Frontiers in Plant Science. Frontiers, 10, p. 55. doi: 10.3389/fpls.2019.00055.

Frichot, E. et al. (2013) ‘Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models’, Molecular Biology and Evolution. Oxford Academic, 30(7), pp. 1687–1699. doi: 10.1093/MOLBEV/MST063.

Frichot, E. et al. (2014) ‘Fast and efficient estimation of individual ancestry coefficients’, Genetics. Genetics, 196(4), pp. 973–983. doi: 10.1534/GENETICS.113.160572/-/DC1.

Frichot, E. and François, O. (2015) ‘LEA: An R package for landscape and ecological association studies’, Methods in Ecology and Evolution. John Wiley & Sons, Ltd, 6(8), pp. 925–929. doi: 10.1111/2041-210X.12382.

Fuchs, S. et al. (2013) ‘Type 2C protein phosphatases in plants’, The FEBS Journal. John Wiley & Sons, Ltd, 280(2), pp. 681–693. doi: 10.1111/J.1742-4658.2012.08670.X.

Gallardo, K. et al. (2014) ‘Sulfate transporters in the plant′s response to drought and salinity: Regulation and possible functions’, Frontiers in Plant Science. Frontiers Media S.A., 5(October), p. 116680. doi: 10.3389/FPLS.2014.00580/BIBTEX.

Gautier, M. et al. (2013) ‘The effect of RAD allele dropout on the estimation of genetic variation within and between populations’, Molecular Ecology. John Wiley & Sons, Ltd, 22(11), pp. 3165–3178. doi: 10.1111/MEC.12089.

Van Gelder, K. et al. (2021) ‘A central role for polyprenol reductase in plant dolichol biosynthesis’, Plant science : an international journal of experimental plant biology. Plant Sci, 303. doi: 10.1016/J.PLANTSCI.2020.110773.

Göldel, B. et al. (2016) ‘Impacts of large herbivores on spinescence and abundance of palms in the Pantanal, Brazil’, Botanical Journal of the Linnean Society. Oxford Academic, 182(2), pp. 465–479. doi: 10.1111/BOJ.12420.

Golldack, D., Lüking, I. and Yang, O. (2011) ‘Plant tolerance to drought and salinity: Stress regulating transcription factors and their functional significance in the cellular transcriptional network’, Plant Cell Reports. Springer, 30(8), pp. 1383–1391. doi: 10.1007/S00299-011-1068-0/METRICS.

Götz, S. et al. (2008) ‘High-throughput functional annotation and data mining with the Blast2GO suite’, Nucleic Acids Research. Oxford Academic, 36(10), pp. 3420–3435. doi: 10.1093/NAR/GKN176.

Guo, Z. H. et al. (2022) ‘Interactions between plant lipid-binding proteins and their ligands’, Progress in Lipid Research. Pergamon, 86, p. 101156. doi: 10.1016/J.PLIPRES.2022.101156.

Habsyah, T. S. et al. (2021) ‘Distribution of polyprenol and dolichol in oil palm genotype (Elaeis guineensis) involving lipase activity’, Biodiversitas Journal of Biological Diversity. Society for Indonesian Biodiversity, 22(2), pp. 830–837. doi: 10.13057/BIODIV/D220237.

He, R. et al. (2018) ‘Overexpression of 9-cis-epoxycarotenoid dioxygenase cisgene in grapevine increases drought tolerance and results in pleiotropic effects’, Frontiers in Plant Science. Frontiers Media S.A., 9, p. 360507. doi: 10.3389/FPLS.2018.00970/BIBTEX.

Hoorn, C. et al. (2010) ‘Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity’, Science. American Association for the Advancement of Science, 330(6006), pp. 927–931. doi: 10.1126/SCIENCE.1194585/SUPPL_FILE/HOORN.SOM.PDF.

Huang, X. et al. (2022) ‘The integrated genomics of crop domestication and breeding', Cell. Elsevier, 185(15), pp. 2828–2839. doi: 10.1016/J.CELL.2022.04.036.

Huang, Y. et al. (2018) '9-cis-epoxycarotenoid dioxygenase 3 regulates plant growth and enhances multi-abiotic stress tolerance in rice', Frontiers in Plant Science. Frontiers Media S.A., 9, p. 328147. doi: 10.3389/FPLS.2018.00162/BIBTEX.

Janzen, D. H. and Martin, P. S. (1982) 'Neotropical Anachronisms: The Fruits the Gomphotheres Ate', Science, New Series, 215(4528), pp. 19–27. Available at: http://www.jstor.org/stable/1688516 (Accessed: 15 March 2018).

Jin, H. et al. (2002) 'NPK1, an MEKK1-like Mitogen-Activated Protein Kinase Kinase Kinase, Regulates Innate Immunity and Development in Plants', Developmental Cell. Cell Press, 3(2), pp. 291–297. doi: 10.1016/S1534-5807(02)00205-8.

Junk, W. J. et al. (2006) 'Biodiversity and its conservation in the Pantanal of Mato Grosso, Brazil', Aquatic Sciences. Birkhauser Verlag AG, 68(3), pp. 278–309. doi: 10.1007/S00027-006-0851-4/METRICS.

Kang, J. et al. (2011) 'Plant ABC Transporters', The Arabidopsis Book / American Society of Plant Biologists. American Society of Plant Biologists, 9, p. e0153. doi: 10.1199/TAB.0153.

Kato-Noguchi, H. (2023) 'Defensive Molecules Momilactones A and B: Function, Biosynthesis, Induction and Occurrence', Toxins 2023, Vol. 15, Page 241. Multidisciplinary Digital Publishing Institute, 15(4), p. 241. doi: 10.3390/TOXINS15040241.

Khoso, M. A. et al. (2022) 'WRKY transcription factors (TFs): Molecular switches to regulate drought, temperature, and salinity stresses in plants', Frontiers in Plant Science. Frontiers Media SA, 13, p. 1039329. doi: 10.3389/FPLS.2022.1039329.

Lee, B. H. et al. (2014) 'The Arabidopsis thaliana GRF-INTERACTING FACTOR gene family plays an essential role in control of male and female reproductive development', Developmental Biology. Academic Press, 386(1), pp. 12–24. doi: 10.1016/J.YDBIO.2013.12.009.

Lentz, D. L. (1990) 'ACROCOMIA MEXICANA: PALM OF THE ANCIENT MESOAMERICANS', J. Ethllobiol, 10(2), pp. 183–194.

Levis, C. et al. (2018) 'How people domesticated Amazonian forests', Frontiers in Ecology and Evolution. Frontiers Media S. A, 5(JAN), p. 299700. doi: 10.3389/FEVO.2017.00171/BIBTEX.

Li, J. et al. (2021) ‘Fatty acid desaturases in the chloroplast and endoplasmic reticulum promote susceptibility to the green peach aphid myzus persicae in arabidopsis thaliana’, Molecular Plant-Microbe Interactions. American Phytopathological Society, 34(6). doi: 10.1094/MPMI-12-20-0345-R/ASSET/IMAGES/LARGE/MPMI-12-20-0345-RF8.JPEG.

de Lima, N. E. et al. (2018) ‘A review of the palm genus Acrocomia: Neotropical green gold’, Organisms Diversity & Evolution. Springer Berlin Heidelberg, pp. 1–11. doi: 10.1007/s13127-018-0362-x.

de Lima, N. E., Meerow, A. W. and Manfrin, M. H. (2020) ‘Genetic structure of two Acrocomia ecotypes (Arecaceae) across Brazilian savannas and seasonally dry forests’, Tree Genetics and Genomes. Springer, 16(4), pp. 1–12. doi: 10.1007/S11295-020-01446-Y/FIGURES/3.

Lin, T. F. et al. (2016) ‘OBE3 and WUS Interaction in Shoot Meristem Stem Cell Regulation’, PloS one. PLoS One, 11(5). doi: 10.1371/JOURNAL.PONE.0155657.

Liu, J. et al. (2013) ‘CAROTENOID CLEAVAGE DIOXYGENASE 7 modulates plant growth, reproduction, senescence, and determinate nodulation in the model legume Lotus japonicus’, Journal of Experimental Botany. Oxford Academic, 64(7), pp. 1967–1981. doi: 10.1093/JXB/ERT056.

Lorenzi, H. et al. (2010) Flora Brasileira Arecaceae (Palmeiras), Instituto Plantarum De Estudos Da Flora Ltda. Nova Odessa: Instituto Plantarum.

Lotterhos, K. E. and Whitlock, M. C. (2015) ‘The relative power of genome scans to detect local adaptation depends on sampling design and statistical method’, Molecular Ecology. John Wiley & Sons, Ltd, 24(5), pp. 1031–1046. doi: 10.1111/MEC.13100.

Lu, Y. et al. (2020) ‘Coordination between GROWTH-REGULATING FACTOR1 and GRF-INTERACTING FACTOR1 plays a key role in regulating leaf growth in rice’, BMC Plant Biology, 20(1). doi: 10.1186/s12870-020-02417-0.

Luu, K., Bazin, E. and Blum, M. G. B. (2017) ‘pcadapt: an R package to perform genome scans for selection based on principal component analysis’, in Molecular Ecology Resources. doi: 10.1111/1755-0998.12592.

Manel, S. et al. (2018) ‘Predicting genotype environmental range from genome–environment associations’, Molecular Ecology. John Wiley & Sons, Ltd, 27(13), pp. 2823–2833. doi: 10.1111/MEC.14723.

Martínez-Andújar, C. et al. (2011) ‘Induction of 9-cis-epoxycarotenoid dioxygenase in Arabidopsis thaliana seeds enhances seed dormancy’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 108(41), pp. 17225–17229. doi: 10.1073/PNAS.1112151108/SUPPL_FILE/PNAS.201112151SI.PDF.

Melo, W. A. et al. (2020) ‘Selective Sweeps Lead to Evolutionary Success in an Amazonian Hyperdominant Palm’, Frontiers in Genetics. Frontiers Media S.A., 11. doi: 10.3389/FGENE.2020.596662/FULL.

Mishra, S., Upadhyay, S. and Shukla, R. K. (2017) ‘The role of strigolactones and their potential cross-talk under hostile ecological conditions in plants’, Frontiers in Physiology. Frontiers Media S.A., 7(JAN), p. 691. doi: 10.3389/FPHYS.2016.00691/BIBTEX.

Morcote-Ríos, G. and Bernal, R. (2001) ‘Remains of palms (palmae) at archaeological sites in the new world: A review’, Botanical Review. New York Botanical Garden Press, 67(3), pp. 309–350. doi: 10.1007/BF02858098/METRICS.

Motoike, S. Y. et al. (2013) A cultura da macaúba: implantação e manejo de cultivos racionais. 1st edn. Viçosa, Minas Gerais, Brazil: Universidade Federal de Viçosa.

Nascimento, A. D. P. Do et al. (2016) ‘Extraction of Acrocomia intumescens Drude oil with supercritical carbon dioxide: Process modeling and comparison with organic solvent extractions’, The Journal of Supercritical Fluids. Elsevier, 111, pp. 1–7. doi: 10.1016/J.SUPFLU.2016.01.013.

Nascimento, L. F. do et al. (2020) ‘Associated evolution of fruit size, fruit colour and spines in Neotropical palms’, Journal of Evolutionary Biology. Oxford Academic, 33(6), pp. 858–868. doi: 10.1111/JEB.13619.

Noack, L. C. et al. (2022) ‘A nanodomain-anchored scaffolding complex is required for the function and localization of phosphatidylinositol 4-kinase alpha in plants’, The Plant Cell. Oxford Academic, 34(1), pp. 302–332. doi: 10.1093/PLCELL/KOAB135.

Pankin, A. et al. (2018) ‘Targeted resequencing reveals genomic signatures of barley domestication’, New Phytologist. John Wiley & Sons, Ltd, 218(3), pp. 1247–1259. doi: 10.1111/NPH.15077.

Perrot, T., Pauly, M. and Ramírez, V. (2022) ‘Emerging Roles of β-Glucanases in Plant Development and Adaptative Responses’, Plants 2022, Vol. 11, Page 1119.

Multidisciplinary Digital Publishing Institute, 11(9), p. 1119. doi: 10.3390/PLANTS11091119.

Petereit, J. et al. (2022) ‘Pangenomics and Crop Genome Adaptation in a Changing Climate’, Plants. Multidisciplinary Digital Publishing Institute (MDPI), 11(15). doi: 10.3390/PLANTS11151949.

Poland, J. A. et al. (2012) ‘Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach’, PLoS ONE. Edited by T. Yin. Public Library of Science, 7(2), p. e32253. doi: 10.1371/journal.pone.0032253.

R Core Team (2016) ‘R Development Core Team’, R: A Language and Environment for Statistical Computing. doi: http://www.R-project.org.

Ramírez-Núñez, C. et al. (2019) ‘MULTIDIRECTIONAL INTERPOLATION of LIDAR DATA from SOUTHERN VERACRUZ, MEXICO: IMPLICATIONS for EARLY OLMEC SUBSISTENCE’, Ancient Mesoamerica. Cambridge University Press. doi: 10.1017/S0956536118000263.

Resende, R. T. et al. (2020) ‘Data-based agroecological zoning of Acrocomia aculeata: GIS modeling and ecophysiological aspects into a Brazilian representative occurrence area’, Industrial Crops and Products. Elsevier, 154, p. 112749. doi: 10.1016/J.INDCROP.2020.112749.

Rodríguez, M. F. and Aschero, C. A. (2005) ‘Acrocomia chunta (Arecaceae) raw material for cord making in the Argentinean Puna’, Journal of Archaeological Science. Academic Press, 32(10), pp. 1534–1542. doi: 10.1016/J.JAS.2005.04.016.

Rupwate, S. D. and Rajasekharan, R. (2012) ‘Plant phosphoinositide-specific phospholipase C An insight’, Plant Signaling & Behavior, 7(10), pp. 1281–1283. doi: 10.4161/psb.21436.

Saeed, W., Naseem, S. and Ali, Z. (2017) ‘Strigolactones biosynthesis and their role in abiotic stress resilience in plants: A critical review’, Frontiers in Plant Science. Frontiers Media S.A. doi: 10.3389/fpls.2017.01487.

Saiga, S. et al. (2012) ‘Control of embryonic meristem initiation in Arabidopsis by PHD-finger protein complexes’, Development (Cambridge, England). Development, 139(8), pp. 1391–1398. doi: 10.1242/DEV.074492.

Salgado, F. F. et al. (2022) ‘The early response of oil palm (Elaeis guineensis Jacq.)

plants to water deprivation: Expression analysis of miRNAs and their putative target genes, and similarities with the response to salinity stress', Frontiers in Plant Science. Frontiers Media S.A., 13, p. 970113. doi: 10.3389/FPLS.2022.970113/BIBTEX.

Sanjinez-Argandoña, E. J. and Chuba, C. A. M. (2011) 'Caracterização biométrica, física e química de frutos da palmeira bocaiuva Acrocomia aculeata (Jacq) Lodd', Revista Brasileira de Fruticultura. Sociedade Brasileira de Fruticultura, 33(3), pp. 1023–1028. doi: 10.1590/S0100-29452011000300040.

Schreiber, M., Stein, N. and Mascher, M. (2018) 'Genomic approaches for studying crop evolution', Genome Biology 2018 19:1. BioMed Central, 19(1), pp. 1–15. doi: 10.1186/S13059-018-1528-8.

Schwartz, S. H., Qin, X. and Loewen, M. C. (2004) 'The biochemical characterization of two carotenoid cleavage enzymes from Arabidopsis indicates that a carotenoid-derived compound inhibits lateral branching', The Journal of biological chemistry. J Biol Chem, 279(45), pp. 46940–46945. doi: 10.1074/JBC.M409004200.

Shen, G. et al. (2022) 'Plant GDSL Esterases/Lipases: Evolutionary, Physiological and Molecular Functions in Plant Development', Plants 2022, Vol. 11, Page 468. Multidisciplinary Digital Publishing Institute, 11(4), p. 468. doi: 10.3390/PLANTS11040468.

Shimura, K. et al. (2007) 'Identification of a biosynthetic gene cluster in rice for momilactones', The Journal of biological chemistry. J Biol Chem, 282(47), pp. 34013–34018. doi: 10.1074/JBC.M703344200.

da Silva, J. M. C. et al. (2018) 'The Caatinga: Understanding the challenges', Caatinga: The Largest Tropical Dry Forest Region in South America. Springer International Publishing, pp. 3–19. doi: 10.1007/978-3-319-68339-3_1/COVER.

Sorefan, K. et al. (2003) 'MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea', Genes & development. Genes Dev, 17(12), pp. 1469–1474. doi: 10.1101/GAD.256603.

Suraninpong, P. et al. (2023) 'Monitoring Drought Tolerance in Oil Palm: Choline Monooxygenase as a Novel Molecular Marker', Plants. Multidisciplinary Digital Publishing Institute (MDPI), 12(17), p. 3089. doi: 10.3390/PLANTS12173089/S1.

Takáč, T., Novák, D. and Šamaj, J. (2019) 'Recent advances in the cellular and developmental biology of phospholipases in plants', Frontiers in Plant Science. Frontiers Media

S.A., 10, p. 432553. doi: 10.3389/FPLS.2019.00362/BIBTEX.

Ting, N. C. et al. (2016) ‘Fine-mapping and cross-validation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm’, BMC Genomics. BioMed Central Ltd., 17(1), pp. 1–17. doi: 10.1186/S12864-016-2607-4/FIGURES/5.

Uchôa, L. R. et al. (2022) ‘Lizards (Reptilia: Squamata) from the Caatinga, northeastern Brazil: Detailed and updated overview’, Vertebrate Zoology 72: 599-659. Senckenberg Gesellschaft für Naturforschung, 72, pp. 599–659. doi: 10.3897/VZ.72.E78828.

Vargas-Carpintero, R. et al. (2021) ‘Acrocomia spp.: neglected crop, ballyhooed multipurpose palm or fit for the bioeconomy? A review’, Agronomy for Sustainable Development 2021 41:6. Springer, 41(6), pp. 1–35. doi: 10.1007/S13593-021-00729-5.

Vargas-Carpintero, R. et al. (2022) ‘A Collaborative, Systems Approach for the Development of Biomass-Based Value Webs: The Case of the Acrocomia Palm’, Land 2022, Vol. 11, Page 1748. Multidisciplinary Digital Publishing Institute, 11(10), p. 1748. doi: 10.3390/LAND11101748.

Vianna, Suelen Alves et al. (2017) ‘Biometric Characterization of Fruits and Morphoanatomy of the Mesocarp of Acrocomia Species (Arecaceae)’, International Journal of Biology. doi: 10.5539/ijb.v9n3p78.

Vianna, Suelen A et al. (2017) ‘Leaf anatomy of Acrocomia (Arecaceae): an additional contribution to the taxonomic resolution of a genus with great economic potential’, Plant Systematics and Evolution, 303(2), pp. 233–248. doi: 10.1007/s00606-016-1369-4.

Vianna, S. A. et al. (2021) ‘Morphological characterization and productivity estimates of Acrocomia totai Mart. (Arecaceae) – a sustainable alternative of extractivism and cultivation’, Revista Brasileira de Fruticultura. Sociedade Brasileira de Fruticultura, 43(1), p. e-730. doi: 10.1590/0100-29452021730.

Villavicencio, N. A. et al. (2016) ‘Combination of humans, climate, and vegetation change triggered Late Quaternary megafauna extinction in the Última Esperanza region, southern Patagonia, Chile’, Ecography. John Wiley & Sons, Ltd, 39(2), pp. 125–140. doi: 10.1111/ECOG.01606.

Walther, G. R. et al. (2007) ‘Palms tracking climate change’, Global Ecology and Biogeography, 16(6), pp. 801–809. doi: 10.1111/J.1466-8238.2007.00328.X.

Wang, L. et al. (2020) ‘Genes, pathways and networks responding to drought stress in oil palm roots’, Scientific Reports 2020 10:1. Nature Publishing Group, 10(1), pp. 1–13. doi: 10.1038/s41598-020-78297-z.

Wu, G. et al. (2015) ‘ENHANCED DISEASE RESISTANCE4 Associates with CLATHRIN HEAVY CHAIN2 and Modulates Plant Immunity by Regulating Relocation of EDR1 in Arabidopsis’, The Plant Cell. Oxford Academic, 27(3), pp. 857–873. doi: 10.1105/TPC.114.134668.

Xuan, T. D. et al. (2016) ‘Allelopathic momilactones A and B are implied in rice drought and salinity tolerance, not weed resistance’, Agronomy for Sustainable Development. Springer-Verlag France, 36(3), pp. 1–8. doi: 10.1007/S13593-016-0383-9/TABLES/2.

Yao, Y. et al. (2022) ‘Evolutionary Origin of the Carotenoid Cleavage Oxygenase Family in Plants and Expression of Pepper Genes in Response to Abiotic Stresses’, Frontiers in Plant Science. Frontiers Media S.A., 12, p. 3169. doi: 10.3389/FPLS.2021.792832/BIBTEX.

Zhang, J., Zhang, J. and Peters, R. J. (2020) ‘Why are momilactones always associated with biosynthetic gene clusters in plants?’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 117(25), pp. 13867–13869. doi: 10.1073/PNAS.2007934117/ASSET/59E73FFE-F483-4549-9CE1-0D755FE833E3/ASSETS/GRAPHIC/PNAS.2007934117FIG01.JPEG.

Zhang, L. and Jim enez-G omez, J. M. (2020) ‘Functional analysis of FRIGIDA using naturally occurring variation in Arabidopsis thaliana’. doi: 10.1111/tpj.14716.

5. Capitulo II

Population structure of *Acrocomia aculeata* in Central America: Genomic diversity and biogeographic patterns of an open canopy palm in the Isthmus

Key words: Arecaceae, GBS, Neotropic, SNP, Population genetics, genomic

Abstract

Focusing in the North American gene pool, Central America has a diverse range of plant species, including the neotropical palm *Acrocomia aculeata* (also known as Coyol or Macaúba). This work presents the genomic diversity and biogeographic patterns of *A. aculeata* in isthmus. We use a double-digest genotyping by sequencing (GBS) to discover single nucleotide polymorphisms (SNPs). A total of 259 samples were collected in Guatemala, Honduras, Nicaragua, Costa Rica, and Panama. Our results showed three primary genomic groups: Mesoamerican, Costa Rican, and Panamanian, each with separate subpopulations with differing degrees of genetic diversity. The Mesoamerican group had the highest diversity and distinctive genetic variation, possibly reflecting past migration routes from South America. Biogeographic barriers, like the Nicaraguan depression and the Talamanca mountain chain, influenced genetic diversity. Ecological niche modeling highlighted the suitability of Central American lowlands for A. aculeata. Overall, this study sheds light on *A. aculeata* evolutionary history and population structure, with a focus on South America as a source region and the significance of biogeographic barriers in shaping genetic divergence. These findings are critical for conservation and breeding techniques aiming at sustaining *A. aculeata* populations' genetic diversity and resilience in the face of environmental changes and human-caused consequences.

Introduction

Central America is one of the world's five plant diversity hotspots. Recent research has revealed insights into the evolutionary history and community formation of palm species in the isthmus. The biodiversity patterns in Central America respond to the geological connection processes between North America and South America. Palms are adequate indicators of diversification processes, research indicates that many palm species have experienced local diversification rather than moving from neighboring regions in the region (Cano *et al.*, 2022). *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. (known as Coyol or Macaúba), is an open-canopy palm widespread across Central and South America, and is a great plant model for understanding the dynamics of the Bactridinae palm subtribe irradiation across the continent.

Acrocomia aculeata is distinguished by its single, spiny trunk (stipe), persisting fragments of fallen leaf sheaths, and a yellow, coccoid drupe high in oil (Fig 1) (de Lima *et al.*, 2018). Its distribution extends from the subtropical to tropical regions of northern Mexico, to Central America (Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, and Panama), and South America (de Lima *et al.*, 2018). It is found in six of the seven hypothesized Neotropical palm biogeographic groups, mostly at transitions between macro ecoregions such as tropical rainforests (wet broadleaf forest) and drier ecosystems such as savannas, grasslands, dry broadleaf forests, and xeric shrublands. *A. aculeata* is divided into two major gene pools: a North American group and a South American group (Díaz *et al.*, 2021). Currently, pre-breeding programs focus on South American genotypes, leaving a considerable knowledge gap regarding species in North America, particularly in Central America. Additionally, the region has been historically neglected in scientific research on biodiversity topics and biological records, making research and development challenging (Morales-Marroquín et al. 2022).

Central America's complex geological history, highlighted by land emergence and tectonic movements have influence *A. aculeata* evolutionary trajectory. The interaction of colonization routes from South America and northward migrations during periods of land exposure is thought to contribute to the genetic diversity observed in Central American species (Gutiérrez-García and Vázquez-Domínguez, 2013). During the Miocene to Early Pliocene (24-5 Ma), land exposure resulted in the formation of most islands, and land masses conforming the isthmus. By the Late Miocene (9 Ma), the first links between North and South America were established (Bagley and Johnson, 2014). Central America underwent many faults, resulting in four major tectonic elements: Maya, Chortis, Chorotega, and Chocó blocks (Fig 2). Geologic events in the region resulted in similar genetic differentiation patterns within each tectonic block. The upper part of Central America was characterized by genetic structure and divergence, with numerous refugia and the appearance of cryptic species. Mid-Central America is distinguished by large population differentiation and volcanic activity, whereas the Panamian region is distinguished by north-south migration, which results in significant species variety and speciation (Gutiérrez-García and Vázquez-Domínguez, 2013).

Furthermore, interactions with pre-Columbian civilizations, such as landscape modification, have influenced the distribution and genetic structure of native crops in the region. Amerindians utilized *Acrocomia* both ritually and as a food source (Clement, 1999a; Morcote-Ríos and Bernal, 2001). The earliest archaeological evidence of its use in North America dates back to the Pre-Olmec and Olmec civilizations in Veracruz, southern Mexico (approximately 4000 BP) (Ramírez-Núñez et al. 2019). *Acrocomia aculeata* is considered an incipiently domesticated crop (Clement *et al.*, 2021).

Genetic and phenotypic variations observed in wild populations of *Acrocomia aculeata* underscore its adaptability to environmental gradients, further highlighting its significance as a target species for genetic studies. Ecologically, *Acrocomia aculeata* thrives in dry, open canopy forest areas with seasonal fires, demonstrating its resilience to diverse ecological niches across the Neotropics (de Lima *et al.*, 2018). Its presence in transition zones between habitats and its association with anthropogenic landscapes suggest promising potential for oil production and the bioeconomy (Vargas-Carpintero *et al.*, 2021). The fruit pulp oil extraction yielded 65% w/w from populations in Minas Gerais, Brazil (Colombo *et al.*, 2018) and 53.6% from Costa Rican populations (Alfaro-Solís *et al.*, 2020).

This study assessed genome-wide diversity in natural populations of *Acrocomia aculeata* across Central America. We used double-digest genotyping by sequencing (GBS) to identify single nucleotide polymorphisms (SNPs). Our analysis describes the genomic population structures for understanding the palm's natural history in the isthmus. This research aims to generate novel insights into the evolution of the *Acrocomia aculeata* North American gene pool and contribute to the domestication and breeding of this promising tropical palm for sustainable vegetable oil production, the bioeconomy, and biofuels.


Adult A. aculeata palms in different ecoregions
A
B
C
D
Adult dwarf phenotypes
E
F
G
H
Trunk morphology
I
J
K
L
A. aculeata in milpa agroecosystem
M
N
Bunch and leaf morphology
O
P
A. aculeata natural populations
Q
R

Fig 1. Characteristics of adult plants of *Acrocomia aculeata* in Central America. A – D: Adult palms in different ecoregions. E – H: Adult dwarf phenotypes. I – L: Truck morphology. M – N: *A. aculeata* in milpa agroecosystems. O – P: Bunch and leaf morphology. Q – R: *A. aculeata* natural populations.

Materials and methods

Plant material and DNA isolation

We sampled leaves from natural populations across Central America (Guatemala, Honduras, Nicaragua, Costa Rica and, Panama) for a total of 259 samples (Table S1). The collections of these samples were registered in each country the National Council of Biodiversity Patrimony. The leaves were dehydrated using silica gel and stored in paper bags at -20 °C. Following Doyle and Doyle's protocol, we extracted whole genomic DNA from 50 mg leaf samples. Agarose gel electrophoresis (1% w/v) with GelRed stain (Sigma-Aldrich) was used to assess DNA quality and integrity. We quantified and normalized DNA concentrations to 30 ng/µL using the dsDNA BR Assay quantification kit for the Qubit3 fluorometer (Invitrogen).

GBS libraries and SNP discovery

Genomic libraries were prepared following the protocol of genotyping-by-sequencing using two restriction enzymes (ddGBS) as described by Poland *et al.* (Poland *et al.*, 2012). The combination of *MseI* and *PstI* (New England Biolabs) was used for the genomic library construction. The ddGBS library was quantified through RT-PCR on the CFX 384 Touch Real Time PCR (BioRad) equipment using a KAPA Library Quantification kit (KAPA Biosystems, cat. KK4824), and the fragments' profiles were inspected using the Agilent DNA 1000 Kit on a 2100 Bioanalyzer (Agilent Technologies). The library was sequenced in an Illumina NextSeq2000 using the Illumina NextSeq1000/2000 kit, with single-end and 150 bp configurations.

The general sequencing quality and the presence of adapters were assessed with FASTQC (Andrews, 2010). The SNP discovery was performed separately for each species following the *de novo* pipeline of the program Stacks v.1.42 (Catchen *et al.*, 2013) with similar filtering criteria. Due to the presence of adapters, sequences were trimmed to 90 bp. Trimming, quality control (removal of sequences with uncalled bases and with Phred scores <10), and demultiplexing were performed with the module *process_radtags*. For each sample, loci were assembled using the module *ustacks* considering minimum sequencing depth (-m) of 3, and distance between reads from the same loci (-N) of 2. A catalog of loci across samples was built with the module *cstacks*, considering the distance between locus (-n) of 2, and loci of the samples were compared to the catalog using *sstacks*. The loci with lower probabilities (-lnl_lim 10) were discarded using the module *rxstacks*. For the three species, candidate SNPs were identified using the module *populations* considering a minimum depth of 3, minor allele frequency of 0.01, the presence of SNP in at least 75% of the samples in each the sampling locations. To avoid explicit linkage only a single SNP was retained per GBS locus. Quality metrics (mean sequencing depth per locus and per sample, percentage of missing data per sample) were assessed with VCFTools (Danecek *et al.*, 2011). Additionally, samples with more than 50% of missing genotypes were removed from the final data set, resulting in 259 samples for *A.aculeata*. We perform a data set imputation in TASSEL using LinkImpute (LD kNNi) imputation algorithm (Glaubitz *et al.*, 2014).

Populations genomic diversity and structure analyses

We used both putatively neutral and under-selection single nucleotide polymorphisms (SNPs) for the analysis of genomic diversity. The hierfstat package (Goudet, 2005) was employed in the R 4.1.0 platform (R Core Team, 2016) to calculate values for H_0 observed heterozygosity, H_E expected heterozygosity, f inbreeding coefficient, and their respective ranges, using 1000 bootstraps. Additionally, we utilized the adegenet 2.1.1 package (Jombart and Ahmed, 2011) and the poppr package (Kamvar, Tabima and Grünwald, 2014) in the R 4.2.1 platform (R Core Team 2016) to determine *A* allele number, *AR* allelic richness, *Ap* private alleles.

Population structure analyses were also conducted using both, neutral and under selection loci. The adegenet v. 2.1.6 package (Jombart, 2022) within the R 4.2.1 platform was utilized to perform the Discriminant Analysis of Principal Components – DAPC (Jombart and Ahmed, 2011), in addition to the methods mentioned earlier (R Core Team, 2016). Genetic groups based on Central America biogeographic zones were considered during the DAPC analysis, and the optimal number of retained principal components was determined using the α-score. This technique allowed for the generation of scatter/density plots illustrating the distribution of populations. The hierfstat software (Goudet, 2005) was employed to compute Wright's F-statistics (Fixation index F_{ST}, F_{IT}, and F_{IS}), also for the pairwise F_{ST} values.

Heatmaps illustrating the pairwise F_{ST} results were generated using the heatmaply package (Galili *et al.*, 2018) in R version 4.2.1 (R Core Team, 2016).

Ecological niche modeling of *Acrocomia aculeata* in Central America

The distribution and occurrence data for *Acrocomia aculeata* was extracted from the following databases: GBIF (https://www.gbif.org/), the Missouri Botanical Garden (https://www.missouribotanicalgarden.org/)), and the New York Botanical Garden (https://www.nybg.org/), using the BIEN package (Maitner et al., 2018) for the R 4.1.0 platform (R Core Team, 2016).

After obtaining the data, we initially filtered the coordinates using the CoordinateCleaner package (Zizka *et al.*, 2019) on the R platform version 4.1.0. Subsequently, we added an occurrence point corresponding to each natural population site surveyed in this study (Table S1). Given *Acrocomia aculeata*'s extensive distribution across the continent, only coordinates from Central America were used. Finally, we employed a minimum of five occurrences for modeling, and spatial rarefaction was applied at twice the model cell resolution using the ENMTML package (Andrade, Velazco and De Marco Júnior, 2020) on the R 4.1.0 platform.

We employed the same 19 bioclimatic variables with a resolution of 2.5 arcminutes from the WorldClim database during the pre-processing stage of the models. To reduce collinearity between variables, Principal Component Analysis (PCA) was utilized, resulting in 19 principal components (PCs) and a correlation matrix. We selected PCs that collectively explained at least 95% of the total variance, generating a new set of variables derived from PCA (Destro *et al.*, 2020), also created using the ENMTML package. Subsequently, utilizing the "BUFFER" method within the ENMTML software, the model was fitted to a region defined by the polygon formed by the occurrence points. Next, pseudo-absences were identified using the "env_const" method. These pseudo-absences are ecologically constrained to areas with reduced fitness values predicted by a Bioclim model (Booth *et al.*, 2014). Thirty percent of the data was reserved for testing, while seventy percent was utilized for training to enhance model accuracy. Additionally, we employed a bootstrap approach, repeating the process ten times.

Several algorithms were employed during the processing step to predict the species' present suitability. Four algorithms were used to create the ecological niche models among these three strategies: (1) Bioclim (Booth *et al.*, 2014), which considered presence only; (2) Random Forest (Breiman, 2001); (3) Support Vector Machine (SVM) (Guo et al., 2005), which considered both presence and absence algorithms; and (4) MaxEnt (Phillips & Dudík, 2008), which considered presence and background algorithms. The models' quality was assessed using an analysis of the True Skill Statistics (TSS) (Allouche et al., 2006) and Area Under the Curve (AUC) (Metz, 1986) metrics.

Based on the suitability models, ensembles (Araújo and New, 2006) were developed for the post-processing stage. This method can help with better planning and enables more robust decision making based on the responses from the models (Araújo and New 2006). The average between the models was examined using the MEAN technique.

Results

SNP discovery and genomic diversity in *A. aculeata* in Central America

The sequencing of the ddGBS library for *A. aculeata* generated a total of 746,344,923 reads. After quality-control filtering, we identified 1523 SNPs. The mean depth per-sample was: 20.12X (SD ± 13.39X, 0.04% of missing data after imputation). The population genomic analysis of *Acrocomia aculeata* showed genetic diversity patterns and population structure across Central America. Three major genomic groups were identified: Mesoamerican, Costa Rican, and Panamanian (Fig 2). Within these groups, distinct subpopulations exhibited varying levels of genetic differentiation and diversity (Table 1). Central American *A. aculeata* populations exhibited moderate to low levels of heterozygosity, indicating recent historical demographic events, genetic drift, and potentially limited gene flow due to regional conservation status and predatory practices in land use.

The Mesoamerican genomic group included individuals from Guatemala to the west region of Nicaragua (Fig 2). Within this group, regional subpopulations exhibited subtle genetic differences. For instance, individuals from the Guatemalan displayed lower observed heterozygosity (H_0) compared to other Mesoamerican subpopulations, indicating potential genetic isolation or drift in this region. In contrast, individuals from Nicaraguan West (NCW) exhibited relatively high levels of genetic diversity, as evidenced by higher Ho, lower *f* coefficient, and allelic richness (*AR*), suggesting historical connectivity or admixture. Furthermore, Guatemalan South (GTS) have more private alleles than all other populations in Central America. The highest value of private alleles within the major genomic groups was from the Mesoamerican (*Ap* = 2062) suggesting unique genetic variation within this population and more genomic diversity. The fixation index (*f* = 0.08538), indicated moderate genetic differentiation.

The Costa Rican genomic group including populations of *Acrocomia aculeata* from Nicaraguan East (NCE) to Costa Rican South (CRS) showed moderate levels of genetic diversity and differentiation. Subpopulations from different regions of Costa Rica exhibited varying degrees of genetic differentiation, with individuals from the Nicoya Peninsula (CRNY) showing signs of genetic isolation or reduced gene flow compared to other Costa Rican subpopulations. The Costa Rican group has the lowest number of private alleles (*Ap* = 397). The fixation index (*f* = 0.094699) was slightly higher indicating genetic differentiation compared to the Mesoamerican and the Panamanian groups.

The Panamanian genomic group exhibited relatively lower levels of genetic diversity compared to Mesoamerican and Costa Rican groups. The subpopulations from Osa Peninsula (CROS – Costa Rica) to Panama East (PNE) displayed varying levels of genetic differentiation, with individuals from the Osa Peninsula (CROS) exhibiting the lowest *Ho* and *AR*. Individuals from Panamanian East (PNE) displayed higher genetic diversity, suggesting potential historical admixture or gene flow within the South American genepool.

Table 1. Genetic diversity statistics of *Acrocomia aculeata* populations from 16 Central American locations. Parameters estimated from ddGBS data (259 individuals and 1523 SNPs). H_0 observed heterozygosity, H_E expected heterozygosity, *A* allele number, *AR* allelic richness, *Ap* private alleles, *f* inbreeding coefficient (Cockerham and Weir, 1993). The complete list of samples and their description can be found in Table S1.

	Major Genomic groups	***N***	H_0	H_E	***A***	***AR***	***Ap***	***f***
	Mesoamerican	94	0.047058	0.048696	2154	1.205316	2062	0.08538
	Costa Rican	70	0.043828	0.039826	1979	1.167094	397	0.094699
	Panamanian	95	0.032116	0.02544	2026	1.155196	607	0.088576
	Sub-regions							
Mesoamerican	Guatemalan North – GTN	12	0.041835	0.032593	1590	1.033074	38	-0.13111
	Guatemalan South – GTS	22	0.041701	0.031596	1747	1.031831	109	-0.11676
	Honduran West – HNW	5	0.047268	0.039236	1675	1.040129	19	-0.16178
	Honduran Center – HNC	19	0.046811	0.04185	1791	1.041988	77	-0.00056
	Honduran North – HNN	12	0.044155	0.037156	1734	1.037462	68	-0.04707
	Nicaraguan West – NCW	24	0.057246	0.050121	1809	1.050272	71	-0.0334
Costa Rican	Nicaraguan East – NCE	10	0.039073	0.041978	1721	1.041828	41	0.212433
	Costa Rican North West – CRNW	27	0.04471	0.037747	1814	1.037881	99	0.027469
	Nicoya Peninsula – CRNY	7	0.039878	0.032479	1670	1.033047	18	-0.09954
	Costa Rican North – CRN	13	0.049796	0.040672	1693	1.041039	25	-0.13439
	Costa Rican Center – CRC	10	0.041577	0.032814	1695	1.033288	46	-0.12769
	Costa Rica South – CRS	3	0.041803	0.030515	1618	1.032436	6	-0.41578
Panamanian	Osa Peninsula – CROS	48	0.028257	0.019821	1779	1.019912	90	0.022495
	Panamanian West – PNW	11	0.034333	0.025256	1607	1.025765	19	-0.0906
	Panamanian Center – PNC	23	0.032026	0.023074	1657	1.02328	69	0.043116
	Panamanian East – PNE	13	0.052754	0.048793	1685	1.048971	101	-0.00409

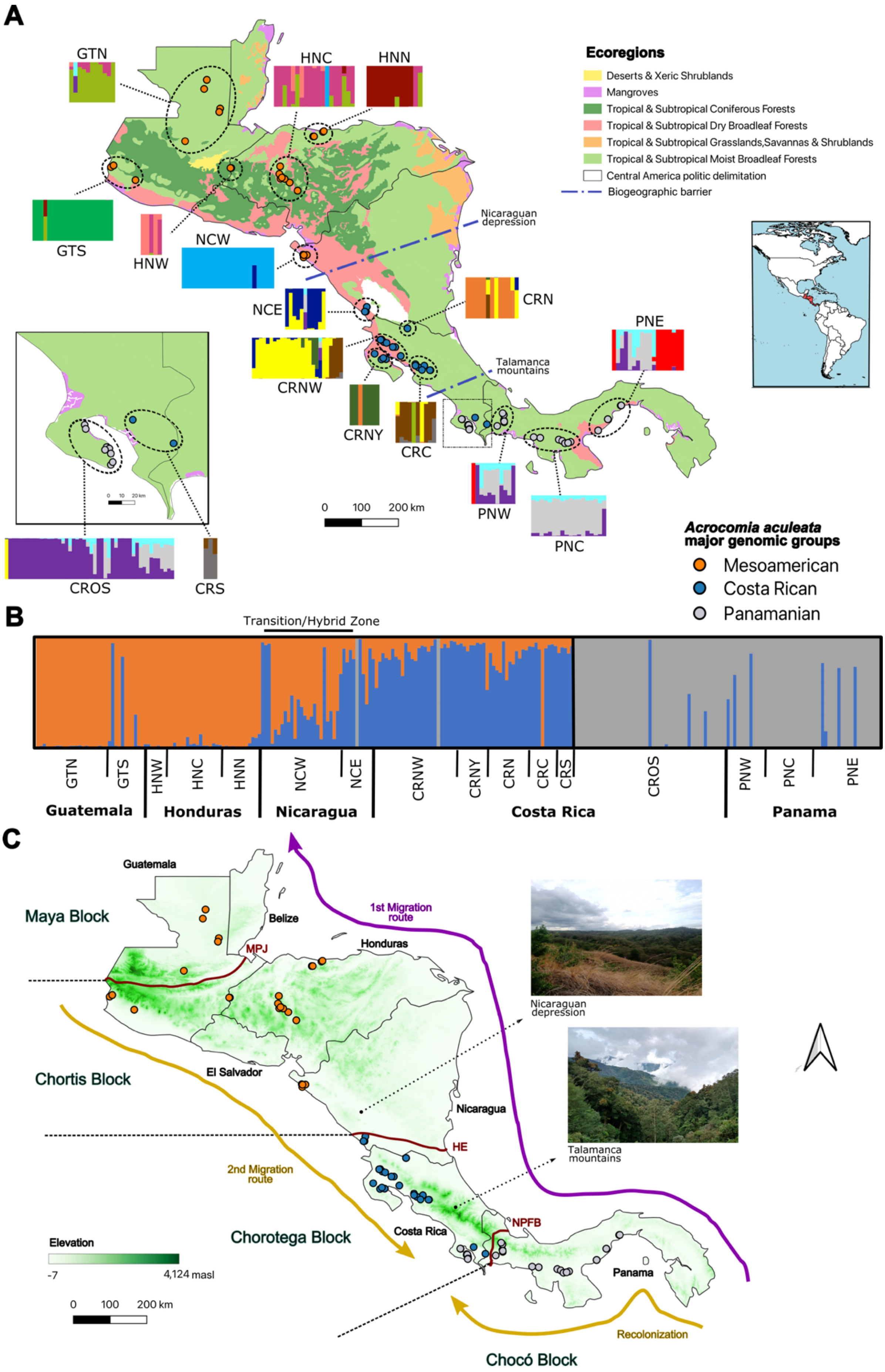
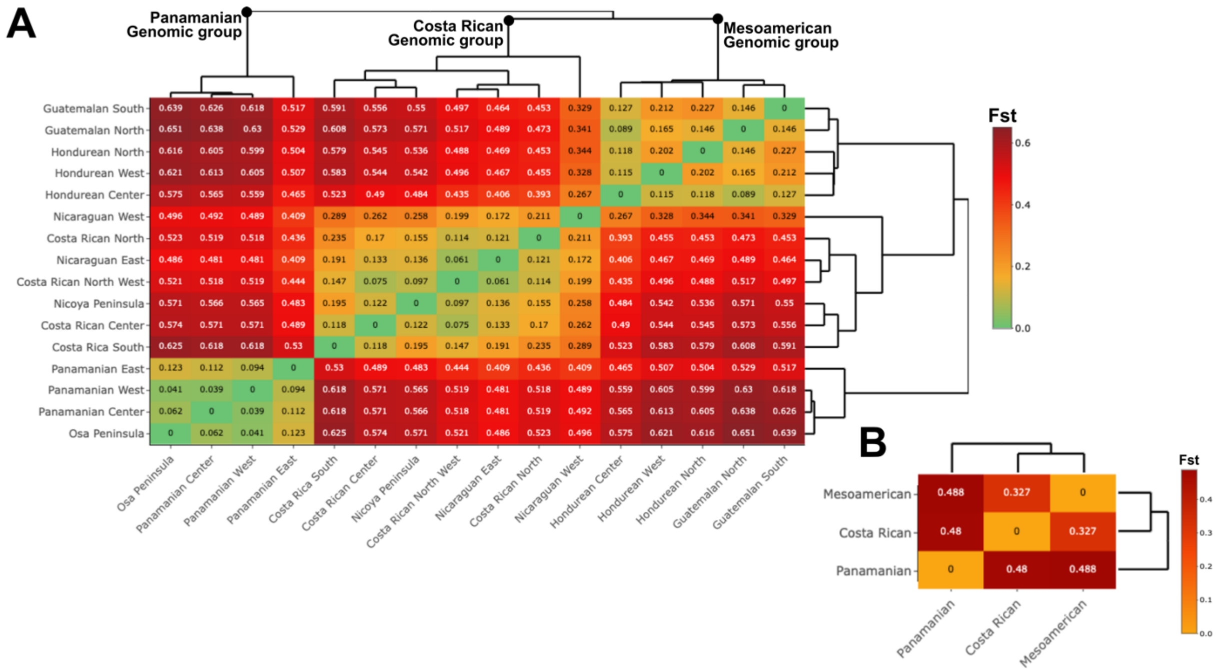

A
GTN
HNC
HNN
GTS
HNW
NCW
NCE
CRN
CRNW
CRNY
CRC
PNE
PNW
PNC
CROS
CRS
Nicaraguan depression
Talamanca mountains
0 100 200 km
Ecoregions
Deserts & Xeric Shrublands
Mangroves
Tropical & Subtropical Coniferous Forests
Tropical & Subtropical Dry Broadleaf Forests
Tropical & Subtropical Grasslands,Savannas & Shrublands
Tropical & Subtropical Moist Broadleaf Forests
Central America politic delimitation
Biogeographic barrier
Acrocomia aculeata major genomic groups
Mesoamerican
Costa Rican
Panamanian
B
Transition/Hybrid Zone
GTN
GTS
HNW
HNC
HNN
NCW
NCE
CRNW
CRNY
CRN
CRC
CRS
CROS
PNW
PNC
PNE
Guatemala
Honduras
Nicaragua
Costa Rica
Panama
C
Guatemala
Maya Block
Belize
MPJ
Honduras
1st Migration route
Chortis Block
El Salvador
Nicaragua
HE
2nd Migration route
Chorotega Block
Costa Rica
NPFB
Panama
Recolonization
Chocó Block
Nicaraguan depression
Talamanca mountains
Elevation
-7
4,124 masl
0 100 200 km

Fig 2. Map of sampling locations for *Acrocomia aculeata* along with major genetic divergences among Central America suggested by DAPC. A. Central America Map showing ecoregions, each bar plot in every subpopulation show the genetic clusters represented by different colors. The color of each population point corresponds to the major genomic group. **B.** Bar plot showing the 3 major genetic clusters within subpopulations in Central America based on DAPC. Acronyms follow Table 1 Sub-regions. **C.** Tectonic blocks of Central America along with biogeographic barriers for *Acrocomia aculeata*. Tectonic limits: MPJ: Motagua–Polochic–Jocotán fault system. HE: Hess escarpment. NPFB: North Panama fracture belt.

Population structuration of *Acrocomia aculeata* in Central America

The pairwise F_{ST} values provide insights into the genetic differentiation among *Acrocomia aculeata* populations across Central America. For instance, populations within the same genomic group generally exhibit lower F_{ST} values compared to those between different one. This suggests that isolation by distance and biogeographic barriers plays a significant role in shaping the genetic structure of *A. aculeata* Central American populations. Three major genomic groups were identified: Mesoamerican, Costa Rican, and Panamanian (Fig 3). For example, the F_{ST} values between populations within the Mesoamerican genomic group, such as Honduran North and Honduran West, tend to be relatively lower (0.202), indicating a moderate level of genetic similarity. Conversely, populations separated by larger geographic distances tend to exhibit higher F_{ST} values, indicating greater genetic differentiation. Moreover, some populations exhibit notable levels of genetic differentiation despite being within the same genomic group. The F_{ST} values between Nicaraguan East and Costa Rican South (0.289) indicate relatively higher genetic differentiation. This could be attributed to a transition zone between the Costa Rican genomic group that is in between of the two others.



	Osa Peninsula	Panamanian Center	Panamanian West	Panamanian East	Costa Rica South	Costa Rican Center	Nicoya Peninsula	Costa Rican North West	Nicaraguan East	Costa Rican North	Nicaraguan West	Hondurean Center	Hondurean West	Hondurean North	Guatemalan North	Guatemalan South
Guatemalan South	0.639	0.626	0.618	0.517	0.591	0.556	0.55	0.497	0.464	0.453	0.329	0.127	0.212	0.227	0.146	0
Guatemalan North	0.651	0.638	0.63	0.529	0.608	0.573	0.571	0.517	0.489	0.473	0.341	0.089	0.165	0.146	0	0.146
Hondurean North	0.616	0.605	0.599	0.504	0.579	0.545	0.536	0.488	0.469	0.453	0.344	0.118	0.202	0	0.146	0.227
Hondurean West	0.621	0.613	0.605	0.507	0.583	0.544	0.542	0.496	0.467	0.455	0.328	0.115	0	0.202	0.165	0.212
Hondurean Center	0.575	0.565	0.559	0.465	0.523	0.49	0.484	0.435	0.406	0.393	0.267	0	0.115	0.118	0.089	0.127
Nicaraguan West	0.496	0.492	0.489	0.409	0.289	0.262	0.258	0.199	0.172	0.211	0	0.267	0.328	0.344	0.341	0.329
Costa Rican North	0.523	0.519	0.518	0.436	0.235	0.17	0.155	0.114	0.121	0	0.211	0.393	0.455	0.453	0.473	0.453
Nicaraguan East	0.486	0.481	0.481	0.409	0.191	0.133	0.136	0.061	0	0.121	0.172	0.406	0.467	0.469	0.489	0.464
Costa Rican North West	0.521	0.518	0.519	0.444	0.147	0.075	0.097	0	0.061	0.114	0.199	0.435	0.496	0.488	0.517	0.497
Nicoya Peninsula	0.571	0.566	0.565	0.483	0.195	0.122	0	0.097	0.136	0.155	0.258	0.484	0.542	0.536	0.571	0.55
Costa Rican Center	0.574	0.571	0.571	0.489	0.118	0	0.122	0.075	0.133	0.17	0.262	0.49	0.544	0.545	0.573	0.556
Costa Rica South	0.625	0.618	0.618	0.53	0	0.118	0.195	0.147	0.191	0.235	0.289	0.523	0.583	0.579	0.608	0.591
Panamanian East	0.123	0.112	0.094	0	0.53	0.489	0.483	0.444	0.409	0.436	0.409	0.465	0.507	0.504	0.529	0.517
Panamanian West	0.041	0.039	0	0.094	0.618	0.571	0.565	0.519	0.481	0.518	0.489	0.559	0.605	0.599	0.63	0.618
Panamanian Center	0.062	0	0.039	0.112	0.618	0.571	0.566	0.518	0.481	0.519	0.492	0.565	0.613	0.605	0.638	0.626
Osa Peninsula	0	0.062	0.041	0.123	0.625	0.574	0.571	0.521	0.486	0.523	0.496	0.575	0.621	0.616	0.651	0.639

	Panamanian	Costa Rican	Mesoamerican
Mesoamerican	0.488	0.327	0
Costa Rican	0.48	0	0.327
Panamanian	0	0.48	0.488

Fig 3. Dendrogram and heatmap based on fixation index values F_{ST} comparing *Acrocomia aculeata* sampling locations. A. F_{ST} was calculated using the whole set of markers (1523 SNP) shows separation between the Mesoamerican, Costa Rican and, Panamanian subpopulations. **B.** F_{ST} shows separation between the Mesoamerican, Costa Rican and, Panamanian groups.

According to discriminant analysis of principal components (DAPC), also three principal genomic clusters appear: Mesoamerican, Costa Rican, and Panamanian (Fig 1 and 4). The clustering also indicates a transition/hybrid zone between lineages from the Mesoamerican and the Costa Rican genomic groups with substantial levels of admixture. The transition zone in the Nicaraguan West populations (Mesoamerican) to the Costa Rican genomic group was further confirmed by F_{ST} analysis (Figure 3) that showed that this population is more related to the Mesoamerican genomic group than with the Costa Rican contrasting with the DAPC. The structuration of the other subpopulations is consistent with the results found in the pairwise F_{ST}.

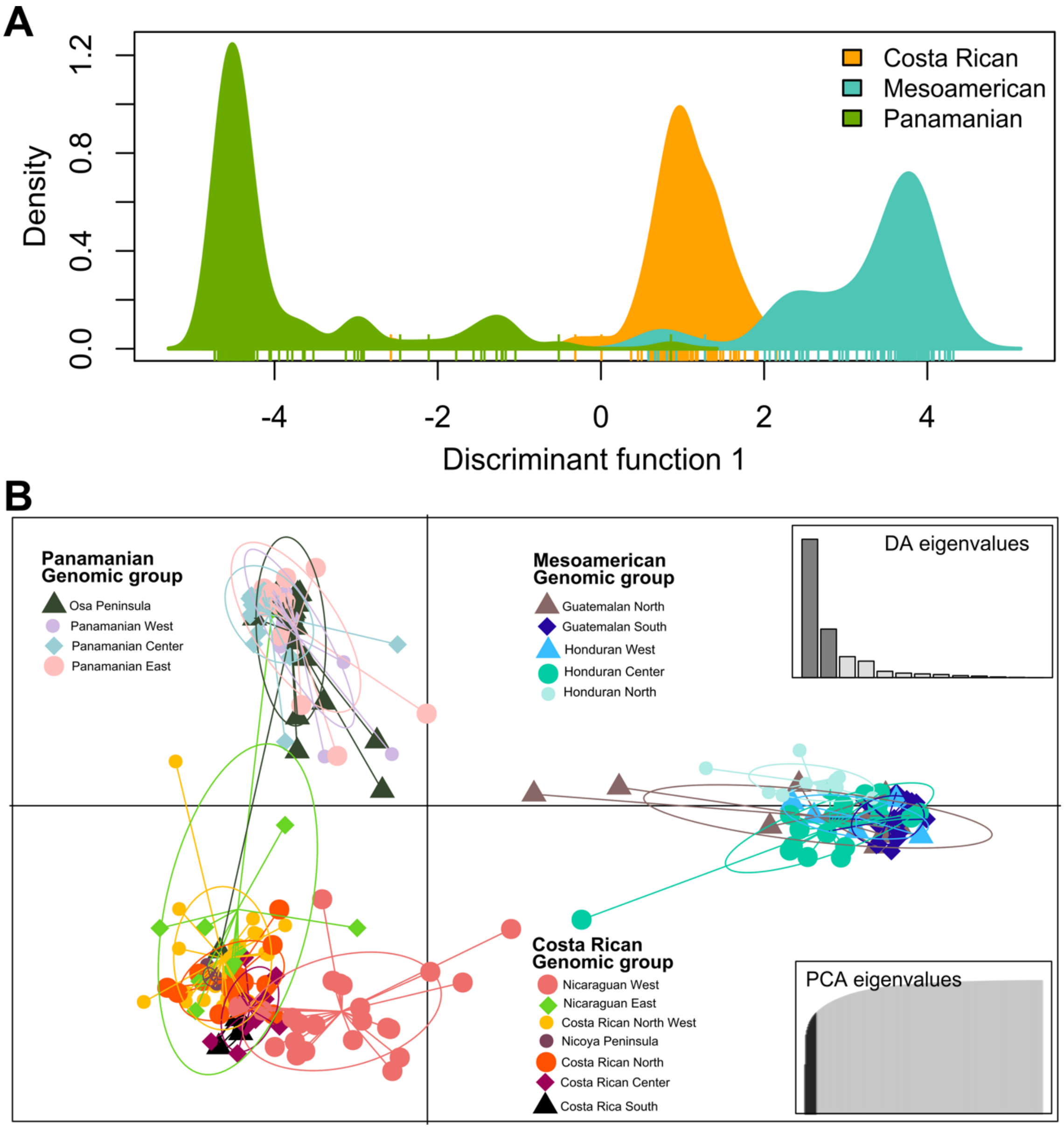


Fig 4. Discriminant analysis of principal components (DAPC) for *Acrocomia aculeata* populations in Central America . A. Density plot based on 1523 makers showing the separation between the Mesoamerican, Costa Rican and, Panamanian groups. **B.** Scatter plot showing the separation between Mesoamerican, Costa Rican and, Panamanian subpopulations.

Ecological niche modeling of *Acrocomia aculeata* in Central America

The final models' TSS ranged from 0.8 to 0.5, and their AUC values ranged from 0.9 to 0.8 (Table S2) meaning that the model is suitable for predicting the potential distribution of *Acrocomia aculeata* in Central America. The ENMs' mean using the MEA method was used to perform the ensemble. Because it produced the greatest TSS values and the highest AUC values (Fig 5), this technique was regarded as the final consensus for the suitability values. Central American lowlands were the most suitable areas for the species.

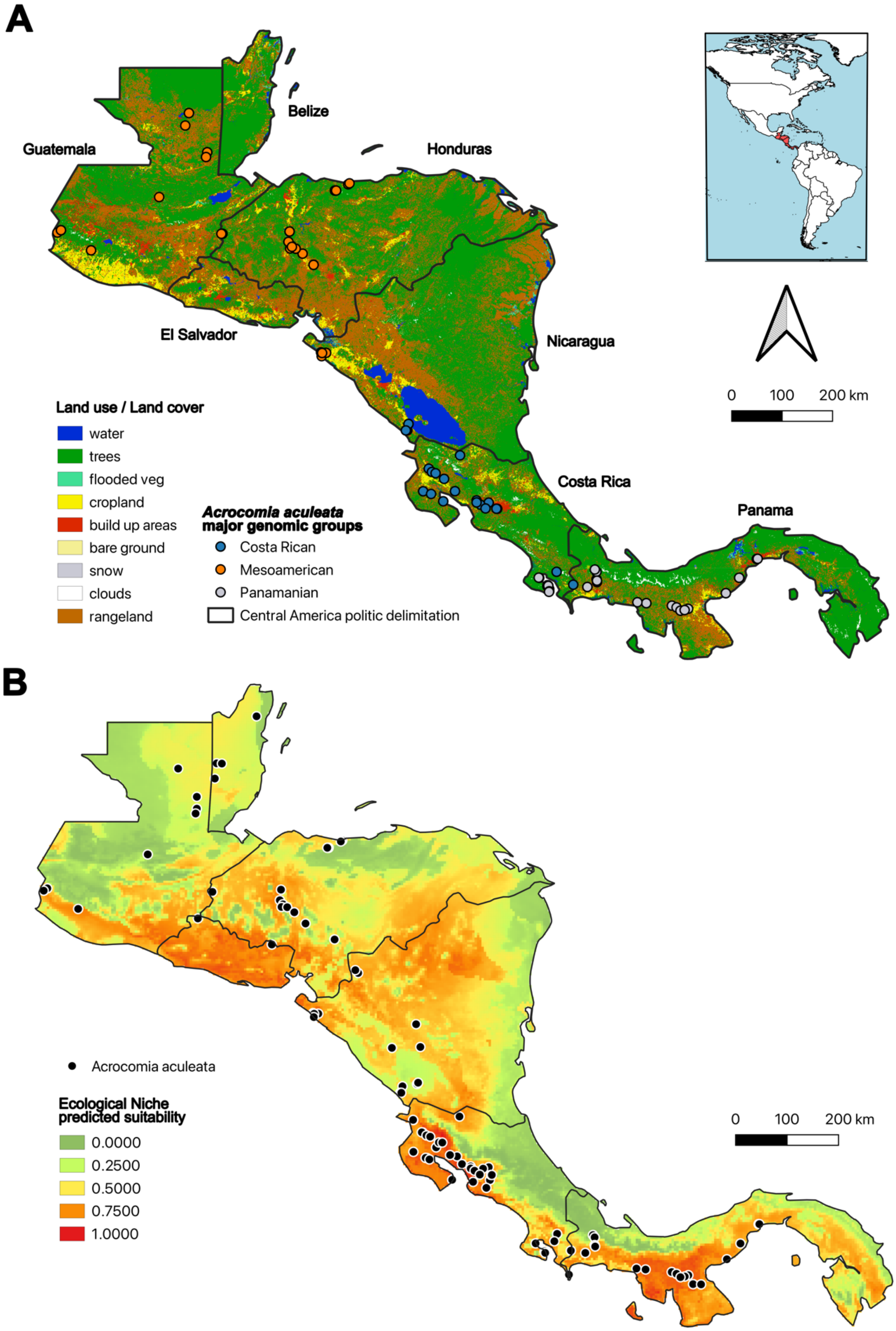

A
Guatemala
Belize
Honduras
El Salvador
Nicaragua
Costa Rica
Panama
0
100
200 km
Land use / Land cover
water
trees
flooded veg
cropland
build up areas
bare ground
snow
clouds
rangeland
Acrocomia aculeata major genomic groups
Costa Rican
Mesoamerican
Panamanian
Central America politic delimitation
B
Acrocomia aculeata
Ecological Niche predicted suitability
0.0000
0.2500
0.5000
0.7500
1.0000
0
100
200 km

Fig 5. Land used map and ecologic niche modeling of *Acrocomia aculeata* in Central America . A. Land use of Central America using Sentinel-2 10m Land Use/Land Cover Time Series. **B.** Ensembles with the means of the algorithms used to model the potential distribution of *Acrocomia aculeata* in Central America.

Discussion

The objective of population genomics is to elucidate the microevolutionary processes on how populations are connected and share a natural history. Central America is considered a biodiversity hotspot, and we used *Acrocomia aculeata* as a model for understanding how broad-range neotropical open canopy plant biodiversity is distributed in the isthmus lowlands. Major geological events have occurred in Central America since the Miocene, including land emergence connecting North and South America. This research marks the first significant sampling effort of an open canopy lowland plant species covering all of Central America for understanding is wide genome diversity using SNP markers. Below, we discuss the most interesting findings regarding the patterns and implications of genomic diversity of *Acrocomia aculeata* in Central America, along with their biological and evolutionary significance and implications for conservation and breeding efforts.

Acrocomia aculeata in Central America, biogeographic patterns and genomic structure of the North American Group

Species of *Acrocomia* inhabit the lowlands (0 – 1700 masl) of tropical America, which extend from Northern Mexico to South America. Their habitats shift from moist broadleaf forests found in tropical and subtropical rainforests to xeric shrublands and savannah, which include dry broadleaf forests found in arid and semi-arid environments (Fig 1). Their wide range suggests that they are able to adapt to various soil and climate conditions. *A. aculeata* is divided into two major gene pools: a North American group and a South American group (Díaz et al. 2021; Morales-Marroquin et al. 2024 *in press*). Our research focus on the North American group where three major genomic groups were identified: a Mesoamerican, a Costa Rican, and a Panamanian (Fig 2). Similar patterns in genetic structure and phylogenomic studies are also present in other palms genera such as *Chamaedorea*, *Geonoma* (Cano *et al.*, 2022) and other clade species like entomofauna (Cano, Schuster and Morrone, 2018; Beza-Beza, Jiménez-Ferbans and McKenna, 2021), arthropods (Crews and Esposito, 2020), herpetofauna (Saldarriaga-Córdoba *et al.*, 2017), Anura (Wang, Crawford and Bermingham, 2008; Mendoza-Henao *et al.*, 2020), ichtyofauna (Matamoros *et al.*, 2015), and parasites (Yisrael *et al.*, 2024). The major biogeographic barriers for *Acrocomia aculeata* in Central America are the Nicaraguan depression and the Talamaca mountain system (Fig 2).

The origins and diversification of *Acrocomia* remain uncertain. The tropical lowlands of South America are proposed as the genus's center of origin, supported by the highest number of *Acrocomia* species, considerable phenotypic diversity, and the oldest fossil records found in Sarataren (State of Pará – Brazil). This hypothesis suggests an irradiation and dispersal from South America through Central America (Morcote-Ríos and Bernal, 2001), a pattern observed in other palms from the Bactridinae subtribe and other plant species (Gentry, 1982; Eiserhardt *et al.*, 2011; Pérez-Escobar *et al.*, 2019; Cano *et al.*, 2022). Phylogenomic studies indicate that the middle of the Miocene (20 -15 million years ago) marked the irradiation of the *Acrocomia* ancestor (Eiserhardt *et al.*, 2011; William J Baker and Couvreur, 2013; Cano *et al.*, 2022), during which the common ancestor was present in both North and South America.

Our analysis detected higher diversity and more private alleles in the Mesoamerican genomic group (Table 1). This could be explain with the colonization routes of *Acrocomia* from South America to Central America (Fig 2). The first dispersal event from South America going to the Maya block (Southern Mexico-Guatemala) the latter the terrestrial corridor disappear. Then, the Mesoamerican group serve as a source region for colonizing southward as terrestrial habitats started to progressively emerge back forming the connection between South America and North America (Gutiérrez-García and Vázquez-Domínguez, 2013; Cano *et al.*, 2022). Additionally, during the Miocene, a separate colonization effort from South America to Central America took place concurrently with the construction of the Panama Isthmus. Based on the floristic character of the region, Gentry postulated more than thirty years ago that South America is an important source of lineages in Central America (Gentry, 1982; Pérez-Escobar *et al.*, 2019). This pattern is evident in the contrasting Fst and genomic structure of the Panamanian group.

Rich geological and paleontological data reveal a striking temporal coincidence between this diversification phase and the proposed progressive closure of the Central American Seaway, which separated South and Central America until approximately 15–10 Ma (Montes *et al.*, 2015). The genetic structure observed in *Acrocomia aculeata* suggests that land emergence connecting these previously separated landmasses was also a major trigger for diversification in Central America, where lineages arriving from the south diversified extensively. Biogeographic barriers play a major role in the genetic divergence of palm (Eiserhardt *et al.*, 2013).

In conclusion, our study sheds light on the genetic diversity, biogeographic patterns, and evolutionary history of *Acrocomia aculeata* natural populations in Central America. By elucidating the genomic structure and colonization routes, we contribute to a deeper understanding of the microevolutionary processes shaping biodiversity in the isthmus lowlands. Our findings underscore the significance of South America as a source region for Central American lineages and highlight the role of biogeographic barriers and land emergence in driving genetic divergence and diversification. Furthermore, conservation and breeding efforts should prioritize preserving habitat connectivity and genetic resources across the region to ensure the resilience and long-term viability of *Acrocomia aculeata* populations in the face of environmental changes and anthropogenic pressures.

Acknowledgements

We acknowledge the financial support from the São Paulo Research Foundation (FAPESP—grants: 2021/10319-0, 2019/20307-0, 2014/23591-7) and the National Council for Scientific and Technological Development (CNPq—grant #431226/2018-0). We also thank the Coordination for the Improvement of Higher Education Personnel of Brazil (CAPES) for the granted scholarship Ph.D. to JM-M.

Supporting information

Table. S1 ***Acrocomia* species samples**. Geographical location and biome of the Acrocomia aculeata samples in Central America. (Excel spreadsheet). https://drive.google.com/drive/folders/1x8oQInx-hu77PPLNU4mfhAAFn7k6fElR?usp=drive_link

Table. S2 Evaluation table with the algorithms and ensembles applied in the Ecological Niche Modeling of *Acrocomia aculeata* in Central America. BIO = Bioclim; MEA = ensemble with mean values of the models; MXS = MaxEnt; RDF = Random Forests; SVM = Support Vector Machine. AUC = Area Under the Curve, and TSS = True Skill Statistics.

Sp	Algorithm	Threshold	Partition	AUC	TSS	AUC_SD	TSS_SD
Acrocomia_aculeata	BIO	MAX_TSS	BOOT	0.943103	0.886207	0.03729	0.07458
Acrocomia_aculeata	MEA	MAX_TSS	BOOT	0.921998	0.717241	0.027655	0.045399
Acrocomia_aculeata	MXS	MAX_TSS	BOOT	0.802021	0.551724	0.037098	0.082886
Acrocomia_aculeata	RDF	MAX_TSS	BOOT	0.943876	0.758621	0.0302	0.072696
Acrocomia_aculeata	SVM	MAX_TSS	BOOT	0.882283	0.668966	0.017268	0.033314

References

Alfaro-Solís, J. D. et al. (2020) 'Acrocomia aculeata fruits from three regions in Costa Rica: an assessment of biometric parameters, oil content and oil fatty acid composition to evaluate industrial potential', Agroforestry Systems. Springer Science+Business Media B.V., 94(5), pp. 1913–1927. doi: 10.1007/S10457-020-00511-8/TABLES/8.

Allouche, O., Tsoar, A. and Kadmon, R. (2006) 'Assessing the accuracy of species distribution models : prevalence , kappa and the true skill statistic (TSS)', Journal of Applied Ecology, 43(6), pp. 1223–1232. doi: 10.1111/j.1365-2664.2006.01214.x.

Andrade, A. F. A. de, Velazco, S. J. E. and De Marco Júnior, P. (2020) 'ENMTML: An R package for a straightforward construction of complex ecological niche models', Environmental Modelling and Software, 125, pp. 1–11. doi: 10.1016/j.envsoft.2019.104615.

Andrews, S. (2010) 'FastQC: A Quality Control Tool for High Throughput Sequence Data'.

Araújo, M. B. and New, M. (2006) 'Ensemble forecasting of species distributions', Trends in Ecology and Evolution, 22(1), pp. 42–47. doi: 10.1016/j.tree.2006.09.010.

Bagley, J. C. and Johnson, J. B. (2014) 'Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas', Biological Reviews. John Wiley & Sons, Ltd, 89(4), pp. 767–790. doi: 10.1111/BRV.12076.

Baker, W. J. and Couvreur, T. L. P. (2013) 'Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages', Journal of Biogeography. Edited by W. D. Kissling, 40(2), pp. 286–298. doi: 10.1111/j.1365-2699.2012.02794.x.

Beza-Beza, C. F., Jiménez-Ferbans, L. and McKenna, D. D. (2021) 'Historical biogeography of New World passalid beetles (Coleoptera, Passalidae) reveals Mesoamerican tropical forests as a centre of origin and taxonomic diversification', Journal of Biogeography. John Wiley & Sons, Ltd, 48(8), pp. 2037–2052. doi: 10.1111/JBI.14134.

Booth, T. H. et al. (2014) 'Bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies', Diversity and Distributions, 20(1), pp. 1–9. doi: 10.1111/ddi.12144.

Breiman, L. (2001) 'Random Forests', Machine Learning, 45, pp. 5–32. doi:

https://doi.org/10.1023/A:1010933404324.

Cano, Á. et al. (2022) ‘Recent and local diversification of Central American understorey palms’, Global Ecology and Biogeography. John Wiley & Sons, Ltd. doi: 10.1111/GEB.13521.

Cano, E. B., Schuster, J. C. and Morrone, J. J. (2018) ‘Phylogenetics of Ogyges Kaup and the biogeography of Nuclear Central America (Coleoptera, Passalidae)’, ZooKeys 737: 81-111. Pensoft Publishers, 737(737), pp. 81–111. doi: 10.3897/ZOOKEYS.737.20741.

Catchen, J. et al. (2013) ‘Stacks: an analysis tool set for population genomics’, Molecular Ecology, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.

Clement, C. R. (1999) ‘1492 and the loss of amazonian crop genetic resources. I. The relation between domestication and human population decline’, Economic Botany, 53(2), pp. 188–202. doi: 10.1007/BF02866498.

Clement, C. R. et al. (2021) ‘Disentangling Domestication from Food Production Systems in the Neotropics’, Quaternary 2021, Vol. 4, Page 4. Multidisciplinary Digital Publishing Institute, 4(1), p. 4. doi: 10.3390/QUAT4010004.

Cockerham, C. C. and Weir, B. S. (1993) ‘Estimation of Gene Flow from F-Statistics’, Evolution. Oxford University Press (OUP), 47(3), p. 855. doi: 10.2307/2410189.

Colombo, C. A. et al. (2018) ‘Macauba: a promising tropical palm for the production of vegetable oil’, OCL. doi: 10.1051/ocl/2017038.

Crews, S. C. and Esposito, L. A. (2020) ‘Towards a synthesis of the Caribbean biogeography of terrestrial arthropods’, BMC Evolutionary Biology 2020 20:1. BioMed Central, 20(1), pp. 1–27. doi: 10.1186/S12862-019-1576-Z.

Danecek, P. et al. (2011) ‘The variant call format and VCFtools’, Bioinformatics. Oxford Academic, 27(15), pp. 2156–2158. doi: 10.1093/BIOINFORMATICS/BTR330.

Destro, G. F. G. et al. (2020) ‘Climate suitability as indicative of invasion potential for the most seized bird species in Brazil’, Journal for Nature Conservation, 58(August). doi: 10.1016/j.jnc.2020.125890.

Díaz, B. G. et al. (2021) ‘Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species’, PLoS ONE, 16(7 July), pp. 1–24. doi: 10.1371/journal.pone.0241025.

Eiserhardt, W. L. et al. (2011) ‘Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences’, Taxon, 60(2), pp. 485–498.

Eiserhardt, W. L. et al. (2013) ‘Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents’, Scientific Reports, 3(1), p. 1164. doi: 10.1038/srep01164.

Galili, T. et al. (2018) ‘heatmaply: an R package for creating interactive cluster heatmaps for online publishing’, Bioinformatics. Oxford Academic, 34(9), pp. 1600–1602. doi: 10.1093/BIOINFORMATICS/BTX657.

Gentry, A. H. (1982) ‘Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny?’, Annals - Missouri Botanical Garden, 69(3), pp. 557–593. doi: 10.2307/2399084.

Glaubitz, J. C. et al. (2014) ‘TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline’, PLoS ONE. Edited by N. A. Tinker. Public Library of Science, 9(2), p. e90346. doi: 10.1371/journal.pone.0090346.

Goudet, J. (2005) ‘hierfstat, a package for r to compute and test hierarchical F-statistics’, Molecular Ecology Notes. John Wiley & Sons, Ltd, 5(1), pp. 184–186. doi: 10.1111/J.1471-8286.2004.00828.X.

Guo, Q., Kelly, M. and Graham, C. H. (2005) ‘Support vector machines for predicting distribution of Sudden Oak Death in California’, Ecological Modelling, 182, pp. 75–90. doi: 10.1016/j.ecolmodel.2004.07.012.

Gutiérrez-García, T. A. and Vázquez-Domínguez, E. (2013) ‘Consensus between genes and stones in the biogeographic and evolutionary history of Central America’, Quaternary Research. Cambridge University Press, 79(3), pp. 311–324. doi: 10.1016/J.YQRES.2012.12.007.

Jombart, T. (2022) ‘An introduction to adegenet 2.1.6’. London: Imperal College London, p. 79.

Jombart, T. and Ahmed, I. (2011) ‘adegenet 1.3-1: New tools for the analysis of genome-wide SNP data’, Bioinformatics. doi: 10.1093/bioinformatics/btr521.

Kamvar, Z. N., Tabima, J. F. and Grünwald, N. J. (2014) ‘Poppr : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction’, PeerJ, 2, p. e281. doi: 10.7717/peerj.281.

de Lima, N. E. et al. (2018) ‘A review of the palm genus Acrocomia: Neotropical green gold’, Organisms Diversity & Evolution. Springer Berlin Heidelberg, pp. 1–11. doi: 10.1007/s13127-018-0362-x.

Maitner, B. S. et al. (2018) ‘The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database’, Methods in Ecology and Evolution, 9(2), pp. 373–379. doi: 10.1111/2041-210X.12861.

Matamoros, W. A. et al. (2015) ‘Derivation of the freshwater fish fauna of Central America revisited: Myers’s hypothesis in the twenty-first century’, Cladistics. John Wiley & Sons, Ltd, 31(2), pp. 177–188. doi: 10.1111/CLA.12081.

Mendoza-Henao, A. M. et al. (2020) ‘Phylogeny-based species delimitation and integrative taxonomic revision of the Hyalinobatrachium fleischmanni species complex, with resurrection of H. viridissimum (Taylor, 1942)’, Systematics and Biodiversity. Taylor & Francis, pp. 464–484. doi: 10.1080/14772000.2020.1776781.

Metz, C. E. (1986) ‘ROC methodology in radiologic imaging.’, Investigative radiology, 21(9), pp. 720–733.

Montes, C. et al. (2015) ‘Middle Miocene closure of the Central American Seaway’, Science, 348(6231), pp. 226–229. doi: 10.1126/science.aaa2815.

Morales-Marroquín, J. A. et al. (2022) ‘Biodiversity Research in Central America: A Regional Comparison in Scientific Production Using Bibliometrics and Democracy Indicators’, Frontiers in Research Metrics and Analytics. Frontiers, 7, p. 44. doi: 10.3389/frma.2022.898818.

Morcote-Ríos, G. and Bernal, R. (2001) ‘Remains of palms (palmae) at archaeological sites in the new world: A review’, Botanical Review. New York Botanical Garden Press, 67(3), pp. 309–350. doi: 10.1007/BF02858098/METRICS.

Pérez-Escobar, O. A. et al. (2019) ‘The Origin and Diversification of the Hyperdiverse Flora in the Chocó Biogeographic Region’, Frontiers in Plant Science. Frontiers Media S.A., 10, p. 468280. doi: 10.3389/FPLS.2019.01328/BIBTEX.

Phillips, S. J. and Dudík, M. (2008) ‘Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation’, Ecography. John Wiley & Sons, Ltd (10.1111), 31(2), pp. 161–175. doi: 10.1111/j.0906-7590.2008.5203.x.

Poland, J. A. et al. (2012) ‘Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach’, PLoS ONE. Edited by T. Yin. Public Library of Science, 7(2), p. e32253. doi: 10.1371/journal.pone.0032253.

R Core Team (2016) ‘R Development Core Team’, R: A Language and Environment for Statistical Computing. doi: http://www.R-project.org.

Saldarriaga-Córdoba, M. et al. (2017) ‘Phylogeography of the Central American lancehead Bothrops asper (SERPENTES: VIPERIDAE)’, PLOS ONE. Public Library of Science, 12(11), p. e0187969. doi: 10.1371/JOURNAL.PONE.0187969.

Vargas-Carpintero, R. et al. (2021) ‘Acrocomia spp.: neglected crop, ballyhooed multipurpose palm or fit for the bioeconomy? A review’, Agronomy for Sustainable Development 2021 41:6. Springer, 41(6), pp. 1–35. doi: 10.1007/S13593-021-00729-5.

Wang, I. J., Crawford, A. J. and Bermingham, E. (2008) ‘Phylogeography of the Pygmy Rain Frog (Pristimantis ridens) across the lowland wet forests of isthmian Central America’, Molecular Phylogenetics and Evolution. Academic Press, 47(3), pp. 992–1004. doi: 10.1016/J.YMPEV.2008.02.021.

Yisrael, L. et al. (2024) ‘Expanding the phylogeography and connectivity of Perkinsus species across North and Central America’, Diversity and Distributions. John Wiley & Sons, Ltd, 30(3), p. e13807. doi: 10.1111/DDI.13807.

Zizka, A. et al. (2019) ‘Coordinate Cleaner : Standardized cleaning of occurrence records from biological collection databases’, Methods in Ecology and Evolution, 10, pp. 744–751. doi: 10.1111/2041-210X.13152.

6. Conclusão

Depois de explorar vários aspectos da genética populacional, dos processos de adaptação e da possível domesticação das espécies de *Acrocomia* no continente Americano, nossos estudos apresentam uma compreensão abrangente da estrutura genética, da história evolutiva e dos mecanismos adaptativos do gênero. Por meio de técnicas genômicas avançadas, como genotipagem por meio de SNP e análises de sinais adaptativas, estudamos as relações entre diferentes espécies de *Acrocomia* e sua adaptação a diversas condições ambientais em biomas neotropicais. Foi observado um alto grau de diversidade nas espécies de *Acrocomia* de vários biomas neotropicais. Essa variação pode ter surgido da história natural da cultura, do processo de domesticação e de seu cultivo em diversos ecossistemas. É plausível que alguns dos sinais de seleção estejam associados a características agronômicas desejáveis, que podem ser cruciais para aprimoramento do melhoramento dessas palmeiras.

A macaúba surge como uma cultura promissora devido ao seu potencial multiuso, oferecendo perspectivas na produção de óleo, biocombustíveis, alimentos e indústria farmacêutica. Notavelmente, o óleo extraído de *A. aculeata* apresenta qualidade comparável à do óleo de palma (*Elaeis guineensis*), exigindo menos recursos hídricos e demonstrando adaptabilidade a ambientes semiáridos e áridos e a ecossistemas degradados. Na última década houve um progresso significativo nas pesquisas, principalmente na germinação de sementes, produção de mudas e sistemas agrícolas integrados utilizando *Acrocomia*, embora o melhoramento molecular ainda seja necessário (Colombo et al., 2018; Vargas-Carpintero et al., 2021). Priorizar o sequenciamento e a anotação dos genomas das três espécies economicamente mais importantes do gênero é crucial para futuros esforços em seu processo de domesticação e melhoramento. É necessário enfrentar desafios como a interação genótipo-ambiente, a prospecção da biodiversidade para a criação de ideótipos de melhoramento fenotípico (plantas modelo) e o gerenciamento da cultura. É necessário compreender melhor a diversidade genética e fenotípica do grupo genômico da América do Norte. A maioria dos esforços de pesquisa concentrou-se apenas no pool genético da América do Sul. Na América Central, nossa pesquisa revelou três grupos genômicos primários em *A. aculeata*: Mesoamericano, Costa-riquenho e Panamenho.

Nossos resultados revelam uma estruturação genética significativa dentro e entre as espécies em toda sua distribuição, destacando a influência da dispersão, das barreiras biogeográficas e das interações ecológicas históricas na evolução de *Acrocomia*. Observamos assinaturas genéticas distintas associadas à adaptação a fatores ambientais específicos, incluindo variáveis climáticas, resistência a patógenos e respostas ao estresse. Notavelmente, as espécies de *Acrocomia* apresentam potencial de resistência às mudanças climáticas, especialmente em ecossistemas semiáridos, o que as torna candidatas promissoras para a agricultura sustentável e a produção de biocombustíveis. Além disso, nosso estudo esclarece as características moleculares da possível domesticação de espécies de *Acrocomia*, enfatizando a importância dos recursos genéticos e dos genes candidatos associados a características agronômicas desejáveis, como teor de óleo, biossíntese de carotenoides e tolerância ao estresse.

A Macaúba é uma cultura valiosa que oferece oportunidades para futuros setores e, ao mesmo tempo, contribui para o desenvolvimento sustentável e a conservação da biodiversidade. Ao priorizar os esforços de pesquisa colaborativa em programas público-privados, conservação genética e apoio a políticas, diferentes partes interessadas podem liderar o potencial da diversidade genética e da capacidade de adaptação das espécies de *Acrocomia* para enfrentar os desafios globais, incluindo a segurança alimentar, a produção de energia renovável e a resistência climática em agroecossistemas.

7. Referências

Abreu, A. G. et al. (2012) ‘The genetic structure and mating system of Acrocomia aculeata (Arecaceae)’, Genet Mol Biol. Brazilian Journal of Genetics, 35(1), pp. 119–121. doi: 10.1590/s1415-47572012005000002.

Abreu, I. S. et al. (2011) ‘First karyotype, DNA C-value and AT/GC base composition of macaw palm (Acrocomia aculeata, Arecaceae) - a promising plant for biodiesel production’, Australian Journal of Botany. CSIRO PUBLISHING, 59(2), p. 149. doi: 10.1071/BT10245.

Aires, G. C. M. and de Carvalho Junior, R. N. (2023) ‘Potential of Supercritical Acrocomia aculeata Oil and Its Technology Trends’, Applied Sciences 2023, Vol. 13, Page 8594. Multidisciplinary Digital Publishing Institute, 13(15), p. 8594. doi: 10.3390/APP13158594.

Alfaro-Solís, J. D. et al. (2020) ‘Acrocomia aculeata fruits from three regions in Costa Rica: an assessment of biometric parameters, oil content and oil fatty acid composition to evaluate industrial potential’, Agroforestry Systems. Springer Science+Business Media B.V., 94(5), pp. 1913–1927. doi: 10.1007/S10457-020-00511-8/TABLES/8.

Araújo, M. R. G. et al. (2017) ‘Fine-scale spatial genetic structure and gene flow in Acrocomia aculeata (Arecaceae): Analysis in an overlapping generation’, Biochemical Systematics and Ecology. Pergamon, 71, pp. 147–154. doi: 10.1016/J.BSE.2017.02.005.

Baker, William J. and Couvreur, T. L. P. (2013) ‘Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography’, Journal of Biogeography. Edited by W. D. Kissling, 40(2), pp. 274–285. doi: 10.1111/j.1365-2699.2012.02795.x.

Baker, William J and Couvreur, T. L. P. (2013) ‘Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages’, Journal of Biogeography. Edited by W. D. Kissling, 40(2), pp. 286–298. doi: 10.1111/j.1365-2699.2012.02794.x.

Bedoya, C. A. et al. (2017) ‘Genetic diversity and population structure of native maize populations in Latin America and the Caribbean’, PLOS ONE. Public Library of Science, 12(4), p. e0173488. doi: 10.1371/JOURNAL.PONE.0173488.

Bitocchi, E. et al. (2012) ‘Mesoamerican origin of the common bean (Phaseolus vulgaris L.) is revealed by sequence data’, Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 109(14), pp. E788–E796. doi:

10.1073/PNAS.1108973109/SUPPL_FILE/ST02.DOC.

Bora, P. S. and Rocha, R. V. M. (2004) 'Frutos macaíba: Composición de aminoácidos e ácidos graxos de froitos', Ciencia y Tecnologia Alimentaria, 4(3), pp. 158–162. doi: 10.1080/11358120409487755.

Cámara-Leret, R., Fortuna, M. A. and Bascompte, J. (2019) 'Indigenous knowledge networks in the face of global change', Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1821843116.

Cano, Á. et al. (2022) 'Recent and local diversification of Central American understorey palms', Global Ecology and Biogeography. John Wiley & Sons, Ltd. doi: 10.1111/GEB.13521.

Catchen, J. et al. (2013) 'Stacks: an analysis tool set for population genomics', Molecular Ecology, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.

Cauz-Santos, L. A. et al. (2017) 'The Chloroplast Genome of Passiflora edulis (Passifloraceae) Assembled from Long Sequence Reads: Structural Organization and Phylogenomic Studies in Malpighiales', Frontiers in Plant Science. Frontiers Media SA, 8, p. 334. doi: 10.3389/FPLS.2017.00334.

Ciconini, G. et al. (2013) 'Biometry and oil contents of Acrocomia aculeata fruits from the Cerrados and Pantanal biomes in Mato Grosso do Sul, Brazil', Industrial Crops and Products. Elsevier, 45, pp. 208–214. doi: 10.1016/J.INDCROP.2012.12.008.

Cláudio, L. et al. (2017) 'Genetic diversity and structure of... GENETIC DIVERSITY AND STRUCTURE OF MACAW PALM: IMPLICATIONS FOR GENETIC VARIABILITY SAMPLING 1', Revista Árvore, 41(5), p. 410508. doi: 10.1590/1806-90882017000500008.

Clement, C. R. (1999a) '1492 and the loss of amazonian crop genetic resources. I. The relation between domestication and human population decline', Economic Botany, 53(2), pp. 188–202. doi: 10.1007/BF02866498.

Clement, C. R. (1999b) '1492 and the loss of amazonian crop genetic resources. II. crop Biogeography at contact', Economic Botany, 53(2), pp. 203–216. doi: 10.1007/BF02866499.

Clement, C. R. et al. (2021) 'Disentangling Domestication from Food Production Systems in the Neotropics', Quaternary 2021, Vol. 4, Page 4. Multidisciplinary Digital Publishing Institute, 4(1), p. 4. doi: 10.3390/QUAT4010004.

Coelho, N. H. P. et al. (2018) 'Understanding genetic diversity, spatial genetic structure, and mating system through microsatellite markers for the conservation and sustainable use of Acrocomia aculeata (Jacq.) Lodd. Ex Mart.', Conservation Genetics. Springer Netherlands, 19(4), pp. 879–891. doi: 10.1007/s10592-018-1061-z.

Colombo, C. A. et al. (2018) 'Macauba: a promising tropical palm for the production of vegetable oil', OCL. doi: 10.1051/ocl/2017038.

Cosner, M. E., Raubeson, L. A. and Jansen, R. K. (2004) 'Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes', BMC Evolutionary Biology. BioMed Central, 4, p. 27. doi: 10.1186/1471-2148-4-27.

Couto, E. G. de O. et al. (2024) 'Genome-Wide Association Insights into the Genomic Regions Controlling Oil Production Traits in Acrocomia aculeata (neotropical native palm)', bioRxiv. Cold Spring Harbor Laboratory, p. 2024.01.17.576138. doi: 10.1101/2024.01.17.576138.

Díaz, B. G. et al. (2021) 'Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species', PLoS ONE, 16(7 July), pp. 1–24. doi: 10.1371/journal.pone.0241025.

Eiserhardt, W. L. et al. (2011) 'Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences', Taxon, 60(2), pp. 485–498.

Eiserhardt, W. L. et al. (2013) 'Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents', Scientific Reports, 3(1), p. 1164. doi: 10.1038/srep01164.

Eiserhardt, W. L., Couvreur, T. L. P. and Baker, W. J. (2017) 'Plant phylogeny as a window on the evolution of hyperdiversity in the tropical rainforest biome', New Phytologist. Wiley/Blackwell (10.1111), 214(4), pp. 1408–1422. doi: 10.1111/nph.14516.

Elshire, R. J. et al. (2011) 'A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species', PLoS ONE. Edited by L. Orban. Public Library of Science, 6(5), p. e19379. doi: 10.1371/journal.pone.0019379.

Escudero, M. et al. (2014) 'Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in Carex (Cyperaceae)', Molecular Phylogenetics and Evolution. Academic Press, 79, pp. 359–367. doi: 10.1016/J.YMPEV.2014.06.026.

Francisconi, A. F. et al. (2023) 'Complete chloroplast genomes of six neotropical palm species, structural comparison, and evolutionary dynamic patterns', Scientific Reports 2023 13:1. Nature Publishing Group, 13(1), pp. 1–15. doi: 10.1038/s41598-023-44631-4.

Glaubitz, J. C. et al. (2014) 'TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline', PLoS ONE. Edited by N. A. Tinker. Public Library of Science, 9(2), p. e90346. doi: 10.1371/journal.pone.0090346.

Gutiérrez-García, T. A. and Vázquez-Domínguez, E. (2013) 'Consensus between genes and stones in the biogeographic and evolutionary history of Central America', Quaternary Research. Cambridge University Press, 79(3), pp. 311–324. doi: 10.1016/J.YQRES.2012.12.007.

Henderson, A., Galeano-Garces, G. and Bernal, R. (1995) 'Field Guide to the Palms of the Americas', Nord. J. Bot. doi: 10.1007/BF02910409.

Janzen, D. H. and Martin, P. S. (1982) 'Neotropical Anachronisms: The Fruits the Gomphotheres Ate', Science, New Series, 215(4528), pp. 19–27. Available at: http://www.jstor.org/stable/1688516 (Accessed: 15 March 2018).

Kim, K. et al. (2015) 'Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species', Scientific Reports. Nature Publishing Group, 5(1), pp. 1–13. doi: 10.1038/srep15655.

Kissling, W. D. et al. (2012) 'Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide', Proceedings of the National Academy of Sciences, 109(19), pp. 7379–7384. doi: 10.1073/pnas.1120467109.

Lentz, D. L. (1990) 'ACROCOMIA MEXICANA: PALM OF THE ANCIENT MESOAMERICANS', J. Ethllobiol, 10(2), pp. 183–194.

Levis, C. et al. (2018) 'How people domesticated Amazonian forests', Frontiers in Ecology and Evolution. Frontiers Media S. A, 5(JAN), p. 299700. doi: 10.3389/FEVO.2017.00171/BIBTEX.

de Lima, N. E. et al. (2018) 'A review of the palm genus Acrocomia: Neotropical green gold', Organisms Diversity & Evolution. Springer Berlin Heidelberg, pp. 1–11. doi: 10.1007/s13127-018-0362-x.

Lorenzi, H. et al. (2010) Flora Brasileira Arecaceae (Palmeiras), Instituto Plantarum De Estudos Da Flora Ltda. Nova Odessa: Instituto Plantarum.

Martínez-Alberola, F. et al. (2013) 'Balanced Gene Losses, Duplications and Intensive Rearrangements Led to an Unusual Regularly Sized Genome in Arbutus unedo Chloroplasts', PLoS ONE. Edited by G. G. Vendramin. Public Library of Science, 8(11), p. e79685. doi: 10.1371/journal.pone.0079685.

Mengistu, F., Motoike, S. and Cruz, C. (2016) 'Molecular Characterization and Genetic Diversity of the Macaw Palm Ex Situ Germplasm Collection Revealed by Microsatellite Markers', Diversity, 8(4), p. 20. doi: 10.3390/d8040020.

Morcote-Ríos, G. and Bernal, R. (2001) 'Remains of palms (palmae) at archaeological sites in the new world: A review', Botanical Review. New York Botanical

Garden Press, 67(3), pp. 309–350. doi: 10.1007/BF02858098/METRICS.

Motoike, S. Y. et al. (2013) A cultura da macaúba: implantação e manejo de cultivos racionais. 1st edn. Viçosa, Minas Gerais, Brazil: Universidade Federal de Viçosa.

Neiva, D. S. et al. (2016) ‘Acrocomia emensis (Arecaceae) genetic structure and diversity using SSR molecular markers’. doi: 10.4238/gmr.15017785.

Nucci, S. M. et al. (2008) ‘Development and characterization of microsatellites markers from the macaw’, Molecular Ecology Resources. John Wiley & Sons, Ltd (10.1111), 8(1), pp. 224–226. doi: 10.1111/j.1471-8286.2007.01932.x.

O’Dea, A. et al. (2016) ‘Formation of the Isthmus of Panama’, Science Advances, 2(8), pp. 1–11. doi: 10.1126/sciadv.1600883.

Ramírez-Núñez, C. et al. (2019) ‘MULTIDIRECTIONAL INTERPOLATION of LIDAR DATA from SOUTHERN VERACRUZ, MEXICO: IMPLICATIONS for EARLY OLMEC SUBSISTENCE’, Ancient Mesoamerica. Cambridge University Press. doi: 10.1017/S0956536118000263.

Roncal, J. et al. (2013) ‘Cenozoic colonization and diversification patterns of tropical American palms: evidence from Astrocaryum (Arecaceae)’, Botanical Journal of the Linnean Society, 171(1), pp. 120–139. doi: 10.1111/j.1095-8339.2012.01297.x.

de Santana Lopes, A. et al. (2018) ‘The complete plastome of macaw palm [Acrocomia aculeata (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae’, Planta, 247(4), pp. 1011–1030. doi: 10.1007/s00425-018-2841-x.

de Santana Lopes, A. et al. (2019) ‘The plastomes of Astrocaryum aculeatum G. Mey. and A. murumuru Mart. show a flip-flop recombination between two short inverted repeats’, Planta. Springer Berlin Heidelberg, pp. 1–18. doi: 10.1007/s00425-019-03217-z.

Shahbandeh, M. (2024) Consumption of vegetable oils worldwide from 2013/14 to 2023/2024, Statista. Available at: https://www.statista.com/statistics/263937/vegetable-oils-global-consumption/ (Accessed: 29 April 2024).

Tangphatsornruang, S. et al. (2011) ‘Characterization of the complete chloroplast genome of Hevea brasiliensis reveals genome rearrangement, RNA editing sites and phylogenetic relationships’, Gene, 475(2), pp. 104–112. doi: 10.1016/j.gene.2011.01.002.

‘The Plant List’ (2013) Choice Reviews Online. doi: 10.5860/choice.48-6278.

Vargas-Carpintero, R. et al. (2021) ‘Acrocomia spp.: neglected crop, ballyhooed multipurpose palm or fit for the bioeconomy? A review’, Agronomy for Sustainable Development 2021 41:6. Springer, 41(6), pp. 1–35. doi: 10.1007/S13593-021-00729-5.

Vargas-Carpintero, R. et al. (2022) ‘A Collaborative, Systems Approach for the Development of Biomass-Based Value Webs: The Case of the Acrocomia Palm’, Land 2022, Vol. 11, Page 1748. Multidisciplinary Digital Publishing Institute, 11(10), p. 1748. doi: 10.3390/LAND11101748.

Vianna, Suelen Alves et al. (2017) ‘Biometric Characterization of Fruits and Morphoanatomy of the Mesocarp of Acrocomia Species (Arecaceae)’, International Journal of Biology. doi: 10.5539/ijb.v9n3p78.

Vianna, Suelen A et al. (2017) ‘Leaf anatomy of Acrocomia (Arecaceae): an additional contribution to the taxonomic resolution of a genus with great economic potential’, Plant Systematics and Evolution, 303(2), pp. 233–248. doi: 10.1007/s00606-016-1369-4.

Vianna, S. A. et al. (2021) ‘Morphological characterization and productivity estimates of Acrocomia totai Mart. (Arecaceae) – a sustainable alternative of extractivism and cultivation’, Revista Brasileira de Fruticultura. Sociedade Brasileira de Fruticultura, 43(1), p. e-730. doi: 10.1590/0100-29452021730.

Yardeni, G. et al. (2016) ‘Reproductive isolation between populations of Iris atropurpurea is associated with ecological differentiation’, Annals of Botany. Oxford University Press, 118(5), pp. 971–982. doi: 10.1093/aob/mcw139.

8. Anexos

Produção científica pelo aluno sobre o tema de pesquisa.

Morales-Marroquín, JA; Solis, R; Pinheiro, J; Zucchi, MI. (2022) **Biodiversity Research in Central America: A Regional Comparison in Scientific Production Using Bibliometrics and Democracy Indicators**. *Frontiers in Research Metrics and Analytics*. Frontiers, 7, p. 44. https://doi.org/10.3389/frma.2022.898818

Morales-Marroquín, JA; Pinheiro, J; Zucchi, MI. (2023) **Biodiversity and Climate policy challenges in Central America towards natural resource governance**. *United Nations IATT (Inter-Agency Task Team)* call for science policy briefs for the STI forum 2023 (Multi-Stakeholder Forum on Science, Technology, and Innovation for the Sustainable Development Goals). **Document:** https://tinyurl.com/mr34568z

Video presentation: https://www.youtube.com/watch?v=ZT01rnC6oBM

Francisconi A*; **Morales-Marroquín JA***; Cauz-Santos L; Zucchi, MI. et al. (2023) **Complete chloroplast genomes of six neotropical palm species, structural comparison, and evolutionary dynamic patterns**. *Scientific Reports* 2023 13:1. Nature Publishing Group, 13(1), pp. 1–15. https://doi.org/10.1038/s41598-023-44631-4 **Both authors contributed equally*

Colaborações com o grupo de pesquisa em palmeiras:

Couto, E; **Morales-Marroquín, JA**; Zucchi, MI. et al. (2024) **Genome-Wide Association Insights into the Genomic Regions Controlling Oil Production Traits in Acrocomia aculeata (neotropical native palm).** *Pre-print in BioRxiv*. https://doi.org/10.1101/2024.01.17.576138

Ferreira Costa, M; **Morales-Marroquín, JA**; Zucchi, MI. et al. (2023) **Population Genomics and Conservation of Copernicia alba: a Non-Timber Palm of Economic Importance in the Brazilian Chaco.** *Plant Molecular Biology Reporter* 2023. Springer, pp. 1–11. https://doi.org/10.1007/s11105-023-01419-0

Ferreira Costa M; **Morales-Marroquín, JA**; Zucchi, MI. et al. (2022) **Population genomics of the neotropical palm Copernicia prunifera (Miller) H. E. Moore: implications for conservation.** *PLoS ONE.* 17(11): e0276408. https://doi.org/10.1371/journal.pone.0276408

Díaz-Hernández, B. G.; Colombo, C; **Morales-Marroquín, JA** et al. (2023) **Assessing the genetic vulnerability of Macaúba palm [Acrocomia aculeata (Jacq.) Lodd. ex Mart.] through the mating system and genetic diversity of open-pollinated progenies**. *Annals of Applied Biology.* John Wiley & Sons, Ltd. https://doi.org/10.1111/aab.12879

Francisconi A; Cauz-Santos, L; **Morales-Marroquín JA**; Van den Berg, C; Zucchi, MI et al. (2022) **Complete chloroplast genomes and phylogeny in three Euterpe palms (E. edulis, E. oleracea and E. precatoria) from different Brazilian biomes**. *PLoS ONE.* Public Library of Science, 17(7), p. e0266304. https://doi.org/10.1371/journal.pone.0266304

Araya-Valverde, E; Arnáez-Serrano, […] **Morales-Marroquín, JA**; (2022) **Genetic diversity and structure of Acrocomia aculeata (Jacq.) Lodd. ex Mart. (Arecaceae) using microsatellite DNA markers in Costa Rica**. *Genetic Resources and Crop Evolution*. Springer Nature B.V. https://link.springer.com/article/10.1007/s10722-022-01501-8

UNICAMP
COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br
IB

DECLARAÇÃO

Em observância ao **§5º do Artigo 1º da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada "***VARIAÇÕES GENÔMICAS ASSOCIADAS À ADAPTAÇÃO EM PALMEIRAS DO GÊNERO ACROCOMIA NO NEOTRÓPICO***", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: ________________________________
Nome do(a) aluno(a): Dr. Jonathan Andre Morales Marroquín

Assinatura: M.I. Zucchi
Nome do(a) orientador(a): Profa. Dr. Maria Imaculada Zucchi

Data: 05/12/2024

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **"VARIAÇÕES GENÔMICAS ASSOCIADAS À ADAPTAÇÃO EM PALMEIRAS DO GÊNERO ACROCOMIA NO NEOTRÓPICO"**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 05 de dezembro, 2024

Assinatura : ______________________________
Nome do(a) autor(a): **Jonathan Andre Morales Marroquín**
RG n.° G239454F

M.I. Zucchi
Assinatura : ______________________________
Nome do(a) orientador(a): **Maria Imaculada Zucchi**
RG n.° 198102136