UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

MARCONES FERREIRA COSTA

GENÔMICA POPULACIONAL E MODELAGEM DE NICHO ECOLÓGICO DAS PALMEIRAS NEOTROPICAIS CARANDÁ E CARNAÚBA: IMPLICAÇÕES PARA CONSERVAÇÃO

POPULATION GENOMICS AND ECOLOGICAL NICHE MODELING OF THE NEOTROPICAL PALMS CARANDÁ AND CARNAÚBA: IMPLICATIONS FOR CONSERVATION

CAMPINAS-SP
2023

MARCONES FERREIRA COSTA

GENÔMICA POPULACIONAL E MODELAGEM DE NICHO ECOLÓGICO DAS PALMEIRAS NEOTROPICAIS CARANDÁ E CARNAÚBA: IMPLICAÇÕES PARA CONSERVAÇÃO

POPULATION GENOMICS AND ECOLOGICAL NICHE MODELING OF THE NEOTROPICAL PALMS CARANDÁ AND CARNAÚBA: IMPLICATIONS FOR CONSERVATION

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

Thesis presented to the Biology Institute of the University of Campinas in a partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Plant Genetics and Genetic Breeding

Orientadora: Prof.ª Dr.ª Maria Imaculada Zucchi

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO MARCONES FERREIRA COSTA E ORIENTADA PELA PROFESSORA MARIA IMACULADA ZUCCHI

CAMPINAS-SP

2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

Costa, Marcones Ferreira, 1989-
C823p Population genomics and ecological niche modeling of the neotropical palms carandá and carnaúba : implications for conservation / Marcones Ferreira Costa. – Campinas, SP : [s.n.], 2023.

Orientador: Maria Imaculada Zucchi.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Palmeira. 2. Carandá. 3. Carnaubeira. 4. Genômica. 5. Polimorfismo de nucleotídeo único. I. Zucchi, Maria Imaculada, 1971-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações Complementares

Título em outro idioma: Genômica Populacional e modelagem de nicho ecológico das palmeiras neotropicais carandá e carnaúba : implicações para conservação
Palavras-chave em inglês:
Palms
Copernicia alba
Carnauba palm
Genomics
Single nucleotide polymorphism
Área de concentração: Genética Vegetal e Melhoramento
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
Maria Imaculada Zucchi [Orientador]
Anete Pereira de Souza
Cleber Juliano Neves Chaves
Fernanda Amato Gaiotto
Miklos Maximiliano Bajay
Data de defesa: 18-01-2023
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-8210-2673
- Currículo Lattes do autor: http://lattes.cnpq.br/9173274166161129

Campinas, 18 de Janeiro de 2023

EXAMINATION COMMITTEE

Comissão examinadora

Profa. Dra. Maria Imaculada Zucchi

Profa. Dra. Anete Pereira de Souza

Prof. Dr. Cleber Juliano Neves Chaves

Profa. Dra. Fernanda Amato Gaiotto

Prof. Dr. Miklos Maximiliano Bajay

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 Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia.

ACKNOWLEDGMENT

I would like to express my gratitude to God for his goodness and mercy, for sustaining and leading me in every moment of my life. Up to this point the LORD has helped us!

I am profoundly grateful to my family: my parents (Agenor and Antônia), my sister (Patrícia), my brother (Marcos Vinícius, *in memoriam*) for all the affection, understanding and emotional and financial support offered since the beginning of my academic journey. I thank my nieces (Priscilla, Maria Paula and Maria Júlia) and my nephew (Leandro) for the moments of relaxation and lightheartedness.

I thank the Federal University of Piauí for granting the 4-year license so that I could dedicate myself to my doctorate. Thanks also to the State University of Campinas and to the Graduate Program in Genetics and Molecular Biology, for the opportunity to carry out the doctorate.

I thank the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) process 431226/2018-0 and 311811/2019-1 for funding this work. This work was carried out with support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Funding Code 001.

My sincere thanks to the professors Dra. Ângela C. A. Lopes and Dra. Regina L. Ferreira-Gomes who have been with me since graduation. I am also extremely grateful to my advisor Dra. Maria I. Zucchi my advisor for her trust, patience and for sharing her professional experience. I wish that this partnership can continue.

I also thank Dr. Fábio A. Vieira, Dr. Alessandro A. Pereira, Dr. Carlos E. A. Batista, MSc Jonathan A. Morales-Marroquín, MSc Maurício H. Vancine for the help they gave me during the development of this thesis with the experiments and data analysis

I also would like to thank all Conservation Genetics and Genomics´s team for friendship and support in Piracicaba.

“Do nothing from selfish ambition or conceit, but in humility count others more significant than yourselve...For it is God who works in you, both to will and to do of his good pleasure.” (Philippians 2:3-13)

RESUMO

As palmeiras (Arecaceae) é uma das mais diversas famílias de monocotiledôneas (cerca de 2.600 espécies) e exibem uma ampla distribuição nos trópicos e apresentam relevância socioetnobotânica para as populações locais. Dentre as palmeiras de importância socioeconômica, destacam-se a *Copernicia alba* e *Copernicia prunifera*, conhecidas como Carandá e Carnaúba. Carandá é uma palmeira neotropical amplamente distribuída na região do Chaco possui uma ampla gama de usos populares com destaque para a exploração de seu caule na construção civil. Carnaúba possui grande importância econômica e cultural para o Nordeste do Brasil, sendo o pó cerífero obtido a partir do processamento de suas folhas, o principal produto de exploração e comercialização. A intensa exploração dessas palmeiras pelas práticas extrativistas, como corte dos estipes do Carandá e o método de extração da cera de Carnaúba, provocam efeitos adversos sobre a manutenção dessas palmeiras, contribuindo para redução no tamanho populacional e perda da diversidade genética. Além disso, não há estudos sobre o impacto das mudanças climáticas na distribuição dessas espécies no neotrópico. Nesse contexto, o presente estudo tem como objetivo avaliar a genômica populacional, estrutura genética e o impacto das mudanças climáticas no Carnadá e na Carnaúba. Essa tese é inovadora, uma vez que é pioneira na utilização de modelos de distribuição de espécie para predizer áreas climaticamente adequadas para essas palmeiras em cenários atuais e futuros na América do Sul, bem como na avaliação da suscetibilidade dessas espécies diante das mudanças climáticas. Ademais, é a primeira pesquisa a determinar a estrutura e a diversidade genética de populações de carandá e carnaúba via marcadores SNPs (*Single Nucleotide Polymorphism*) utilizando a técnica de genotipagem por sequenciamento (GBS). Com base nos resultados de genômica populacional, observou-se que as populações de Carandá e Carnaúba apresentam altos níveis de diversidade genética. O coeficiente de endogamia foi negativo para todas as populações de ambas as espécies, indicando excesso de heterozigotos. Ambas as espécies apresentaram níveis baixos a moderados de diferenciação genética. Os modelos de nicho ecológico (ENMs) utilizados para prever as áreas climáticas adequadas para a potencial ocorrência dessas espécies em cenários atuais e futuros indicaram que o Carandá e a Carnaúba apresentam diferentes respostas às mudanças climáticas. Os modelos mostram uma redução na distribuição potencial do Carandá, enquanto para Carnaúba observou-se um aumento nas áreas climaticamente adequadas. De modo geral, os resultados destacam a necessidade de esforços para conservação das populações de Carandá e Carnaúba. Entre as opções de manejo sustentável estão as estratégias de conservação *in situ* e *ex situ* e a implantação de unidades de cultivo ou sistemas agroflorestais visando à redução das práticas extrativista. As informações obtidas nesta tese

contribuirão para a conservação da diversidade genética dessas palmeiras, aumentando as chances de persistirem ao longo do tempo.

Palavras-chave: Arecaceae, *Copernicia alba*, *Copernicia prunifera*, Genômica da conservação, Marcadores SNPs.

ABSTRACT

Palms (Arecaceae) is one of the most diverse families of monocots (ca 2600 species). They are widely distributed in the tropics and have socioeconomic relevance for local populations. Among palm trees of socioeconomic importance are *Copernicia alba* and *Copernicia prunifera*, known as Carandá and Carnaúba, respectively. Carandá is a neotropical palm widely distributed in the Chaco region and has a range of popular uses, especially the exploitation of its stem for construction. Carnaúba has economic and cultural importance in Northeast Brazil, as the waxy powder obtained from processing its leaves is the primary product of exploitation and commercialization. The intense exploitation of these palms by extractive practices, such as cutting off the Carandá stipes and the carnaúba wax extraction method, adversely affects their maintenance, contributing to a reduction in population size and loss of genetic diversity. To our knowledge, there are no studies on the impact of climate change on the distribution of these species in the neotropics. This study aimed to evaluate the population genomics, genetic structure, and impact of climate change on carnadá and carnaúba populations. This thesis is innovative since it pioneers the use of species distribution models to predict climatically suitable areas for these palms in current and future scenarios in South America and evaluation of the susceptibility of these species to climate change. To our knowledge, this is the first study to determine the structure and genetic diversity of Carandá and Carnaúba populations with single nucleotide polymorphism (SNP) markers using genotyping by sequencing (GBS). Based on the population genomics results, the Carandá and Carnaúba populations exhibited high levels of genetic diversity. The inbreeding coefficient was negative for all populations of both species indicating excess heterozygotes. Both species showed low to moderate levels of genetic differentiation. Ecological niche models (ENMs) are used to predict suitable climate areas for the potential occurrence of these species under current. Future scenarios indicated that Carandá and Carnaúba show different responses to climate change. The models show a reduction in the potential distribution of caranda. For Carnaúba, an increase in climatically suitable areas was observed. Overall, the results highlight the need for efforts to conserve Carandá and Carnaúba populations. Among the sustainable management options are *in-situ* and *ex-situ* conservation strategies and the implementation of cultivation units or agroforestry systems aimed at reducing extractive practices. The information obtained in this thesis will contribute to the conservation of the genetic diversity of these palms, increasing their chances of persistence over time.

Keywords: Arecaceae, *Copernicia alba*, *Copernicia prunifera*, Conservation genomics, SNPs markers.

SUMMARY

GENERAL INTRODUCTION

The Arecaceae (palm trees) family includes ~2,600 species in 181 genera, of which 48 are native to Brazil and occur in transitional biomes between the Amazon, Cerrado, and Caatinga (Barreto et al. 2019). The country was originally Pindorama, land of many palms in the Tupi Guarani language (Barreto et al. 2019; Souza et al. 2020).

Palms are biologically important because of their relatively high biological richness, density, and biomass among the flora (Pulido-Silva et al. 2022). They have important economic, social, and ethnobotanical roles for the locals, who use them as a source of food and in the production of fuel, medicine, fiber, shelter, and handicrafts (Campos et al. 2019).

Among the main palms with socioeconomic relevance are *Copernicia alba* and *Copernicia prunifera*, popularly known as Carandá and Carnaúba, respectively. *C. prunifera*, a species native to the Brazilian semiarid region, is called the "tree of life" by the sertanejos owing to its multiple uses, highlighting the extraction of the waxy powder used as raw material in the electronics and cosmetics industries (Ximenes-Neto et al. 2019).

C. prunifera is restricted to northeastern Brazil, whereas *C. alba* occurs in the Brazilian Chaco and Pantanal floodplains of Mato Grosso do Sul (Silva, 2018). It is a key species for frugivores during the dry season and serves as a shelter and nesting ground for mammals and birds. In addition, the Carandá presents a diversity of uses by local ethnic groups. The stem is used for building houses. The fruits are used for food, fodder, and oil. (Moraes, 2014).

These palms are often intensively exploited in their areas of occurrence in an extractive manner without the management and stimulation of forest fragmentation, which contributes to the loss of genetic diversity, reduced gene flow, and disadvantageous changes in the frequency of alleles with adaptive value (Reis et al. 2011; Vieira et al. 2015). Due to anthropic actions, it is assumed that some of the genetic variability of the *C. alba* and *C. prunifera* populations has already been lost. Therefore, it is necessary to research to quantify and characterize the genetic diversity between and within populations to define efficient strategies for the conservation and sustainable use of these species.

Several techniques exist for conducting genetic diversity studies, including morphological, biochemical, and molecular markers of random amplified polymorphism (RAPD), amplified fragment length polymorphism (AFLP), inter simple sequence repeats (ISSR), and simple sequence repeats (SSR) (Nadeem et al. 2018). With the advent of next-generation sequencing (NGS), it is possible to identify thousands of single-nucleotide

polymorphism (SNP)-type molecular markers across the genome, enabling genome-wide analysis of genetic diversity (Xia et al. 2019).

Genotyping by sequencing (GBS) is an important technique for the identification and discovery of SNPs in plants (Torkamaneh et al. 2018). GBS is a suitable method for population studies, germplasm characterization, breeding, and genetic mapping. This simple, low-cost, and high-throughput technique can be applied to genotyping models and non-model species (Elshire et al., 2011).

There are few studies on genetic diversity in *C.prunifera*. One can mention the research of Vieira et al. (2015), Pinheiro et al. (2017), and Fajardo et al. (2018) with ISSR markers. However, for *C. alba* no work of this nature has been published thus far. It is noteworthy that there are no studies regarding the diversity and population structure of *C. prunifera* and *C. alba* using SNPs-type markers.

In addition to knowledge of genetic diversity, data on the potential distribution of species are important for the establishment of conservation strategies and the prediction of suitable habitats in future scenarios of climate change. Species distribution models (SDMs) have been used to predict the impacts of climate change on biodiversity and are one of the essential tools for proposing priority areas for conservation (Esser et al. 2019; Bello et al. 2020).

It is worth emphasizing that there are no ecological niche modeling studies for *C. alba*. Santos et al. (2021) estimated the geographical distribution of *C. prunifera* and identified the environmental variables that influence its distribution. However, the model has only one algorithm, MaxEnt, and is limited to Brazil.

This study provides information on genetic diversity and population genomics of these species with SNP markers obtained using the GBS technique. This is the first study to use GBS with NGS technology to obtain SNP markers for these palms. *C. prunifera* populations from the Restinga and Caatinga biomes and *C. alba* populations from the Brazilian Chaco were analyzed. Another point of this research is the identification of loci with outlier patterns associated with adaptation to different environmental conditions in *C. prunifera* since populations in two different ecosystems and their geographical distribution will be sampled. In addition, the climatic niche of these palms was estimated. The effects of climate change on the potential geographic distribution in South America were evaluated.

This thesis is organized into three chapters, presented as scientific papers. The first chapter entitled "**Population genomics of the neotropical palm *Copernicia prunifera* (Miller) H. E. Moore: implications for conservation**" was published in the journal Plos One (doi:10.1371/journal.pone.0276408). This study evaluated the diversity and genetic structure of

14 populations of *C. prunifera* in Northeast Brazil using single nucleotide polymorphisms (SNPs) identified by genotyping by sequencing (GBS), providing information for their conservation. In the second chapter, **"Population genomics and genetic structure of *Copernicia alba* Morong ex Morong & Britton in Brazilian Chaco areas"** submitted for publication in the journal Conservation Genetics, we estimated the genetic diversity and evaluated the population structure of *C. alba* in the Brazilian Chaco using SNP markers, providing a scientific basis for conservation and management efforts in native populations of this palm. The third chapter entitled "**Climate change impacts on the *Copernicia alba* and *Copernicia prunifera* (Arecaceae) distribution in South America"** published in the Brazilian Journal of Botany (doi:10.1007/s40415-022-00801-8), used the ecological niche models (ENMs) to predict the climatic niche of the palms *C. alba* and *C. prunifera* in current and future scenarios, besides evaluating the vulnerability of these species to climate change. The last section of the thesis is composed of perspectives and final considerations about the results synthesized in the three chapters presented here.

BIBLIOGRAPHIC REVIEW

Arecaceae Family

The Arecaceae family emerged in the mid-Upper Cretaceous and is considered one of the oldest families of angiosperms and one of the primary groups in the evolution of monocots (Dransfield et al. 2008; Soares et al. 2014). In the tertiary geological period (Eocene), this family was at the apex of its development with wide dissemination throughout the planet, occupying all continents and constituting two-thirds of the arboreal vegetation (Dransfield et al. 2008).

Palm trees have adapted to different climates and soil types. However, in the warm and humid equatorial regions, such as Malaysia, tropical Asia, and equatorial America, they have progressed to a larger number of species (Lorenzi et al. 2010). The remarkable diversity of palms in tropical humid forests has been attributed to high rates of speciation and diversification compared with other habitats. This has been associated with climatic features, such as high humidity and low seasonality (Cassia-Silva et al. 2019). Furthermore, many species thrive in temperate climates and adapt to drought and fire (Bacon et al., 2016). Notably, ~27% of the total palm flora in the Neotropics are represented by species of the genera *Attalea*,

Coccothrinax, *Syagrus*, *Butia*, *Copernicia*, and *Sabal* in seasonally dry habitats. In Brazil, they are an element present in virtually all plant formations (Cassia-Silva et al. 2019).

Palms are recognized as a model group for understanding the evolution of biomes because they are ecologically representative and have a long and rich fossil history and a well-established geographic, phylogenetic, and taxonomic structure (Bacon, 2016). In the Americas, seven biogeographic regions have been proposed as putative centers of diversification: Mexico, the Caribbean, Central America, the Andean region, the Amazon region, the Brazilian central highlands, and the Atlantic Forest (Eiserhardt et al. 2013).

The family has the following characteristics: stipe simple or occasionally branched, often with thorns; leaves petiolate, simple, pinnatipartite or flabelliform, alternate spiral, or rarely dystic. The inflorescence of the panicle type, lax or compact and spiciform, subtended by a bract, is commonly woody. Flowers are usually inconspicuous, unisexual, or rarely bisexual. The fruits are present as drupes or rarely berries, usually with a single seed (Souza and Lorenzi, 2012).

Along grasses (Poaceae) and legumes (Leguminosae), palms play an essential economic and socioeconomic role as they provide a wide variety of products for rural and indigenous populations (Campos et al. 2019; Smith, 2015). Several species have ornamental (*Coccothinax*, *Livistona*, *Pritchardia*, *Sabal*, *Copernicia*, and *Phoenix*), food (*Cocos nucifera*, *Euterpe edulis*, *Euterpe oleracea*), and biodiesel production (*Acrocomia aculeata* and *Elaeis guineensis*) potential.

Copernicia alba Morong ex Morong & Britton

The palm, *Copernicia alba* Morong ex Morong & Britton, popularly known as Carandá, is economically important and has massive ornamental potential (Lorenzi, 2010). It is a species that constitutes monodominant formations called "carandazais" (Figure 1). It is native to the Chaco region. In Brazil, its natural occurrence is restricted to the states of Mato Grosso and Mato Grosso do Sul (Viana, 2019).

Figure 1. Monodominate formations of carandazais in Porto Murtinho, Mato Grosso do Sul, Brazil (Source: Author's personal file, 2023).

It is a solitary palm tree that can reach up to 30 m in height. The trunk can be 36–71 cm in diameter. The crown is rounded with fan-shaped palmate leaves, bluish-green, and petiole over 1.5 m long. The fruits are 1–1.5 cm in diameter, sub-globose, or ellipsoidal. The seeds are 1 cm in diameter, globose-ovoid, and have a sub-basilar elliptical hilum (Silva, 2018).

According to Viana (2019), this species blooms from July to December, with numerous inflorescences departing from the crown of the leaves. Silva (2018) stated that flowering occurs irregularly and discontinuously at any time of the year and may overlap the fruiting periods, which occur from January to May. It shows fast germination in 25 days, a positive factor for planting and degraded area recovery projects.

C. alba is a species of inestimable value for the inhabitants of the Chaco, since they offer multiple uses and ecosystem services, and have a remarkable symbolic meaning. It has numerous uses (food, construction, and ornamental) (Suarez et al. 2020). The stipe of this palm has ideal characteristics for civil construction. In 2013 the volume of *C.alba* exported in Paraguay was about 500,000 kg (Gauto and Stauffer, 2016).

Copernicia prunifera (Miller) HE Moore

Copernicia prunifera (Mill.) H. E. Moore, popularly known as Carnaúba or Carnaúba palm, is an endemic palm tree of northeastern Brazil used historically by the local population as a source of employment and income due to its economic and cultural importance (Almeida et al. 2021). Carnaúba is often found in low-drainage soils, occurring in the river valleys of the

northeast (Parnaíba, Jaguaraibe, Acaraú, Cauípe, Apodi, and São Francisco), where it forms extensive areas that remain flooded at certain times of the year. In addition to the northeast, it can also occur in the states of Tocantins and Goiás (Figure 2).

Figure 2 Monodominate Carnaúba formations on the coast of Piauí, Brazil (Source: Author's personal file, 2023).

Wax is the primary product of economic value extracted from Carnaúba, which has excellent properties that give it great acceptance in the pharmaceutical, automotive, cosmetic, and computer industries, and the conservation of post-harvest fruits (Ferreira et al. 2013). carnaúba wax was obtained by processing the waxy powder formed on the surface of the leaves. The waxy powder is a protective layer that plays an essential role in plant physiology, assists in reducing transpiration, and protects plants against pathogens (Rodrigues et al. 2013). The fan-shaped leaves serve as forage and shelter for humans and other animals and are also used in handicrafts. The stipe-like stem reaches 10–15 m in height, has a diameter between 15 and 25 cm, and can be used in civil and rural construction (Arruda and Calbo, 2004). Fasciculated roots exhibit several pharmacological properties. This palm is known as the tree of life because of the versatility of its form and use (Costa et al. 2022).

Other aspects that can be highlighted in carnaúba wax are the presence of long-chain fatty acids, esters, free alcohols, aliphatic acids, aromatic acids, triterpenes, proteins, and inorganic compounds such as aluminum, calcium, copper, iron, manganese, magnesium,

sodium, and zinc (Rodrigues et al., 2017). Nevertheless, studies have shown that carnaúba wax has promising pharmacological properties in reducing total cholesterol and triglyceride levels and may be beneficial in the treatment of hyperlipidemia and atherosclerosis (Freitas et al. 2019; Silva et al. 2021).

Population genomics

Population genomics is the simultaneous study of numerous loci or genome regions to understand the roles of evolutionary processes such as mutation, genetic drift, gene flow, and natural selection that influence variation among genomes and populations. Population genomics is an area of genetics that can enhance studies in evolutionary genetics, molecular ecology, and conservation biology by facilitating the identification of adaptive molecular variation and estimating important parameters such as population size, migration rates, and phylogenetic relationships (Luikart et al. 2003).

With the advent of next-generation sequencing (NGS), several methods can be used to discover and genotype thousands of markers in any genome of interest, even in populations where little or no genetic information is available (Davey et al. 2011). Notable among these methods are restriction site-associated DNA sequencing (RAD-seq) (Baird et al. 2008) and GBS (Elshire et al. 2011). Minor changes to the original GBS protocol have been suggested (Poland et al. 2012). GBS is fundamentally similar to RAD-Seq, however the selection of 400-500 bp fragments is not performed, and the protocol is simpler and less expensive (Poland et al. 2012).

Genotyping by sequencing methods have several similarities and involve the following steps: a) digestion of multiple genomic DNA samples by restriction enzymes, b) binding of adapters with specific barcode sequences for the identification of individuals, c) mixing of the fragments into a single pool of samples (multiplexing), d) purification and amplification by polymerase chain reaction (PCR), and e) sequencing by NGS of the final set of fragments. The difference between the methods is that the protocol with modifications extends to a system that includes a "rare cut" and a "frequent cut" enzyme, while the original approach uses a single restriction enzyme to capture the genomic sequence between restriction sites (Figure 3).

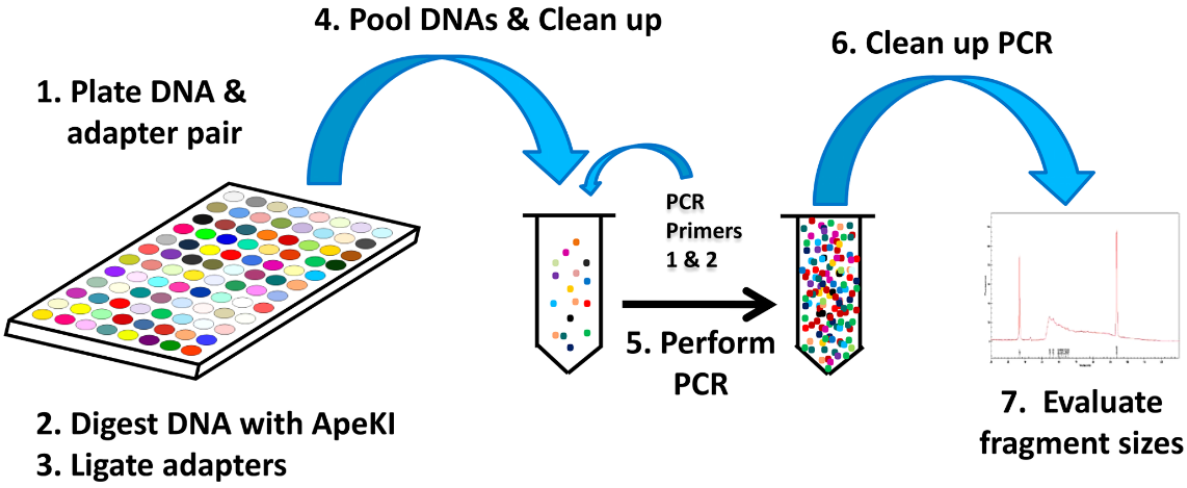


Figure 3. Steps in GBS library construction (Image modified from Elshire et al. 2011/doi: 10.1371/journal.pone.0019379)

Genotyping by sequencing can be used to identify markers of single nucleotide polymorphisms (SNPs) in the genome. These are used in genomic diversity studies in native plants: *Croton tetradenius* (Brito et al. 2021), *Avicennia germinans* (Cruz et al. 2020), *Centrolobium tomentosum* (Cordeiro et al. 2019), *Casearia sylvestris* (Viana et al. 2018), and *Euterpe edulis* (Novello et al. 2017). A large number of SNPs also makes it easier to identify genes or regions that show signs of selection by examining which of the thousands of SNPs shows significant differences between populations (Narum et al. 2013).

Ecological Niche Models

Ecological niche models (ENMs), also known as species distribution models (SDMs), are empirical or mathematical approaches that configure the relationship between species distribution and environmental conditions and are heavily employed to investigate the potential geographic distribution of species (Peterson et al. 2011; Feng et al. 2019).

The advent of computational packages and the use of geographic information system (GIS)-based approaches have considerably aided the development of studies on ENMs in the last decade such that more than 1,000 publications related to this topic are released every year (Zurell et al. 2020). ENMs are useful for mapping invasive species (Bello et al. 2020; Mothes et al. 2019), planning priority areas for conservation (Guisan et al. 2013; Peterman et al. 2013; Sobral-Souza et al. 2018), identifying possible routes of infectious disease spread (Silva et al. 2018; Chalghaf et al., 2018), and predicting the impacts of climate change on the geographic distribution of native and cultivated species (Blach-Overgaard et al. 2015; Liu et al. 2021).

Ecological niche models are classified into two groups: mechanistic (mechanistic modeling) and correlative (correlative modeling). The mechanistic model estimates the potential geographic distribution of the species from its physiological tolerances, while the correlative model statistically establishes an association between the locations of the species and the environmental conditions (Alvarado-Serrano and Knowles, 2014). Correlative models are more feasible and used because environmental and species occurrence data are abundant and readily obtained from online documents and sources, such as the Global Biodiversity Information Facility (GBIF), NatureServe, International Union for Conservation of Nature (IUCN), Neotoma Paleoecology Database and Community, and SpeciesLink (Sillero and Barbosa, 2020; Title and Bemmels, 2018).

For correlative models, biotic and abiotic data are used. Biotic data corresponded to the occurrence records of species obtained from geographic information and virtual databases. The abiotic data are the environmental variables. The 19 most commonly used bioclimatic variables for SDMs are those from WorldClim due to their high resolution, global coverage, and availability for historical climate and future scenarios (Hijmans et al. 2005; Fick and Hijmans 2017).

Several algorithms exist for modeling the ecological niche of species as a function of environmental variables. The choice of an algorithm should be based on the study question and the availability of occurrence data. These algorithms can be classified into three groups: i) envelope methods, such as BioClim, Mahalanobis distance, and ecological-niche factor analysis (ENFA); ii) statistical methods, for example, the generalized additive model (GAM), multivariate adaptive regression splines (MARS), and generalized linear models (GLM); and iii) machine-learning models, such as maximum entropy (MaxEnt), artificial neural networks (ANN), and random forests (Rangel and Loyola, 2012).

The area under the curve (AUC) calculation is one way of interpreting the results obtained by the algorithm and validating the models (Peterson et al. 2011). AUC values between 0.90 and 1.0 are considered excellent, between 0.80 to 0.90 are good, from 0.70 to 0.80 is average, between 0.60 to 0.70 are poor, and below 0.60 are considered very poor. For publications, the acceptable AUC values should be above 0.75.

OBJECTIVES

General objective

The aim of this study was to evaluate the population genomics, genetic structure, and impact of climate change on the Neotropical carnadá and carnaúba palms for the conservation of the species.

Specific objectives

a. To estimate the genetic diversity of carnaúba populations using SNP markers and to identify loci that may be under selection influence (outliers) and indicate differences between Carnaúba populations based on the distinct ecosystems, Restinga and Caatinga.
b. To analyze diversity and genetic structure between Carandá populations, moreover, to identify candidate loci that are influenced by selection across (outliers) the five assessed locations.
c. To evaluate the vulnerability of Carandá and Carnaúba populations to climate change.

CAPÍTULO I

Population genomics of the neotropical palm *Copernicia prunifera* (Miller) H. E. Moore: implications for conservation

Autores: Marcones Ferreira Costa. Jonathan Andre Morales-Marroquín. Carlos Eduardo de Araújo Batista. Alessandro Alves-Pereira. Fábio de Almeida Vieira. Maria Imaculada Zucchi

Published in Plos One

https://doi.org/10.1371/journal.pone.0276408

Abstract

Copernicia prunifera (Miller) H. E. Moore is a palm tree native to Brazil. The products obtained from its leaf extracts are a source of income for local families and the agroindustry. Owing to the reduction of natural habitats and the absence of a sustainable management plan, the maintenance of the natural populations of this palm tree has been compromised. Therefore, this study aimed to evaluate the diversity and genetic structure of 14 *C. prunifera* populations using single nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing (GBS) to provide information that contributes to the conservation of this species. A total of 1,013 SNP markers were identified, of which 84 loci showed outlier behavior and may reflect responses to natural selection. Overall, the level of genomic diversity was compatible with the biological aspects of this species. The inbreeding coefficient (*f*) was negative for all populations, indicating excess heterozygotes. Most genetic variations occurred within populations (77.26%), and a positive correlation existed between genetic and geographic distances. The population structure evaluated through discriminant analysis of principal components (DAPC) revealed low genetic differentiation between populations. The results highlight the need for efforts to conserve *C. prunifera* as well as its distribution range to preserve its global genetic diversity and evolutionary potential.

Introduction

Habitat reduction and deforestation resulting from human activities have had adverse effects on forest populations, contributing to high rates of species extinction, particularly in the Neotropical region [1]. With the exception of some natural areas, most tropical species occur in anthropogenic landscapes, where the previously continuous forest has now been reduced to smaller and isolated patches [2]. This modification of landscape composition and structure leads to habitat fragmentation, contributing to the loss of alleles, reduction of heterozygosity, and increase in inbreeding [3].

The palm tree *Copernicia prunifera* (Mill.) H. E. Moore (Arecaceae; subfamily: Coryphoideae), known as carnaúba, generally forms monodominant populations known as "carnaubais" [4]. The species has multiple inflorescences, which are made up of yellowish and hermaphroditic flowers [5]. Flowering is more intense between November and February, and the fruiting period is between January and March [6]. Fruits are likely dispersed by sanhaçu-do-coqueiro (*Tangara palmarum*) [5].

C. prunifera is endemic to the Caatinga biome [7], which is one of the largest seasonally dry tropical forest areas in South America [8]. The Caatinga is an exclusively Brazilian biome, covering an area of approximately 900,000 km^2 in northeast Brazil. The climate in this region is characterized by a long dry season with irregular rainfall, representing a xeric, semi-deciduous shrubland and forest vegetation [9]. This palm tree also grows in the Restinga region, which contains vegetation of the coastal plain under marine influence and established on sandy soil composed of physiognomic variations from the beach towards the interior of the coastal plain [10, 11]. Caatinga and Restinga are the major vegetation units in northeast Brazil [12].

Local populations use this species as a source of employment and income. Leaf extraction is responsible for sustaining several families during the period of drought that extends from July to December in northeast Brazil [13]. The fruits serve as food for animals, the stems can be used in the construction of houses, and the fasciculated roots have medicinal properties [14]. Due to its versatility and usefulness, this palm tree is known as "the tree of life." [15]. The main product of economic value obtained from this tree is carnaúba wax, which is extracted from young leaves and is of interest in the pharmaceutical and automotive industries [16]. Carnaúba populations suffer from intense exploitation because the method used to extract carnaúba wax, which consists of practically removing all the leaves of the plant to obtain ceriferous powder [17].

However, unsustainable harvesting practices and the absence of sustainable management programs pose major threats to the long-term survival of this palm tree species. *C. prunifera* populations show signs of intense exploitation with visible signs of anthropogenicity, such as fire, extraction and cutting of leaves, soil impacted by livestock, and low or absence of regeneration (Fig. 1). In addition, anthropogenic disturbances in the last century, mainly due to deforestation, agricultural expansion, and modernization of agriculture, have led to a rapid decline in these populations [17].

Fig. 1 Signs of anthropization in the evaluated populations of *C.prunifera*. a) wood cutting; b) leaf extraction and cutting; c) cattle raising; and d) burns.

The maintenance of genetic diversity is a powerful conservation strategy for preserving the adaptive potential of species in neotropical regions [18]. In addition to configuring the ability of species to adapt to various changing environments, genetic diversity is the driving force behind evolution and speciation. [19]. Consequently, maintenance of genetic diversity within populations ensures that the species can remain biologically active and adaptable to structural changes caused by anthropogenic actions [20, 21].

The genetic diversity and structure of forest populations evaluated based on molecular markers is a widely used strategy in conservation genetics [22, 23]. With the advent of next-generation sequencing, it is possible now to identify thousands of molecular markers of single nucleotide polymorphism (SNP) throughout the genome. This provides a genomic approach to evaluating genetic diversity [24]. A larger SNP sample size facilitates the

identification of regions that show signs of selection and can serve as a starting point for the identification of adaptive differences between populations, which is fundamental for optimizing biological conservation efforts [25, 26]. These markers enable the identification of outlier and neutral loci. Specifically, outlier loci show differentiated behavior regarding genetic variation and offer an opportunity to evaluate local adaptation patterns; neutral loci are similarly affected by the demographic and evolutionary history of populations [27].

Genetic diversity studies based on molecular markers of natural populations of *C. prunifera* in tropical areas such as Caatinga and Restinga are still scarce [28, 29, 30-17]. In addition, no studies on *C. prunifera* have applied next-generation sequencing technology for data acquisition in population genomics. Due to the importance of this neotropical palm tree for local communities and considering the rapid and recent increases in the exploitation of its populations, the present study employed next-generation sequencing to evaluate the genetic diversity and structure of 14 natural populations of *C. prunifera* in two environments (Caatinga and Restinga) in Brazil using SNP markers to provide information that can help in the design of efficient strategies for the conservation and sustainable use of this species.

Material and Methods

Plant material and DNA extraction

In the present study, 160 individual plants from 14 populations of *C. prunifera* were evaluated. Out of the samplings collected, 10 populations came from the Caatinga (RUS, LGP, SER, MACZ, MACE, JUC, APD, IPG, MOS, and MAT) and four from the Restinga (ICA, SMG, AR1, and AR2) regions in the states of Ceará and Rio Grande do Norte, Brazil (Fig. 2 and S1 Table). The distance between the plants evaluated within the 14 populations was 15–20 m, with a minimum height of 6–10 m; regenerating and young plants were not collected. The IPG population is composed of a different type of carnaúba, known as “white carnaúba,” which is phenotypically distinct from the “common carnaúba” due to the presence of a light stipe, smaller fruits, and the absence of thorns in the petiole in addition to limited occurrence in the region [14].

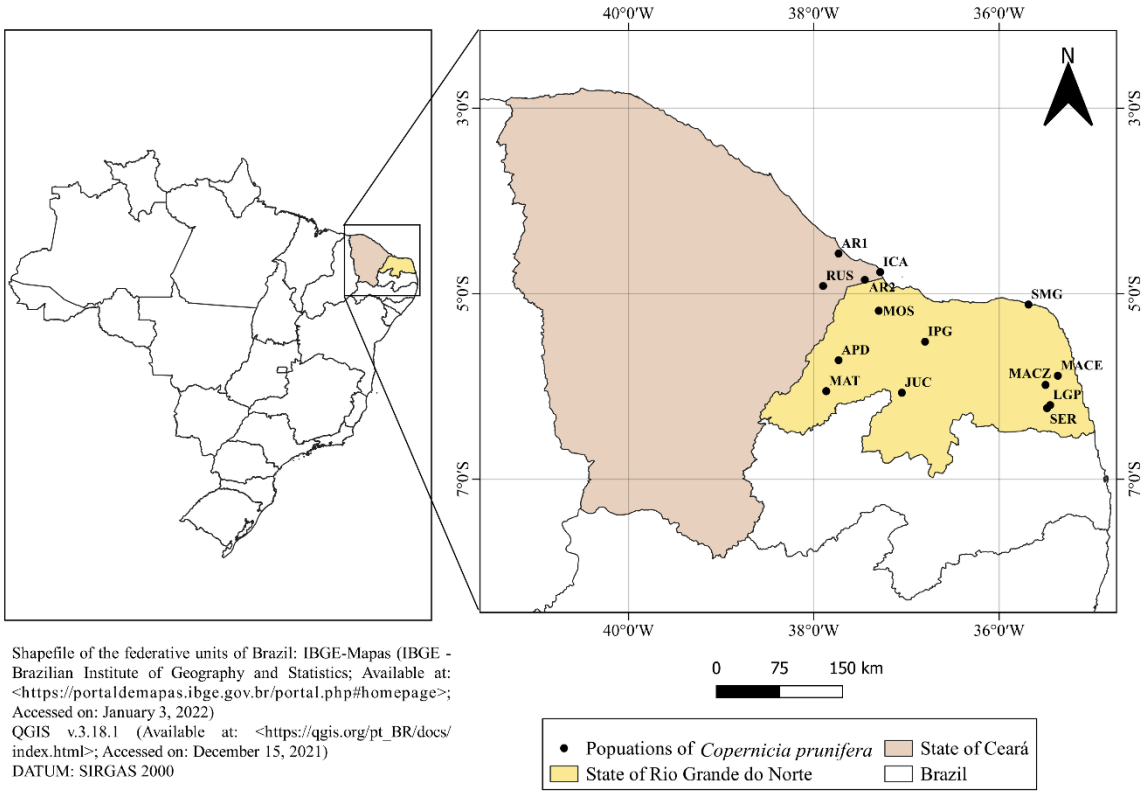


Fig.2 Map of the collection sites of *C. prunifera* populations in the states of Ceará and Rio Grande do Norte, Brazil. Distribution map of the evaluated populations was drawn using the software QGIS v3.18.1. (Open Access Geographic Information System, https://qgis.org/pt_BR/site). This figure is licensed under CC BY 4.0.

Small pieces of leaves were cut using a tree trimmer, placed in plastic tubes containing 2 mL of hexadecyltrimethylammonium bromide (CTAB 2X), labeled, and stored in a freezer at -20 °C until DNA extraction. This study was conducted according to the recommendations of the Brazilian Ministry of the Environment and registered in the National System of Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN; *Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado*) with the number A411583.

Genomic DNA was extracted from the processed leaves according to the protocol described by Doyle and Doyle [31]. DNA quality was evaluated in a 1% agarose gel and stained with SYBR Safe™ (Life Technologies Corporation) for visualization under ultraviolet light, using lambda phage DNA of known concentrations as a reference. Quantification of the samples

was performed using a Qubit 3.0 fluorometer with the dsDNA BRKitt (Life Technologies), and the DNA was standardized to a concentration of 30 ng.μl^{-1}

GBS library preparation and high-performance sequencing

To obtain SNPs, genomic libraries were developed using the genotyping-by-sequencing technique (GBS) with two restriction enzymes, according to the protocol described by Poland et al. [32] with modifications. First, 7 µl of genomic DNA from each sample was digested at 37 °C for 12 h using the restriction enzymes *NsiI* and *MspI*. Subsequently, 0.02 µM barcode-specific adapters for Illumina technology were ligated to the ends of the digested fragments. Binding reaction was performed at 22 °C for 2 h, 65 °C for 20 min, 10 °C indefinitely. After the adapters were ligated, the samples were purified using a QIAquick PCR Kit (Qiagen). The library was enriched by PCR (Polymerase Chain Reaction) using the following amplification program: 95 °C for 30 s, followed by 16 cycles of 95 °C for 10 s, 62 °C for 20 s, and 72 °C for 30 s, and ending at 72 °C for 5 min. Finally, the library was purified using a QIAgen® QIAquick PCR Purification Kit. The Agilent DNA 12000 kit and Agilent® 2100 Bioanalyzer System were used to verify the average size of the DNA fragments. Sequencing was performed using the Illumina® HiSeq 2500 Mid Output Kit v4 (50 cycles) (Illumina Inc., San Diego, CA, USA) in a single-end configuration.

Identification of SNPs

The identification of SNPs was performed using Stacks software v.1.42 [33, 34]. The first step comprised filtering and demultiplexing with the *process_ radtag* module. In the absence of a reference genome for *C. prunifera*, the *DeNovo* Stacks pipeline was used, starting with the *ustacks* module to identify putatively homologous read stacks (putative loci). This step was performed for each sample separately using the following parameters: minimum stack depth (-m = 3) and maximum distance between stacks (-M = 2). The loci of each sample were grouped into a catalog using the *cstacks* module, allowing a maximum distance of two nucleotides (-n 2) between the loci of each sample. Loci with lower probability values (--lnl_lim -10) were eliminated using the *rxstacks* correction module. Finally, the *population* module was used to filter the SNP markers using the following parameters: only one marker per sequenced tag, frequency of least frequent allele (MAF $\geq$ 0.01), minimum stack depth 3X, and minimum occurrence in 75% of saplings in each population.

Loci determination under selection

Two complementary tests were performed, pcadapt and fsthet, were performed to detect outlier loci (hypothetically under selection). The pcadapt method [35] was used to identify loci associated with the genetic structure revealed by a principal component analysis (PCA), that is, without any underlying genetic model. The analysis was performed using the pcadapt package [35] on the R platform [36] by retaining the first eight principal components of the PCA and considering the loci with q-values ≤ 0.1 as outlier SNPs. The fsthet method [37] was used to identify loci with F_{ST} values that were excessively high or low compared with what was expected under neutrality. The analysis was performed using the fsthet package [37] on the R platform [36] by considering the loci below or above the 95% confidence intervals constructed with 1000 bootstraps for the expected relationship between H_E and F_{ST} as outlier SNPs. This test was performed by considering the estimates of F_{ST} in two different scenarios: i) comparing the Restinga and Caatinga populations and ii) comparing the samples of the white morphotype with those of conventional morphotype. The final set of SNP markers hypothetically under selection consisted of loci identified as outliers in at least two of the three tests performed. Thus, the outlier SNP loci may reflect the action of selection on different types of vegetation.

Sequences containing outlier SNPs were searched with the BLASTX tool against the genomic data set of the National Center for Biotechnology Information (NCBI) using blast2go [38]. This analysis was performed to identify the similarities between the protein-coding data deposited in the NCBI database and the loci with outlier SNPs identified. For sequences with significant BLASTX hits, the functional annotation associated with characterized and/or described coding sequences was performed using the gene ontology system (GO terms). GO terms summarize information on cellular components, molecular functions, and biological processes in which the gene products are involved.

Population genomic analyses

Genetic diversity was estimated based on the number of alleles (A), number of private alleles (A*p*), observed heterozygosity (H_O), and expected heterozygosity (H_E). Inbreeding coefficients (*f*) were also estimated, and their confidence intervals were obtained using 1000 bootstraps. Estimates of diversity and inbreeding were obtained using the diveRsity [39] and PoPPr [40] packages of the R software [36]. The distribution of genetic variation within and between populations of *C. prunifera* was evaluated using analysis of molecular

variance (AMOVA), and its significance was tested with 10,000 permutations using the PoPPr program [40].

Genetic differentiation was estimated using pairwise F_{ST} values with confidence intervals of 1000 bootstraps, using the diveRsity package [39] of R software [36]. The population structure was evaluated using discriminant analysis of principal components (DAPC) with Adegenet [41,42] for R software [36]. This analysis was performed for neutral loci, and *priori* groups were defined from the 14 sampling sites. DAPC does not presuppose the underlying population genetic processes (e.g., binding equilibrium and Hardy–Weinberg equilibrium) common to other methods used to detect population structure, and as it is based on principal component analysis, this method can analyze genomic datasets relatively efficiently [43].

Genetic relationships and divergence between individuals were investigated by constructing a dendrogram generated based on the distance of Nei using the neighbor-joining method [44]. The final dendrograms were formatted using MEGA version 7 [45].

Results

Identification of SNPs and determination of loci under selection

Sequencing of the genomic libraries resulted in 566,922,165 reads, and after quality control, the total number of reads retained was 397,047,980. In total, 1,013 SNPs (average depth of 21X) were identified. A total of 391 outlier SNPs were identified using the pcadapt method, 70 using the fsthet method to compare morphotypes, and 47 using the fsthet method to compare vegetation types (Fig. 3). Of these, 84 SNP markers were identified in at least two of the three tests and were considered hypothetically under selection, whereas the other 929 markers were considered as neutral loci. Among the outlier loci, 55 were putatively under positive selection and 29 were putatively under balancing selection. Only six outlier loci were found in the sequences similar to the annotated proteins (S2 Table). Considering the results of the GO terms, the most frequent annotations for these proteins were the molecular functions of “binding” and “catalytic activity,” and the biological process of cell metabolism (S1 Fig.).

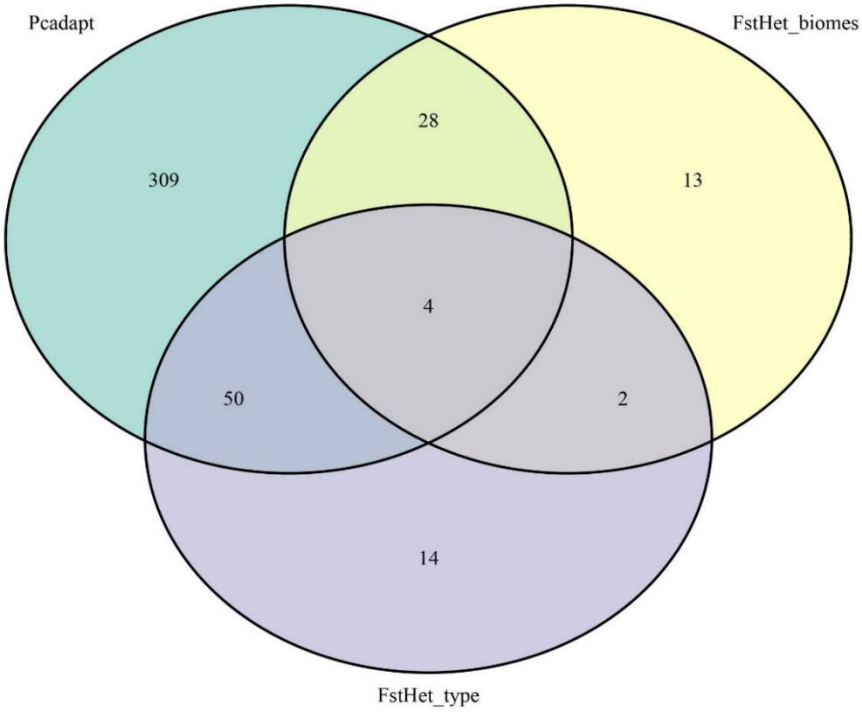


Fig. 3 Venn diagram with the number of outlier loci detected for the fsthet and Pcadapt tests with the overlap between them.

Population genomic analyses

The genomic diversity estimates were based on 929 neutral SNP markers. The number of alleles (A) ranged from 1,059 to 1,497. The IPG population (white carnaúba) had the lowest number of alleles, probably because of the small sample size of the population. Expected heterozygosity (H_E) ranged from 0.201 to 0.265 (Table 1). The APD population had higher genetic diversity (H_E = 0.265) and the highest number of private alleles (Ap = 40) compared with that of the other populations. The inbreeding coefficients (*f*) were similar and negative for all populations, indicating an excess of heterozygotes.

Table 1. Estimates of genomic diversity and inbreeding based on 929 neutral SNP markers for populations of *C. prunifera*.

Population	N	A	AP	H_O	H_E	*f*	*f* (95% CI)
ICA	7	1439	2	0.375	0.257	-0.367	(-0.473 - -0.308)
AR1	17	1497	35	0.387	0.260	-0.391	(-0.439 - -0.357)
AR2	11	1484	3	0.388	0.258	-0.454	(-0.526 - -0.394)
RUS	21	1495	27	0.383	0.260	-0.287	(-0.556 - -0.326)
SMG	10	1353	16	0.370	0.235	-0.272	(-0.34 - -0.211)
LGP	6	1059	5	0.378	0.229	-0.553	(-0,642 - -0.473)
SER	10	1235	6	0.381	0.246	-0.438	(-0,556 - -0.326)
MACZ	10	1462	9	0.387	0.259	-0.405	(-0,461- -0.365)
MACE	14	1227	7	0.369	0.201	-0.102	(-0,126- -0.081)
JUC	12	1466	3	0.376	0.247	-0.447	(-0,564 - -0.38)
APD	12	1317	40	0.372	0.265	-0.238	(-0,383 - -0.0127)
IPG	5	1249	1	0.375	0.224	-0.123	(-0,209 - -0.049)
MOS	18	1404	28	0.380	0.255	-0.112	(-0.174 - -0.065)
MAT	7	1361	8	0.376	0.237	-0.353	(-0.437- -0.306)

Number of individuals (N); total number of alleles (A); number of private alleles (Ap); observed heterozygosity (H_O); expected heterozygosity (H_E); inbreeding coefficient (*f*); *f* (CI of 95%) = lower and upper confidence interval of 95% of the inbreeding coefficients.

F_{ST} values 0–0.05 and 0.05–0.15 indicate low and moderate genetic differentiation, respectively, whereas values > 0.15 indicate high differentiation (Hartl, Clark 1997). In the present study, F_{ST} estimates suggested low to high genetic differentiation between populations of *C. prunifera* (Table 2). In general, there was a greater differentiation between the population from MACE and those from the other sites (F_{ST} ranged from 0.118 to 0.20). In addition, the SMG and IPG populations showed moderate levels of differentiation. A low genetic structure was observed for the populations from LGP and SER (0.18) and AR1 and AR2 (0.19), suggesting a genetic flow between these localities.

Table 2. Estimates of pairwise F_{ST} between populations of *C. prunifera* (lower diagonals). Upper diagonals contain the lower and upper limits of the confidence interval.

	IPG	APD	AR2	JUC	LGP	MACE	MACZ	MAT	MOS	RUS	SER	AR1	ICA	SMG
IPG		(0.05 - 0.14)	(0.02 - 0.11)	(0.05 - 0.20)	(0.05 - 0.17)	(0.17 - 0.27)	(0.03 - 0.12)	(0.06 - 0.19)	(0.04 - 0.14)	(0.03 - 0.13)	(0.05 - 0.16)	(0.04 - 0.13)	(0.02 - 0.13)	(0.05 - 0.16)
APD	0.092		(0.01 - 0.07)	(0.01 - 0.12)	(0.03 - 0.12)	(0.12 - 0.19)	(0.02 - 0.08)	(0.06 - 0.13)	(0.02 - 0.08)	(0.01 - 0.06)	(0.02 - 0.09)	(0.02 - 0.07)	(0.00 - 0.09)	(0.07 - 0.14)
AR2	0.053	0.036		(0.02 - 0.11)	(0.03 - 0.10)	(0.13 - 0.20)	(0.01 - 0.05)	(0.03 - 0.10)	(0.01 - 0.06)	(0.00 - 0.04)	(0.02 - 0.08)	(0.00 - 0.04)	(-0.03 - 0.03)	(0.04 - 0.09)
JUC	0.109	0.062	0.057		(0.03 - 0.15)	(0.11 - 0.18)	(0.01 - 0.09)	(0.01 - 0.15)	(0.05 - 0.14)	(0.03 - 0.11)	(0.02 - 0.10)	(0.03 - 0.11)	(0.02 - 0.12)	(0.06 - 0.14)
LGP	0.096	0.064	0.054	0.072		(0.10 - 0.16)	(0.02 - 0.09)	(0.06 - 0.15)	(0.06 - 0.13)	(0.04 - 0.09)	(-0.01 - 0.07)	(0.02 - 0.08)	(0.01 - 0.11)	(0.05 - 0.13)
MACE	0.210	0.150	0.162	0.144	0.123		(0.09 - 0.15)	(0.17 - 0.23)	(0.15 - 0.21)	(0.13 - 0.17)	(0.09 - 0.15)	(0.12 - 0.16)	(0.14 - 0.21)	(0.12 - 0.17)
MACZ	0.063	0.046	0.027	0.042	0.051	0.118		(0.04 - 0.10)	(0.03 - 0.09)	(0.01 - 0.05)	(0.02 - 0.08)	(0.02 - 0.06)	(0.00 - 0.07)	(0.03 - 0.08)
MAT	0.113	0.087	0.058	0.073	0.094	0.195	0.063		(0.06 - 0.12)	(0.04 - 0.09)	(0.06 - 0.12)	(0.04 - 0.10)	(0.03 - 0.10)	(0.08 - 0.15)
MOS	0.077	0.047	0.032	0.084	0.086	0.172	0.050	0.084		(0.02 - 0.07)	(0.05 - 0.10)	(0.03 - 0.07)	(0.00 - 0.07)	(0.06 - 0.11)
RUS	0.068	0.033	0.016	0.060	0.057	0.148	0.030	0.054	0.041		(0.04 - 0.09)	(0.01 - 0.04)	(0.00 - 0.05)	(0.05 - 0.11)
SER	0.094	0.053	0.045	0.054	0.018	0.118	0.046	0.080	0.069	0.063		(0.03 - 0.07)	(0.02 - 0.08)	(0.05 - 0.10)
AR1	0.074	0.041	0.019	0.062	0.043	0.138	0.034	0.064	0.045	0.019	0.048		(0.01 - 0.07)	(0.05 - 0.10)
ICA	0.060	0.036	-0.003	0.058	0.050	0.171	0.028	0.055	0.032	0.022	0.044	0.029		(0.05 - 0.11)
SMG	0.090	0.101	0.064	0.098	0.081	0.142	0.048	0.102	0.079	0.077	0.070	0.073	0.072	

The low genetic divergence suggested by the pairwise F_{ST} was also observed in DAPC, which retained 28.7% of the total variation in the first two principal components (Fig. 4). This analysis also showed greater genetic differentiation of the population from MACE in comparison with that of others in addition to pointing out an overlap between individuals from almost all populations, especially AR1, AR2, and RUS.

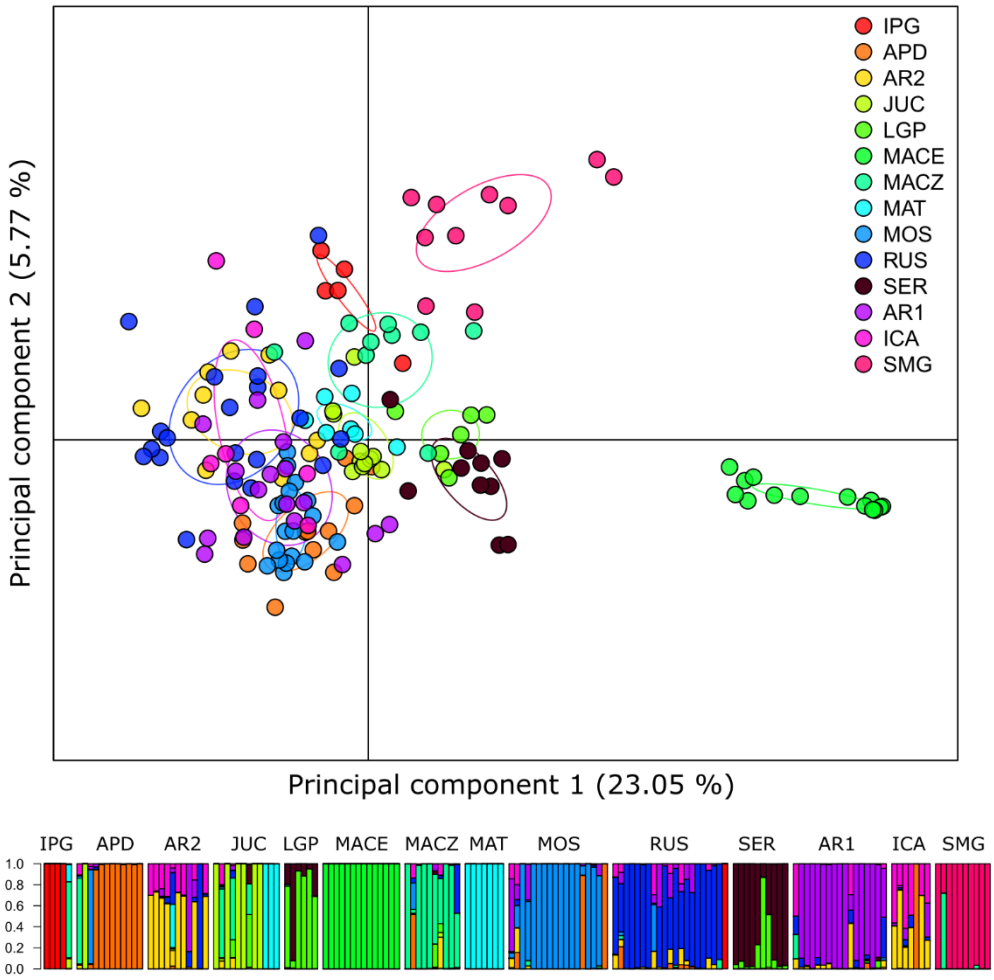


Fig. 4 a) Discriminant analysis of principal components (DAPC) representing the genetic structure of *C. prunifera* populations based on 929 SNPs. b) Bar graph representing the coefficients of DAPC, where each bar delimits one individual.

Analysis of molecular variance (AMOVA) indicated that most of the variation was found within populations (77.26%), and the genetic differentiation between populations was high and significant ($\varphi = 0.227$) (Table 3). The Mantel test revealed a positive and significant correlation between the geographical and genetic distances based on the F_{ST} values ($r = 0.0612$; $p = 0.002$).

Table 3. Analysis of molecular variance (AMOVA) based on 929 neutral SNP markers for fourteen natural populations of *C. prunifera*.

	Degrees of freedom	Sum of the squares	Mean square	Variance	Percentage	F_{ST} global (φ)	*p*-value
Between the populations	13	3307.963	254.45867	17.32471	22.7366	0.227366	0.00005
Within the population	146	8595.417	58.87272	58.87272	77.2634		
Total	159	11903.38	74.86402	76.19743			

According to the dendrogram (Fig. 5), the MACE population was the most genetically distant, corroborating the results observed in the DAPC population. Individual saplings from the LGP and SER populations exhibited similar levels of genetic similarity. In addition, there was a clear distinction among the three groups: the first group was formed by the populations from LGP, SER, MACE, SMG, MACZ, JUC, and MAT; the second group consisted of AR2, ICA, IPG (white carnaúba), and MOS; and the third group consisted of RUS, AR1, and APD. Populations from AR2 and ICA had the highest bootstrap value, which indicates a statistically well-supported grouping.

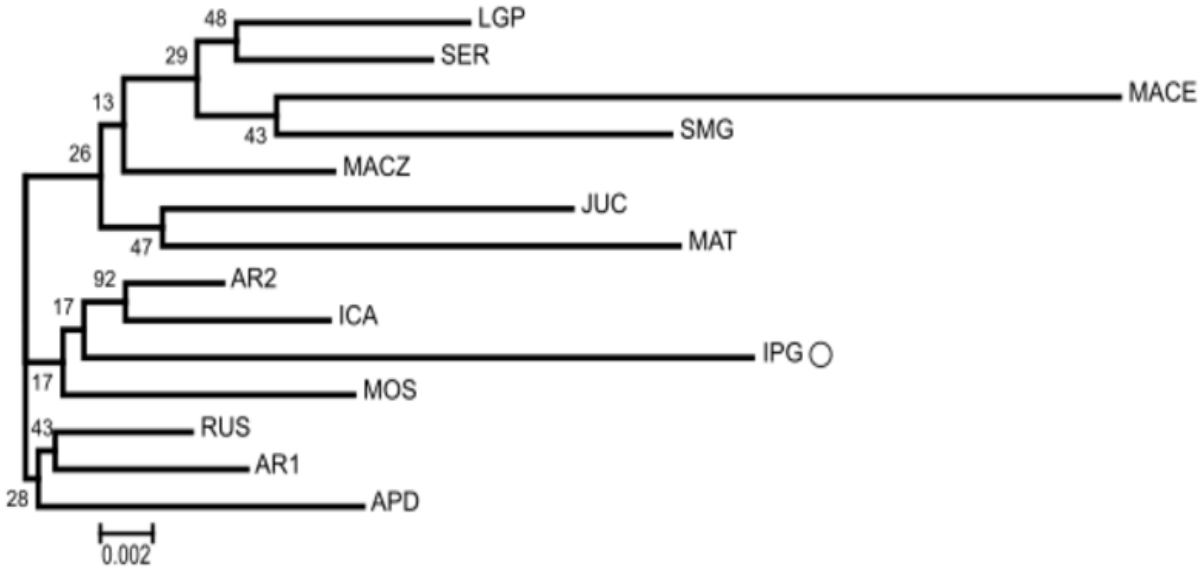


Fig. 5 Dendrogram obtained by the Neighbor-Joining method based on SNP markers for the 14 populations of *C. prunifera*.

When the genetic diversity and structure of the *C. prunifera* populations were evaluated based on the type of vegetation (Caatinga and Restinga), similar levels of genetic diversity were observed (Table 4).

Table 4. Estimates of genomic diversity and inbreeding based on 929 neutral SNP markers for populations of *C. prunifera*, considering the different types of vegetation (Caatinga and Restinga) and morphotypes (common carnaúba and white carnaúba).

Population	**N**	**A**	**AP**	**H_O**	**H_E**	***f***	***f* (95% CI)**
Caatinga	126	1802	253	0.379	0.270	-0.400	(-0.417 - -0.382)
Restinga	34	1603	54	0.380	0.266	-0.432	(-0.468 - -0.397)
Common carnaúba	155	1855	607	0.348	0.271	-0.400	(-0.417 - -0.384)
White carnaúba	5	1249	1	0.345	0.224	-0.123	(-0.209 - -0.048)

Number of individuals (N); total number of alleles (A); number of private alleles (Ap); observed heterozygosity (H_O); expected heterozygosity (H_E); inbreeding coefficient (*f*); *f* (CI of 95%) = lower and upper confidence interval of 95% of the inbreeding coefficients.

The populations from the Caatinga had the largest number of private alleles (Ap = 253) compared to the Restinga populations and this result is probably associated with the sample size. The inbreeding coefficients (*f*) are both similar and negative. In addition, the F_{ST} estimates suggested low genetic differentiation between the Caatinga and Restinga populations (F_{ST} = 0.008). When considering only two morphotypes of *C. prunifera* (white carnaúba and common carnaúba), similarities were observed in the estimates of diversity in addition to low genetic differentiation (F_{ST} = 0.008) (Table 4). Analysis of molecular variance (AMOVA) among the vegetation types (Caatinga and Restinga) produced small genetic differentiation ($\varphi = 0.024$). It revealed 97.522% of the genetic variation within the vegetation types whereas, 2.478% of the total genetic variation was observed between types of vegetation (Table 5).

Table 5. Molecular analysis of variance (AMOVA) considering the Caatinga and Restinga for the populations of *C.prunifera*.

	Degrees of freedom	Sum of the squares	Mean square	Variance	Percentage	F_{ST} global (φ)	*p*-value
Between types of vegetation	1	175.234	175.23395	1.886185	2.478	0.02478074	0.00005
Within types of vegetation	158	11728.146	74.22877	74.22877	97.522		
Total	159	11903.38	74.86402	76.114954			

Discussion

Loci Putatively under Selection

The large number of SNP markers obtained in this study allowed for the identification of loci with deviations from the expected neutral behavior, which are putatively under selection (outlier loci). The identification of outlier loci is an important step in understanding local adaptation and evaluating the evolutionary potential of a species [46]. The palm tree *C. prunifera* has no annotated reference genome, and probably for this reason, most sequences with outlier loci are similar to uncharacterized proteins. Regarding the results obtained from the annotation, most loci are associated with genes involved in metabolic processes, which have been regularly found under selection in a variety of organisms because the gene functionality correlates with environmental stressors [47].

Interestingly, some annotated loci were associated with genes of transposable elements (S2 Table). According to Gogvadze and Buzdin [48], transposable elements promote changes in the genome, which is an important evolutionary mechanism for the adaptation of organisms to changes in environmental conditions. This is expected in *C. prunifera* because the palm trees grow in different environments such as seasonally flooded areas in the semi-arid region [49]. In addition, outlier loci may be associated with environmental differences in the collection sites, especially as sampling areas are scattered over the Restinga and Caatinga.

It is important to highlight that the analyses performed in this study are unable to indicate associations between genomic and functional variation; therefore, it is not possible to associate generic molecular functions or biological processes with any adaptive traits involved in the diversification of the evaluated populations. Therefore, studies with larger sample sizes with better representation of the different geographical habitats are needed to generate

information on the evolution and diversification of *C. prunifera.* Small sample sizes belonging to populations with relatively small geographical distances, which enable gene flow to quickly spread new adaptations to surrounding areas, reduce the capacity to detect recent evolutionary changes [50]. However, the identified outlier loci can be used as candidates in association mapping studies. Thus, integrative approaches of association genetics, genome-wide scans, and measurements of phenotype selection are necessary to understand the adaptive nature of a given allele [51].

Genetic diversity, inbreeding, and structure

Genetic diversity is one of the three classes of biodiversity recognized as a global conservation priority and plays a decisive role in conservation efforts. Genetic diversity has a substantial effect on both individual fitness and the adaptive capacity of the population, playing a vital role in maintaining the capacity of species to withstand various biotic and abiotic stressors and evolve under altered environmental conditions [52]. The present study provides the first estimates based on SNPs for genetic diversity in *C. prunifera*. The GBS approach used in this study produced a large number of SNP loci for the genomic evaluation of this palm tree without the need for a reference genome. This has resulted in robust estimates of diversity and patterns of genetic structure.

The results of genetic diversity and population structure were similar based on the results of the analysis according to population (among the 14 localities), type of vegetation (Caatinga and Restinga), and morphotype (common carnaúba and white carnaúba). In all situations, the populations showed a negative f value, suggesting limited inbreeding with reduced self-pollination capacity under environmental conditions. Therefore, individual plants are less related than expected under conditions of random mating. Genetic diversity and population structure is influenced by biological characteristics of the species, including the mating system [53]. Therefore, the reproductive biology of this species may explain the observed patterns of genetic variation. The mating system of *C. prunifera* is mixed and preferably allogamous [5], which favors the crossing between unrelated individuals. Thus, inbreeding coefficients are reduced, and the maintenance of genetic diversity within populations is ensured.

Although the evaluated populations were susceptible to anthropogenic threats, it is possible that they had high genetic diversity (H_E). High levels of genetic diversity led to an increase in long-term survival of a species; therefore, a strong positive correlation exists between heterozygosity and population fitness, which is important for populations to adapt to

new environmental conditions [54]. This high level of diversity is expected in forest species that are largely not domesticated as a result of local adaptation and neutral evolutionary processes in heterogeneous environments [22].

The identification of private alleles is useful for genetic conservation [55]. In the present study, the populations from APD (Ap = 40), AR1 (Ap = 35), and MOS (Ap = 28) had the highest number of private alleles and diversity was not found in the other localities; therefore, these populations deserve special management because the levels of private alleles are indicative of individual fitness and explain the evolutionary potential of populations and their ability to adapt to the adverse environmental conditions [21]. Therefore, this information can be used to increase the genetic representation in germplasm banks. and to convey the need to explore seed collection *in situ* to ensure future replacement.

Genetic variation in plant species is strongly affected by several historical and demographic factors, including geographic distribution, life form, and population size [56]. The results of AMOVA showed that most of the genetic diversity was found within *C. prunifera* populations (Table 3). Similarly, Santos et al. [17] analyzed the genetic differentiation of this palm tree in the northeast region of Brazil and found that 62.86% of molecular variance was accounted for by differences within populations. These results agree with those of different studies conducted on forest species that reproduce by allogamy, seeing as these species have maintained most of their genetic variability within populations [57].

Genetic structure analyses indicated that the 14 collection sites did not belong to a single homogeneous population, and the geographically closest populations showed low values of pairwise F_{ST} and overlap in the DAPC. Greater genetic similarity was found between the populations from LGP and SER and between AR1 and AR2. In addition, low genetic differentiation was observed when the populations were evaluated according to vegetation type and morphotype.

The low level of global genetic differentiation found between the populations studied here (supported by F_{ST}, cluster analysis, and DAPC) and the higher proportion of genetic diversity within populations with only fewer partitions between them could result from the combined effect of different factors, such as cross rate, reproductive system, and high genetic flow rate in this species. The Mantel test corroborates this result. Since geographically close populations tend to be genetically similar, this indicates a pattern of isolation by distance. However, the MACE population had the lowest level of diversity (H_E = 0.201) and the highest degree of structuring, being the most genetically divergent population compared with the others.

This differentiation was supported by the F_{ST} value, which is an indirect estimator of the population connectivity between subpopulations (Table 2). The observed grouping can be explained by geographical barriers such as roads that surround the population. Furthermore, the population from MACE is more isolated than the other population groups. A limitation in dispersal ability is expected to be a key factor in geographical isolation, which probably favors low levels of gene flow between MACE and other populations. Another characteristic of this population is that it corresponds to a small carnaubal in terms of the number of plants in an area of approximately 0.9 hectares.

Implications for Conservation

Conservation genomics is an extension of conservation genetics that seeks to apply genomic techniques to the practical management of natural populations [58]. In this context, evaluations of genetic variations in the entire genome are powerful approaches to gain an understanding of the processes that lead to molecular diversification and inform effective management and conservation strategies [59]. However, application in real-time has been slow and a persistent gap exists between theory and practice.

In Brazil, the legislation that guides forest management does not clearly describe the importance of genetic evaluation within natural populations; therefore, information that seeks to associate genetic data with the formulation of sustainable management plans is unfortunately not mentioned [60]. Although *C. prunifera* is not listed as an endangered species, the expansion of agricultural activities over time has contributed to a reduction in its natural population [17]. Therefore, conservation measures are necessary to minimize the additional loss of alleles and to ensure the maintenance of genetic resources.

Conserving genetic diversity within a population should be the cornerstone of any conservation strategy aimed at ensuring the long-term persistence of species and habitats [61]. *In situ* and *ex situ* conservation strategies are considered promising alternatives for the conservation of forest genetic resources (FGR) and aim to maintain the genetic diversity of species over time, preserving the evolutionary processes and adaptive potential of populations [62]. Although *ex situ* approaches have the potential to conserve much of the biological diversity, they do have a limitation of being more suited and efficient for conservation in plants that have orthodox seeds. Therefore, *in situ* conservation is recommended for *C. prunifera* because this species contains recalcitrant seeds [63]. However, active management, including the establishment of *in vivo* seed banks and the promotion of natural regeneration, can prevent

the decrease of population size, loss of genetic variability, and ensure long-term conservation [17].

The high genetic diversity observed in the evaluated populations of *C. prunifera* indicates the need for large areas of land dedicated to *in situ* conservation for capturing the existing genetic diversity of these populations. F_{ST} values estimated in the present study could help in recommending the optimal number of populations for sampling, including populations that had the highest estimates of diversity and the largest number of private alleles.

C. prunifera exploitation is an important source of employment and income for local communities in the semi-arid region of Brazil. In this context, the rational management of palm tree products should be a principal strategy in the efforts to conserve the natural habitats of the species. Another strategy aimed at conservation and sustainable use would be the development of a community and family forest management (CFFM) plan [64], which consists of the planning and management of actions and appropriate techniques for the sustainable use of forest resources aimed at traditional communities and family farmers [65].

In addition, practical measures aimed at successful plant regeneration, such as the pause of extractive activity during reproductive periods and the introduction of rotation cycles for leaf harvesting in the explored areas, need to be implemented for the sustainable management of carnaúba. However, the current social and economic conditions of workers employed in the activity of extraction and production of carnaúba wax must be considered. Workers in poorer areas need to be provided additional support, including investments, to maintain the balance between socioeconomic demand and conservation, which would pave the way to a more sustainable supply of resources while reducing the pressure of uncontrolled harvesting.

A third approach would be to preserve populations and divergent genetic groups identified in this study throughout their geographic distribution range through effective long-term genetic and ecological monitoring, stimulating the development of ecological corridors between fragments and natural forests, and avoiding the reduction of genetic variability. In addition, interdisciplinary programs that study different aspects of *C. prunifera* populations (e.g., habitat quality, impact of extractive activity on individuals, and genetic diversity) throughout their distribution would be fundamental for the successful implementation of species conservation management.

References

1. Chacón-Vargas K, García-Merchán VH, Sanín MJ. From keystone species to conservation: conservation genetics of wax palm Ceroxylon quindiuense in the largest wild populations of Colombia and selected neighboring ex situ plant collections. Biodivers Conserv. 2020; 29: 283-302. doi: 10.1007/s10531-019-01882-w.

2. Soares LASS, Cazetta E, Santos LR, França DDS, Gaiotto FA. Anthropogenic disturbances eroding the genetic diversity of a threatened Palm tree: A multiscale approach. Frontiers in Genetics. 2019; 10:1090. doi: 10.3389/fgene.2019.01090.

3. Santos AS, Cazetta E, Dodonov P, Faria D, Gaiotto FA. Landscape-scale deforestation decreases gene fow distance of a keystone tropical palm, *Euterpe edulis* Mart (Arecaceae). Ecol Evol. 2016; 6:6586–6598. doi.org: 10.1002/ece3.2341.

4. Arruda GMT, Calbo MER. Effects of flooding on carnaúba growth, gas exchange and root porosity (*Copernicia prunifera* (Mill.) H.E. Moore). Acta Bot Bras .2004; 18(3):219–224 doi: 10.1590/S0102-33062004000200002.

5. Silva RAR, Fajardo CG, Vieira FA. Mating system and intrapopulational genetic diversity of *Copernicia prunifera* (Arecaceae): a native palm from Brazilian semiarid. Genet Mol Res. 2017; 16 (3) 1: 12. doi: 10.4238/gmr16039764.

6. Rocha TGF, Silva RAR, Dantas EX, Vieira FA. Phenology of *Copernicia prunifera* (Arecaceae) in a caatinga area of Rio Grande do Norte. Cerne. 2015; 21:673–681. doi: 10.1590/01047760201521041758

7. Tomchinsky B, Ming LC. As plantas comestíveis no Brasil dos séculos XVI e XVII segundo relatos de época. Rodriguésia. 2019; 70:1–16. doi: 10.1590/2175-7860201970040

8. Siegmund-Schultze M. A multi-method approach to explore environmental governance: A case study of a large, densely populated dry forest region of the neotropics. Environ Dev Sustain. 2021; 23(2):1539–1562. doi: 10.1007/s10668-020-00635-y.

9. Silva JMC, Leal IR, Tabarelli M. Caatinga: the largest tropical dry forest region in South America. Springer, New York, p 482 2017.

10. Correia BEF, Alencar MM, Almeida- Júnior E B. Lista florística e formas de vida da vegetação de uma restinga em Alcântara, litoral ocidental do Maranhão, Nordeste do Brasil. Revista Brasileira de Geografia Física. 2020; 13(5):2198-2211.

11. Correia BEF, de Almeida EB, Zanin M. Key Points about North and Northern Brazilian Restinga: a Review of Geomorphological Characterization, Phytophysiognomies Classification, and Studies' Tendencies. Bot Rev. 2020; 86(3), 329-337. doi: 10.1007/s12229-020-09230-2.

12. Nascimento ELDL, Maia LC, Caceres MEDS, Lücking R. Phylogenetic structure of lichen metacommunities in Amazonian and Northeast Brazil. Ecol Res. 2021; 36(3), 440-463. doi: 10.1111/1440-1703.12206.

13. Almeilda JAS, Feitosa NA, Sousa LDC, Silva RNO, Morais RF, Monteiro JM, Sousa Júnior J R. Use, perception, and local management of *Copernicia prunifera* (Miller) HE Moore in rural communities in the Brazilian Savanna. J Ethnobiol and Ethnomed. 2021; 17: 1-13. doi: 10.1186/s13002-021-00440-5.

14. Sousa RFD, Silva RAR, Rocha TGF, Santana JADS, Vieira FDA. Etnoecologia e etnobotânica da palmeira carnaúba no semiárido brasileiro. Cerne. 2015; 21:587-594.

15. Costa MF, Francisconi AF, Vancine MH, Zucchi MI (2022) Climate change impacts on the *Copernicia alba* and *Copernicia prunifera* (Arecaceae) distribution in South America. Braz J Bot. 2022; 45:807-818. doi: 10.1007/s40415-022-00801-8.

16. Ferreira CDS, Nunes JAR, Gomes RLF. Manejo de corte das folhas de *Copernicia prunifera* (Miller) HE Moore no Piauí. Revista Caatinga. 2013; 26: 25-30.

17. Santos, JRM, Vieira FA, Fajardo CG, Brandão MM, Silva RAR, Jump AS. Overexploitation and anthropogenic disturbances threaten the genetic diversity of an economically important neotropical palm. Biodivers Conserv. 2021; 30: 2395–2413. doi: 10.1007/s10531-021-02200-z.

18. Mijangos JL, Pacioni C, Spencer PB, Craig MD. Contribution of genetics to ecological restoration. Mol Ecol. 2015; 24: 22-37. doi: 10.1111/mec.12995.

19. Luo Q, Li F, Yu L, Wang L, Xu G, Zhou Z. Genetic diversity of natural populations of *Taxus mairei*. Conserv Genet. 2021; 1-12. doi: 10.1007/s10592-021-01403-9.

20. Duarte JF, de Carvalho D, de Vieira FA. Genetic conservation of *Ficus bonijesulapensis* R.M. Castro in a dry forest on limestone outcrops. Biochem Syst Ecol. 2015; 59:54–62. doi: 10.1016/j.bse.2015.01.008.

21. Siqueira MVBM, Bajay MM, Grando C, Campos JB, Toledo JAM, Domingues GT, Macrini C, Tambarussi EV, Brancalion PHS, Rodrigues RR, Pinheiro JB, Zucchi, M. I. Genetic diversity of reintroduced tree populations of Casearia sylvestris in Atlantic forest restoration sites. For Ecol Manag. 2021; 502:119703. doi: 10.1016/j.foreco.2021.119703.

22. Nunes VV, Silva-Mann R, Souza JL, Calazans CC. Geno-phenotypic diversity in a natural population of *Hancornia speciosa* Gomes: implications for conservation and improvement. Genet Resour Crop Evol. 2021; 68: 2869–2882. doi: 10.1007/s10722-021-01160-1.

23. Díaz BG, Zucchi MI, Alves-Pereira A, de Almeida CP, Moraes ACL, Vianna SA, et al. Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species. PLoS One. 2021;16: 1–24. doi:10.1371/journal.pone.0241025.

24. Xia W, Luo T, Zhang W, Mason AS, Huang D, Huang X, et al. Development of high-density snp markers and their application in evaluating genetic diversity and population

structure in *Elaeis guineensis*. Front Plant Sci. 2019; 1–11. doi: 10.3389/fpls.2019.00130.

25. Supple MA, Shapiro B. Conservation of biodiversity in the genomics era. Genome biology. 2019; 19: 1-12. doi: 10.1186/s13059-018-1520-3.

26. Dalapicolla J, Alves R, Jaffé R, Vasconcelos S, Pires ES, Nunes GL, et al. Conservation implications of genetic structure in the narrowest endemic quillwort from the Eastern Amazon. Ecol Evol. 2021; 11: 10119-10132. doi: 10.1002/ece3.7812

27. Rajora OP. Population Genomics. Concepts, Approaches and Applications. Cham: Springer. 2019.

28. Vieira FA, Sousa RF, Fajardo CG, Brandão MM. Increased relatedness among the neighboring plants from seedling to adult stages in carnaúba wax palm. Genet. Mol. Res. 2016; 15: 1-10. doi: 10.4238/gmr15049100.

29. Pinheiro LG, Chagas KPT, Freire ASM, Ferreira MC, Fajardo CG, Vieira FA. Anthropization as a determinant factor in the genetic structure of *Copernicia prunifera* (Arecaceae). Genet Mol Res. 2017; 16(3). doi: 10.4238/gmr16039768

30. Fajardo CG, Silva RAR, Chagas KPT, Vieira FA. Genetic and phenotypic association of the carnauba palm tree evaluated by inter-simple sequence repeat and biometric traits. Genet Mol Res. 2018; 17(3). doi: 10.4238/gmr18018.

31. Doyle JJ, Doyle JL. Isolation ofplant DNA from fresh tissue. Focus. 1990; 12(13):39–40.

32. Poland JA, Brown PJ, Sorrells ME, Jannink J-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one. 2012; 7(2): e32253. doi: 10.1371/journal.pone.0032253 PMID: 22389690.

33. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes genom genet. 2011; 1(3):171–82. doi: 10.1534/g3.111.000240 PMID: 22384329.

34. Catchen JM, Hohenlohe P, Bassham S, Amores A Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013; 22(11):3124–3140.

35. Luu K, Bazin E, Blum MG. pcadapt: an R package to perform genome scans for selection based onprincipal component analysis. Mol Ecol Resour. 2017; 17(1):67–77. doi: 10.1111/1755-0998.12592 PMID: 27601374.

36. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2018.

37. Flanagan SP, Jones AG. Constraints on the F_{ST}–heterozygosity outlier approach. J Hered. 2017; 108(5):561–73. doi:10.1093/jhered/esx048 PMID: 28486592.

38. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High- through-put functional annotation and data mining with the Blsat2Go suite. Nucleic Acids Res. 2008; 36(10):3420–3435. doi: 10.1093/nar/gkn176.

39. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodo¨hl PA. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in ecology and evolution. 2013; 4(8):782–8.

40. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in genetics. 2015; 6:208. doi: 10.3389/fgene. 2015.00208 PMID: 26113860.

41. Jombart T, Ahmed I. adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011; 27(21):3070–1. doi: 10.1093/bioinformatics/btr521 PMID: 2192612.

42. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010; 11:94.

43. Miller JM, Cullingham CI, Peery RM. The infuence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity.2020; 124(5): 269-280. doi:10 .1038/s41437-020-0348-2.

44. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 1978; 89(3):583–90. PMID: 17248844.

45. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33(7):1870–1874. doi: 10.1093/molbev/msw054.

46. Brancalion PH, Oliveira GC, Zucchi MI, Novello M, van Melis J, Zocchi SS et al. Phenotypic plasticity and local adaptation favor range expansion of a Neotropical palm. Ecol Evol. 2018; 8:7462–7475. doi :10.1002/ece3.4248.

47. Rhode C, Bester-van der Merwe AE, Roodt-Wilding R. An assessment of spatio-temporal genetic variation in the South African abalone (*Haliotis midae*), using SNPs: implications for conservation management. Conserv Genet. 2017; 18: 17-31. doi: 10.1007/510592-016-0879-5.

48. Gogvadze E, Buzdin A. Retroelements and their impact on genome evolution and functioning. Cell Mol Life Sci CMLS. 2009; 66(23):3727–3742. doi: 10.1007/s00018-009-0107-2.

49. Holanda SJ, Araújo FSD, Gallão MI, Medeiros Filho S. Impacto da salinidade no desenvolvimento e crescimento de mudas de carnaúba (*Copernicia prunifera* (Miller) HE Moore). Revista Brasileira de Engenharia Agrícola e Ambiental. 2011; 15: 47-52.

50. Cordeiro EMG, Macrini CM, Sujii PS, Schwarcz KD, Pinheiro JB, Rodrigues RR, et al. Diversity, genetic structure, and population genomics of the tropical tree Centrolobium tomentosum in remnant and restored Atlantic forests. Conserv Genet. 2019;20: 1073–1085. doi:10.1007/s10592-019-01195-z.

51. Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet 12:767–780. doi: 10.1038/nrg3015.

52. Oliveira IS, Machado T, Banci KRDS, Almeida-Santos SM, Silva MJDJ. Genetic variability, management, and conservation implications of the critically endangered Brazilian pitviper *Bothrops insularis*. Ecol Evol. 2020; 10:12870-12882. doi: 10.1002/ece3.6838.

53. Kireta D, Christmas MJ, Lowe AJ, Breed MF. Disentangling the evolutionary history of three related shrub species using genome-wide molecular markers. Conserv Genet. 2019; 20:1101–1112. doi: 10.1007/s10592-019-01197-x.

54. Stojni´c S, Avramidou EV, Fussi B, Westergren M, Orlovi´c S, Matovi´c B, Trudi´c B, Kraigher H, Aravanopoulos FA, Konnert M. Assessment of genetic diversity and population genetic structure of norway spruce (*Picea abies* (L.) Karsten) at its southern lineage in Europe: Implications for conservation of forest genetic resources. Forests. 2019; 10:258. doi.org: 10.3390/f10030258.

55. Laviola BG, Santos A, Rodrigues EV, Teodoro LPR, Teodoro PE, Rosado TB, Guimarães CG, Carson LDH, Conceição LDHCS. Structure and genetic diversity of macauba [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] approached by SNP markers to assist breeding strategies. Genet Resour Crop Evol. 2021; 1-13. doi: 10.1007/s10722-021-01295-1.

56. Crispim BDA, Fernandes JDS, Bajay MM, Zucchi MI, Batista CEDA, Vieira MDC, Barufatti A. Genetic Diversity of *Campomanesia adamantium* and Its Correlation with Land Use and Land Cover. Diversity. 2021; 13: 160. https://doi.org/10.3390/d13040160

57. Tong Y, Durka W, Zhou W, Zhou L, Yu D, Dai L. Ex situ conservation of *Pinus koraiensis* can preserve genetic diversity but homogenizes population structure. For Ecol Manag. 2020; 465:117820. doi: 10.3390/f11010039.

58. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17:81. doi: 10.1038/nrg.2015.28

59. Ferrante JA, Smith CH, Thompson LM, Hunter ME. Genome-wide SNP analysis of three moose subspecies at the southern range limit in the contiguous United States. Conser Genet. 2021; 1-13. doi: 10.1007/s10592-021-01402-w

60. Pádua JAR, Rocha LF, Brandão MM, Vieira FA, Carvalho D. Priority areas for genetic conservation of *Eremanthus erythropappus* (DC.) MacLeish in Brazil. Genet Resour Crop Evol. 2021; 68(6): 2483-2494. doi: 10.1007/s10722-021-01144-1.

61. Fussi B, Westergren M, Aravanopoulos FA, Baier R, Kavaliauskas D, Finžgar D, Alizoti P, Božič G, Avramidou E, Konnert MW, Kraigher H. Forest genetic monitoring: an overview of concepts and defnitions. Environm Monit Assessm. 2016; 188:493–504. doi: 10.1007/s10661-016-5489-7.

62. Fady B, Aravanopoulos FA, Alizoti P, Mátyás C, von Wühlisch G, Westergren M, Belletti P, Cvjetkovic B, et al. Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. For Ecol Manage. 2016; 375:66–75. doi: 10.1016/j.foreco.2016.05.015.

63. Araújo LHB, Silva RAR, Dantas EX, Sousa RF, Vieira FA. Germinação de sementes da *Copernicia prunifera*: biometria, pré-embebição e estabelecimento de mudas. Enciclopédia Biosfera. 2013; 9(17):1517–1528.

64. Brasil. Portaria Nº 313, de 30 de dezembro de 2019. Brasília, 2019.

65. Miranda K, Neto MA, Sousa R, Coelho R. Manejo Florestal Sustentável em Unidades de Conservação de uso comunitário na Amazônia. Sociedade & Natureza. 2020; 32: 778-792. doi: 10.14393/SN-v32-2020-51621

Supporting information

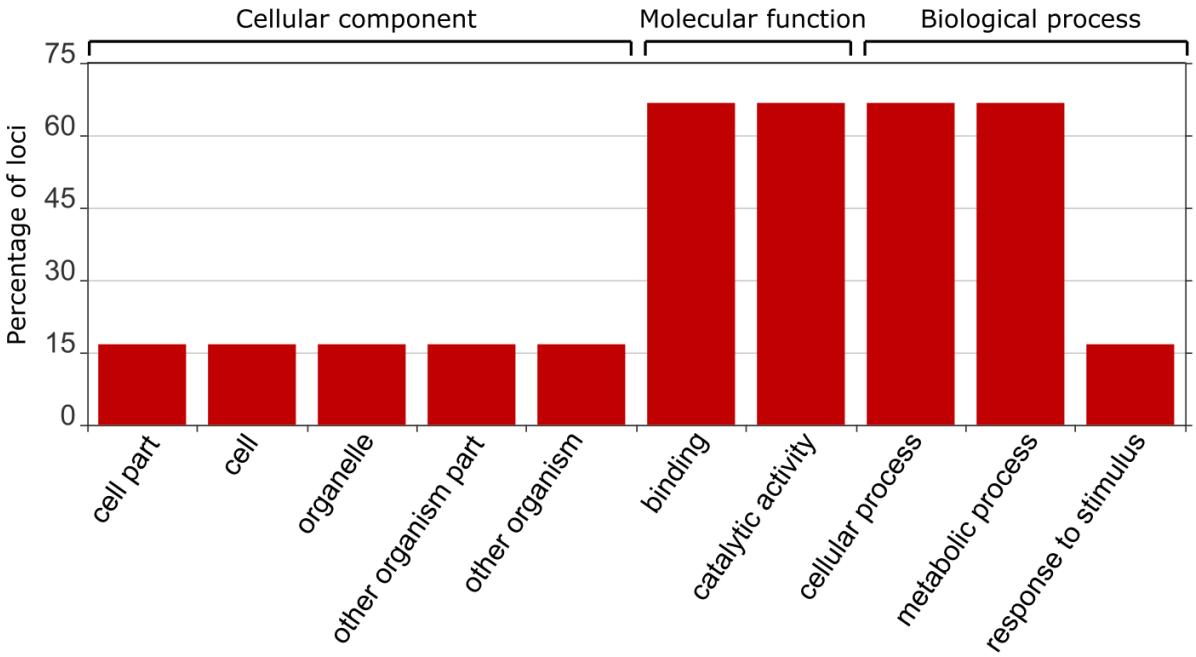
S1 Table. Collection sites of the evaluated populations of *Copernicia prunifera*.
(DOCX)

Population	Initials	Stade	Latitude	Longitude	Biome
Icapuí	ICA	Ceará	4.766	37.283	Restinga
Aracati 1	AR1	Ceará	4.566	37.733	Restinga
Aracati 2	AR2	Ceará	4.85	37.45	Restinga
Russas	RUS	Ceará	4.916	37.9	Caatinga
São Miguel do Gostoso	SMG	Rio Grande do Norte	5.116	35.683	Restinga
Lagoa de Pedras	LGP	Rio Grande do Norte	6.2	35.45	Caatinga
Serrinha	SER	Rio Grande do Norte	6.233	35.483	Caatinga
Macaíba (Zumbi)	MACZ	Rio Grande do Norte	5.983	35.5	Caatinga
Macaíba (EAJ)	MACE	Rio Grande do Norte	5.883	35.366	Caatinga
Jucurutu	JUC	Rio Grande do Norte	6.066	37.05	Caatinga
Apodi	APD	Rio Grande do Norte	5.716	37.733	Caatinga
Ipanguaçu	IPG	Rio Grande do Norte	5.516	36.8	Caatinga
Mossoró	MOS	Rio Grande do Norte	5.183	37.3	Caatinga
Martins	MAT	Rio Grande do Norte	6.05	37.866	Caatinga

S2 Table. Similarity with proteins and Gene Ontology classifications obtained in blast2go for outlier SNPs putatively under selection in carnaúba (*Copernicia prunifera*). H_E = expected heterozygosity of the locus; F_{ST} = genetic divergence among groups of accessions estimated based on the locus; e-value = number of hits expected by chance (E = x 10); Sim (%) = BLASTX percentage of similarity between SNP tags and annotated proteins.
(DOCX)

Locus	H_E	F_{ST}	e-value	Sim (%)	Molecular function	Biological process
178120	0.495	0.114	3.41E-10	96.15	Nucleic acid binding	DNA Biosynthesis
113155	0.322	0.253	5.94 E-07	79.5	Retrotransposon	

158078	0.452	0.130	1.21E-14	88.65	ADP binding	Cellular defense
319537	0.122	0.070	3.88 E-8	88.82	Nucleic acid binding	DNA Biosynthesis
26734	0.172	0.105	3.88E-8	83.95	Nucleic acid binding	DNA Biosynthesis
162642	0.046	0.024	2.17E-6	85.6	Retrotransposon	



S1 Fig. Genetic ontology assignment graph (GO). GO Annotations are summarized into three main categories: cellular location, biological process and molecular function for carnaúba (*Copernicia prunifera*)

(TIF)

CAPÍTULO II

Population genomics and genetic structure of *Copernicia alba* Morong ex Morong & Britton in Brazilian Chaco areas

Autores: Marcones Ferreira Costa. Jonathan Andre Morales-Marroquín. Carlos Eduardo de Araújo Batista. Alessandro Alves-Pereira. Fábio de Almeida Vieira. Maria Imaculada Zucchi

Submitted for publication in Biodiversity and Conservation

Abstract

Chaco is the largest remaining continuous tropical dry forest on the American continent and has one of the highest deforestation rates in the world. *Copernicia alba* Morong ex Morong & Britton is a typical species of Chaco forests and has been subjected to predatory extraction in its areas of occurrence. Because of anthropic activities and intense extractivism it is assumed that there is a reduction in the genetic diversity of *C.alba* populations. Therefore, this study evaluated the diversity and genetic structure of five populations of *C. alba* in the Brazilian Chaco, using single-nucleotide polymorphism (SNP) markers in order to generate information for the management and conservation of this species. Populations showed high levels of genetic diversity (H_E). The inbreeding coefficient (*f*) was negative for all populations, indicating an excess of heterozygotes. Moderate genetic differentiation was observed among populations. Despite these results, conservation efforts should be adopted to avoid the loss of genetic variation. Conservation *ex situ* is pointed out as an important strategy for the preservation of *C. alba* populations, prioritizing sampling throughout the distribution area. The information obtained in this study will contribute to the conservation of genetic diversity, increasing the chances of the natural populations of *C. alba* persisting over time.

Keywords: Carandá. Conservation genetics. Palm tree. Pantanal. Population genetics. SNP.

Introduction

Chaco is the largest remaining continuous segment of tropical dry forest in South America, with approximately 800,000 km², occurring in Argentina, Paraguay, Brazil and Bolivia (Alves et al. 2018). In Brazil, Chaco is located mainly in the Pantanal region and comprises about 70,000 km^2 as a parallel strip to the Paraguay River, in the state of Mato Grosso do Sul (Fava et al. 2020).

Chaco vegetation has recently turned into a global deforestation hotspot resulting from the exploitation of natural resources and agricultural expansion, mainly driven by large-scale, market-oriented agribusiness, has converted more than 20% of the Chaco's forests to pastures and croplands between 1985 and 2015 (Vallejos et al. 2015; Baumann et al. 2017; Basualdo et al. 2018). This situation poses challenges to forest species in the Chaco region, since this scenario affects the genetic diversity of populations.

Among the forest species with greater socioeconomic visibility of this biome, *Copernicia alba* Morong ex Morong & Britton, popularly known in Portuguese as 'Carandá', stands out. *C. alba* is a widely distributed neotropical palm tree and establishes monodominant formations known as 'carandazais' (Arrúa; Negrelle, 2014). This palm tree has a wide range of popular uses with emphasis on exploitation of its stem (stipe), which has wood with high strength and indicated for the manufacture of poles, fences and walls. Industrially, stipes are used in the manufacture of parquet and laminate flooring (Negrelle; Degen-Naumamm, 2012). From its leaves one can obtain wax of great quality, slightly inferior to wax from 'carnaúba' (*Copernicia prunifera* (Miller) HE Moore), and it can be exported under the classification of Paraguayan 'carnaúba' (Sammartino 2010). Its leaves also have high cellulose content, being potentially suitable for the production of polymeric compounds, and the fiber is used in handicrafts (Sammartino 2010).

Due to the pressure for cutting the stipes, this species is continuously extracted from its places of occurrence (Arrúa; Negrelle 2014). The predatory extraction performed without respecting the natural dynamics of substitution causes serious damage, leading to a reduction in population size, which causes a decrease in the evolutionary potential and increases the probability of local extinction due to the effects of inbreeding (González et al. 2020). Therefore, knowledge about how genetic variability is distributed among populations is essential for the conservation of diversity, and consequently of the evolutionary potential of the species (Wadt et al. 2018). In this context, one of the main objectives in conservation programs, in addition to habitat preservation, is the maintenance of existing levels of genetic diversity in populations,

considering that the reduction of genetic variability can decrease the ability of populations to adapt to environmental changes and increase susceptibility to biotic pressures such as pests and diseases (Baldoni et al. 2020).

Molecular markers provide direct and more accurate estimates of the genetic diversity existing in the populations and is a reliable alternative for the knowledge of genetic variation and in elucidating the genetic relationships between and within populations (Lanes et al. 2016). Currently, one of the most used molecular markers to estimate plant diversity and genetic structure are the single-nucleotide polymorphism (SNP) markers (Alves-Pereira et al. 2019). These markers allow identifying the variation of a single nucleotide in an automated way and at a moderate cost, thus enabling the analysis of genetic diversity on a genomic scale (Brito et al. 2021).

Genomic approaches such as genotyping by sequencing (GBS) allow the identification of a large number of SNPs throughout the genome of an individual organism, which facilitates the identification of markers that show signals of selection (outlier loci), for example when determining the set of SNPs that has greater genetic differentiation (F_{ST} outliers) between populations (LeBlanc et al. 2020). Although the analyses of outlier loci are biased for the detection of a single locus with strong signals of selection on more subtle polygenic adaptation, they can serve as a starting point to identify adaptive differences between populations (Nielsen et al. 2020). Once identified, these loci can be examined separately to gain insights into adaptive selection in a population and highlight potential candidate genes for future studies (Hoban et al. 2016). Therefore, population genomic analyses provide an opportunity to explore adaptive signatures and neutral genetic variation of natural populations, producing information that supports the conservation and sustainable use of genetic resources (Laviola et al. 2021).

Studies on genetic diversity using SNP markers have not yet been reported in the literature for the species *C. alba*; in addition, there is a lack of information on the impact of anthropic actions on patterns of genetic variation of native populations of this palm tree. In this context, the present study aimed to evaluate genetic diversity and estimate the population structure of *C. alba* in the Brazilian Chaco, using SNP markers. Furthermore, outlier SNPs associated with adaptive responses (outlier loci) were identified for the populations assessed. This information will provide a scientific basis for conservation and management efforts in native populations of *C. alba*.

Material and Methods

Study area and sampling

Botanical material was collected in localities with ‘carandazais’ distributed in the Brazilian Chaco, in the municipalities of Corumbá and Porto Murtinho, Mato Grosso do Sul, Brazil (Fig. 1). In all, 75 individuals were sampled in five localities: Amolar (Amo), Castelo (Cas), Nabileque (Nhe), Corumbá (Cor) and Porto Murtinho (Por). Depending on the size and color of the stipe, two phenotypes are recognized for *C. alba*: black (taller and black stipe) and white (shorter and white stipe) (Cisz 2011). The population of Nhe consisted only of black Carandá, while the other populations analyzed were formed by white Carandá.

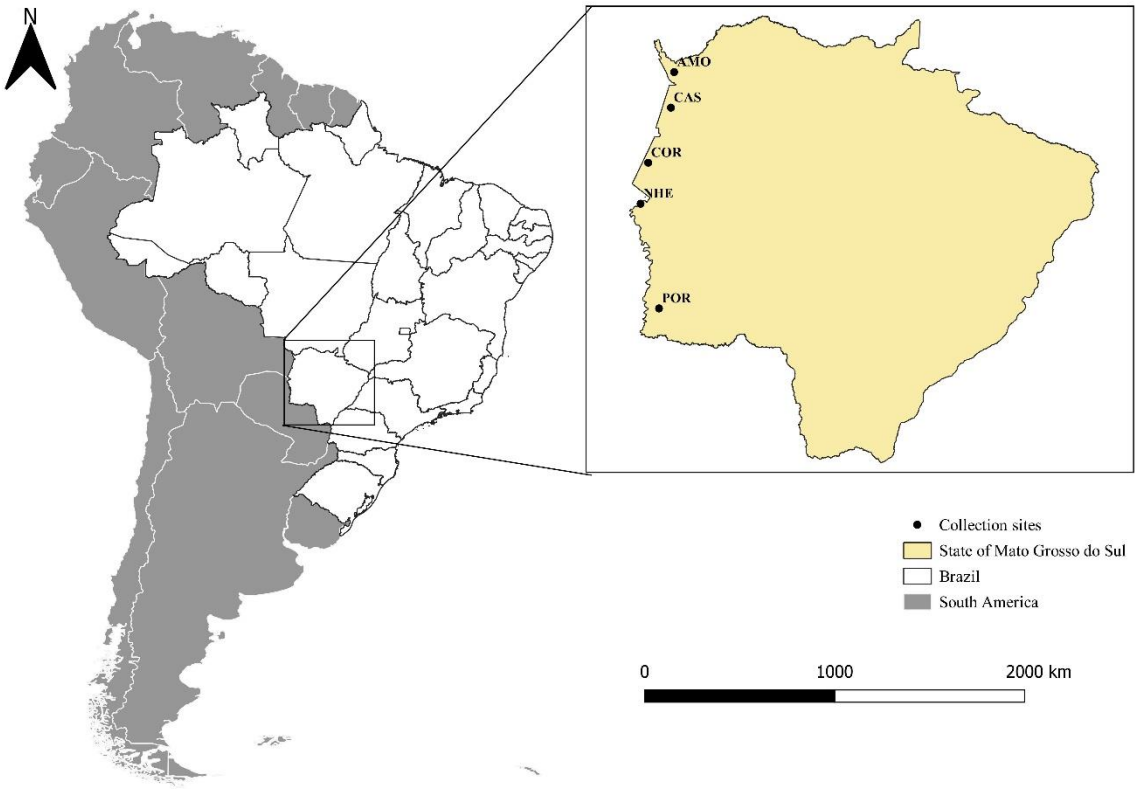


Fig.1 Map of the collection sites of *C. alba* populations in Mato Grosso do Sul, Brazil.

The collected leaves were stored in silica in gel for rapid dehydration (Chase, Hills 1991). Genomic DNA was extracted from leaf material using the protocol described by Doyle & Doyle (1987). DNA quality was evaluated in 1% agarose gel stained with SYBR Safe™ (Life Technologies Corporations) for visualization under ultraviolet light using lambda phage

DNA of known concentrations as a reference. The samples were quantified with a Qubit 3.0 fluorometer with the dsDNA BR kit (Life Technologies), and the DNA samples were normalized to a concentration of 30 ng.μl^{-1}. This study was carried out according to the recommendations of the Brazilian Ministry of the Environment and registered in the National System of Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN) with the number A411583.

GBS library preparation and high-throughput sequencing

To obtain SNPs, genomic libraries were developed from the genotyping by sequencing (GBS) technique with two restriction enzymes, according to the protocol described by Poland et al. (2012), with modifications. We digested 7 μl of genomic DNA from each sample at 37° C for 12 h with the restriction enzymes NsiI and MspI. Subsequently, 0.02 μM of Illumina technology-specific barcode adapters were connected to the ends of the digested fragments. The ligation reaction was performed at 22° C for 2 h; 65° C for 20 min; 10° C indefinitely. After binding of the adapters, the samples were purified with the QIAquick PCR Kit (Qiagen). The library was enriched by PCR (Polymerase Chain Reaction) using the following amplification program: 95° C for 30 s, followed by 16 cycles of 95° C for 10 s, 62° C for 20 s, 72° C for 30 s, ending at 72° C for 5 min. Finally, the library was purified using the QIAquick PCR Purification Kit from QIAgen. The Agilent DNA 12,000 kit and 2100 Bioanalyzer System equipment (Agilent) was used to verify the average size of the DNA fragments. Sequencing was performed using the Illumina® HiSeq 2500 Mid Output Kit v4 (50 cycles) (Illumina Inc., San Diego, CA, USA) in a single-end configuration.

Identification of SNPs

The identification of SNPs was performed with the aid of Stacks v.1.42 software (Catchen et al. 2011, 2013). The first step comprised filtering and demultiplexing with the aid of the process_radtags module. Since there is no reference genome for *C. alba*, we used the denovo pipeline of Stacks starting with the ustacks module to identify putatively homologous groups of reads (putative loci). This step was performed for each sample separately with the following parameters: the minimum sequencing depth (-m = 3), the maximum distance between stacks (-M = 2). The loci from each sample were grouped into a catalog using the cstacks module, allowing a maximum distance of two nucleotides (-n 2) between the loci from each sample. The sstacks module was used to compare the information of the loci in each sample

with the loci included in the catalog. Loci with lower probability values (--lnl_lim -10) were eliminated by the rxstacks correction module. Finally, the populations module was used to filter the SNP markers, using the following parameters: only one marker per tag sequenced; least frequent allele frequency (MAF $\geq$ 0.01); minimum sequencing depth 3X; and minimum occurrence in 75% of individuals in each of the five loci.

Determination of outlier loci

Two complementary tests, pcadapt and fsthet, were performed for the detection of outlier loci (hypothetically under selection). The pcadapt method (Luu et al., 2017) was used to identify loci associated with the genetic structure revealed by a principal component analysis (PCA), i.e., without any underlying genetic model. The analysis was performed with the pcadapt package (Luu et al., 2017) on the R platform (R Development Core Team 2016), retaining the first 8 principal components of the PCA, and considering as SNPs outliers the loci with q-values $\leq$ 0.1. The fsthet method (Flanagan, Jones 2017) was used to identify loci with excessively high or low F_{ST} (Wright, 1943) values with respect to that expected under neutrality. The analysis was performed with the fsthet package (Flanagan, Jones 2017) on the R platform (R Development Core Team 2016), considering as SNPs outliers the loci below or above the 95% confidence intervals constructed with 1000 bootstraps for the expected relationship between H_E and F_{ST}. This test was performed considering the F_{ST} estimates in two different scenarios: i) comparing localities and ii) comparing samples with black (Nhe) and white (remaining localities) phenotype. The final set of SNP markers hypothetically under selection consisted of the loci identified as outliers in at least two of the three tests performed. Thus, the outlier SNP loci possibly reflect the performance of selection in different localities or distinct phenotypes of *C. alba*.

Sequences containing outlier SNPs were searched with the BLASTX tool against the National Center for Biotechnology Information (NCBI) genomic dataset using blast2go (Gotz et al. 2008). This analysis was performed aiming to identify the similarities between the proteins deposited in the NCBI database and the loci with outlier SNPs identified for *C.alba*. For the sequences with significant BLASTX hits, the functional annotation associated with the characterized and/or described coding sequences was performed using the gene ontology system (GO terms). GO terms summarize information of cellular components, molecular functions, and biological processes in which gene products are involved (Gene Ontology Consortium 2019).

Population genomic analyses

Genetic diversity was estimated from the number of alleles (A), number of private alleles (Ap), observed heterozygosity (H_O) and expected heterozygosity (H_E). Endogamy coefficients (f) were also estimated, and their confidence intervals obtained with 1000 bootstraps. Diversity and endogamy estimates were obtained using the diveRsity (Keenan et al. 2013) and PopGenKit (Paquette 2012) packages of R (R Development Core Team 2016). The distribution of genetic variation within and among *C. alba* localities was assessed by molecular analysis of variance (AMOVA), and its significance was tested with 10,000 permutations with the poppr program (Kamvar et al. 2014).

Genetic differentiation was estimated by pairwise FST values with 1000 bootstraps confidence intervals using the diveRsity package (Keenan et al. 2013) of R (R Development Core Team 2016). Population structure was assessed using discriminant principal component analysis (DAPC, Jombart et al. 2010) with Adegenet (Jombart, Ahmed 2011) for R (R Development Core Team 2016). This analysis was performed for neutral loci and the a priori groups were defined from the five sampling sites. DAPC does not assume the underlying population genetic processes (e.g., linkage equilibrium, Hardy-Weinberg equilibrium) common to other methods used to detect population structure. Furthermore, since it is based on principal component analysis, this method can analyze genomic data sets relatively efficiently (Miller et al. 2020).

Genetic relationships and divergence among individuals were investigated by constructing a dendrogram by the neighbor-joining method (Saitou and Nei, 1987) generated based on the Nei (1978) distance. The final dendrogram was formatted using MEGA version 7 software (Kumar et al. 2016).

Results

Neutral and outlier SNPs

The sequencing of genomic libraries produced a total of 274,129,903 reads, which after filtering and quality control were reduced to 185,852,474. 1,774 SNPs were identified (mean depth of 16X), of which 1,709 were considered as neutral loci and used to characterize the diversity and genetic structure of *C. alba* populations. 402 outlier SNPs were identified with the *pcadapt* method, 138 with the *fsthet* method comparing the white and black phenotypes, and 82 with the *fsthet* method comparing the localities (Fig. 2). Of these, 65 SNP markers were identified in at least two of the three tests and were considered hypothetically under selection.

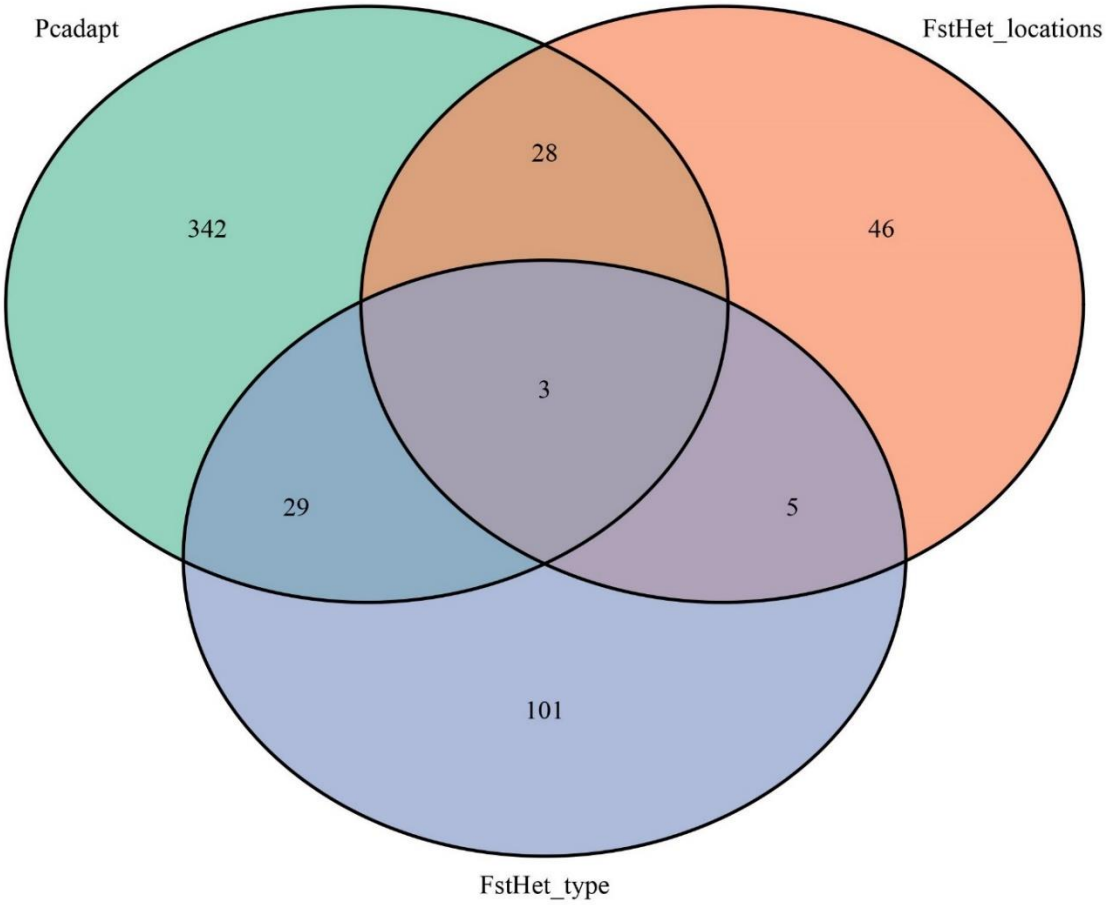


Fig. 2 Venn diagram with the number of outlier loci detected for the fsthet and Pcadapt tests with the overlap between them.

Among the outlier loci, 39 showed F_{ST} values lower than expected and are putatively under balancing selection, whereas 26 outlier loci showed F_{ST} values higher than expected and are putatively under positive selection. Only two outlier loci (263704 and 326121) were present in sequences similar to annotated proteins. The locus 263704 was identified in a sequence similar to a protein of the integral component of membrane type, involved in the metabolism of fucose, while the loco 326121 was identified in a sequence similar to a protein that participates in the DNA repair process. Considering the gene ontology system (GO), the

most frequent annotations for these proteins were the molecular functions “binding” and “catalytic activity”, and the biological process of cellular metabolism (Fig. 3).

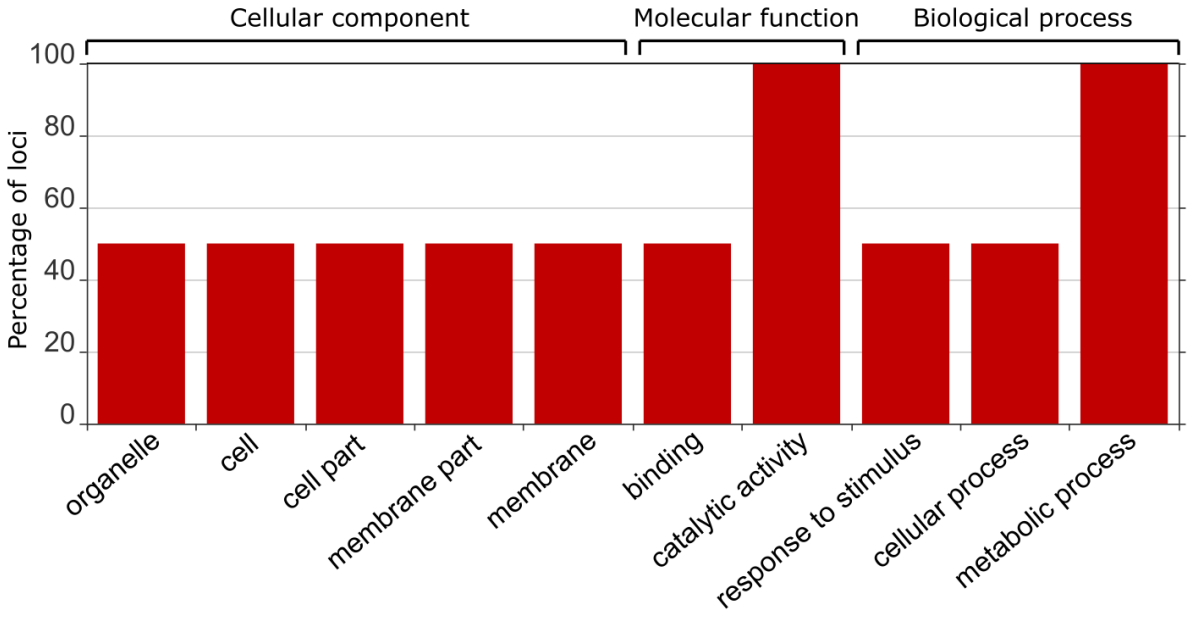


Fig. 3 Genetic ontology assignment graph (GO). GO Annotations are summarized into three main categories: cellular location, biological process and molecular function for Carandá (*C. alba*).

Genetic diversity and structure

The parameters of genetic diversity were similar for the five populations evaluated (Table 1). The observed heterozygosity (H_O) ranged from 0.095 (Cor) to 0.118 (Por), with an average of 0.110. The expected heterozygosity (H_E) ranged from 0.087 (Cor) to 0.114 (Por), with an average of 0.099. The inbreeding coefficients (*f*) were negative, indicating the excess of heterozygotes, in particular, in the localities Nhe and Amo. Regarding the presence of private alleles per population, the populations Por and Cas had the highest number of private alleles, 282 and 185, respectively. The population Nhe had the lowest number of private alleles, and this difference is probably due to the smaller sampling size of the population.

Table 1. Estimates of genomic diversity and inbreeding based on 1,709 neutral SNP markers for *C. alba* populations.

Population	N	A	AP	H_O	H_E	*f*	*f* (95% CI)
Amo	20	2442	92	0.114	0.096	-0.072	(-0.241 - 0.308)
Por	18	2671	282	0.118	0.114	-0.032	(-0.155 - 0.195)
Cas	22	2565	185	0.115	0.105	-0.021	(-0.191 - 0. 136)
Cor	9	2071	49	0.095	0.087	-0.036	(-0.299 - 0.155)
Nhe	6	2232	32	0.110	0.097	-0.096	(-0.397- 0.095)

Number of individuals (N); total number of alleles (A); number of private alleles (Ap); observed heterozygosity (H_O); expected heterozygosity (H_E); inbreeding coefficient (*f*); *f* (95% CI) = lower and upper 95% confidence intervals of the inbreeding coefficients

Analysis of molecular variance (AMOVA) revealed that 12.4% of the total variation was found between localities, while the remaining variation (87.6%) was within the localities (Table 2).

Table 2. Analysis of molecular variance (AMOVA) based on 1,709 neutral SNP markers for five locations of *C. alba*.

	Degrees of freedom	Sum of squares	Mean squares	Variance (%)	Global F_{ST} (φ)	*p*-value
Between populations	13	852,2	231,0	12.4	0.124	0.00005
Within populations	146	4924,8	70,4	87.6		
Total	159	5777,0	80,3			

The overall level of genetic structure described by the global F_{ST} was 0.124 and indicated moderate genetic differentiation. The comparison between the pairwise F_{ST} values showed variation from 0.033 (Cas and Nhe) to 0.078 (Por and Amo/Cor and Amo) (Table 3). These results demonstrate moderate genetic differentiation.

Table 3. Estimates of pairwise F_{ST} between the localities of *C. alba* (lower diagonals). Upper diagonals contain the lower and upper limits of the 95% confidence interval.

	Amo	Por	Cas	Cor	Nhe
Amo		(0.06 - 0.10)	(0.03 - 0.07)	(0.05 - 0.12)	(0.02 - 0.12)
Por	0.078		(0.04 - 0.07)	(0.04 - 0.10)	(0.00 - 0.09)
Cas	0.050	0.057		(0.03 - 0.09)	(0.00 - 0.09)
Cor	0.078	0.063	0.056		(0.00 - 0.13)
Nhe	0.057	0.034	0.033	0.05	

The discriminant analysis of principal components (DAPC), which retained 27.23% of the total variation in the first two principal components (Fig. 4), revealed a small overlap between the localities Amo and Cas, Por and Nhe, and Nhe and Cor.

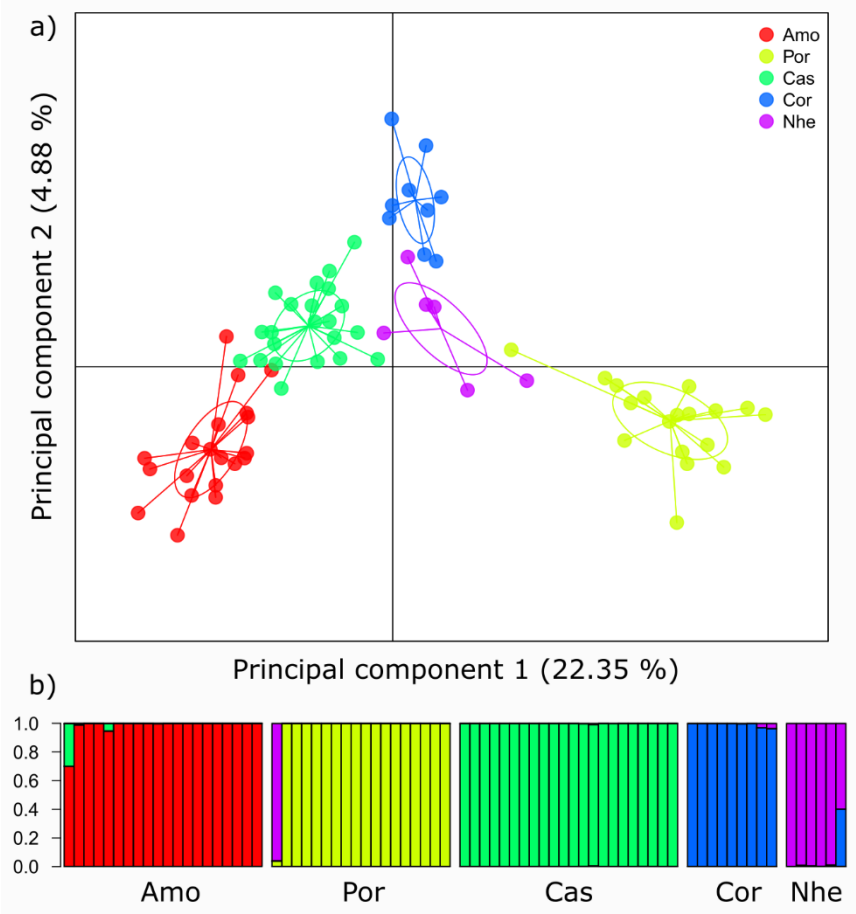


Fig. 4 a) Discriminant analysis of principal components (DAPC) representing the genetic structure of *C. alba* populations based on 1,709 neutral SNPs. b) Bar graph representing the coefficients of DAPC; each bar represents an individual.

In relation to the clustering obtained with the dendrogram (Fig. 5), it is possible to observe the formation of two clusters: the first formed by the populations Amo, Cas and Cor, and the second consisting of Nhe (black Carandá) and Por. Amo and Cas populations have the highest value of bootstrap, which indicates a statistically well supported cluster.

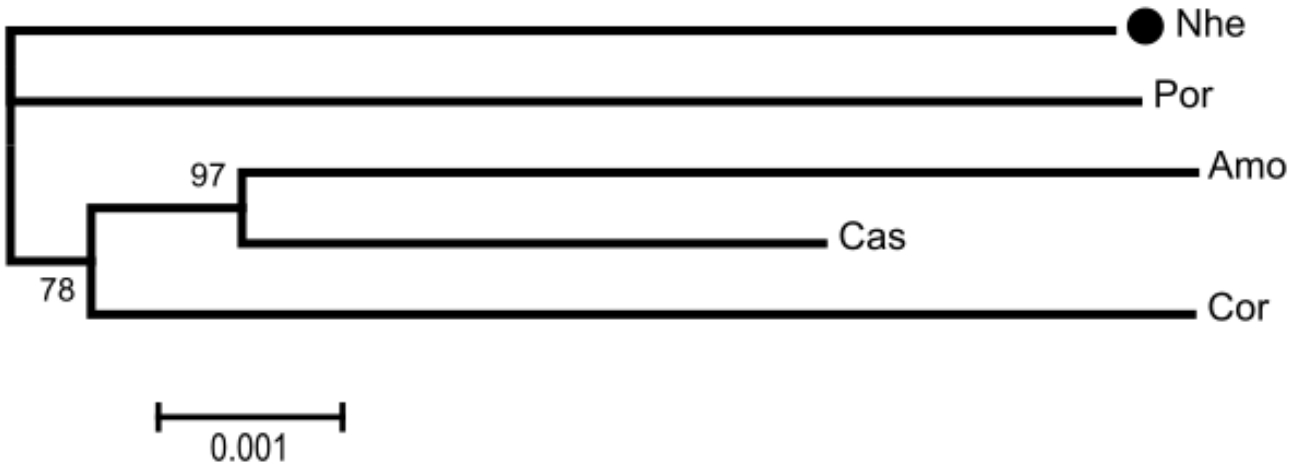


Fig. 5 Dendrogram obtained by the neighbor-joining method based on genetic distances of Nei (1978) for 1,709 neutral SNP markers for the 5 localities of *C. alba*.

When the genetic diversity and structure of *C. alba* was evaluated considering the type of Carandá (black or white), similar levels of genetic diversity were observed (Table 4). The inbreeding coefficients (*f*) were negative. In addition, the F_{ST} estimates suggested moderate genetic differentiation between Carandá types (F_{ST} = 0.017).

Table 4. Estimates of genomic diversity and inbreeding for *C. alba* populations, considering the different phenotypes (white Carandá and black Carandá)

Population	N	A	Ap	H_O	H_E	*f*	*f* (95% CI)
White carandá	69	3365	1165	0.113	0.110	-0.032	(-0.061 - -0.004)
Black carandá	6	2232	32	0.110	0.097	-0.091	(-0.390 - -0.090)

Number of individuals (N); total number of alleles (A); number of private alleles (Ap); observed heterozygosity (H_O); expected heterozygosity (H_E); inbreeding coefficient (*f*); *f* (95% CI) = lower and upper 95% confidence intervals of the inbreeding coefficients.

Discussion

The present work is the first genomic study that uses next-generation sequencing to evaluate the genetic diversity and structure of five populations of *C. alba* in the Brazilian Chaco region by means of SNP markers. Through these markers it was possible to identify outlier and neutral loci. Outlier loci are associated with genetic variation exposed to natural selection and underlying adaptation-related traits and provide valuable insights into the likelihood of populations persisting under certain environmental conditions and local adaptive forces

(Barbosa et al. 2018). Conversely, neutral loci are related to genetic variation that do not directly reflect the action of natural selection, but mainly reflect the interaction between genetic flow and genetic drift (Silva et al. 2020). Therefore, the evaluation of neutral and adaptive genetic variation is fundamental to design appropriate conservation measures (Li et al. 2019).

Although the approaches used by the *pcadapt* and *fsthet* methods have indicated markers as putative candidates under positive selection, no decisive evidence was found for the action of natural selection on the differentiation of the populations evaluated, so it is not possible to associate generic molecular functions or biological processes with any adaptive traits involved in the diversification of the five populations of *C. alba*. One of the explanations can be attributed to the absence of a reference genome for this species, which limits a more refined characterization of the loci hypothetically under selection. However, the outlier loci identified here can be used in future studies with integrative approaches with association genetics to understand the adaptive nature of a given allele (Barret Hoekstra, 2011).

Genetic diversity and population structure can be attributed to multiple factors such as distribution, life forms, reproduction systems, seed dispersal mechanisms, evolutionary history, climatic factors and human interference (Coates et al. 2018). Thus, the evaluation of population genetic diversity has often been used to design conservation strategies, since genetic diversity determines the adaptive potential of populations to environmental changes and supports the persistence of species in the long term (Oyundelger et al. 2021). The levels of genetic diversity (H_E) for *C. alba* were higher than or similar to those observed for other neotropical palm trees (Novello et al. 2018; Díaz et al. 2021; Laviola et al. 2021). These results are satisfactory, given that high levels of genetic diversity lead to an increase in the long-term fitness and survival of a species (Texeira and Uber 2021).

The values of observed heterozygosity (H_O) were slightly higher than those of expected heterozygosity (H_E) in all localities, so the differences found between H_O and H_E estimates explain the negative values of the inbreeding coefficient detected in the populations (*f*). The mating system plays an important role in determining genetic diversity within and between populations of plant species (Garcia-Jacas et al. 2021). There are no studies in the literature on the mating system and modes of dispersal of *C. alba*. However, the low inbreeding coefficients (*f*) are a strong indication that the mating system of this species is predominantly allogamous, which explains the excess of heterozygotes in the populations.

In addition to heterozygosity, some authors have identified the number of private alleles (Ap) as one of the most appropriate parameters for conservation purposes because they may indicate local adaptation (Porth and Kassaby 2014). The results of the present study reveal that the Por and Cas populations have the highest number of private alleles compared to the others, so they deserve special management.

The results of AMOVA reinforce the idea that this species tends to allogamy, since most of the genetic diversity was found within populations. Inbred species are generally characterized by high levels of genetic differentiation between populations, while species with outcrossing breeding mode tend to maintain considerable variability within populations (Gladfelter et al. 2020). According to Chung et al. (2020), although forest species have a variety of breeding systems, they are predominantly outcrossing and generally have life history and ecological traits that maintain high levels of genetic diversity within the population with higher dispersal capacity, which results in high rates of genetic flow and large population sizes.

Generally, trees have low population density, which is beneficial for the long-distance migration of pollen and seeds (Garcia-Jacas et al. 2021). In addition, the anthropogenic action can contribute to long-distance dispersal events. There are examples that humans act as one of the vectors of active and passive plant dispersal, allowing different types of seeds to reach long distances and resulting in introgression (Bullock et al. 2018; Silva et al. 2020). Therefore, we assume that the genetic flow between *C. alba* populations may be related to the movement of pollen by wind, pollinators or humans.

Another important parameter that provides insights for conservation strategies is genetic differentiation (F_{ST}) (Ottewell et al. 2016). F_{ST} values suggest moderate differentiation between the localities of *C. alba*. Furthermore, populations of *C. alba* showed high levels of genetic diversity. Therefore, *C.alba* populations are historically connected and currently maintain gene flow with a high dispersal rate, factors that assist in the conservation of the genetic diversity. Our results are encouraging in the sense that anthropogenic activities have not yet negatively impacted the genetic diversity and population structure of this palm tree. However, the absence of sustainable management plans for *C. alba* may result in the decline of genetic diversity in the future, which may reduce fitness and affect the survival of this species. Therefore, conservation efforts and appropriate management activities for *C. alba* should be designed to ensure the maintenance of its genetic diversity. In addition, the Brazilian Chaco has

lost more than 35.81% of its native cover and no Brazilian Chaco area is designated as a conservation area (Alves et al. 2018)

Forest tree conservation programs include *in situ* conservation efforts aimed at protecting areas with natural populations of the species, or *ex situ* conservation, in which a species is preserved outside its natural habitat with seed banks (Hartvig et al. 2020). Both conservation strategies should preserve genetic diversity throughout the entire geographic distribution of the species, protecting population genetic processes that support its long-term adaptation, maintaining genetic connectivity and reducing the risk of inbreeding (Kireta et al. 2019).

Considering our results, when selecting populations for *ex situ* conservation, one should prioritize those with the highest levels of H_E and Ap, assuming that these populations would harbor high genetic diversity. Therefore, an effort should be made to sample the largest possible number of *C. alba* populations throughout its distribution range in order to encompass the genetic variability of the species. Cas population stood out with one of the highest Ap and the highest degree of genetic structure, so this population has a somewhat different combination of alleles compared to the other populations. In addition, an active management must be carried out, including the establishment of seed banks, as well as the promotion of natural regeneration, avoiding the reduction of population size and loss of genetic variability, thus ensuring the long-term conservation of *C. alba* populations in the Brazilian Chaco.

Conclusion

The natural populations of *C. alba* have high levels of genetic diversity, despite being distributed in the Chaco region, one of the areas with the highest deforestation rates in South America. Moderate structuring between populations and apparently high crossing rates suggest that there is no risk of short-term inbreeding depression, although possible reductions in genetic flow may occur in the future due to habitat loss and fragmentation, if management and conservation programs are not developed for the species. Among the sustainable management options to avoid the loss of genetic variation of these populations, we point out *in situ* or *ex situ* conservation actions, prioritizing sampling throughout their distribution area.

Another important aspect to be considered is the exploitation of *C. alba*, which is an important source of employment and income for the local population (Arrúa; Negrelle, 2014). Therefore, it is necessary to develop appropriate management plans in order to maintain

the levels of genetic diversity. Thus, the implementation of cultivation units or agroforestry systems of the natural populations of *C. alba* may provide a solid basis for reducing extractivism pressure and its respective negative impacts on diversity. In addition, it is important to raise awareness in local communities about the vulnerability of unsustainable extractivism. In this context, future studies with a larger number of populations and associating genetic indicators with knowledge about the mating system and adaptive genetics should be carried out to monitor *C. alba* populations, hence ensuring sustainability, maintenance and perpetuation of this species in the Chaco region.

References

Alves FM, Sartori ÂL, Zucchi MI, Azevedo-Tozzi AM, Tambarussi EV, Alves-Pereira A, de Souza AP (2018) Genetic structure of two *Prosopis* species in Chaco areas: a lack of allelic diversity diagnosis and insights into the allelic conservation of the afected species. Ecol Evol 8:6558–6574. https://doi.org/10.1002/ece3.4137.

Alves-Pereira A, Novello M, Dequigiovanni G, Pinheiro JB, Brancalion PH, Veasey EA, Clement CR, Souza AP, Zucchi, M. I. (2019). Genomic diversity of three Brazilian native food crops based on double-digest restriction site-associated DNA sequencing. Tropical Plant Biology, 12(4), 268-281. https://doi.org/10.1007/s12042-019-09229-z

Arrúa RD, Negrelle RRB (2014) Estructura poblacional, regeneración y producción potencial de cera de *Copernicia alba* Morong ex Morong & Britton en tres sitios de la región del Chaco, Paraguay. Iheringia, Série Botânica 69(2):277-284.

Baldoni AB, Teodoro LPR, Teodoro PE, Tonini H, Tardin FD, Botin AA, Hoogerheide ESS, Botelho SDCC, Lulu J, Farias-Neto AL, et al (2020) Genetic diversity of brazil nut tree (*Bertholletia excelsa* bonpl.) in southern Brazilian Amazon. For Ecol Manag 458:117795. https://doi.org/10.1016/j.foreco.2019.117795.

Barbosa S, Mestre F, White TA, Paupério J, Alves PC, Searle JB (2018) Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors. Mol Ecol 27:3452–65. https://doi.org/10.1111/mec.14806

Basualdo M, Huykman N, Volante JN, Paruelo JM, Piñeiro G (2018) Lost forever? Ecosystem functional changes occurring after agricultural abandonment and forest recovery in the semiarid Chaco forests. Sci Total Environ 650:1537–1546. https://doi.org/10.1016/j.scitotenv.2018.09.001.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet 12:767–780. https://doi.org/10.1038/nrg3015

Baumann M, Israel C, Piquer-Rodríguez M, Gavier-Pizarro G, Volante JN, Kuemmerle T (2017) Deforestation and cattle expansion in the Paraguayan Chaco 1987–2012. Reg Environ Change 17:1179–1191. https://doi.org/10.1007/s10113-017-1109-5.

Brito FA, Zucchi MI, Pinheiro JB, Arrigoni-Blank MF, Dequigiovanni G, Grando C, Batista CE, Alves-Pereira A, Blank AF (2021) Low diversity in the native populations of Croton tetradenius Baill. when using SNP markers: a future crop with an insecticidal activity. Resour Crop Evol 68:3145–3159. https://doi.org/10.1007/s10722-021-01176-7.

Bullock JM, Bonte D, Pufal G, da Silva Carvalho C, Chapman DS, Garcı´a C, Garcı´a D, Matthysen E, Mar Delgado M (2018) Human-mediated dispersal and the rewiring of spatial networks. Trends Ecol Evol. https://doi.org/10.1016/j.tree.2018.09.008.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3 (Bethesda) 1(3):171–182.

Catchen JM, Hohenlohe P, Bassham S, Amores A Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22(11):3124–3140.

Chase MW, Hills HG (1991) Silica gel: an ideal material for field preservation of samples for DNA studies. Taxon 40: 215-220.

Chung MY, Son S, Herrando-Moraira S, Tang CQ, Maki M, Kim YD, Lo´pez-Pujol J, Hamrick JL, Chung MG (2020) Incorporating differences between genetic diversity of trees and herbaceous plants in conservation strategies. Conserv Biol 34:1142–1151. https://doi.org/10.1111/cobi.13467.

Cisz ME (2011) The spatial distribution of *Copernicia alba* (Morong) in the district of Bahía Negra, Paraguay. M.S. thesis, Michigan Technological University, Houghton.

Coates DJ, Byrne M, Moritz C (2018) Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. Front Ecol Evol 6:65. https://doi.org/10.3389/fevo.2018.00165.

Díaz BG, Zucchi MI, Alves-Pereira A, Almeida CP, Moraes ACL, Vianna AS, Azevedo-Filho J, Colombo, C. A. (2021). Genome-wide SNP analysis to assess the genetic population structure and diversity of *Acrocomia* species. PloS one, 16(7), e0241025. https://doi.org/10.1371/journal.pone.0241025

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bull 12:11–15.

Fava WS, Gomes VGN, Lorenz AP, Paggi GM (2020) Cross-amplification of microsatellite loci in the cacti species from Brazilian Chaco. Mol Biol Rep 47: 1535-1542. https://doi.org/10.1007/s11033-019-05064-3.

Flanagan SP, Jones AG (2017) Constraints on the F_{ST}-heterozygosity outlier approach. Journal of Heredity, 108(5):561–573. https://doi.org/10.1093/jhered/esx048.

Garcia-Jacas N, Requena J, Massó S, Vilatersana R, Blanche C, López-Pujol J (2021) Genetic diversity and structure of the narrow endemic *Seseli farrenyi* (Apiaceae): implications for translocation. PeerJ 9:e10521. https://doi.org/10.7717/peerj.10521

Gene Ontology Consortium (2019) The gene ontology resource: 20 years and still GOing strong. Nucleic acids research 47: D330-D338. https://doi.org/10.1093/nar/gky1055.

Gladfelter HJ, Yadav LK, Merkle SA, Wilde HD (2020) Genetic diversity and population structure analysis of *Franklinia alatamaha*, a tree species existing only in cultivation. Tree Genetics & Genomes 16(4):1-9. https://doi.org/10.1007/s11295-020-01455-x

González AV, Gómez-Silva V, Ramírez MJ, Fontúrbel FE (2020) Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity. Conservation Biology 34(3): 711-720. https://doi.org/10.1111/cobi.13422.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High- through-put functional annotation and data mining with the Blsat2Go suite. Nucleic Acids Res 36(10):3420–3435. https://doi.org/10.1093/nar/gkn176.

Hartvig I, So T, Changtragoon S, Tran HT, Bouamanivong S, Ogden R, Seen H, Vieira FG, Turner F, Talbot R, Theilade I, Nielsen LR, Kjær ED (2020). Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): Recommendations for management and sustainable use. Conserv Genet 21(4), 677–692. https://doi.org/10.1007/s10592-020-01279-1.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American naturalist, 188(4), 379–397. https://doi.org/10.1086/688018.

Jombart T, Ahmed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94.

Kamvar ZN, Tabima JF, Grunwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation of population genetics parameters and their associated errors. Methods Ecol Evol 4(8):782–78.

Kireta D, Christmas MJ, Lowe AJ, Breed MF (2019) Disentangling the evolutionary history of three related shrub species using genome-wide molecular markers. Conserv Genet 20:1101–1112. https://doi.org/10.1007/s10592-019-01197-x.

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054.

Lanes ECM, Motoike SY, Kuki KN, Resende MDV, Caixeta ET (2016) Mating system and genetic composition of the macaw palm (*Acrocomia aculeata*): Implications for breeding and genetic conservation programs. J Hered 107:527–536. https://doi.org/10.1093/jhered/esw038

Laviola BG, Santos A, Rodrigues EV, Teodoro LPR, Teodoro PE, Rosado TB, Guimarães CG, Conceição LDHCS (2021) Structure and genetic diversity of macauba [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] approached by SNP markers to assist breeding strategies. Genet Resour Crop Evol 1-13. https://doi.org/10.1007/s10722-021-01295-1.

LeBlanc NM, Gahagan BI, Andrews SN, Avery TS, Puncher GN, Reading BJ, Buhariwalla CF, Curry RA, Whiteley AR, Pavey SA (2020) Genomic population structure of striped bass (*Morone saxatilis*) from the Gulf of St. Lawrence to Cape Fear River. Evol. Appl. 13:1468–1486. https://doi.org/10.1111/eva.12990.

Li YL, Xue DX, Zhang BD, Liu JX (2019) Population genomic signatures of genetic structure and environmental selection in the catadromous roughskin sculpin *Trachidermus fasciatus*. Genome Biol Evol 11:1751–64. https://doi.org/10.1093/gbe/evz118.

Luu K, Bazin E, Blum MG (2017) Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Res 17:67–77. https://doi.org/10.1111/1755-0998.12592.

Miller JM, Cullingham CI, Peery RM (2020) The infuence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity. https://doi.org/10.1038/s41437-020-0348-2.

Negrelle RRB, Degen-Naumannn RL (2012) *Copernicia alba* Morong ex Morong & Britton: aspectos botânicos, ecológicos, etnobotânicos e agronômicos. Visão Acadêmica 13(2): 60-71.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89(3): 583-590. https://doi.org/10.1093/genetics/89.3.583.

Nielsen ES, Beger M, Henriques R, von der Heyden S (2020) A comparison of genetic and genomic approaches to represent evolutionary potential in conservation planning. Biol Conserv 251:108770. https://doi.org/10.1016/j.biocon.2020.108770.

Novello M, Viana JPG, Alves-Pereira A, Silvestre EA, Nunes HF, Pinheiro JB, Brancalion PHS, Zucchi MI (2018) Genetic conservation of a threatened Neotropical palm through community management of fruits in agroforest and second-growth forests. Forest Ecol Manag 407:200–209 https://doi:10.1016/j.foreco.2017.06.059.

Ottewell KM, Bickerton DC, Byrne M, Lowe AJ (2016) Bridging the gap: a genetic assessment framework for population-level threatened plant conservation prioritization and decision-making. Divers Distrib 22:174–188. https://doi.org/10.1111/ddi.12387.

Oyundelger K, Herklotz V, Harpke D, Oyuntsetseg B, Wesche K, Ritz CM (2021) Contrasting effects of local environment and grazing pressure on the genetic diversity and structure of *Artemisia frigida*. Conservation Genetics: 1-16. https://doi.org/10.1007/s10592-021-01375-w.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one 7(2):e32253. pmid:22389690. https://doi.org/10.1371/journal.pone.0032253.

Porth I, El-Kassaby YA (2014) Assessment of the genetic diversity in forest tree populations using molecular markers. Diversity 6:283–295. doi:10.3390/d6020283.

R Development Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical computing. R Foundation for Statistical Computing, Vienna.

Paquette SR (2012) Useful functions for (batch) file conversion and data resampling in microsatellite datasets. https://cran.r-project.org/package=PopGenKit.

Sammartino LEK, Aranguren MI, Reboredo MM (2010) Chemical and mechanical characterization of two south-American plant fibers for polymer reinforcement: caranday palm and phormium. J Appl Polym Sci 115(4): 2236-2245.

Saitou, N., & Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4): 406–425.

Silva AR, Resende-Moreira LC, Carvalho CS, Lanes ECM, Ortiz-Vera MP et al (2020) Range-wide neutral and adaptive genetic structure of an endemic herb from Amazonian Savannas. AoB Plants 12:plaa003. https://doi.org/10.1093/aobpla/plaa003.

Teixeira JC, Huber CD (2021) The infated signifcance of neutral genetic diversity in conservation genetics. Proc Natl Acad Sci USA 118:e2015096118. https://doi.org/10.1073/pnas.2015096118.

Vallejos M, Volante JN, Mosciaro MJ, Vale LM, Bustamante ML, Paruelo JM (2015) Transformation dynamics of the natural cover in the Dry Chaco ecoregion: a plot level geo-database from 1976 to 2012. Journal of Arid Environments 123: 3-11. https://doi.org/10.1016/j.jaridenv.2014.11.009.

Wadt LHO, Faustino CL, Staudhammer CL, Kainer KA, Evangelista JS (2018) Primary and secondary dispersal of *Bertholletia excelsa*: implications for sustainable harvests. For Ecol Manag 415:98–105. https://doi.org/10.1016/j.foreco.2018.02.014.

Wright S (1943) Isolation by distance. Genetics, 28: 114-138

CAPÍTULO III

Climate change impacts on the *Copernicia alba* and *Copernicia prunifera* (Arecaceae) distribution in South America

Autores: Marcones Ferreira Costa. Ana Flávia Francisconi. Maurício Humberto Vancine. Maria Imaculada Zucchi

Published in Brazilian Journal of Botany

https://doi.org/10.1007/s40415-022-00801-8

ABSTRACT

Climate changes are one of the main factors that affect palm trees distribution in the tropics. Among the palm trees with social, economic, and ecological relevance, we highlight the native species, *Copernicia alba* Morong ex Morong & Britton and *Copernicia prunifera* (Miller) H. E Moore. An important strategy for protecting biodiversity is to identity the climate areas that will be suitable for future habitats of the species. In this sense, we used the ecological niche models (ENMs) to predict the suitable climate areas for the potential occurrence of *C. alba* and *C. prunifera* palm trees in current and future scenarios, RCP 4.5 (optimistic) and 8.5 (pessimistic), besides to evaluate these species vulnerability facing the climate changes. Our results predicted the *C. prunifera* habitat would continue to increase over the past years. In the RCP 8.5 scenario, the climate model projected an increase of 23.88% for the *C. prunifera* population between 2050-2070. Also, our results can be used for the application and the establishment of commercial *C. prunifera* plantations. By contrast, the predicted habitat of *C. alba* will decrease 22.2% between 2050-2070, according to the RCP 8.5 scenario. For both *C. prunifera* and *C. alba* species, we observed a low percentage of the potential distribution in protected areas for future scenarios. Therefore, we suggest the creation and maintenance of extensive forestry Protected Areas (PAs) with ecological corridors and the construction of germplasm banks to manage and conserve these two important palm tree species.

Keywords Climate change. Conservation. Carnaúba palm. Caranda. Ecological niche models.

Introduction

Tropical ecosystems are considered one of the greatest treasures and reservoirs of the world's biological diversity, being important sources of ecological services for human beings (Richardson and Pennington 2016; Kissling et al. 2019). However, during the past years, tropical forests have lost approximately half of their original distribution and constantly suffer the impacts caused by climate changes (Sheffield and Wood 2008; Mendoza-González et al. 2013; Bellard et al. 2014; Faurby and Araújo 2018). According to Borges and Loyola (2020) climate change can profoundly impact biodiversity and reduce ecosystem service provisioning. It is expected that these effects will intensify once climate projections suggest an increase in the global temperature of ca. 4.8 °C, putting ecosystems, societies and the economic sectors at risk.

The palm trees (Arecaceae or Palmae), a native of the tropical forests, are considered essential for the tropic's maintenance (Johnson 2011; Eiserhardt et al. 2011; Bacon 2013, Fleming and Kress 2013). Moreover, palm trees can provide a valuable source of biogeography and evolution of the tropical forests, and the vulnerability of ecosystems facing global changes (Blach-Overgaard et al. 2015; Göldel et al. 2015; Kissling et al. 2019). The *Copernicia* genus stands out among the palm trees genera with ecological and socioeconomic importance, representing an essential natural reservoir for geographical distribution studies in the South American dry diagonal (Cássia-Silva et al. 2019; Freitas et al. 2019). In Brazil, this genus encompasses two species: *Copernicia alba* Morong ex Morong & Britton known as caranda and *Copernicia prunifera* (Miller) H. E Moore, popularly known as carnaúba palm.

C. prunifera, known as the "tree of life", is economically significant because of the commercially important wax (carnaúba wax) that covers its leaves, especially younger leaves. The wax produced from its leaves is used in cosmetics, pharmaceutical capsules, electronics, food products, polishing waxes, and coatings (Sousa et al. 2015). The production value of its wax and fibers brings in more than $55 million per year (Santos et al. 2021). *C. alba* often forms monodominant populations known as *carandazais* in the Pantanal in Brazil. This species also occurs in the Chaco of Argentina, Paraguay and Bolivia (Lorenzi 2010). The economic relevance of *C. alba* regards its wood durability, uses in rural constructions, corrals, fences, and as an ornamental plant (Pivetta et al. 2011)

Despite these two closely related species belonging to the same genus, they are spread in different biomes. *C. alba* predominantly occurs in the Pantanal and Chaco region, while *C. prunifera* occurs in the Caatinga and Cerrado biomes, reinforcing the idea that the distribution of palm species is strongly influenced by the climate (Blach-Overgaard et al. 2010; Peterson and Soberón 2012; Ley and Hardy 2014; Velasco et al. 2020). Despite the critical role of these palm trees in tropical ecosystems, studies of their distribution pattern in the future scenarios facing climate changes are still scarce (Göldel et al. 2015; Onstein et al. 2017). In this context, understanding the factors that determine the distribution and dynamics of palm trees diversity is a great challenge. Climate changes are one of the key factors that affect the diversity patterns of these palm trees. The future climate scenarios and their effects on the maintenance of the species is crucial for the development of successful strategies for conservation and mitigating the impact of these changes on these species' biodiversity (Couvreur and Bake 2013; Roncal et al. 2013).

Experimental, mathematical, and empirical models have been developed to predict and to evaluate the impacts of climate changes on biodiversity (Vaz et al. 2015; Vaz and Nabout 2016). Among them, we highlight the Ecological Niche Modeling (ENM). The ENMs have been used to predict the abundance, genetic variability, spatial distribution, species extinction, and biological invasions (Rodríguez et al. 2007; Guillera-Arroita et al. 2015; Bello et al. 2020). Therefore, the ENM is one of the main approaches to predict the climate change impacts and propose strategies and priority areas for conservation (Synes and Osborne 2011; Esser et al. 2019).

Hence, our work aims (a) to evaluate the climate niches of the *C. alba* and *C. prunifera* species; (b) to assess the *C. alba* and *C. prunifera* species vulnerability to climate changes; and (c) to identify presumable strategies for conservation of these palm trees according to the climate change impacts. We used the ENM to examine these two palm tree species' current and future (considering optimistic and pessimistic scenarios) distribution patterns in South America. Our results will help create and establish potential management strategies for conservation of these two socioeconomic important *C. alba* and *C. prunifera* palm trees.

Material and Methods

Species description

C. alba and *C. prunifera* species belong to the Coryphoideae subfamily. Generally, they are solitary, rarely caespitose, with circular tree crown and no visible palm heart (Lorenzi 2010). *C. alba*, the *caranda,* can grow up to 30 m in height with an average of 17-22 cm diameter trunk. This species exhibits early successional characteristics and may tolerate fires; its fruits are edible and provide food for macaws, parrots, and fish. Flowering occurs irregularly and discontinuously at any time of the year and may overlap the fruiting periods from January to May (Araujo and Lobo, 2020).

C. prunifera, the carnaúba palm, can be found in river valleys and in seasonally flooded areas in the semi-arid region of northeastern Brazil, where they generally form monodominant populations known as *carnaubais*. The species is highly resistant to the prolonged absence of water and permanent floods (Arruda and Calbo 2004). It can grow up to 15 m in height with an average of 15-25 cm diameter trunk. Carnaúba palm has many labeliform and palmate leaves, globose cup, and long petiole with spines at the base of the leaf. The species has a mixed mating system that is preferentially allogamous (Silva et al. 2017). Fruits are likely dispersed by the sanhaçu-do-coqueiro (*Tangara palmarum*) and bats (Sousa et al. 2015; Silva et al. 2017).

Occurrence, geographic area, and environmental variables

Two sets are necessary for the ENMs adjustment: the occurrence of species and the environmental variables. We compilated occurrence data of the two palm tree species from the following databases: GBIF (https://www.gbif.org/) and iNaturalist (https://www.inaturalist.org/) using the "occ" function of the spocc package (Chamberlain 2019); speciesLink (http://splink.cria.org.br/), NeoTropTree (http://www.neotroptree.info/), DryFlor (http://www.dryflor.info/), and New York Botanical Garden (https://www.nybg.org/) using "BIEN_occurrence_species" function of the BIEN package (Maitner 2018) in R (version 3.6.1, R Core Team 2019). This information was obtained in September 2019.

After the data integration, we corrected the possible taxonomic and geolocalization errors of these occurrences. For this, we used different filters: (i) taxonomic filter, considering only the species listed at Global Names Resolver (https://resolver.globalnames.org/), using the

"gnr_resolve" function of the taxize package (Chamberlain 2019); (ii) spatial filter, considering only the coordinates that did not localize in the capitals, in the centers of the countries, states or cities, GBIF headquarters, institutions of biodiversity, ocean, urban areas, coordinates and reference systems (in the case of geographical coordinates with Datum WGS-84), with zero as value, with equal longitudinal and latitudinal coordinates, using "clean_coordinates" function of the CoordinateCleaner package (Zizka et al. 2019); (iii) filter of missing data lack – excluding the occurrences with no latitudinal and/or longitudinal information; and (iv) filter of spatial bias – we only filtered one occurrence for each pixel (2.5 arc-minutes or ~5km) of the raster for the variables that contained environmental information, excluding closely or out of the boundary coordinates to reduce the problems of adjustment and evaluation of the models, so that this filter acts on the dimensions where the original bias of occurrence records occurred (Radosavljevic & Anderson 2014, Aiello-Lammens et al. 2015). We found 466 and 430 total occurrences for *C. alba* and *C. prunifera*, respectively. After using the filters, we found 189 and 165 occurrences remains for *C. alba* and *C. prunifera*, respectively. These data are described in detail in Table S1 and S2.

The ENMs are usually adjusted considering that only the climate variable acts in the geographical delimitation of the species (Peterson et al. 2011). At first, we used 19 bioclimatic variables (BIO01-BIO19) variables were obtained from the WorldClim v1.4 dataset (for more details, see http://www.worldclim.org) (Hijmans et al. 2005) (Table S3). The WorldClim dataset uses altitude, temperature, and precipitation to derive climate indices (monthly, quarterly, and annual). These indices represent trends (e.g., mean diurnal temperature range), seasonality (e.g., temperature seasonality), and extremes (e.g., maximum temperature of the warmest month) that are biologically relevant. These databases were used to estimate the niche and the distribution of both species in the current and future scenarios based on the climate change prediction: for the current (integration data between 1960-1990) and for the future (2050-2070). These scenarios are described in terms of Representative Concentration Pathways (RCPs), for both scenarios of CO_2 emission, RCP 4.5 (optimistic) and 8.5 (pessimistic), and for six Global Climate Models (GCMs): ACCESS1-0, CCSM4, HadGEM2-AO, IPSL-CM5A-LR, MIROC-ESM and MRI-CGCM3. These are the main GCMs used in ENM works of climate change predictions for the neotropical region, described Assessment Report of the International Panel of Climate Change (Araújo et al. 2019; Gouveia et al. 2016). The WorldClim data raster were in the GeoTiff format, with the geographical coordinate system ("lat/lon"), Datum WGS-84 and spatial resolution of 2.5 arc-minutes (~5 km). These variables were appropriated for the

Neotropic limit proposed by Morrone (2014) and available by Löwenberg-Neto (2014), at https://sites.google.com/site/biochartis/, adjusted for South America, which would be the historical limit of these species (Barve et al. 2011), using "mask" and "crop" functions of the raster package (Hijmans et al. 2016). Finally, to reduce the dimensionally and collinearity of these variables, we performed the Spearman correlation analysis with the present variables, by adopting only the variables with correlation values of (p >= 0.7). According to Dormann et al (2013) correlation coefficients between predictor variables of p >0.7 is an appropriate indicator for when collinearity begins to severely distort model estimation and subsequent prediction. The variables used were: BIO02 [Mean Diurnal Range (Mean of monthly (max temp - min temp))], BIO03 [Isothermality (BIO2/BIO7) (* 100)], BIO08 [Mean Temperature of Wettest Quarter], BIO15 [Precipitation Seasonality (Coefficient of Variation)], and BIO18 [Precipitation of Warmest Quarter]. The correlation found in our results is available at Table S4 and Fig. S1.

Ecological Niche Models (ENM)

Different mathematical algorithms can produce the niche inference of a species. Generally, these algorithms can be classified into three main groups: (i) only presence, (ii) presence and absence, and (iii) presence and background (Guisan et al. 2017). When different estimated niches are projected in the geographical space (maps), the results predict the potential distribution of one species in a different manner (Qiao et al. 2015). The combination (ensemble) of the predicted results increases the possibility of the prediction improvement, once it considers the uncertain potential distribution of the species (Guisan et al. 2017). Thereby, the ENMs were adjusted using the four following algorithms: Bioclim (Booth et al. 2014), Random Forest (Breiman 2001), Maximum Entropy (MaxEnt; Phillips et al. 2006), and Support Vector Machine (SVM; Guo et al. 2005).

To evaluate the ENMs, we used presence and pseudo-presence data (randomly sampled in all the modeling limits and with the same number of occurrence data for each species). These data were partitioned into 70% for training and 30% for tests. This partition was randomly made to reposition the sample (bootstrap) for each algorithm and for each species, being realized 10 times. For each species, 40 models were predicted for the current scenario (4 algorithms x 10 replicates) and 960 models were predicted for the future scenario (4 algorithms x 10 replicates x 2 periods x 2 scenarios x 6 GCMs). The test data (occurrence and pseudoabsence) were used to calculate the area under the curve (AUC), based on the ROC

(Receiver Operating Characteristic) curve and True Skill Statistic (TSS, Allouche et al. 2006) for the maximization of the sum of sensitivity and specificity (Liu et al. 2013), considering well-adjusted models when they had values above 0.5 (Lawson et al. 2014). The AUC considers the rate of the correct and incorrect forecast (30% of presence and pseudo absence) with several suitability thresholds. To assess the accuracy of predictive distribution models, the AUC values are generally classified in: (1) random forecast (<0.5), (2) poor forecast (0.5 to 0.7); (3) reasonable forecast (0.7 to 0.9); and (4) excellent forecast (>0.9) (Elith et al. 2006; Peterson et al. 2011).

After obtaining the models, we used the ensemble technique for each algorithm and for each GCMs, from the weighted average of the standardized values for each algorithm, using only the models of each replicate and of each algorithm with AUC values higher than 0.75. Models with values higher than their limit are considered as reasonable forecasts (Elith et al. 2006; Peterson et al. 2011). The last step was calculating the threshold values of maximization of the sum of sensitivity and specificity (Liu et al. 2013) to generate models for binary outcomes and consider higher pixels as presence (1) and minor as absence (0) potential values for each species, being 0.44 for *C. alba* and 0.35 for *C. prunifera.*

Lastly, to evaluate the climate change impacts on the species distribution, we used binary maps to identify areas potentially suitable in the present and the future. Thus, we observe: (1) Areas potentially stable, suitable in the present and the future for both periods (Stable [2050 & 2070]); (2) Areas of potential habitat gain that are not suitable in the present but will be in the future for both predicted periods (Gain [2050 & 2070]); (3) Areas of potential habitat gain that are not suitable in the present but will be in the future for at least one of the periods (Gain [2050 | 2070]); (4) Areas of potential habitat loss, which are not suitable in for both predicted periods (Loss [2050 & 2070]); (5) Areas of potential habitat loss, which are not suitable in for at least one of the periods (Loss [2050 | 2070]). We also analyzed the protected areas overlap from Protected Planet (UNEP-WCMC and IUCN 2020, www.protectedplanet.net.), filtered for the IUCN categories of protected areas ("Ia", "Ib", "II", "III", "IV", "Not Applicable", "Not Assigned" and "Not Reported").

All the models were generated in the GeoTiff format, with the geographical coordinate system (“lat/lon”), Datum WGS-84 and spatial resolution of 2.5 arc minutes (~5 km), using R (R Core Team, 2019), “bioclim” and “maxent” functions of the *dismo* package (Hijmans et al. 2012), “randomForest” of the *randomForest* package (Liaw and Wiener 2002),

and “svm” of the *e1071* package (Meyer et al. 2019). Besides, for managing the data and performing the ensembles, we used the *sf* (Pebesma 2018), *raster* (Hijmans 2016) and *tidyverse* packages (Wickham, 2019). All the maps and figures were generated by the *ggplot2* package (Wickham 2016). From the models proposed in different scenarios based on known and projected environmental parameters, we indicate strategies for conservation of two evaluated palm tree species.

The climatic niche overlap

To extrapolate the climatic niche overlap between palm tree analyzed species, we used Schoener’s model (Schoener 1970; Warren et al. 2010), as proposed by Broennimann et al. (2012). First, we reduced the environmental space, based on the 19 bioclimatic variables (only for current scenarios), using the ordination technique (PCA-env *sensu* Broennimann et al. 2012). The PCA-env was calibrated using the combination of the climate information of all the modeling limits. The two first PCA-env axles were gridded into 100 x 100 cells, covering the maximum and minimum values of the data, using the “ecospat.grid.clim.dyn” function of the ecospat package (Broennimann et al. 2018). The first and the second PCA axles captured ~57% and ~17% of the data variation, respectively, totalizing around ~74% of the explanation (Fig. S2). Lastly, we performed the similarity tests for the Schoener D index using the “ecospat.niche.similarity.test” function for analyzing the niche conservatism (alternative = "greater", i.e., the niche overlap is more equivalent/similar than random) between the two studied species, with 999 bootstraps, using the ecospat package (Broennimann et al. 2018). These values range from 0 to 1; values close to 0 indicate low climatic niche overlap (niche differentiation) and values close to 1 indicate higher climatic niche overlap (niche conservatism) (Broennimann et al. 2012).

3 Results

Ecological niche models (ENMs)

The ENMs showed reliable results regarding to the evaluation values. The mean values and standard deviation of AUC were (0.96+0.02) for *C. alba* and (0.92+0.04) for *C. prunifera*, as well as the mean values and standard deviation of TSS were (0.83+0.05) *C. alba* for and (0.76+0.08) for C. *prunifera*, indicating a good prediction of model replicas. We highlight the Random Forest model as the best performance, with 0.962 for *C. alba* and 0.969 for *C. prunifera* (Table S5, Fig. S3 and S4) (Elith et al. 2006; Peterson et al. 2011). The models

predicted distribution areas for both species that matched the rate distribution patterns. (Fig. 1). The generated maps revealed the Chaco, considered the central area for *C. alba* occurrence, as the most suitable region for its species, especially in Argentina, Bolivia, and Paraguay. In Brazil, the *C. alba* distribution is restricted to the Mato Grosso and Mato Grosso do Sul States. For *C. prunifera*, the potential distribution area was the Brazilian Northeast, a dry area with high temperatures and seasonal water deficit. Therefore, the distribution in the current period can be an essential aid for the reforestation programs of the *Copernicia* species.

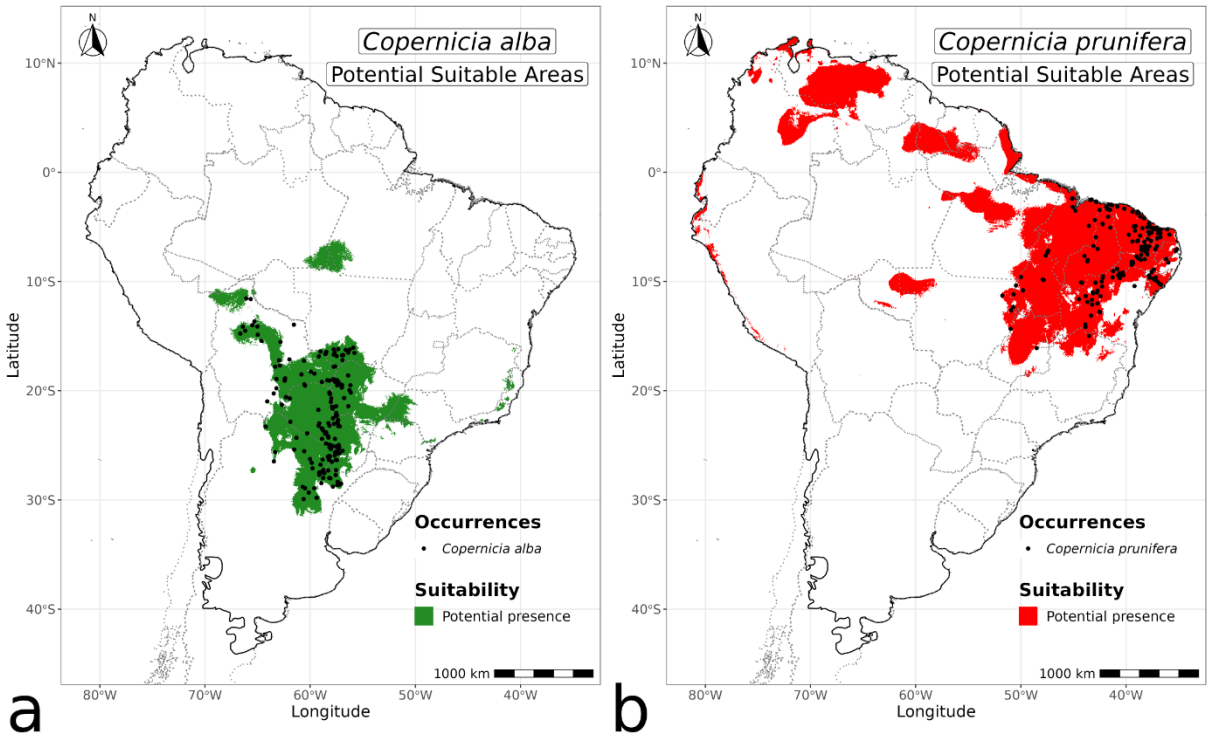


Fig. 1 Occurrences and current potential distribution of *C. alba* (panel a) and *C. prunifera* (panel b) species.

The climatic niche overlap

The PCA-env graphic showed the *C. alba* occurrence area has low isothermality and high-temperature seasonality, while *C. prunifera* occurrence area showed lower values regarding to the seasonality of precipitation. In all simulations, the precipitation variables were the most important to determine the potential areas for the *Copernicia* species (Fig. 2a). We observed a direct influence of the climate factors in the distribution and abundance of both species.

The climatic niche overlap values found in our work indicate low overlap (D = 0.103) (Fig. 2b). We did not observe significant differences in the niche similarity tests of the species. The *C. alba* and *C. prunifera* similarity found was Dsim = 0.132 and p = 0.12 (Fig. 2c), and *C. prunifera* and *C. alba* similarity found was Dsim = 0.04 and p = 0.35 (Fig. 2d). Thus, the observed climatic niche overlap is less than expected and the climatic areas occupied by palm trees are divergent.

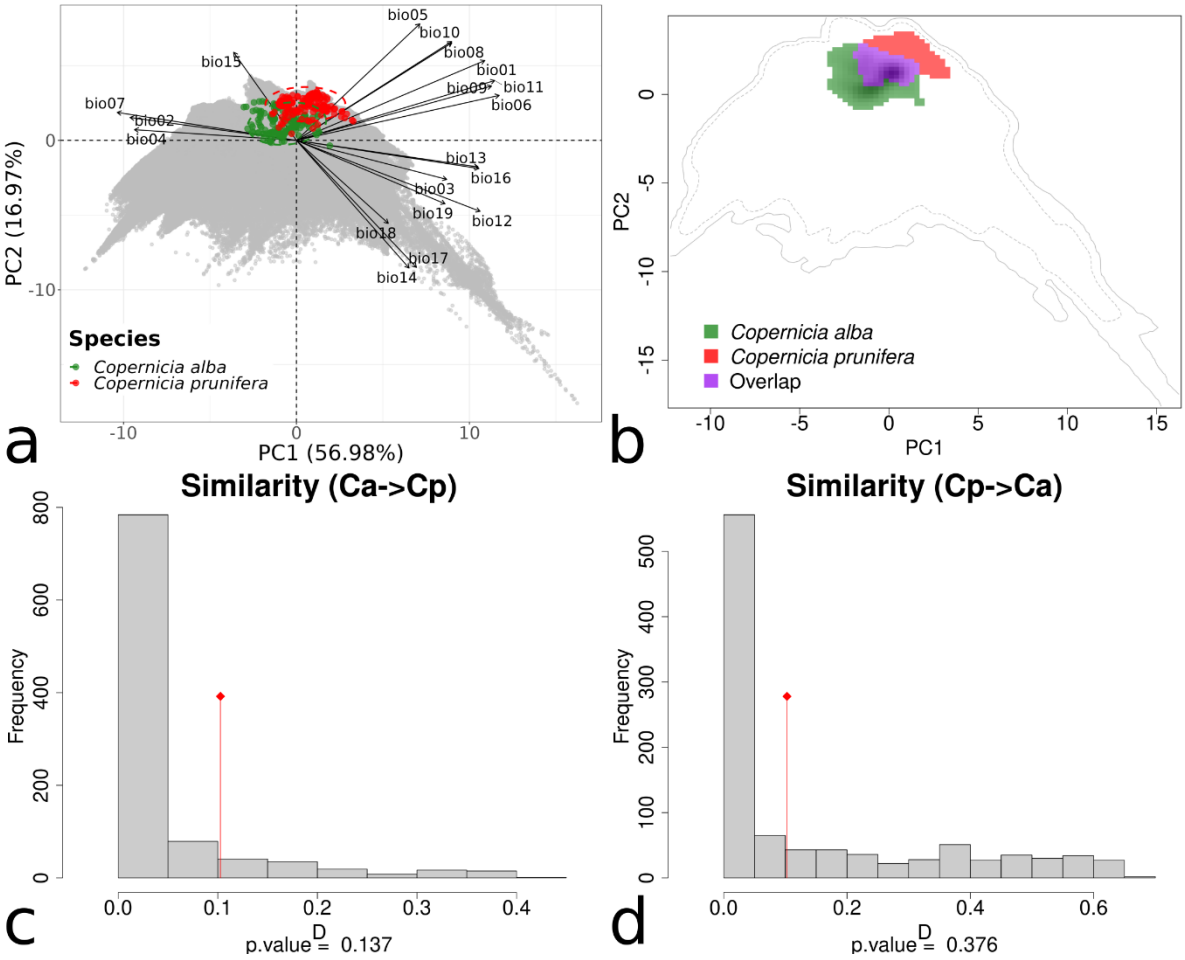


Fig. 2 Environmental niche of *C. alba* (green) and *C. prunifera* (red). a: Biplot represents the first factorial design, which explains 73.95% of the variance. The dots represent the occurrences, and the circles encompass 95% of them. The black arrows indicate the variable direction in the first factorial design. The name of the variables follows the Bioclimatic variable patterns of Hijmans et al. (2005). b: The niche overlap between *C. alba* and *C. prunifera* in the climate space. The green area represents the *C. alba* niche, while the red area represents the *C. prunifera* niche. The magenta area represents the overlap of *C. alba* and *C. prunifera* niches. The pixels shading represents the density in the species occurrences per cell; the solid and

dotted contour lines illustrate the available environment (second design). c and d panels represent the frequency distribution of the overlapped rate of the Schoener's D niche according to the *bootstrap* analysis between *C. alba* and *C. prunifera*, respectively.

Climate change impacts for *C. alba*

We verified the climate changes would have significant impacts on the potential distribution of *C. alba*. Our results obtained by the ENMs suggest that climate changes will promote variation in the suitable total area in future scenarios and change the ideal climate conditions (Fig. 3). For *C. alba*, in both RCP 4.5 (optimistic; Fig. 3a) and RCP 8.5(pessimistic; Fig. 3b) scenarios, it was possible to identify an area reduction in all the distribution regions, with loss more observed in the RCP 8.5 when compared to the RCP 4.5. In RCP 4.5, greenhouse gas emissions will peak in 2040 and then decrease. In the RCP 8.5, the emissions keep increasing over the XXI century (Representative Concentration Pathways – RCP; Meinshausen et al. 2011).

For both predicted periods under the RCP 4.5 emission scenario, 47.8% of the current habitat will keep suitable, while new habitats will increase by 29% (2050 & 2070) and 7.3% (2050 | 2070), the previous habitats will decrease in 11.5% (2050 & 2070) and 4.2% (2050 | 2070). The models for the RCP 8.5 scenario predict the total habit of *C. alba* will decrease 22.3 % (2050 & 2070) and 13.5% (2050 | 2070); around 28.0% will keep suitable with the current weather forecast (Table 1).

This dynamic boosted by the climate in the boundaries will lead to changes in *C. alba* distribution. The reduction of the population distribution of *C. alba* is mainly concentrated at the central portion and Northeast of Bolivia, South of Mato Grosso, Middle East and South of Paraguay, and Northeast of Argentina. Generally, *C. alba* loses part of its current distribution climatic area but will gain few climatically suitable areas in the future. The observed pattern shows that *C. alba* will possibly suffer negative impacts from climate changes.

The models indicate that in future scenarios, the climatic suitability of this species will occur in distinct regions of its current potential distribution. The new climate changes can expose its species to warmer conditions where *C. alba* is not well adapted. According to the climatic models, the *C. alba* populations can undergo to local extinction in the borders where

the climate becomes more severe. When comparing the protected areas, there is a low percentage of the potential area of protected distribution for *C. alba* in both scenarios (Fig. 3a and 3b). However, the optimistic scenario is better in stable areas. Gain areas have low percentages in the protected areas compared to the total gain in 2050 & 2070, and gain in 2050 | 2070. Despite being significant, the loss concentrates out of the protected areas for loss in 2050 & 2070 and in 2050 | 2070. For the pessimistic scenario, there is a general reduction of the stable areas (Table 1)

Climate change impacts for *C. prunifera*

The Caatinga, a biome endemic to Brazil that comprehends the semi-arid region, is considered an area of climate suitability for *C. prunifera* in future climate scenarios. This climatic suitability gradually increases when compared to the current climate to 2050 until 2070 (Fig. 3c and 3d). Our model shows potential areas for the establishment and growth of *C. prunifera*, which significantly contributes to the local market.

In the RCP 4.5 and RCP 8.5 scenarios for *C. prunifera*, we observed growing trends with habitat availability over the years (Fig. 3c and 3d). The RCP 4.5 predicted 61.3% will remain as suitable habitat, while the new habitat will increase 24.7% (2050 & 2070) and 10% (2050 | 2070); the previous habitats will reduce 1.64% (2050 & 2070) and 2.3% (2050 | 2070). For the RCP 8.5, 51% of the predicted climatic suitable area will remain stable. In addition, new habitats will increase 23.8% (2050 & 2070) and 26.9% (2050 | 2070) and the previous suitable habitats will decrease 2.5% (2050 & 2070) and 5.7% (2050 | 2070) (Table 1).

Until 2050, we note the emergence of new climatic suitability regions in the Middle and South of Guyana, Southwest of Suriname, and the East portion of the Rondônia Brazilian State. The model of future projection predicts that in 2070, the Northeast Brazilian region would be potentially more suitable than the current scenario. Moreover, Suriname, Guyana and French Guyana countries can become regions with high climate suitability for *C. prunifera.* These areas are considered key for *C. prunifera* preservation because they will have high climatic suitability even in the worst-case, reinforcing the idea that until 2070 there may be shifts in the geographical distribution of *C. prunifera* in South America. Most of these suitable areas are dry and water deficit regions since the annual precipitation was the variable used to predict its occurrence.

When comparing the protected areas, there is a great difference for *C. prunifera*, with a lower percentage reduction in protected areas for both climate change scenarios, despite a low percentage of the potential distribution in the protected areas (Fig. 3c and 3d). We observed a low percentage of stable areas in the optimistic scenario, higher for gain areas in 2050 & 2070 and 2050 | 2070. The losses are low inside and outside the protected areas for 2050 & 2070 and 2050 | 2070. For the pessimistic scenario, there general maintenance in the climatic stable areas and gain in both 2050 & 2070 and 2050 | 2070 and for the losses in 2050 & 2070 and in 2050 | 2070 (Table 1).

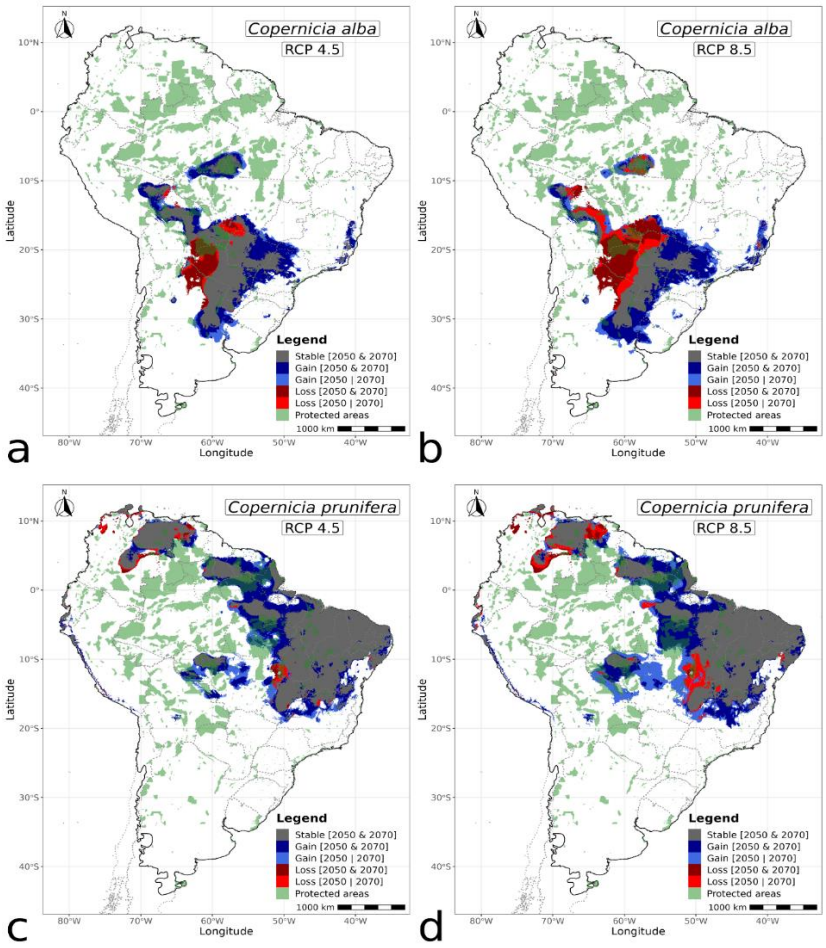


Fig. 3 Potential binary distribution of: a. *C. alba* for the optimistic scenario (RCP 4.5); b. *C. alba* for the pessimistic scenario (RCP 8.5); c. *C. prunifera* for the optimistic scenario (RCP 4.5); d. *C. prunifera* for the pessimistic scenario (RCP 8.5), between 2050-2070. In the maps: Stable (2050 & 2070) represents potential areas, current suitable areas that are maintained in the future; Gain (2050 & 2070) represents areas of potential habitat gain for both predicted periods; Gain (2050 | 2070) represents areas that are current suitable but will be suitable in the future 2050 or 2070; Loss (2050 & 2070) represents areas of potential habitat loss, which will

be suitable for both predicted periods; Loss (2050 | 2070) current suitable areas but will not be suitable in the future 2050 or 2070.

Table 1. Potential distribution of species for the optimistic (RCP 4.5) and pessimistic (RCP 8.5) scenarios, between 2050 and 2070, indicating areas of stability, gain and loss in relation to protected areas.

Species	Scenario	Stability	pa	n	per
C. alba	RCP 4.5	Stable [2050 & 2070]	0	47714	40.3
C. alba	RCP 4.5	Stable [2050 & 2070]	1	8929	7.5
C. alba	RCP 4.5	Gain [2050 & 2070]	0	31174	26.3
C. alba	RCP 4.5	Gain [2050 & 2070]	1	3241	2.7
C. alba	RCP 4.5	Gain [2050/2070]	0	8010	6.8
C. alba	RCP 4.5	Gain [2050/2070]	1	589	0.5
C. alba	RCP 4.5	Loss [2050 & 2070]	0	10682	9
C. alba	RCP 4.5	Loss [2050 & 2070]	1	2920	2.5
C. alba	RCP 4.5	Loss [2050/2070]	0	4511	3.8
C. alba	RCP 4.5	Loss [2050/2070]	1	670	0.6
C. alba	RCP 8.5	Stable [2050 & 2070]	0	29475	24.9
C. alba	RCP 8.5	Stable [2050 & 2070]	1	3728	3.1
C. alba	RCP 8.5	Gain [2050 & 2070]	0	30620	25.9
C. alba	RCP 8.5	Gain [2050 & 2070]	1	2226	1.9
C. alba	RCP 8.5	Gain [2050/2070]	0	8564	7.2
C. alba	RCP 8.5	Gain [2050/2070]	1	1604	1.4
C. alba	RCP 8.5	Loss [2050 & 2070]	0	20788	17.6
C. alba	RCP 8.5	Loss [2050 & 2070]	1	5510	4.7
C. alba	RCP 8.5	Loss [2050/2070]	0	12644	10.7
C. alba	RCP 8.5	Loss [2050/2070]	1	3281	2.8
C.prunifera	RCP 4.5	Stable [2050 & 2070]	0	117451	56.1
C.prunifera	RCP 4.5	Stable [2050 & 2070]	1	10890	5.2
C.prunifera	RCP 4.5	Gain [2050 & 2070]	0	38753	18.5
C.prunifera	RCP 4.5	Gain [2050 & 2070]	1	12909	6.2
C.prunifera	RCP 4.5	Gain [2050/2070]	0	16427	7.9
C.prunifera	RCP 4.5	Gain [2050/2070]	1	4474	2.1
C.prunifera	RCP 4.5	Loss [2050 & 2070]	0	2946	1.4
C.prunifera	RCP 4.5	Loss [2050 & 2070]	1	489	0.2
C.prunifera	RCP 4.5	Loss [2050/2070]	0	4069	1.9
C.prunifera	RCP 4.5	Loss [2050/2070]	1	776	0.4
C.prunifera	RCP 8.5	Stable [2050 & 2070]	0	107054	46.4
C.prunifera	RCP 8.5	Stable [2050 & 2070]	1	10533	4.6
C.prunifera	RCP 8.5	Gain [2050 & 2070]	0	39993	17.3
C.prunifera	RCP 8.5	Gain [2050 & 2070]	1	15066	6.5
C.prunifera	RCP 8.5	Gain [2050/2070]	0	32393	14.0
C.prunifera	RCP 8.5	Gain [2050/2070]	1	6639	2.9
C.prunifera	RCP 8.5	Loss [2050 & 2070]	0	5384	2.3
C.prunifera	RCP 8.5	Loss [2050 & 2070]	1	498	0.2
C.prunifera	RCP 8.5	Loss [2050/2070]	0	12028	5.2
C.prunifera	RCP 8.5	Loss [2050/2070]	1	1124	0.5

*pa: Protected área (0- outside the protected area; 1- inside the protected area); n: number of pixels of the distribution of the present in relation to the scenarios of the future (stable, gain and loss); per: percentage of the number of pixels in relation to the total number of pixels in the distribution of the present in relation to future scenarios (stable, gain and loss).

Discussion

One important strategy to protect these palm trees biodiversity is to identify suitable climate areas that will maintain propitious habitats (Borges and Loyola 2020). According to Urban (2015), the geographical distribution of the species can to move under climate changes. Thus, the species must adapt, disperse, or extinguish. Modeling the potential distribution of *C. alba* and *C. prunifera* is an efficient approach to estimate the climate niche dimension and predict the potential distribution of these palm trees in South America. In this context, the high suitable regions for both species' occurrence are the best places to establish protected areas and population reintroduction.

The ENMs predicted the ongoing climate changes would threaten the *C. alba*. Especially by the exposure to climate conditions to which its species is not well adapted, representing a reduction of the climatically suitable areas that can lead to local extinction. Moreover, models predict that in both future scenarios, there will be a low percentage of the potential distribution of *C. alba* in the protected areas (Fig. 3a and 3b). In this sense, monitoring these populations is essential, mainly because its endemism in Chaco region.

We highlight that the rate of deforestation in the Chaco region is higher when compared to the subtropical seasonal dry forests in the world, reinforcing the urgency to develop strategies at conserving of its ecosystem (Basualdo et al. 2019). Although Chaco is considered an important socioeconomic area, because of its unique diversity resulting from paleoclimate changes and vicariant events (Vallejos et al. 2015), it is not a protected area probably because of the lack of scientific knowledge. Therefore, it is urgent to develop ecological and conservational studies of the Chaco endemic and/or living species.

Moreover, the current protected areas network is ineffective (and will continue to be) to protect *C. alba* area under current and future conditions due to considerable loss of its distribution within the conservation units caused by climate change. The low protection degree and the losses caused by climate changes will lead to *C. alba* vulnerability extinction, reflecting the negligence of protecting the Pantanal and Chaco regions.

Therefore, it is necessary, to develop a systematic conservation planning for the creation and maintenance of great forest extensions of *C. alba* in protected areas, besides the designing and implementation of ecological corridors, which seek to maintain the species through time. Furthermore, palm trees conservation should be carefully planned to avoid a lack of

regeneration under high grazing pressures, as these protected areas allow sustainable management and usage of natural resources (Calambáz-Trochez et al. 2021).

For *C. prunifera*, we observed an expansion of potential distribution in South America, suggesting its species adaptation to global climate changes. This increase in potential climatic distribution can reflect its species adaptation to low precipitation and drought areas globally. Besides, the climatic niche expansion for new regions can result from the adaptative plasticity of *C. prunifera* due to its occurrence in different biomes. Moreover, *C. prunifera* can adapt to disturbed environments (with vegetation removal) known as *ruderal*. It implies the species potential to develop in distinct environments, well adapting to soils with different textures and chemical composition.

The expansion of the climatic distribution of *C. prunifera* calls attention to the favorable areas for the development of commercial plantations due to the great economic and social importance of its species. The main product of exploration and commercialization is the wax powder extracted from their leaves (Silva et al. 2017; Fajardo et al. 2018). Thereby, the climate changes can positively impact the regional productivity of *C. prunifera*, short-term improving the yield for the regions with more suitability, as the Brazilian semi-arid region. Otherwise, promoting *C. prunifera* can help raise awareness of its multiple usages and its historical and cultural significance. Additionally, incentives could increase the demand and create local markets for commercial products derived from its species. However, the *C. prunifera* cultivation in uninhabited areas should be carefully performed. Its successful application will depend on the species' adaptation to a new environment.

Although climate seasonality is a driver of palm species distribution (Eiserhardt et al. 2011), the physical and nutritional quality of the soil, together with the availability of water, also shapes plant species distribution in future scenarios (Emílio et al. 2021; Velazco et al. 2021). Therefore, studies using both climate and edaphic are also essential to generate ENMs with higher predictive power, which adds helpful information for plant distribution.

Regarding niche overlap analysis, the *C. alba* and *C. prunifera* palm trees present disjunct distribution, with low niche overlap. Moreover, previous studies indicate the historical factors related to climate and geological events can modify the habitat suitability, the establishment or persistence of species, contributing to the endemism and disjunct distribution (Hewitt 2000; Fahrig 2003; Bacon et al. 2012; Carvalho et al. 2017).

According to Jaime et al. (2015), niche conservatism occurs with sympatry species, not with species that grow in allopatry or peripatry. This pattern suggests an adaptation of new climate niches followed by the colonization of the Copernicia species in South America. Our results corroborate their hypothesis since we found low values of climatic niche overlap between *C. alba* and *C. prunifera*. Therefore, niche differentiation is more common than niche conservatism, indicating the taxonomic differentiation within the *Copernicia* species is related to adapting to different climates.

Allopatric speciation has probably occurred during the diversification of both species. Allopatry, the most common type of speciation, begins with the emergence of natural barriers or physical limitation that shares the geographical distribution of one species, resulting in geographic isolation, frequently stopping, or decreasing the gene flow (Benítez-Benítez et al. 2018).

As suggested by Bacon et al. (2016), when two or more taxonomic related groups are widely separated geographically, they present a high degree of climate divergence. Thus, the climate differences may have played a fundamental role in the potential distribution of *C. alba* and *C. prunifera*. These palm trees will hardly experience some ecological interaction, as the competitive exclusion. *C. alba* shows low tolerance to dry and high-temperature environments when compared to *C. prunifera*. Consequently, the heat and dry stresses are the limiting factors that reduce the *C. alba* distribution under current and future scenarios.

Our work observed that these palm trees show different responses to climate changes in current and future scenarios. Our models clearly show a reduction in *C. alba* distribution, with higher suitability in the Chaco region. For *C. prunifera,* we observed an increase in the climatically suitable areas. For *C. prunifera*, we observed an increase in the climatically suitable areas. These areas comprehend the Tropical Dry Forests (TDFs), drastically affected by agribusiness expansion, fire forests and desertification. Climate changes can accentuate the loss of the TDFs ecosystems caused by desertification because of the dry weather and land-use practices with no adequate management (Silva et al. 2019; Lucas et al. 2021). Nevertheless, we identified a low potential distribution in the protected areas in future for both species. In this context, our maps provide a detailed comprehension of the climate change impacts over both *Copernicia* species. Our results will supply additional aids that will help the management and conservation national policies, as the creation and maintenance of the

PAs and the development of *in situ* conservation banks, enabling the genetic conservation of *C. alba* and *C. prunifera* species.

References

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38(5): 541-545. https://doi.org/10.1111/ecog.01132.

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

Araújo MB, Anderson RP, Márcia BA, Beale CM, Dormann CF, Early R, Garcia RA, Guisan A, Maiorano L, Naimi B, O'Hara RB, Zimmermann NE, Rahbek C (2019). Standards for distribution models in biodiversity assessments. Science Advances, 5(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858.

Araujo MMV, Lobo FDA (2020) Phenology of *Copernicia alba* in flooded and not flooded environments. Floresta e Ambiente 27 (1): e20170979. https://doi.org/10.1590/2179-8087.097917.

Arruda GMT, Calbo MER (2004) Efeitos da inundação no crescimento, trocas gasosas e porosidade radicular da carnaúba (*Copernicia prunifera* (Mill.) HE Moore). Acta Bot Bras 18: 219-224.

Bacon CD, Baker WJ, Simmons MP (2012) Miocene dispersal drives island radiations in the palm tribe Trachycarpeae (Arecaceae). Syst Biol 61: 426-442.

Bacon CD (2013) Biome evolution and biogeographical change through time. Front Biogeogr 5: 227–231.

Bacon CD, Look SL, Gutiérrez–Pinto N, Antonelli A, Tan HT, Kumar PP, Guan SL, Dransfield J, Baker WJ (2016) Species limits, geographical distribution and genetic diversity in *Johannesteijsmannia* (Arecaceae). J Linn Soc 182: 318-347.

Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Sóberon J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222: 1810-1819

Basualdo M, Huykman N, Volante JN, Paruelo JM, Piñeiro G (2019) Lost forever? Ecosystem functional changes occurring after agricultural abandonment and forest recovery in the semiarid Chaco forests. Sci Total Environ 650: 1537-1546.

Bello C, Cintra ALP, Barreto E, Vancine MH, Sobral-Souza T, Graham CH, Galetti M (2020) Environmental niche and functional role similarity between invasive and native palms in the Atlantic Forest. Biol Invasions 23: 741–754.

Bellard C, Leclerc C, Leroy B, Bakkenes M, Veloz S, Thuiller W, Courchamp F (2014) Vulnerability of biodiversity hotspots to global change. Glob Ecol Biogeogr 23: 1376-1386.

Benítez-Benítez C, Escudero M, Rodríguez-Sánchez F, Martín-Bravo S, Jiménez-Mejías P (2018) Pliocene–Pleistocene ecological niche evolution shapes the phylogeography of a Mediterranean plant group. Mol Ecol 27: 1696-1713.

Blach-Overgaard A, Svenning J, Dransfield J, Greve M, Balslev H (2010) Determinants of palm species distributions across Africa: The relative roles of climate, non-climatic environmental factors, and spatial constraints. Ecography, 33: 380–391. https://doi.org/10.1111/j.1600-0587.2010.06273.x.

Blach-Overgaard A, Balslev H, Dransfield J, Normand S, Svenning JC (2015) Global-change vulnerability of a key plant resource, the African palms. Sci Rep 5: 12611. https://doi.org/10.1038/srep12611.

Borges FJA, Loyola R (2020) Climate and land-use change refugia for Brazilian Cerrado birds. Perspect Ecol Conserv 18: 109-115.

Booth TH, Nix HA, Busby JR, Hutchinson MF (2014). Bioclim: the first species distribution modelling package, its early applications and relevance to most current Maxent studies. Divers Distrib 20: 1-9. https://doi.org/10.1111/ddi.12144.

Breiman L (2001) Random forests. Machine learning 45: 5-32.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Graham CH (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. Glob Ecol Biogeogr 21: 481-497.

Broennimann O (2018) Package 'ecospat', version 3.0. Spatial Ecology Miscellaneous Methods.

Calambáz-Trochez LF, Elías SJ, Melo P, Gurski EM, Thiesen F, Bergmann M (2021) Climate and land-use changes coupled with low coverage of protected áreas threaten palm species in South Brazilian grasslands. Perspect. Ecol. Conserv. 19: 345–353. https://doi.org/10.1016/j.pecon.2021.03.010.

Carvalho CS, Ballesteros-Mejia L, Ribeiro MC, Cortes MC, Santos AS, Collevatti RG (2017) Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. Conserv Genet 18: 467-478.

Cássia-Silva C, Freitas CG, Alves DM, Bacon CD, Collevatti RG (2019) Niche conservatism drives a global discrepancy in palm species richness between seasonally dry and moist habitats. Glob Ecol Biogeogr 28: 814-825.

Chamberlain. spocc: Interface to Species Occurrence Data Sources (2019) URL https://CRAN.R-project.org/package=spocc. R package versio.

Couvreur TL, Baker WJ (2013) Tropical rain forest evolution: palms as a model group. BMC Biol 11: 1-4. https://doi.org/10.1186/1741-7007-11-48.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Eiserhardt WL, Svenning JC, Kissling WD, Balslev H (2011) Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. Ann Bot 108: 1391-1416.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans J R, Huettmann F, Leathwick RJ, Lehmann A, Li J (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

Emilio T, Havale P, Costa, RC (2021) "Intraspecific Variation on Palm Leaf Traits of Co-Occurring Species-Does Local Hydrology Play a Role?" Frontiers in Forests and Global Change 4: 112. https://doi.org/10.3389/ffgc.2021.715266.

Esser LF, Saraiva DD, Jarenkow JA (2019) Future uncertainties for the distribution and conservation of *Paubrasilia echinata* under climate change. Acta Bot Bras 33: 770-776.

Fajardo CG, Silva RAR, Chagas KPT, Vieira FA (2018) Genetic and phenotypic association of the carnauba palm tree evaluated by inter-simple sequence repeat and biometric traits. Genet Mol Res 17. http://dx.doi.org/10.4238/gmr18018.

Faurby S, Araújo MB (2018) Anthropogenic range contractions bias species climate change forecasts. Nat Clim Change 8: 252–256.

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst 34: 487-515.

Fleming TH, Kress WJ (2013) The ornaments of life: coevolution and conservation in the tropics. University of Chicago Press.

Freitas CG, Bacon CD, Souza-Neto AC, Collevatti RG (2019) Adjacency and area explain species bioregional shifts in Neotropical palms. Front. Plant Sci. 10: 55. https://doi.org/10.3389/fpls.2019.00055.

Göldel B, Kissling WD, Svenning JC (2015) Geographical variation and environmental correlates of functional trait distributions in palms (Arecaceae) across the New World. J Linn Soc 179: 602–617.

Gouveia, S. F., Souza-Alves, J. P., Rattis, L., Dobrovolski, R., Jerusalinsky, L., Beltrão-Mendes, R.,Ferrari, S. F. 2016. Climate and land use changes will degrade the configuration of the landscape for titimonkeys in eastern Brazil. Glob. Change Biol. Bioenergy. 22(6), 2003-2012.https://doi.org/10.1111/gcb.13162

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA (2015) Is my species distribution model fit for purpose? Matching data and models to applications. Glob Ecol Biogeogr 24: 276-292.

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models: with applications in R. Cambridge University Press.

Guo Q, Kelly M, Graham CH (2005) Support vector machines for predicting distribution of Sudden Oak Death in California. Ecol Model 182: 75-90.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907-913.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25: 1965-1978.

Hijmans RJ, Phillips S, Leathwick J, Elith J (2012) R package 'dismo': species distribution modeling. Version 0.7-23.

Hijmans RJ (2016) Introduction to the raster package (version 2.5-8).

Jaime R, Alcántara JM, Bastida JM, Rey PJ (2015) Complex patterns of environmental niche evolution in *Iberian columbines* (genus *Aquilegia*, Ranunculaceae). Plant Ecol 8: 457-467.

Johnson, DV (2011) Tropical palms. Univerza v Mariboru, Fakulteta za kmetijstvo in biosistemske vede.

Kissling WD, Balslev H, Baker WJ, Dransfield J, Göldel B, Lim JY, Onstein RE, Svenning JC (2019) PalmTraits 1.0, a species-level functional trait database of palms worldwide. Sci Data 6: 178. https://doi.org/10.1038/s41597-019-0189-0.

Lawson CR, Hodgson JA, Wilson RJ, Richards SA (2014) Prevalence, thresholds and the performance of presence–absence models. *Methods in Ecology and Evolution*, *5*(1), 54-64. https://doi.org/10.1111/2041-210X.12123.

Ley AC, Hardy OJ (2014) Contrasting patterns of gene flow between sister plant species in the understorey of African moist forests–The case of sympatric and parapatric Marantaceae species. Mol Phylogenetics Evol 77: 264-274.

Liaw A,Wiener M (2002) Classification and regression by randomForest. R news, 2: 18-22.

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography, 40(4): 778-789. https://doi.org/10.1111/jbi.12058.

Lorenzi HJ (2010) Flora brasileira: Arecaceae (palmeiras). Instituto Plantarum.

Löwenberg-Neto P (2014) Neotropical region: a shapefile of Morrone's (2014) biogeographical regionalisation. Zootaxa 3802: 300-300.

Lucas FMF, Chagas KPT, Vieira FA (2021) Modeling the spatial distribution of a tropical dry forest tree facing climate change. Rev Bras Cienc Agrar 16: 1-7. https://doi.org/10.5039/agraria.v16i1a8856.

Maitner BS, Boyle B, Casler N, Condit R, Donoghue J, Durán SM, Guaderrama D, Hinchliff CE, Jørgensen PM, Kraft NJB, McGill B, Merow C, Morueta-Holme N, Peet RK, Sandel B, Schildhauer M, Smith SA, Svenning JC, Thiers B, Violle C, Wiser S, Enquist BJ (2018) The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods Ecol Evol 9: 373-379.

Meinshausen M, Smith SJ, Calvin K, Daniel JS, Kainuma ML, Lamarque JF, Matsumoto K, Montzka SA, Raper SC, Riahi K, Thomson AG (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Clim Change 109: 213-41.

Mendoza-González G, Martínez ML, Rojas-Soto OR, Vázquez G, Gallego-Fernández JB (2013) Ecological niche modeling of coastal dune plants and future potential distribution in response to climate change and sea level rise. Glob Chang Biol 19: 2524-2535.

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang CC, Lin CC (2019) Misc functions of the department of statistics, probability theory group (formerly: E1071). Package e1071. TU Wien.

Morrone JJ (2014) Cladistic biogeography of the N eotropical region: identifying the main events in the diversification of the terrestrial biota. Cladistics 30: 202-214.

Qiao H, Soberon J, Peterson AT (2015) No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. Methods in Ecology and Evolution 6: 1126-1136.

Onstein RE, Baker WJ, Couvreur TL, Faurby S, Svenning JC, Kissling WD (2017) Frugivory-related traits promote speciation of tropical palms. Nat Ecol Evol 1: 1903-1911.

Pebesma EJ (2018). Simple features for R: Standardized support for spatial vector data. *R J.*, *10*: 439.

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo M.B (2011) Ecological niches and geographic distributions (MPB-49) (Vol. 49). Princeton University Press.

Peterson AT, Soberón J (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. Nat Conserv 10: 102-107.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231-259.

R Core Team (2019) R Foundation for Statistical Computing. R: A language and environment for statistical computing.

Radosavljevic A, Anderson RP (2014) Making better Maxent models of species distributions: Complexity, overfitting and evaluation. Journal of Biogeography, 41(4): 629-643. https://doi.org/10.1111/jbi.12227.

Richardson JE, Pennington RT (2016) Origin of Tropical Diversity: From Clades to Communities. Front Genet 7, 186. https://doi.org/10.3389/fgene.2016.00186

Roncal J, Kahn F, Millan B, Couvreur TLP, Pintaud JC (2013) Cenozoic colonization and diversification patterns of tropical American palms: Evidence from *Astrocaryum* (Arecaceae). J Linn Soc 171:120–139.

Rodríguez JP, Brotons L, Bustamante J, Seoane J (2007) The application of predictive modelling of species distribution to biodiversity conservation. Divers Distrib 13: 243-251.

Santos JRM, Vieira FA, Fajardo CG, Brandão MM, Silva RAR, Jump AS (2021) Overexploitation and anthropogenic disturbances threaten the genetic diversity of an economically important neotropical palm. Biodivers Conserv 30: 2395–2413. https://doi.org/10.1007/s10531-021-02200-z.

Sheffield J, Wood EF (2008) Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. Clim Dyn 31: 79-105.

Silva RAR, Fajardo CG, Vieira FA (2017) Mating system and intrapopulational genetic diversity of *Copernicia prunifera* (Arecaceae): a native palm from Brazilian semiarid. Genet Mol Res 16(3). http:// dx.doi.org/10.4238 /gmr16039764.

Silva JLSE, Cruz-Neto O, Peres CA, Tabarelli M, Lopes AV (2019) Climate change will reduce suitable Caatinga dry forest habitat for endemic plants with disproportionate impacts on specialized reproductive strategies. PloS one 14: e0217028. https://doi.org/10.1371/journal.pone.0217028.

Schoener TW (1970) Non-synchronous spatial overlap of lizards in patchy habitats. Ecology 51: 408–418.

Sousa RF, Silva RAR, Rocha TGF, Santana JAS, Vieira FA (2015) Ethnoecology and ethnobotany of the palm carnauba wax in Brazilian semi-arid. Cerne 21(4):587–594. https://doi.org/10.1590/01047760201521041764.

Synes NW, Osborne PE (2011) Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. Glob Ecol Biogeogr 20: 904-914.

UNEP-WCMC and IUCN (2020) Protected Planet: The World Database on Protected Areas (WDPA) [Online], September 2020, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.

Urban MC (2015) Accelerating extinction risk from climate change. Science, 348: 571-573

Vallejos M, Volante JN, Mosciaro MJ, Vale LM, Bustamante ML, Paruelo JM (2015) Transformation dynamics of the natural cover in the Dry Chaco ecoregion: a plot level geo-database from 1976 to 2012. J Arid Environ 123: 3-11.

Vaz UL, Cunha HF, Nabout JC (2015) Trends and biases in global scientific literature about ecological niche models. Braz J Biol 75:17-24.

Vaz UL, Nabout JC (2016) Using ecological niche models to predict the impact of global climate change on the geographical distribution and productivity of *Euterpe oleracea* Mart. (Arecaceae) in the Amazon. Acta Bot Bras 30: 290-295.

Velazco SJE, Svenning JC, Ribeiro BR, Laureto LMO (2021) On opportunities and threats to conserve the phylogenetic diversity of Neotropical palms. Divers Distrib 27(3): 512-523. https://doi.org/10.1111/ddi.13215.

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33: 607-611.

Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M (2019) Welcome to the Tidyverse. J Open Source Softw 4:1686. https://doi.org/10.21105/joss.01686.

Wickham H, Chang W, Wickham MH (2016) Package 'ggplot2'. Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2:1-189.

Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Schard R, Antonelli A (2019) CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods Ecol Evol 10: 744-751.

Supplementary data

Additional Supporting Information may be found in the online version of this article.

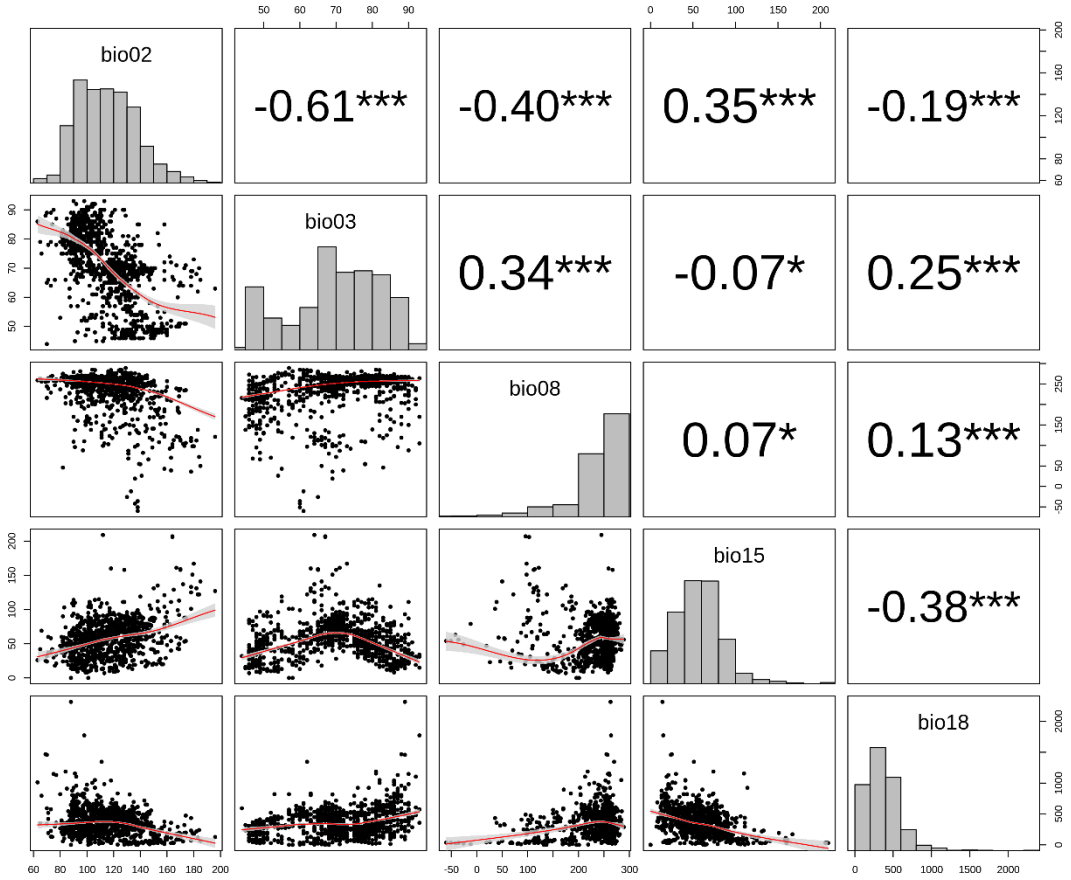


Fig. S1. Scatter plot of non-related variables, their frequency distribution and Spearman correlation coefficients.

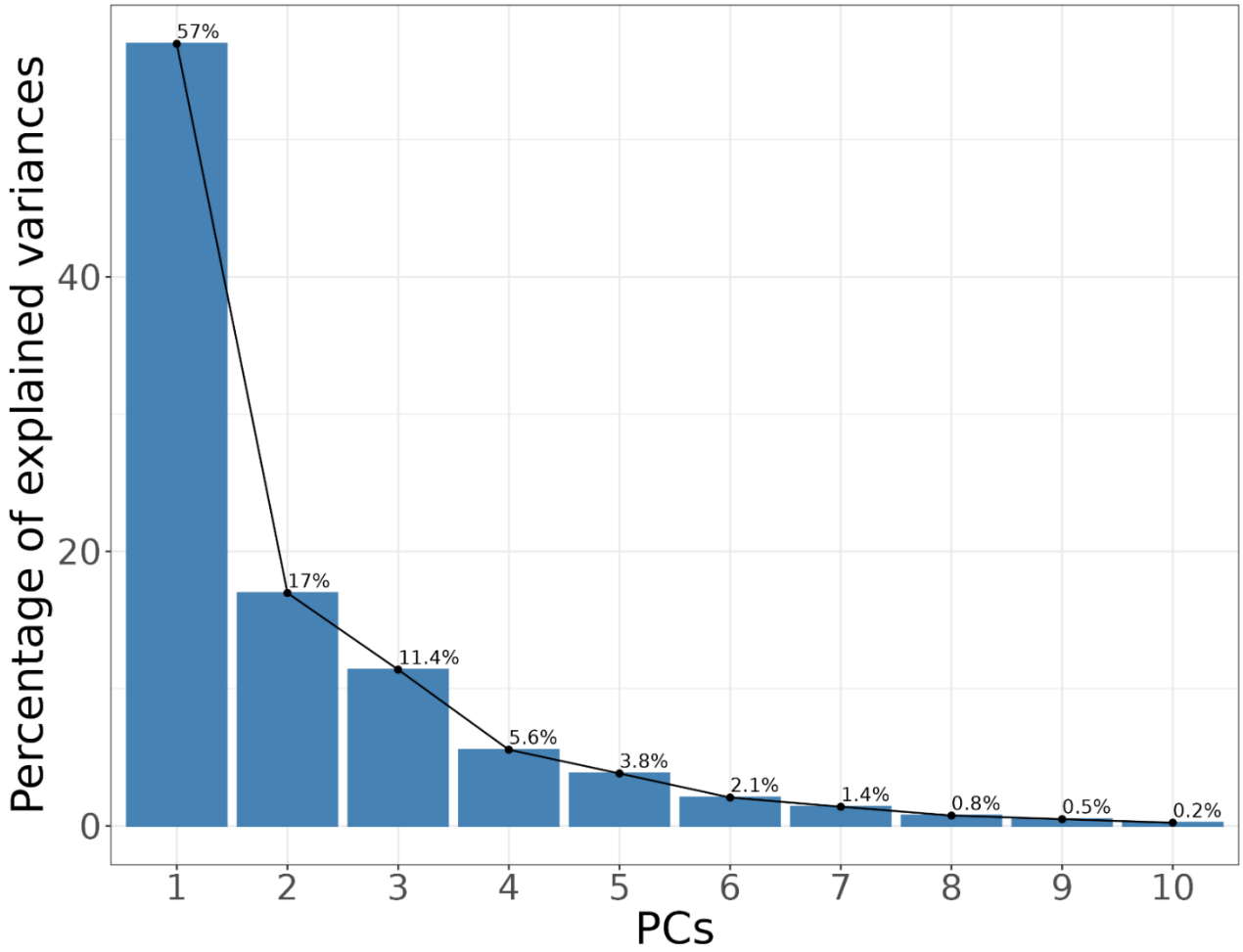


Fig. S2. Screen plot with the explanation percentage of each axis of PCA-env based on the occurrence of *C. alba* and *C. prunifera*. Only 10 axes are showed in the figure.

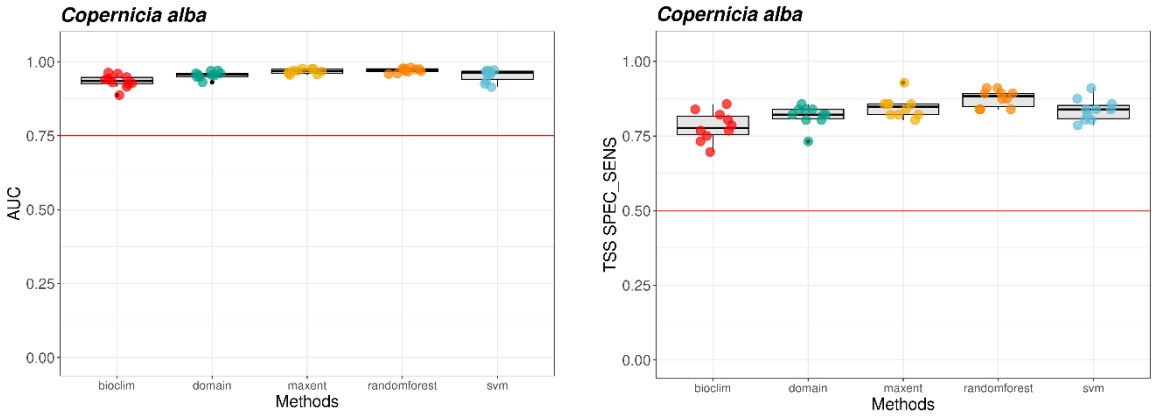


Fig. S3. Graphical representation showing the AUC values of all the generated models for *C. alba*. The red lines represents the cut-off value considering the ensemble models (AUC > 0.75 and TSS > 0.5).

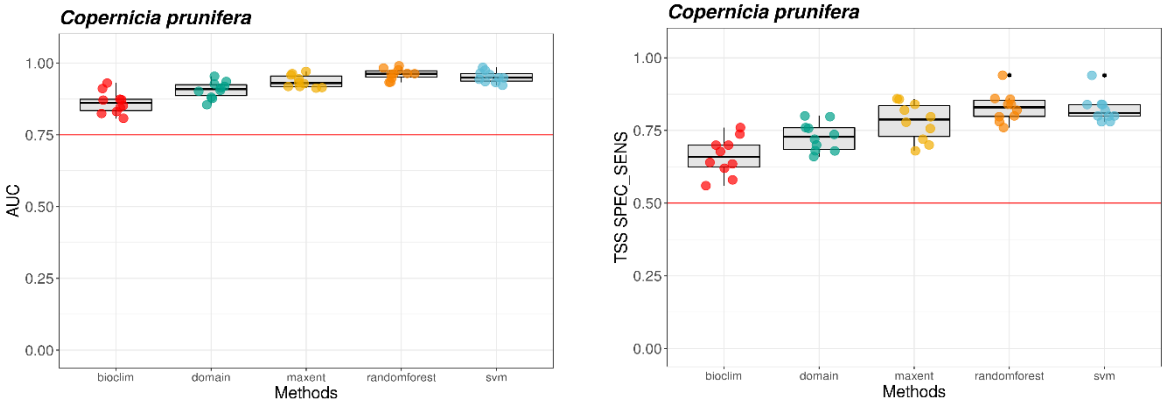


Fig. S4. Graphical representation showing the AUC values of all the generated models for *C. prunifera*. The red lines represents the cut-off value considering the ensemble models (AUC > 0.75 and TSS > 0.5).

Table S1. List of species on the Global Names Resolver used as taxonomy filters.

User supplied name	Submitted name	Matched name	Data source title	Score
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	NCBI	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	Freebase	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	Encyclopedia of Life	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	CU*STAR	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	uBio NameBank	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	Arctos	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera*	Open Tree of Life Reference Taxonomy	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E.Moore	Catalogue of Life	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E. Moore	ITIS	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H. E. Moore	GRIN Taxonomy for Plants	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* Mill.	Union 4	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Miller) H. Moore	The Interim Register of Marine and Nonmarine Genera	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E.Moore H.E.Moore (Mill.)	GBIF Backbone Taxonomy	0.988

Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H. E. Moore	Catalogue of Vascular Plant Species of Central and Northeastern Brazil	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.Moore	Wikipedia in EOL	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E. Moore	USDA NRCS PLANTS Database	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Miller) H. Moore	BioLib.cz	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E. Moore	Tropicos - Missouri Botanical Garden	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E.Moore	nlbif	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H.E.Moore	The International Plant Names Index	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Mill.) H. E. Moore	uBio NameBank	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* Mill.	uBio NameBank	0.988
Copernicia prunifera	*Copernicia prunifera*	*Copernicia prunifera* (Miller) H. Moore	uBio NameBank	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba*	NCBI	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba*	Freebase	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba*	Encyclopedia of Life	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba*	CU*STAR	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba*	uBio NameBank	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba*	Arctos	0.988

Copernicia alba	*Copernicia alba*	*Copernicia alba*	Open Tree of Life Reference Taxonomy	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong	Catalogue of Life	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	ITIS	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong	GRIN Taxonomy for Plants	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong, 1893	Union 4	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	The Interim Register of Marine and Nonmarine Genera	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong	GBIF Backbone Taxonomy	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	Catalogue of Vascular Plant Species of Central and Northeastern Brazil	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong.	Wikipedia in EOL	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	USDA NRCS PLANTS Database	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	BioLib.cz	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	Tropicos - Missouri Botanical Garden	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong	nlbif	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong	The International Plant Names Index	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong	uBio NameBank	0.988

Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong ex Morong & Britton	uBio NameBank	0.988
Copernicia alba	*Copernicia alba*	*Copernicia alba* Morong, 1893	uBio NameBank	0.988

Table S2. Occurrences used for predicted model adjustments. The abbreviations mean: GBIF (Global Biodiversity Information Facility) and MO (Tropical Specimen Data)

Species	**Longitude**	**Latitude**	**Year**	**Base**
Copernicia alba	-56.91666	-25.5	1978	MO
Copernicia alba	-65.5	-14	1979	GBIF
Copernicia alba	-66.4	-14.13333	1979	GBIF
Copernicia alba	-58.898027	-23.443197	1980	GBIF
Copernicia alba	-59.2	-23.2	1980	GBIF
Copernicia alba	-58.85	-23.46667	1980	GBIF
Copernicia alba	-57.66666	-25.91666	1982	GBIF
Copernicia alba	-58.5	-23.8	1985	MO
Copernicia alba	-57	-19.666667	1985	GBIF
Copernicia alba	-57.43333	-25.18333	1987	GBIF
Copernicia alba	-57.333333	-16.183333	1987	GBIF
Copernicia alba	-57.616666	-19.166666	1987	speciesLink
Copernicia alba	-60.58	-17.25	1987	GBIF
Copernicia alba	-60.48333	-19.53333	1989	GBIF
Copernicia alba	-57.43333	-25.13333	1989	GBIF

Copernicia alba	-57.25	-25.13333	1989	MO
Copernicia alba	-62.88333	-17.7	1989	MO
Copernicia alba	-59.6	-18.43	1989	GBIF
Copernicia alba	-61.51666	-25.38333	1990	MO
Copernicia alba	-57.3	-25.13333	1990	MO
Copernicia alba	-63.4666667	-20.25	1992	MO
Copernicia alba	-57.56666	-25.6	1992	MO
Copernicia alba	-57.6533	-19.0092	1992	speciesLink
Copernicia alba	-57.4	-25.61666	1992	GBIF
Copernicia alba	-57.33333	-23.23333	1993	GBIF
Copernicia alba	-57.5	-25.75	1993	GBIF
Copernicia alba	-56.24	-16.23	1993	GBIF
Copernicia alba	-57.61666	-25.63333	1993	MO
Copernicia alba	-57.18333	-23.26666	1993	MO
Copernicia alba	-58.46666	-23.86666	1993	MO
Copernicia alba	-62.41666	-18.91666	1993	MO
Copernicia alba	-56.6228	-16.2567	1993	speciesLink
Copernicia alba	-62.38	-18.87	1993	GBIF
Copernicia alba	-57.34138	-22.6725	1994	MO
Copernicia alba	-58.17833	-24.92805	1994	MO

Copernicia alba	-57.4	-22.9	1994	GBIF
Copernicia alba	-57.316667	-19.2	1994	GBIF
Copernicia alba	-57.92111	-25.03222	1994	MO
Copernicia alba	-57.38333	-22.96666	1994	MO
Copernicia alba	-58.55	-16.65	1994	MO
Copernicia alba	-57.83333	-25.08333	1995	GBIF
Copernicia alba	-57.8333333	-28.7833333	1995	MO
Copernicia alba	-59.13111	-22.98527	1996	MO
Copernicia alba	-57.991	-20.704	1997	GBIF
Copernicia alba	-58.167	-20.155	1997	GBIF
Copernicia alba	-58.104	-20.307	1997	GBIF
Copernicia alba	-57.917	-21.467	1997	GBIF
Copernicia alba	-58.31666	-26.05	1998	MO
Copernicia alba	-57.8075	-18.94472	1998	MO
Copernicia alba	-58.85	-27.38333	1998	GBIF
Copernicia alba	-57.448333	-22.285833	2000	GBIF
Copernicia alba	-57.5189	-22.1744	2000	GBIF
Copernicia alba	-59.16805	-16.85055	2000	GBIF
Copernicia alba	-64.226667	-23.286111	2001	GBIF
Copernicia alba	-58.30264	-24.15467	2002	GBIF

Copernicia alba	-58.580833	-23.937778	2003	GBIF
Copernicia alba	-58.199444	-25.972778	2004	GBIF
Copernicia alba	-64.0822223	-20.9758334	2005	MO
Copernicia alba	-40.33	-9.349444	2007	GBIF
Copernicia alba	-57.043194	-19.487833	2007	GBIF
Copernicia alba	-57.003247	-19.483528	2007	GBIF
Copernicia alba	-57.018611	-19.576667	2009	GBIF
Copernicia alba	-57.3722545	-28.3325998	2009	iNaturalist
Copernicia alba	-58.20718	-26.334566	2012	GBIF
Copernicia alba	-56.994366	-19.579887	2013	GBIF
Copernicia alba	-56.9557	-19.496693	2013	GBIF
Copernicia alba	-56.992577	-19.712526	2013	GBIF
Copernicia alba	-56.974068	-19.484476	2013	GBIF
Copernicia alba	-58.8892691308	-27.4239242033	2015	iNaturalist
Copernicia alba	-59.846134	-26.602495	2015	iNaturalist
Copernicia alba	-57.520123	-25.101083	2015	iNaturalist
Copernicia alba	-58.263674	-24.296227	2015	iNaturalist
Copernicia alba	-57.2109724961	-28.541668769	2016	iNaturalist
Copernicia alba	-57.761372	-17.802539	2016	iNaturalist
Copernicia alba	-57.2199254099	-28.5929293327	2016	iNaturalist

Copernicia alba	-58.732021266	-27.2491986826	2016	iNaturalist
Copernicia alba	-57.19652585	-28.5464891	2017	iNaturalist
Copernicia alba	-56.2621450424	-19.9392453443	2017	iNaturalist
Copernicia alba	-56.8504283333	-16.7408916667	2017	iNaturalist
Copernicia alba	-57.224634	-25.752912	2017	iNaturalist
Copernicia alba	-57.4111099243	-28.6318435669	2018	iNaturalist
Copernicia alba	-58.3665665406	-19.0366907105	2018	iNaturalist
Copernicia alba	-58.67872	-27.5500216667	2018	iNaturalist
Copernicia alba	-57.7594	-16.6925	2018	speciesLink
Copernicia alba	-57.4323366667	-19.5049333333	2019	iNaturalist
Copernicia alba	-57.0658	-19.4328	NA	NeoTropTree
Copernicia alba	-57.4419	-19.0297	NA	NeoTropTree
Copernicia alba	-57.83470154	-21.59939957	NA	DryFlor
Copernicia alba	-58.8778	-24.4058	NA	NeoTropTree
Copernicia alba	-56.92470169	-17.19330025	NA	DryFlor
Copernicia alba	-58.2647	-22.9683	NA	NeoTropTree
Copernicia alba	-56.3664	-21.4133	NA	NeoTropTree
Copernicia alba	-57.46	-26.3511	NA	NeoTropTree
Copernicia alba	-57.7708	-19.015	NA	NeoTropTree
Copernicia alba	-58.03	-25.34	NA	GBIF

Copernicia alba	-58.1611	-20.2703	NA	NeoTropTree
Copernicia alba	-57.36667	-25.63333	NA	GBIF
Copernicia alba	-56.17169952	-19.49749947	NA	DryFlor
Copernicia alba	-62.41999817	-19.07999992	NA	DryFlor
Copernicia alba	-59.93000031	-18.25	NA	DryFlor
Copernicia alba	-63.2075	-19.7753	NA	NeoTropTree
Copernicia alba	-63.33060074	-25.62700081	NA	DryFlor
Copernicia alba	-58.6481	-27.5369	NA	NeoTropTree
Copernicia alba	-57.4831	-16.8269	NA	NeoTropTree
Copernicia alba	-64.9836	-14.8678	NA	NeoTropTree
Copernicia alba	-57.56667	-22.28333	NA	GBIF
Copernicia alba	-56.9714	-16.7483	NA	NeoTropTree
Copernicia alba	-58.1689	-23.2025	NA	GBIF
Copernicia alba	-57.5394	-27.2731	NA	NeoTropTree
Copernicia alba	-60.555	-19.4164	NA	NeoTropTree
Copernicia alba	-57.56000137	-26.34000015	NA	DryFlor
Copernicia alba	-57.28333	-25.31667	NA	GBIF
Copernicia alba	-55.8972	-16.0536	NA	NeoTropTree
Copernicia alba	-60.70069885	-28.79870033	NA	DryFlor
Copernicia alba	-57.2744	-24.9756	NA	NeoTropTree

Copernicia alba	-59.1375	-27.5	NA	NeoTropTree
Copernicia alba	-56.39250183	-18.60110092	NA	DryFlor
Copernicia alba	-58.115	-26.02	NA	NeoTropTree
Copernicia alba	-57.73333	-23.03333	NA	GBIF
Copernicia alba	-57.8578	-21.6514	NA	NeoTropTree
Copernicia alba	-57.6364	-19.1781	NA	NeoTropTree
Copernicia alba	-65.3011	-13.6622	NA	NeoTropTree
Copernicia alba	-55.7847	-16.5319	NA	NeoTropTree
Copernicia alba	-56.3767	-16.5075	NA	NeoTropTree
Copernicia alba	-60.6225	-29.9181	NA	NeoTropTree
Copernicia alba	-61.3036	-24.3133	NA	NeoTropTree
Copernicia alba	-57.53279877	-21.41749954	NA	DryFlor
Copernicia alba	-61.9367	-20.7025	NA	NeoTropTree
Copernicia alba	-60.8239	-18.5131	NA	NeoTropTree
Copernicia alba	-58.15000153	-19.06999969	NA	DryFlor
Copernicia alba	-57.8261	-22.5258	NA	NeoTropTree
Copernicia alba	-57.7561	-16.2008	NA	NeoTropTree
Copernicia alba	-57.5814	-27.7153	NA	NeoTropTree
Copernicia alba	-59.0217	-16.4147	NA	NeoTropTree
Copernicia alba	-58.5	-24.83333	NA	GBIF

Copernicia alba	-66.0819	-11.5486	NA	NeoTropTree
Copernicia alba	-59.60020065	-28.92480087	NA	DryFlor
Copernicia alba	-57.6444	-24.4075	NA	NeoTropTree
Copernicia alba	-65.6489	-11.6075	NA	NeoTropTree
Copernicia alba	-59.3786	-29.7994	NA	NeoTropTree
Copernicia alba	-58.7597	-27.9811	NA	NeoTropTree
Copernicia alba	-57.5142	-22.0336	NA	NeoTropTree
Copernicia alba	-58.4597	-16.44	NA	NeoTropTree
Copernicia alba	-60.5617	-25.5411	NA	NeoTropTree
Copernicia alba	-57.6414	-19.1075	NA	NeoTropTree
Copernicia alba	-57.62250137	-17.77249908	NA	DryFlor
Copernicia alba	-61.9722	-17.1344	NA	NeoTropTree
Copernicia alba	-62.6953	-21.2817	NA	NeoTropTree
Copernicia alba	-60.13380051	-29.32559967	NA	DryFlor
Copernicia alba	-57.88359833	-21.49810028	NA	DryFlor
Copernicia alba	-62.95000076	-17.21999931	NA	DryFlor
Copernicia alba	-61.5458	-13.9561	NA	NeoTropTree
Copernicia alba	-64.6281	-15.4331	NA	NeoTropTree
Copernicia alba	-59.13	-27.7	NA	GBIF
Copernicia alba	-61.8817	-22.8472	NA	NeoTropTree

Copernicia alba	-57.9381	-26.5086	NA	NeoTropTree
Copernicia alba	-56.14279938	-20.16110039	NA	DryFlor
Copernicia alba	-58.77000046	-23.62999916	NA	DryFlor
Copernicia alba	-60.2886	-23.9019	NA	NeoTropTree
Copernicia alba	-58.8914	-19.1983	NA	NeoTropTree
Copernicia alba	-66.6469	-14.7686	NA	NeoTropTree
Copernicia alba	-62.81999969	-15.52999973	NA	DryFlor
Copernicia alba	-63.4517	-26.4692	NA	NeoTropTree
Copernicia alba	-58.5	-16.5	NA	DryFlor
Copernicia alba	-58.93	-28.46	NA	GBIF
Copernicia alba	-66.1792	-14.5047	NA	NeoTropTree
Copernicia alba	-59.77700043	-27.10770035	NA	DryFlor
Copernicia alba	-59.2217	-21.7475	NA	NeoTropTree
Copernicia alba	-56.7678	-20.6664	NA	NeoTropTree
Copernicia alba	-60.48	-28.93	NA	GBIF
Copernicia alba	-63.20000076	-18.93000031	NA	DryFlor
Copernicia alba	-62.3025	-20.5844	NA	NeoTropTree
Copernicia alba	-57.5807991	-19.02779961	NA	DryFlor
Copernicia alba	-59.0181	-26.7758	NA	NeoTropTree
Copernicia alba	-57.3764	-25.2261	NA	NeoTropTree

Copernicia alba	-60.52999878	-19.43000031	NA	DryFlor
Copernicia alba	-58.17	-25.17	NA	GBIF
Copernicia alba	-58.0775	-28.0233	NA	NeoTropTree
Copernicia alba	-63.2222	-18.915	NA	NeoTropTree
Copernicia alba	-57.05	-28.52	NA	GBIF
Copernicia alba	-59.9328	-18.2525	NA	NeoTropTree
Copernicia alba	-58.2794	-27.2969	NA	NeoTropTree
Copernicia alba	-58.0336	-21.0297	NA	NeoTropTree
Copernicia alba	-58.18	-27.99	NA	GBIF
Copernicia alba	-57.6728	-16.4631	NA	NeoTropTree
Copernicia alba	-58.8114	-26.3044	NA	NeoTropTree
Copernicia alba	-60.0122	-26.2561	NA	NeoTropTree
Copernicia alba	-65.01999664	-14.11999989	NA	DryFlor
Copernicia alba	-63.33	-17.8	NA	GBIF
Copernicia prunifera	-42.766667	-11.016667	1977	GBIF
Copernicia prunifera	-42.7	-10.833333	1984	GBIF
Copernicia prunifera	-51.75	-11.283333	1985	GBIF
Copernicia prunifera	-43.833333	-11.333333	1988	GBIF
Copernicia prunifera	-44.475833	-11.618611	1988	GBIF
Copernicia prunifera	-38.228056	-6.759444	1992	GBIF

Copernicia prunifera	-41.520278	-3.458056	2004	GBIF
Copernicia prunifera	-40.591111	-9.691667	2007	GBIF
Copernicia prunifera	-43.233333	-11.683333	2007	GBIF
Copernicia prunifera	-43.133333	-11.083333	2007	GBIF
Copernicia prunifera	-41.279103	-9.613661	2008	speciesLink
Copernicia prunifera	-47.483055	-7.3438888	2008	speciesLink
Copernicia prunifera	-42.3015	-10.0012	2009	GBIF
Copernicia prunifera	-42.188333	-9.526667	2009	GBIF
Copernicia prunifera	-40.6525	-9.562222	2009	GBIF
Copernicia prunifera	-38.898333	-3.679167	2009	GBIF
Copernicia prunifera	-39.368167	-8.192722	2010	speciesLink
Copernicia prunifera	-39.330278	-8.307222	2010	speciesLink
Copernicia prunifera	-38.5925	-8.012056	2012	GBIF
Copernicia prunifera	-40.387439	-9.362231	2012	GBIF
Copernicia prunifera	-40.333575	-9.274358	2012	GBIF
Copernicia prunifera	-38.211111	-6.996667	2012	GBIF
Copernicia prunifera	-40.597789	-9.471692	2012	GBIF
Copernicia prunifera	-40.728778	-9.434	2013	GBIF
Copernicia prunifera	-38.434814	-7.163875	2013	GBIF
Copernicia prunifera	-38.140939	-6.981106	2013	GBIF

Copernicia prunifera	-40.673611	-9.458583	2013	GBIF
Copernicia prunifera	-38.998389	-7.700667	2013	speciesLink
Copernicia prunifera	-37.783056	-5.65	2014	GBIF
Copernicia prunifera	-37.919444	-6.095556	2014	GBIF
Copernicia prunifera	-42.696033	-7.0104	2014	GBIF
Copernicia prunifera	-38.255	-6.370833	2014	GBIF
Copernicia prunifera	-38.637222	-4.039722	2014	GBIF
Copernicia prunifera	-36.720278	-5.580278	2014	GBIF
Copernicia prunifera	-36.720278	-5.939722	2014	GBIF
Copernicia prunifera	-37.8575	-5.464722	2014	speciesLink
Copernicia prunifera	-41.69875	-3.372167	2015	GBIF
Copernicia prunifera	-42.83558333	-2.73373333	2015	iNaturalist
Copernicia prunifera	-44.8122496499	-3.5153573541	2016	iNaturalist
Copernicia prunifera	-45.116556	-2.421333	2017	GBIF
Copernicia prunifera	-40.294972	-9.184583	2017	GBIF
Copernicia prunifera	-41.495453	-5.062361	2017	GBIF
Copernicia prunifera	-41.866944	-3.175	2017	GBIF
Copernicia prunifera	-45.1059830556	-2.3653197222	2017	iNaturalist
Copernicia prunifera	-39.397083	-3.61325	2017	speciesLink
Copernicia prunifera	-41.773875	-3.367925	2017	speciesLink

Copernicia prunifera	-41.879583	-3.166611	2017	speciesLink
Copernicia prunifera	-37.7828916	-4.8378331	2018	iNaturalist
Copernicia prunifera	-35.9729959723	-8.2648044219	2018	iNaturalist
Copernicia prunifera	-43.380003	-2.519996	2018	speciesLink
Copernicia prunifera	-44.280178	-10.079572	2018	speciesLink
Copernicia prunifera	-39.240258	-4.9479	2018	speciesLink
Copernicia prunifera	-45.077111	-3.175675	2018	speciesLink
Copernicia prunifera	-41.318778	-3.000167	2018	speciesLink
Copernicia prunifera	-38.0867	-6.7844	2018	speciesLink
Copernicia prunifera	-37.3367317021	-5.1944106183	2019	iNaturalist
Copernicia prunifera	-37.0472221375	-6.76222229	2019	iNaturalist
Copernicia prunifera	-43.1575	-10.9797	NA	NeoTropTree
Copernicia prunifera	-41.4997	-3.4328	NA	NeoTropTree
Copernicia prunifera	-38.8608	-6.2547	NA	NeoTropTree
Copernicia prunifera	-37.5558	-9.1392	NA	NeoTropTree
Copernicia prunifera	-41.6389	-2.9397	NA	NeoTropTree
Copernicia prunifera	-43.2278	-12.1039	NA	NeoTropTree
Copernicia prunifera	-40.8444	-8.8517	NA	NeoTropTree
Copernicia prunifera	-47.92219925	-9.816699982	NA	DryFlor
Copernicia prunifera	-47.39110184	-7.187799931	NA	DryFlor

Copernicia prunifera	-40.2517	-9.4197	NA	NeoTropTree
Copernicia prunifera	-37.7769	-6.7706	NA	NeoTropTree
Copernicia prunifera	-40.825	-3.1403	NA	NeoTropTree
Copernicia prunifera	-39.1878	-10.4117	NA	NeoTropTree
Copernicia prunifera	-37.59780121	-5.074200153	NA	DryFlor
Copernicia prunifera	-37.5578	-5.5025	NA	NeoTropTree
Copernicia prunifera	-42.88529968	-10.11830044	NA	DryFlor
Copernicia prunifera	-42.22779846	-4.713600159	NA	DryFlor
Copernicia prunifera	-37.74250031	-9.646699905	NA	DryFlor
Copernicia prunifera	-37.0314	-6.0547	NA	NeoTropTree
Copernicia prunifera	-43.2664	-2.4992	NA	NeoTropTree
Copernicia prunifera	-42.5036	-3.4475	NA	NeoTropTree
Copernicia prunifera	-47.6267	-7.62	NA	NeoTropTree
Copernicia prunifera	-43.5431	-4.9031	NA	NeoTropTree
Copernicia prunifera	-37.0411	-9.9839	NA	NeoTropTree
Copernicia prunifera	-35.8664	-5.7025	NA	NeoTropTree
Copernicia prunifera	-37.9586	-5.0192	NA	NeoTropTree
Copernicia prunifera	-37.07310104	-5.054999828	NA	DryFlor
Copernicia prunifera	-38.9561	-3.735	NA	NeoTropTree
Copernicia prunifera	-37.39720154	-9.744999886	NA	DryFlor

Copernicia prunifera	-37.1578	-9.6203	NA	NeoTropTree
Copernicia prunifera	-38.2881	-4.2169	NA	NeoTropTree
Copernicia prunifera	-38.5461	-7.5625	NA	NeoTropTree
Copernicia prunifera	-42.9664	-4.3669	NA	NeoTropTree
Copernicia prunifera	-37.8119	-4.6858	NA	NeoTropTree
Copernicia prunifera	-37.965	-7.5158	NA	NeoTropTree
Copernicia prunifera	-38.8375	-7.8731	NA	NeoTropTree
Copernicia prunifera	-42.2711	-10.455	NA	NeoTropTree
Copernicia prunifera	-42.4264	-11.3372	NA	NeoTropTree
Copernicia prunifera	-43.6653	-11.8633	NA	NeoTropTree
Copernicia prunifera	-41.2153	-9.0481	NA	NeoTropTree
Copernicia prunifera	-43.3242	-13.1167	NA	NeoTropTree
Copernicia prunifera	-41.69670105	-4.067200184	NA	DryFlor
Copernicia prunifera	-43.68330002	-8.168100357	NA	DryFlor
Copernicia prunifera	-36.9394	-5.6244	NA	NeoTropTree
Copernicia prunifera	-39.7392	-5.7186	NA	NeoTropTree
Copernicia prunifera	-43.84560013	-14.16390038	NA	DryFlor
Copernicia prunifera	-38.6556	-5.8039	NA	NeoTropTree
Copernicia prunifera	-37.3606	-8.905	NA	NeoTropTree
Copernicia prunifera	-36.8283	-5.3378	NA	NeoTropTree

Copernicia prunifera	-37.39360046	-5.144199848	NA	DryFlor
Copernicia prunifera	-42.72249985	-12.01080036	NA	DryFlor
Copernicia prunifera	-49.7989	-10.8083	NA	NeoTropTree
Copernicia prunifera	-38.0767	-7.3228	NA	NeoTropTree
Copernicia prunifera	-43.6331	-8.0294	NA	NeoTropTree
Copernicia prunifera	-37.8092	-9.4486	NA	NeoTropTree
Copernicia prunifera	-40.6003	-9.2342	NA	NeoTropTree
Copernicia prunifera	-47.4661	-7.2597	NA	NeoTropTree
Copernicia prunifera	-42.455	-12.8133	NA	NeoTropTree
Copernicia prunifera	-43.52140045	-14.98219967	NA	DryFlor
Copernicia prunifera	-50.7331	-12.3281	NA	NeoTropTree
Copernicia prunifera	-37.2417	-5.2081	NA	NeoTropTree
Copernicia prunifera	-44.18	-10.1144	NA	NeoTropTree
Copernicia prunifera	-42.8358	-2.6708	NA	NeoTropTree
Copernicia prunifera	-43.9483	-7.6564	NA	NeoTropTree
Copernicia prunifera	-37.2622	-9.8572	NA	NeoTropTree
Copernicia prunifera	-40.1611	-3.3314	NA	NeoTropTree
Copernicia prunifera	-50.4564	-10.3653	NA	NeoTropTree
Copernicia prunifera	-39.1394	-5.3458	NA	NeoTropTree
Copernicia prunifera	-43.8683	-13.9311	NA	NeoTropTree

Copernicia prunifera	-38.7781	-3.7053	NA	NeoTropTree
Copernicia prunifera	-36.97890091	-5.303899765	NA	DryFlor
Copernicia prunifera	-36.6872	-10.2947	NA	NeoTropTree
Copernicia prunifera	-35.25439835	-7.177199841	NA	DryFlor
Copernicia prunifera	-39.4436	-8.2011	NA	NeoTropTree
Copernicia prunifera	-47.8208	-9.8603	NA	NeoTropTree
Copernicia prunifera	-38.4294	-5.1161	NA	NeoTropTree
Copernicia prunifera	-39.3783	-6.3256	NA	NeoTropTree
Copernicia prunifera	-42.9275	-10.0928	NA	NeoTropTree
Copernicia prunifera	-50.6792	-11.1703	NA	NeoTropTree
Copernicia prunifera	-39.4522	-8.4611	NA	NeoTropTree
Copernicia prunifera	-36.66139984	-10.27579975	NA	DryFlor
Copernicia prunifera	-36.1742	-7.9397	NA	NeoTropTree
Copernicia prunifera	-38.2228	-7.9575	NA	NeoTropTree
Copernicia prunifera	-38.92829895	-6.300000191	NA	DryFlor
Copernicia prunifera	-42.865	-5.9447	NA	NeoTropTree
Copernicia prunifera	-35.1594	-7.0358	NA	NeoTropTree
Copernicia prunifera	-36.70080185	-5.251699924	NA	DryFlor
Copernicia prunifera	-38.5247	-7.0992	NA	NeoTropTree
Copernicia prunifera	-43.0886	-6.7539	NA	NeoTropTree

Copernicia prunifera	-36.7194	-5.2417	NA	NeoTropTree
Copernicia prunifera	-48.4944	-16.0992	NA	NeoTropTree
Copernicia prunifera	-50.9731	-14.3328	NA	NeoTropTree
Copernicia prunifera	-38.8883	-3.5981	NA	NeoTropTree
Copernicia prunifera	-38.1764	-6.7183	NA	NeoTropTree
Copernicia prunifera	-36.8564	-10.1717	NA	NeoTropTree
Copernicia prunifera	-38.5328	-6.8267	NA	NeoTropTree
Copernicia prunifera	-42.2194	-9.6569	NA	NeoTropTree
Copernicia prunifera	-39.0719	-4.9183	NA	NeoTropTree
Copernicia prunifera	-49.985	-9.5883	NA	NeoTropTree
Copernicia prunifera	-37.2078	-7.0375	NA	NeoTropTree
Copernicia prunifera	-37.9544	-7.0978	NA	NeoTropTree
Copernicia prunifera	-42.50719833	-12.76580048	NA	DryFlor
Copernicia prunifera	-36.0711	-5.2269	NA	NeoTropTree
Copernicia prunifera	-41.3933	-2.9683	NA	NeoTropTree
Copernicia prunifera	-43.8286	-11.3383	NA	NeoTropTree
Copernicia prunifera	-37.6114	-6.2753	NA	NeoTropTree
Copernicia prunifera	-42.2189	-4.7631	NA	NeoTropTree
Copernicia prunifera	-50.9228	-12.6369	NA	NeoTropTree

Table S3. The bioclimatic variables for niche suitability modeling of *C.alba* e *C.prunifera* in South America.

Variable	Description	Source
BIO01	Annual mean temperature	Worldclim
BIO02	Mean diurnal range (mean of monthly (max temp - min temp))	Worldclim
BIO03	Isothermality (bio2/bio7) (* 100)	Worldclim
BIO04	Temperature seasonality (standard deviation * 100)	Worldclim
BIO05	Max temperature of warmest month	Worldclim
BIO06	Min temperature of coldest month	Worldclim
BIO07	Temperature annual range (bio5-bio6)	Worldclim
BIO08	Mean temperature of wettest quarter	Worldclim
BIO09	Mean temperature of driest quarter	Worldclim
BIO10	Mean temperature of warmest quarter	Worldclim
BIO11	Mean temperature of coldest quarter	Worldclim
BIO12	Annual precipitation	Worldclim
BIO13	Precipitation of wettest month	Worldclim
BIO14	Precipitation of driest month	Worldclim
BIO15	Precipitation seasonality (coefficient of variation)	Worldclim
BIO16	Precipitation of wettest quarter	Worldclim
BIO17	Precipitation of driest quarter	Worldclim
BIO18	Precipitation of warmest quarter	Worldclim
BIO19	Precipitation of coldest quarter	Worldclim

Table S4. Correlation between the environmental variables in the current scenario for the Neotropics limit proposed by Morrone (2014).

	BIO01	BIO02	BIO03	BIO04	BIO05	BIO06	BIO07	BIO08	BIO09	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19
BIO01																			
BIO02	-0.62																		
BIO03	0.63	-0.58																	
BIO04	-0.69	0.61	-0.91																
BIO05	0.65	-0.02	0.03	-0.16															
BIO06	0.95	-0.79	0.75	-0.78	0.45														
BIO07	-0.73	0.88	-0.87	0.85	-0.05	-0.88													
BIO08	0.85	-0.4	0.35	-0.39	0.77	0.72	-0.45												
BIO09	0.96	-0.7	0.7	-0.75	0.53	0.97	-0.8	0.72											
BIO10	0.89	-0.46	0.36	-0.39	0.82	0.78	-0.48	0.95	0.8										
BIO11	0.98	-0.66	0.72	-0.79	0.56	0.97	-0.79	0.77	0.98	0.81									
BIO12	0.65	-0.63	0.65	-0.75	0.21	0.73	-0.71	0.4	0.7	0.43	0.71								
BIO13	0.7	-0.56	0.65	-0.75	0.3	0.75	-0.69	0.43	0.75	0.47	0.75	0.91							
BIO14	0.26	-0.61	0.36	-0.36	-0.13	0.41	-0.5	0.16	0.31	0.17	0.29	0.67	0.41						
BIO15	0	0.38	-0.06	0.08	0.18	-0.12	0.2	0.03	-0.04	0.02	-0.01	-0.38	-0.05	-0.84					
BIO16	0.7	-0.55	0.65	-0.75	0.31	0.74	-0.68	0.43	0.74	0.47	0.75	0.93	0.99	0.42	-0.07				
BIO17	0.31	-0.63	0.39	-0.41	-0.09	0.45	-0.53	0.18	0.36	0.2	0.34	0.72	0.47	0.99	-0.84	0.48			
BIO18	0.09	-0.16	0.2	-0.24	-0.1	0.11	-0.2	0.12	0.04	0	0.1	0.53	0.38	0.53	-0.37	0.4	0.54		
BIO19	0.59	-0.74	0.55	-0.61	0.15	0.71	-0.7	0.35	0.67	0.45	0.63	0.8	0.68	0.77	-0.53	0.69	0.8	0.26	

Table S5. Values of AUCs and TSSs mean values and standard deviation for ten replicates of each algorithm for *C. alba* and *C. prunifera* species.

Species	**Algorithm**	**AUC average**	**AUC standard deviation**	**TSS average**	**TSS standard deviation**
Copernicia alba	Bioclim	0.93	0.02	0.78	0.05
Copernicia alba	MaxEnt	0.97	0.01	0.85	0.04
Copernicia alba	Random Forest	0.97	0.01	0.88	0.03
Copernicia alba	SVM	0.95	0.02	0.84	0.04
Copernicia prunifera	Bioclim	0.86	0.04	0.66	0.07
Copernicia prunifera	MaxEnt	0.94	0.02	0.78	0.07
Copernicia prunifera	Random Forest	0.97	0.02	0.83	0.05
Copernicia prunifera	SVM	0.96	0.02	0.82	0.05

PERSPECTIVES AND FINAL CONSIDERATIONS

The first two chapters described in this thesis correspond to pioneering studies on the population genomics of two palm species of the genus *Copernicia* (Carandá and Carnaúba) using the GBS technique. The third chapter was the first to assess the climatic niches of these palms and predict vulnerability to climate change in South America.

Therefore, this chapter provides a complete overview of the population genomics of these species in Brazil, generating information that can help design efficient strategies for the conservation and sustainable use of these palms. The genetic diversity levels of Carnaúba populations from the Caatinga and Restinga biomes were similar, indicating low to moderate level of genetic structuring. Therefore, there is no genetic separation between Carnaúba from Caatinga and Restinga or reproductive subdivision. Moderate levels of genetic differentiation were observed in the Carandá populations of the Brazilian Chaco region, indicating no significant genetic differentiation. In general, the Carandá and Carnaúba populations are genetically similar and can be managed as a single unit. Furthermore, the populations of the two species are historically connected with high dispersal, preventing the loss of genetic diversity.

Regarding outlier SNPs (under selection), 65 and 84 putatively adaptive loci were detected using Fsthet and PCAdapt approaches for Carandá and Caraúba, respectively. For the Carandá populations, two outlier loci were present in sequences similar to the annotated proteins, whereas six outliers were detected in the Carnaúba populations. The lack of a reference genome for the species has limited a more refined characterization of these loci. Therefore, further studies using integrative approaches, such as associative mapping, are important for understanding the adaptive nature of these loci.

Importantly, the analyses performed in this thesis were unable to associate generic molecular functions or biological processes with adaptive features involved in the diversification of Carandá and carnaúba populations. Therefore, phylogeographic studies are required to generate information on the evolution and diversification of these palms, given that the genus *Copernicia* is an important group for studying species distribution in the dry diagonal. Carnaúba occurs in Caatinga, Cerrado, and Restinga. However, in this thesis populations from the Cerrado were not evaluated.

Ecological niche models (ENMs) predicted that climate change affects the geographic distribution of Carandá and carnaúba in Neotropical ecosystems. Furthermore, these palms show different responses to future climate change. Therefore, climate projections for 2070 show

a reduction in the range of Carandá and increasing habitat availability for carnaúba. However, it should be emphasized that this study only analyzed the climatic niche of these species. Thus, studies using climate and edaphic data should be developed to generate ENMs with greater predictive power, since these palms occur in periodically flooded regions.

The results obtained in Chapters I, II, and III are promising for the conservation of Carandá and Carnaúba populations in Brazil. This information provides subsidies that will assist in developing national policies aimed at the management and conservation of genetic diversity and reducing the pressure of extractivism and its negative impacts.

Although the impact of human disturbance was observed in most Carandá and Carnaúba populations, they managed to maintain a high degree of genetic diversity and low levels of inbreeding. According to conservation genetic principles, habitat fragmentation and anthropogenic activities are expected to reduce genetic diversity, increase genetic differentiation and inbreeding levels. Although these palms are in an area subject to fragmentation and other disturbances, they apparently have mechanisms to maintain genetic diversity. According to Lowe et al. (2015) tree species have different strategies for maintaining genetic diversity levels in fragmented environments. The first strategy is extensive gene flow via pollen and/or seed. The second is due to the long-lived nature of trees and the existence of overlapping generations in unique locations that serve to slow the loss of genetic diversity. Third is the presence of mobile pollinators such as birds, which tend to buffer the negative impact of fragmentation, in the case of Carnaúba the main disperser is birds (*Thraupis palmarum*)

However, if this level of disturbance persists, it may lead to further population decline with unpredictable consequences on future levels of diversity. Therefore, some of the conservation actions that should be implemented or established are: (I) creation and maintenance of forest protection areas with ecological corridors; (II) development of *in situ* e *ex situ* conservation banks with the promotion of natural regeneration, avoiding reduction in population size and loss of genetic variability; (III) reinsertion of the plantations in the original natural populations, increasing the number of herds and preserving genetic diversity; and (IV) Development of planted forests (productive Carnaúba forests) to supply the demand by producers and reduce the pressure of extractivism on natural populations.

Finally, this thesis revealed that the Carnaúba populations of Northeastern Brazil and the Carandá populations of Brazilian Chaco have high genetic diversity, with low to moderate

genetic differentiation between populations. This information should be considered in developing a best management practice plan for the sustainable extraction of these palms.

REFERENCES

ALMEILDA, J.A.S. et al. Use, perception, and local management of *Copernicia prunifera* (Miller) HE Moore in rural communities in the Brazilian Savanna. **Journal of Ethnobiology and Ethnomedicine**, v. 17, n. 1, p. 1-13, 2021.

ALVARADO-SERRANO, D.F.; KNOWLES, L.L. Ecological niche models in phylogeographic studies: applications, advances and precautions. **Molecular ecology resources**, v. 14, n. 2, p. 233-248, 2014.

ARRUDA, G.M.T.; CALBO, M.E.R. Efeitos da inundação no crescimento, trocas gasosas e porosidade radicular da carnaúba (*Copernicia prunifera* (Mill.) HE Moore). **Acta Botanica Brasilica**, v. 18, n. 2, p. 219-224, 2004.

BACON, C.D. et al. Species limits, geographical distribution and genetic diversity in *Johannesteijsmannia* (Arecaceae). **Botanical Journal of the Linnean Society**, v. 182, n. 2, p. 318-347, 2016.

BAIRD, N.A. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PloS one**, v. 3, n. 10, p. e3376, 2008.

BARRETO, H.N.; PARISE, C.K.; ALMEIDA, E.B. The Cocais Forest Landscape. In: **The Physical Geography of Brazil**. Springer, Cham, 2019. p. 151-167.

BELLO, C. et al. Environmental niche and functional role similarity between invasive and native palms in the Atlantic Forest. **Biological invasions**, v. 23, n. 3, p. 741-754, 2021.

BLACH-OVERGAARD, A. et al. Global-change vulnerability of a key plant resource, the African palms. **Scientific reports**, v. 5, n. 1, p. 1-10, 2015.

BOOTH, T. H. et al. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. **Diversity and Distributions**, v. 20, n. 1, p. 1-9, 2014.

BREIMAN, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

BRITO, F.A. et al. Low diversity in the native populations of *Croton tetradenius* Baill. when using SNP markers: a future crop with an insecticidal activity. **Genetic Resources and Crop Evolution**, p. 1-15, 2021.

BROENNIMANN, O. et al. Measuring ecological niche overlap from occurrence and spatial environmental data. **Global ecology and biogeography**, v. 21, n. 4, p. 481-497, 2012.

CAMPOS, J.L.A. et al. Socioeconomic Factors and Cultural Changes Explain the Knowledge and Use of Ouricuri Palm (*Syagrus coronata*) by the Fulni–ô Indigenous People of Northeast Brazil. **Economic Botany**, v. 73, n. 2, p. 187-199, 2019.

CARVALHO, C.S. et al. Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. **Conservation Genetics**, v. 18, n. 2, p. 467-478, 2017.

CÁSSIA-SILVA, C. et al. Niche conservatism drives a global discrepancy in palm species richness between seasonally dry and moist habitats. **Global Ecology and Biogeography**, v. 28, n. 6, p. 814-825, 2019.

CATCHEN, J. et al. Stacks: an analysis tool set for population genomics. **Molecular ecology**, v. 22, n. 11, p. 3124-3140, 2013.

CHALGHAF, B. et al. Ecological niche modeling predicting the potential distribution of *Leishmania* vectors in the Mediterranean basin: impact of climate change. **Parasites & vectors**, v. 11, n. 1, p. 1-9, 2018.

CHAMBERLAIN, S.A.; BOETTIGER, Carl. **R Python, and Ruby clients for GBIF species occurrence data**. PeerJ Preprints, 2017.

CORDEIRO, E.M.G et al. Diversity, genetic structure, and population genomics of the tropical tree *Centrolobium tomentosum* in remnant and restored Atlantic forests. **Conservation Genetics**, v. 20, n. 5, p. 1073-1085, 2019.

COSTA, M. F. et al. Climate change impacts on the *Copernicia alba* and *Copernicia prunifera* (Arecaceae) distribution in South America. **Brazilian Journal of Botany**, p. 1-12, 2022.

CRUZ, M.V. et al. Molecular responses to freshwater limitation in the mangrove tree *Avicennia germinans* (Acanthaceae). **Molecular ecology**, v. 29, n. 2, p. 344-362, 2020.

DAVEY, J.W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, v. 12, n. 7, p. 499-510, 2011.

DRANSFIELD, J. et al. A new Coryphoid palm genus from Madagascar. **Botanical Journal of the Linnean Society**, v. 156, n. 1, p. 79-91, 2008.

DOYLE, J.J.; DOYLE, J.L. **A rapid DNA isolation procedure for small quantities of fresh leaf tissue**. 1987.

EISERHARDT, W.L. et al. Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents. **Scientific reports**, v. 3, n. 1, p. 1-8, 2013.

ELSHIRE, Robert J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, v. 6, n. 5, p. e19379, 2011.

EMILIO, T. et al. Embolism resistance in petioles and leaflets of palms. **Annals of botany**, v. 124, n. 7, p. 1173-1183, 2019.

ESSER, L.F.; SARAIVA, D.D.; JARENKOW, J.A. Future uncertainties for the distribution and conservation of *Paubrasilia echinata* under climate change. **Acta Botanica Brasilica**, v. 33, n. 4, p. 770-776, 2019.

FAJARDO, C. G. et al. Genetic and phenotypic association of the carnauba palm tree evaluated by inter-simple sequence repeat and biometric traits. **Genetics and Molecular Research**, v. 17, n. 3, 2018.

FENG, X. et al. Collinearity in ecological niche modeling: Confusions and challenges. **Ecology and Evolution**, v. 9, n. 18, p. 10365-10376, 2019.

FERREIRA, C.S.; NUNES, J.A.R.; GOMES, R.L.F. Manejo de corte das folhas de *Copernicia prunifera* (Miller) HE Moore no Piauí. **Revista Caatinga**, v. 26, n. 2, p. 25-30, 2013.

FICK, S.E.; HIJMANS, R.J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International journal of climatology**, v. 37, n. 12, p. 4302-4315, 2017.

FREITAS, C.A.S. et al. Carnauba wax uses in food–A review. **Food chemistry**, v. 291, p. 38-48, 2019.

GAUTO, M.I.; STAUFFER, F.W. **Botánica económica de las palmeras de Paraguay: familia'Arecaceae'**. 2016.

GUISAN, Antoine et al. Predicting species distributions for conservation decisions. **Ecology letters**, v. 16, n. 12, p. 1424-1435, 2013.

GUO, Q.; KELLY, M.; GRAHAM, C.H. Support vector machines for predicting distribution of Sudden Oak Death in California. **Ecological modelling**, v. 182, n. 1, p. 75-90, 2005.

HIJMANS, R.J. et al. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology: A Journal of the Royal Meteorological Society**, v. 25, n. 15, p. 1965-1978, 2005.

JOMBART, T.; AHMED, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. **Bioinformatics**, v. 27, n. 21, p. 3070-3071, 2011.

KARGER, D.N. et al. Climatologies at high resolution for the earth's land surface areas. **Scientific data**, v. 4, n. 1, p. 1-20, 2017.

LORENZI, Harri . (2010). **Flora Brasileira – Arecaceae (Palmeiras)**. Nova Odessa: Instituto Plantarum. 368 p.

LOWE, A. J. et al. The resilience of forest fragmentation genetics—no longer a paradox—we were just looking in the wrong place. **Heredity**, v. 115, n. 2, p. 97-99, 2015.

LIU, L. et al. Modeling habitat suitability of *Houttuynia cordata* Thunb (Ceercao) using MaxEnt under climate change in China. **Ecological Informatics**, p. 101324, 2021.

LUIKART, G. et al. The power and promise of population genomics: from genotyping to genome typing. **Nature reviews genetics**, v. 4, n. 12, p. 981-994, 2003.

MAITNER, B. S. et al. The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. **Methods in Ecology and Evolution**, v. 9, n. 2, p. 373-379, 2018.

MEYER, D et al. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. **R package version**, p. 1.7-1, 2019.

MOTHES, C. C. et al. Evaluating ecological niche model accuracy in predicting biotic invasions using South Florida's exotic lizard community. **Journal of biogeography**, v. 46, n. 2, p. 432-441, 2019.

MORAES, M. et al. Palmas útiles de Bolivia, Colombia, Ecuador y Perú. **Cosecha Palmas en el Noroeste Suramérica Bases Científicas Para Su Manejo Conserv; Balslev, H., Macia, M., Navarrete, H., Eds**, p. 87-102, 2014.

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

NARUM, Shawn R. et al. Genotyping-by-sequencing in ecological and conservation genomics. **Molecular ecology**, v. 22, n. 11, p. 2841, 2013.

NOVELLO, M. et al. Genetic conservation of a threatened Neotropical palm through community-management of fruits in agroforests and second-growth forests. **Forest Ecology and Management**, v. 407, p. 200-209, 2018.

PETERMAN, W.E.; CRAWFORD, J.A.; KUHNS, A.R. Using species distribution and occupancy modeling to guide survey efforts and assess species status. **Journal for Nature Conservation**, v. 21, n. 2, p. 114-121, 2013.

PETERSON, A. T. et al. **Ecological niches and geographic distributions (MPB-49)**. Princeton University Press, 2011.

PHILLIPS, S. J. et al. Opening the black box: An open-source release of Maxent. **Ecography**, v. 40, n. 7, p. 887-893, 2017.

PINHEIRO, L. G. et al. Anthropization as a determinant factor in the genetic structure of *Copernicia prunifera* (Arecaceae). **Genetics and Molecular Research**, v. 16, n. 3, 2017.

POLAND, J.A. et al. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. **PloS one**, v. 7, n. 2, p. e32253, 2012.

PULIDO-SILVA, M.T. et al. Richness, Endemism and Floristic Affinities of the Palms of Mexico. **The Botanical Review**, p. 1-25, 2022.

RANGEL, T.F.; LOYOLA, R.D. Labeling ecological niche models. **Natureza Conservação**, v. 10, p. 119-126, 2012.

REIS, R.G.E. et al. Emergência e qualidade de mudas de *Copernicia prunifera* em função da embebição das sementes e sombreamento. **Revista Caatinga**, v. 24, n. 4, p. 43-49, 2011.

RODRIGUES, L.C. et al. Conhecimento e uso da carnaúba e da algaroba em comunidades do Sertão do Rio Grande do Norte, Nordeste do Brasil. **Revista Árvore**, v. 37, n. 3, p. 451-457, 2013.

RODRIGUES, P.A.S. et al. Hypolipidemic activity of P-methoxycinnamic diester (PCO-C) isolated from *Copernicia prunífera* against Triton WR-1339 and hyperlipidemic diet in mice. **Environmental toxicology and pharmacology**, v. 56, p. 198-203, 2017.

SANTOS, J.R.M. **Modelagem de nicho ecológico e diversidade genética de *Copernicia prunifera* (Miller) HE Moore (Arecaceae): implicações para a conservação**. 2017. Dissertação de Mestrado. Brasil.

SILVA, Rosa Helena. **Estabelecimento de protocolo para uso sustentável de *Copernicia alba* MORONG EX MORONG & BRITTON na bacia do rio paraguai, MS**. 2018. Tese de Doutorado. Brasil.

SMITH, Nigel. **Palms and People in the Amazon**. Springer, 2014.

SILVA, A.M.F. et al. Spatial prediction of risk areas for vector transmission of *Trypanosoma cruzi* in the State of Paraná, southern Brazil. **PLoS neglected tropical diseases**, v. 12, n. 10, p. e0006907, 2018.

SILVA, J.Y.G. et al. Chemical characterization of p-hydroxycinnamic diesters extracted from *Copernicia prunifera* and attenuation of biomarkers in C57BL/6J diabetic mice. **Food Bioscience**, v. 41, p. 100961, 2021.

SILLERO, N.; BARBOSA, A.M. Common mistakes in ecological niche models. **International Journal of Geographical Information Science**, v. 35, n. 2, p. 213-226, 2021.

SOARES, K.P. et al. Palmeiras (Arecaceae) no Rio Grande do Sul, Brasil. **Rodriguésia**, v. 65, n. 1, p. 113-139, 2014.

SOBRAL-SOUZA, T. et al. Efficiency of protected areas in Amazon and Atlantic Forest conservation: A spatio-temporal view. **Acta Oecologica**, v. 87, p. 1-7, 2018.

SOUZA, F.G. et al. Brazilian fruits of Arecaceae family: an overview of some representatives with promising food, therapeutic and industrial applications. **Food Research International**, p. 109690, 2020.

SOUZA, V. C.; LORENZI, H. Botânica sistemática: guia ilustrativo para identificação das famílias de fanerógamas nativas e exóticas no Brasil, baseado em APG III. **Nova Odessa, SP: Instituto Plantarum**, 2012.

SUAREZ, Maria Eugenia et al. **Etnobotánica de las palmeras *Copernicia alba* Morong y *Trithrinax schizophylla* Drude en el Gran Chaco y alrededores: una revisión**. 2020.

TITLE, P.O.; BEMMELS, J.B. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. **Ecography**, v. 41, n. 2, p. 291-307, 2018.

TEAM, R. C. 2019. **R Foundation for Statistical Computing**. R: A language and environment for statistical computing.

TORKAMANEH, D.; BOYLE, B.; BELZILE, F. Efficient genome-wide genotyping strategies and data integration in crop plants. **Theoretical and applied genetics**, v. 131, n. 3, p. 499-511, 2018.

URBAN, M.C. Accelerating extinction risk from climate change. **Science**, v. 348, n. 6234, p. 571-573, 2015.

VIANA, E.O. Pós-colheita de sementes e produção de mudas de *Copernicia*. 2019. Tese de Doutorado. Brasil.

VIEIRA, F.A. et al. Diversidade genética de *Copernicia prunifera* com o uso de marcadores moleculares ISSR. **Revista Brasileira de Ciências Agrárias**, v. 10, n. 4, p. 525-531, 2015.

VIANA, J.P.G. et al. Genomic diversity is similar between Atlantic Forest restorations and natural remnants for the native tree *Casearia sylvestris* Sw. **PloS one**, v. 13, n. 3, p. e0192165, 2018.

XIA, W. et al. Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. **Frontiers in plant science**, v. 10, p. 130, 2019.

XIMENES NETO, F.R.G.; CRISPIM, F.S.P.; BRAGA, P.E.T. Processos produtivos de trabalhadores rurais no extrativismo da palha de carnaúba. **Interações (Campo Grande)**, v. 20, n. 4, p. 1263-1273, 2019.

ZIZKA, A. et al. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. **Methods in Ecology and Evolution**, v. 10, n. 5, p. 744-751, 2019.

ZURELL, D. et al. A standard protocol for reporting species distribution models. **Ecography**, v. 43, n. 9, p. 1261-1277, 2020.

ATTACHMENTS

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 "**Population genomics and ecological niche modeling of the neotropical palms carandá and carnaúba: implications for conservation**", 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 Marcones Ferreira Costa.
Nome do aluno: Marcones Ferreira Costa

Assinatura MFZucchi
Nome da Orientadora: Maria Imaculada Zucchi

Data: 18 de Janeiro de 2023.

DECLARAÇÃO

As cópias de artigos de minha autoria ou de minha coautoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Tese de Doutorado "**Population genomics and ecological niche modeling of the neotropical palms carandá and carnaúba: implications for conservation**", não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 18 de Janeiro de 2023.

Assinatura Marcones Ferreira Costa.
Nome do aluno: Marcones Ferreira Costa

Assinatura MIZucchi
Nome da Orientadora: Maria Imaculada Zucchi