

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
PÓS-GRADUAÇÃO EM GENÉTICA E MELHORAMENTO**

FRANCINE ALVES NOGUEIRA DE ALMEIDA

**POPULATION GENOMICS AND
PHYLOGEOGRAPHY OF *Euterpe edulis***

ALEGRE
2023

FRANCINE ALVES NOGUEIRA DE ALMEIDA

POPULATION GENOMICS AND PLHYLOGEOGRAPHY
OF *Euterpe edulis*

Thesis presented to the Postgraduate
Genetics and Improvement Programme
of the Center for Agricultural Sciences
and Engineering of the Federal
University of Espírito Santo.

Advisor: Dr Marcia Flores da Silva
Ferreira

Co-advisor: Dr. Christine D. Bacon

Co-advisor: Dr Adésio Ferreira

ALEGRE
2023

ACKNOWLEDGMENT

I would like to thank my advisor Dr. Marcia Flores and my co-supervisor Adésio Ferreira, for their guidance and opportunity to participate in this work of such importance, which enchanted me so much.

To Dr. Christine Bacon for hosting at the University of Gothenburg and co-supervising this work. I am grateful for all the attention and motivation.

To the entire team at the University of Gothenburg and Gothenburg Global Biodiversity Center who were very receptive and allowed me to participate in several meetings and enriching discussions.

To my colleagues for collecting samples Aléxia, José Guilherme, Guilherme, Ramon and Wagner.

To Lucas Vieira for his collaboration in the analysis and discussion of this work.

To my friends from the Laboratory of Genetics and Improvement and Biometrics, Guilherme, Vinicius, Gilza, Jhonatas and Diego for their support in difficult times and for the laughs they provided.

To the friends I made in Gothenburg, Alex, Rute, Lucas, Milena, Adele, Anne-Sophie, Daniela, and Adrian. And so many others who made homesickness bearable.

I would like to thank my colleagues and technicians from the Biometry and Genetics and Improvement laboratories at the UFES Campus de Alegre, for their contributions and good times.

I would like to thank FAPES and VALE for funding this project, CAPES for the grant and the Graduate Program in Genetics and Improvement at UFES. I would like to thank CNPq for the scholarship to carry out the Sandwich Doctorate. And to all farmers who provided samples.

POPULATION GENOMICS AND PLHYLOGEOGRAPHY OF *Euterpe edulis*

ABSTRACT

DE ALMEIDA, F. A. N. POPULATION GENOMICS AND PLHYLOGEOGRAPHY OF *Euterpe edulis*. 2023. Thesis (PhD in Genetics and Improvement) - Postgraduate Genetics and Improvement, UFES, Espírito Santo. Brazil.

The Atlantic Forest is one of the five priority global biodiversity hotspots for conservation. Along its distribution, the vegetation presents different characteristics due to biogeographic variation. The palm *Euterpe edulis* Mart., popularly known as Juçara, is a key species in this biome, with an important ecological role. It is a highly valued food source and is used for a variety of purposes, including the production of juice, jelly, and cosmetics. Knowledge about the diversity and genetic structure are important for the management and conservation of this species and explain the diversity pattern in Atlantic Forest of Brazil. Therefore, this work intends to understand the evolutionary and historical processes that led to genetic diversity and the genetic pattern of current distribution of species in this biome. For this purpose, three types of molecular markers (SNP, Silico-DarT and SSR) were evaluated to estimate the diversity and genetic structure of *E. edulis* populations collected along the Atlantic Forest of Brazil. Making it possible to choose the marker that would best answer our questions. Then, species distribution models over the last 130,000 years were used to correlate seven biogeographical variables, related to temperature and precipitation, with genetic differentiation between populations and species distribution. Thus, it was possible to test whether there is influence of adaptive selection, geographical distance, and climatic stability on the genetic pattern of *Euterpe edulis* populations. Our results suggest that SNP and Silico-DarT markers are effective for assessing population structure, but SSR are better able to detect diversity between samples. We show that in addition to genetic drift, natural selection is also acting on the population structure of *E. edulis*. Additionally, several SNPs with selection signals were observed in genes associated with constitutive and adaptive traits. Ecological niche models show a decline in areas suitable for *E. edulis* over the last 130,000 years and that the current pattern of genetic diversity of *E. edulis* is a result of geographic distance between populations and little related to resistance isolation. In conclusion, this study is very relevant for *E. edulis* conservation programs and evolutionary studies of other species that occur in the Atlantic Forest.

Keywords: genetic diversity; juçara; molecular markers; niche models; palm; rainforest.

SUMMARY

1 INTRODUCTION	7
2 LITERATURE REVIEW	9
2.1 Genetics diversity in the Atlantic Forest	9
2.2 Genomic characterization study of the <i>Euterpe edulis</i>	10
2.3 Molecular markers	16
2.3.1 Microsatellites	17
2.3.2 Silico-DArT	18
2.3.3 SNP	20
3 JUSTIFICATIONS	21
4 AIM	22
4.1 Specific aims:	22
5 REFERENCE	23
6 CHAPTER I	30
Different molecular markers influence genetic diversity parameters of <i>Euterpe edulis</i>	30
6.1 INTRODUCTION	31
6.2 MATERIAL AND METHODS	33
6.2.1 Sampling information	33
6.2.3 DNA Extraction	34
6.2.4 SSR markers	34
6.2.5 SNP and Silico-DArT markers	34
6.2.6 Data analysis	35
6.3 RESULTS	37
6.4 DISCUSSION	46
6.5 CONCLUSION	48
6.6 REFERENCES	49
6.7 SUPPLEMENTARY DATA	55
7 CHAPTER II	59
Population genomics and phylogeography of <i>Euterpe edulis</i> : A comprehensive study of genetic diversity and adaptation	59
7.1 INTRODUCTION	60
7.2 MATERIAL AND METHODS	62

7.2.1 Population sampling	62
7.2.2 SNP genotyping	63
7.2.3 Genome scans for selection footprints	64
7.2.4 Demographic dynamics	64
7.2.5 Genetic diversity and population structure	65
7.2.6 Candidate SNPs and annotation analysis	66
7.2.8 Predictors variables	67
7.3 RESULTS	68
7.3.1 SNP genotyping	68
7.3.2 Genome scans for selection footprints	68
7.3.3 Demographic dynamics	69
7.3.4 Genetic diversity and population structure analyses	71
7.3.5 Functional annotation of SNPs with selection signal	75
7.4 DISCUSSION	78
7.5 CONCLUSION	82
7.6 REFERENCES	83
7.7 SUPPLEMENTARY DATA	86

1 INTRODUCTION

The Atlantic Forest is a tropical forest biome that extends to the east of Brazilian territory, reaching part of Argentina and Paraguay (MATA ATLÂNTICA, 2011; MYERS et al., 2000). The connection of this biome is maintained through ecological processes such as migrations and gene flow (CARVALHO; et al., 2015). The different biotic and abiotic conditions allowed the diversification of organisms and the establishment of centers of endemism (DA SILVA; CASTELETI, 2003; MYERS et al., 2000; RIBEIRO et al., 2011). Despite the fragmentation of this biome, it continues to be part of the list of five priority global hotspots for conservation (DE LIMA et al., 2020; MYERS et al., 2000). Understanding biodiversity is necessary to develop effective conservation and restoration actions for this biome.

In this sense, several studies use Atlantic Forest species to make ecological and phylogeographic inferences, which investigate lineages in time and space, and contribute to understanding the evolutionary history of populations based on historical events (CARNAVAL et al., 2014; CARVALHO; et al., 2015). The Arecaceae family, represented by palm trees, encompasses 2,400 species and about 183 genera, which are considered models for studying the evolutionary history of Neotropical regions (BAKER; COUVREUR, 2013; DRANSFIELD et al., 2014). The genus *Euterpe* has seven species, distributed from Central America to southeastern Brazil and Bolivia. (PICHARDO-MARCANO et al., 2019).

In Brazil, five species are found, one of the most frequent being *Euterpe edulis* Martius (Juçara), which is widely distributed in the Atlantic Forest and in part of the Cerrado. (LORENZI, 2010). Due to the extension of the Atlantic Forest, this species occurs in several biogeographic regions, based on the composition of climate, altitude, and biodiversity (DA SILVA; CASTELETI, 2003). *Euterpe edulis* is considered a key species in the Atlantic Forest biome as it is an important food source for many frugivorous species (SILVA; REIS, 2018; BRASIL FLORA, 2022).

Species that was once dominant in this ecosystem, is currently classified as vulnerable on the list of endangered species, in the Livro Vermelho Flora do Brasil, due to human actions and habitat fragmentation (SILVA; REIS, 2018; BRASIL FLORA, 2022). It is an allogamous, diploid species with chromosome number $2n=36$ and DNA content estimated by flow cytometry of approximately 4 gigabases (Gb) (OLIVEIRA; PADILHA, 2016).

Juçara was intensively exploited for the extraction of palm hearts for commercialization until 1998 when it became illegal (REIS et al., 2000; SCHULZ et al., 2016). Cutting the meristem of the trunk to remove the palm heart causes the death of the plant, which caused a significant. Therefore, as an alternative source of great economic and sustainable potential, due to the growth of acai consumption, the Juçara fruit started to be increasingly consumed by the market, increasing its conservation status (BARROSO; REIS, 2010; CARLA et al., 2015). Juçara fruits have a high nutritional value, and studies point to antioxidant and anti-inflammatory properties (KANG et al., 2012; LIMA et al., 2012; OLIVEIRA; BICUDO, 2014). In addition, it can still be used as a base for the manufacture of cosmetics, in the pharmaceutical industry and as a nutritious source of bioactive compounds (CARPIN et al., 2020; FRANCO et al., 2019).

The processes responsible for the origin of the diversity of *E. edulis* species remain partially unknown. In the work by Pereira et al. (2022) detected two genetic groups for *E. edulis* throughout the Atlantic Forest, as well as detected in other species, which may be related to the evolutionary history of this biome. The heterogeneity of environmental conditions influences the distribution of species and may play an important role in structuring plant diversity, as tested in the work by Carvalho et al. (2017) with microsatellites. Furthermore, natural selection may be shaping the current patterns and distribution of genetic variation associated with limited gene flow in isolated areas.

Most studies were carried out with effort in a single state and suggest that this diversity can be explained by the evolutionary history of the Atlantic Forest (CARVALHO; et al., 2015; CARVALHO et al., 2017), by biotic and abiotic factors (seed dispersers and biogeographical differences) (BRANCALION et al., 2018; CARVALHO et al., 2016; EDUARDO SÍCOLI SEOANE et al., 2005) and anthropic factors (CERQUEIRA et al., 2022; SANTOS et al., 2016; SOARES et al., 2019). Understanding how populations are organized and which processes influence the dynamics, diversity and functions of the ecosystem is a critical point for the management and conservation of this species. of the species.

In these studies, molecular markers based on DNA are used. These markers represent all molecular phenotypes obtained from an expressed gene, or from a specific segment of DNA in expressed or non-expressed regions of the genome (FERREIRA; GRATTAPAGLIA, 1998). There are several different types of molecular markers that can be used in population

genetics studies, including microsatellites (SSR), single nucleotide polymorphisms (SNPs) and markers that use methodology of genome complexity reduction by restriction enzymes, per example the Silico Diversity Arrays Technology (Silico-DArT) (ADU et al., 2021; MATTHIES et al., 2012; SANSALONI et al., 2020).

Microsatellites are short, repeated sequences of DNA that are highly variable in length and can be used to identify and track specific genes within a population. SNPs are single base pair changes in the DNA sequence that occur at a relatively high frequency in the genome and can be used to track specific genetic characteristics within a population. Silico-DArT markers are a type of molecular marker that is based on restriction fragment length and are derived from high-throughput sequencing data (TURCHETTO-ZOLET et al., 2017).

Due to different mutation rates and mechanisms, genome-wide distribution patterns, and biological functions, the use of each type of marker can lead to substantially different results when testing specific hypotheses (FISHER; FAVREAU, 1991). In view of the above, this work intends to expand the knowledge about the genetic diversity of the *E. edulis* species comparing different types of markers and using ecological niche modeling approach. In order to select the molecular marker better discriminate populations of *E. edulis* and understand the evolutionary and historical processes that led to the genetic diversity and current distribution of the species in the Atlantic Forest of Brazil. The results of this study may be used in *E. edulis* conservation programs and evolutionary studies of other species that occur in the Atlantic Forest.

2 LITERATURE REVIEW

2.1 Genetics diversity in the Atlantic Forest

The Atlantic Forest is a biodiversity hotspot in South America. This biome, originally covering over 1 million km² from northern Argentina to eastern Brazil, has suffered extensive fragmentation due to human activities such as agriculture, urbanization, and deforestation. Today, only around 12-15% of its original extent remains, with most of this forest being fragmented into isolated fragments (RIBEIRO et al., 2011).

Fragmentation of the Atlantic Forest has serious implications for its biota, as it increases the risk of species extinction, alters ecosystem processes, and decreases the resistance of the forest to climate change (COSTA; LEITE, 2013). To mitigate these impacts and conserve the Atlantic Forest, it is crucial to understand the patterns of species and lineage diversity in this biome, as well as the processes driving these patterns. This information is essential for the development of effective conservation strategies and for ensuring the long-term persistence of this important biome and its unique biota (RULL; CARNAVAL, 2020).

The Atlantic Rainforest of Brazil is a biodiverse region that harbors a high level of species richness and endemism (MYERS et al., 2000). Higher species richness is often found in the more topographically complex coastal regions of Rio de Janeiro and São Paulo (MERCIER et al., 2023; RULL; CARNAVAL, 2020). Many of studies show the biological communities in the southern portion of the Atlantic Forest are distinct from those in the north (PEREIRA et al., 2022). Areas of endemism have been recognized within the forest, with a nested pattern of low-dispersal organisms having contiguous areas of endemism that fit within more inclusive endemism centers of high-dispersal groups (NOVELLO et al., 2018; VALE et al., 2018). The fauna and flora of the forest have been compartmentalized into five main areas of endemism: Pernambuco, Coastal Bahia, Central Bahia, Serra do Mar, and Paraná/Araucaria (CARNAVAL et al., 2009; DASILVA; PINTO-DA-ROCHA; DESOUZA, 2015).

While there are similar geographic patterns of species and genetic diversity across Atlantic Forest organisms, the diversification processes are not the same. Climate change, historical connections between the Atlantic Forest and other neotropical biomes, Pleistocene refugia and geographic barriers have been identified as key drivers of diversification in the Atlantic Forest biota (BACCI et al., 2021; CARDOSO et al., 2000; MERCIER et al., 2023). Population genomics and phylogeographic studies have provided insight into the main patterns of lineage divergence and population structure in the Atlantic Forest. These studies have revealed that both climate change and geographic barriers have played a role in shaping the diversification of the biota in this region (AVISE, 2013).

2.2 Genomic characterization study of the *Euterpe edulis*

Some of the specific areas that genomic characterization studies may focus on include population genomics and phylogeography. These studies aim to understand the genetic

variation within and among populations of organisms, with the goal of understanding the mechanisms that drive population divergence and speciation, as well as the factors that shape the distribution and diversity of species (AVISE, 2013; ZOLET; et al., 2013).

Population genomics is an interdisciplinary field that combines population genetics, evolutionary biology, and genomics to study the genetic variation within and among populations of organisms (OTTENBURGHS et al., 2019). It uses high-throughput sequencing technologies to generate large amounts of genetic data, which can be used to infer population structure, demographic history, and selective pressures on genomes (AGUIAR-MELO et al., 2019).

Phylogeography, on the other hand, is a subfield of population genomics that combines phylogenetics and biogeography to infer the historical processes that have shaped the genetic variation and distribution of populations (AVISE, 2013). It uses genetic data along with information about geographical and ecological factors to infer past movements and interactions of populations, as well as the effects of historical events such as glaciations, range expansions, and founder effects (OTTENBURGHS et al., 2019). Together, population genomics and phylogeography provide a tool for understanding the evolutionary dynamics of populations, including how populations and species have changed over time and how they are related to each other (GOODWIN; MCPHERSON; MCCOMBIE, 2016; MCCORMACK et al., 2013).

There have been several studies conducted on the genomic characterization of *Euterpe edulis* (Fig. 1), some of which are in Table 1. They have used molecular markers, such as amplified fragment length polymorphism (AFLP), microsatellites and single nucleotide polymorphisms (SNPs), to evaluate the levels of genetic diversity, differentiation, and structure among populations.



Figure 1: Image of adult individual of *Euterpe edulis* Mart. (Arecaceae).

Table 1: Population genomics studies of *Euterpe edulis*.

Article Reference	Type of Marker Used	Purpose of the Study	Number of Samples	Number of sites	Region
Cardoso et al. (2000)	AFLP	Genetic diversity	150	11	North, Southeast, and South
Conte et al. (2003)	AFLP	Genetic diversity	~425	1 (permanent plots, in a natural population)	Southern
Gaiotto et al. (2003)	Microsatellites	Genetic diversity	583	2	Midwest
Seoane et al. (2005)	Microsatellites	Population structure	34 isolated population and 43 continuous populations	2	Southeast
Conte et al. (2008)	AFLP and Microsatellites	Genetic structure	50	4	Southern
Article Reference	Type of Marker Used	Purpose of the Study	Number of Samples	Number of sites	Region

Carvalho et al. (2015)	Microsatellites	Population genetics and historical demographic events	463	16	Southeast and Midwest
Carvalho et al. (2016)	Microsatellites	Microevolution	545	19	Southeast
Konzen and Martins (2017)	Isozyme	Genetic diversity	160 each site	5	Southeast and South
Novello et al., (2018)	Microsatellites	Population genetic	189 juvenile palms and 172 adults	10	
Carvalho et al. (2017)	Microsatellites	Population genetic and ecological niche modeling	229	57	North, Southeast and South
Santos et al. (2016)	Microsatellites	Gene flow and pollen dispersal	268 reproductive, 148 juvenile plants, and the embryos of 149 fruits	1	South
Brancalion et al. (2018)	SNP	Phenotypic plasticity and local adaptation	55	3	Southeastern
Soares et al. (2019)	Microsatellites	Genetic diversity	544	17	North
Moraes et al. (2020)	Microsatellites	Population genetic	208 progenies	21 matrices	Southeast
Carvalho et al. (2021)	Microsatellites	Genetic diversity and microevolution	1,330 endocarps	15	Southeast
Carvalho et al. (2020)	Microsatellites	Genetic diversity and population structure	160	4	Southeast
Coelho et al. (2020)	Microsatellites	Genetic Diversity	250	8	North and Midwest
Cerqueira et al. (2022)	Microsatellites	Genetic diversity and physiological traits	462	2	North
Pereira et al. (2022)	Microsatellites	Genetic diversity	527	26	North, Southeast, Midwest and South

These studies have shown that *E. edulis* has high levels of genetic diversity within populations and moderate levels of differentiation among populations (CARDOSO et al., 2000; PEREIRA et al., 2022). In addition, they estimated the diversification of *E. edulis* was

around mid-Miocene (PICHARDO-MARCANO et al., 2019). The results suggest that *E. edulis* has a strong capacity to adapt to different environmental conditions, which is an important factor for the conservation of the species (BRANCOLIN et al., 2018). Additionally, studies on the population genetics of *E. edulis* have investigated the effects of anthropic activities, such as deforestation, defaunation and palm heart harvesting, on the genetic diversity of populations (CARVALHO et al 2016; CERQUEIRA et al., 2022). The results of these studies showed that habitat fragmentation for this species has a significant impact on the genetic diversity and structure of *E. edulis* populations (SEOANE et al., 2005; CARVALHO et al., 2016; CERQUEIRA et al., 2022). This occurs because it increases the isolation between populations by decreasing gene flow.

One study discusses the impact of animal defaunation, or the loss of animal species, on the microevolution of *Euterpe edulis*. The study found that defaunation led to changes in the plant's reproductive behavior, such as increased selfing, and reduced the amount of genetic diversity in the population. These changes have implications for the survival and resilience of palm species in the face of environmental stressors, such as climate change (CARVALHO et al., 2021, 2016; GALETTI et al., 2013b). In addition, a difference in the size of the *E. edulis* seed was noticed due to the size of the beak of the dispersers. Which leads us to wonder if there is influence of adaptive selection on the genetic diversity of the species.

Another study explores the effects of climate stability and human impacts on the genetic diversity and conservation status of the palm species *Euterpe edulis* in the Atlantic Forest of Brazil (CARVALHO et al., 2017). The authors found that populations in areas with more stable historical climates had greater genetic diversity, while those in areas with more recent human disturbances had lower diversity. These findings suggest that climate stability and anthropogenic impacts are important factors to consider when developing conservation strategies for the species. (CARVALHO et al., 2017).

On this issue of climate stability, there are many other works that report the influence of biogeographical factors on the genetic diversity of species in the Atlantic Forest (BACCI et al., 2021; CARNAVAL et al., 2014). Highlighting that the species have different genetic groupings geographically dividing this biome into north and south. A population structure study of *E. edulis* shows exactly this break (Fig.2), but no hypothesis was tested (PEREIRA, et al. 2022).

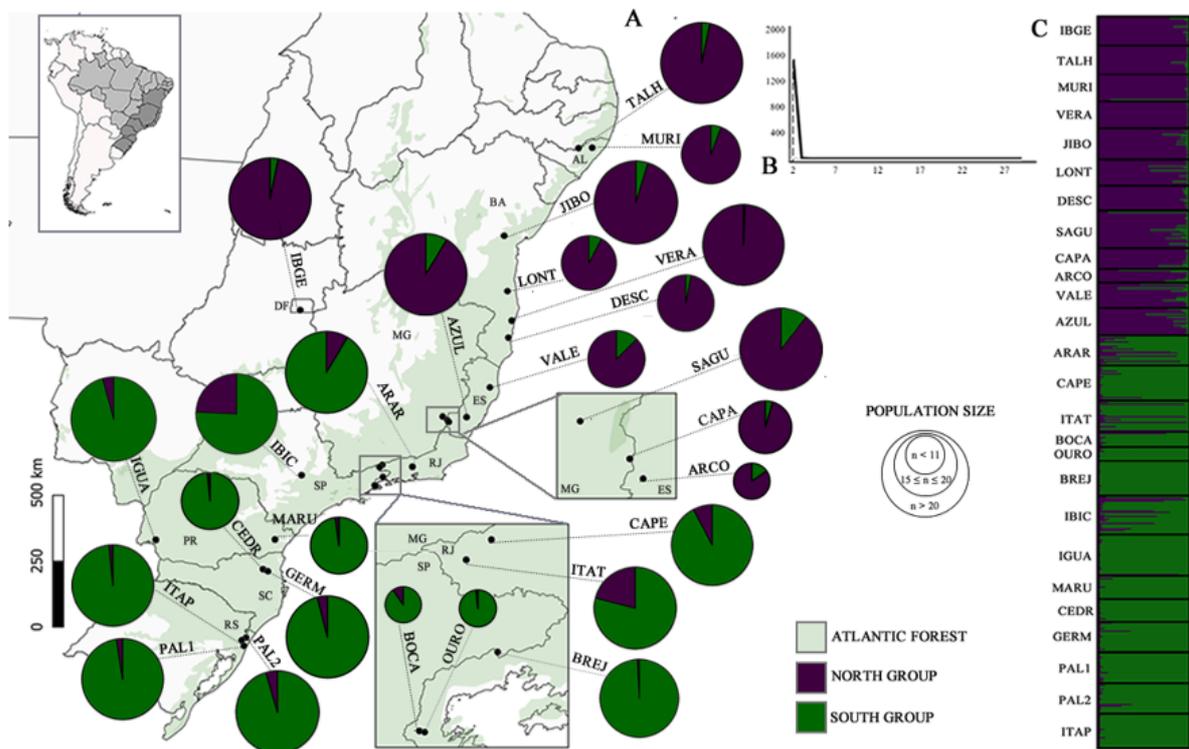


Figure 2: Clustering analysis and geographical distribution of two geographic clusters in *Euterpe edulis*. (A) Geographical distribution in the Atlantic Forest of Brazil (light greenish-gray) and lineage assignments of populations to North group (dark purple) and South group (dark green). (PEREIRA et al., 2022)

In conclusion, studies on the population genetics of *E. edulis* have provided valuable insights into the genetic diversity and structure of populations of this important species. These studies have highlighted the importance of conservation efforts to protect the genetic diversity of *E. edulis*, particularly in the face of anthropogenic activities such as deforestation and harvest. Overall, these studies suggest that *Euterpe edulis* has a high degree of genetic diversity and may have undergone changes in population size and structure in the Atlantic Forest over time. However, works that contemplate a larger sampling along the Atlantic Forest and using different types of markers are necessary to fully understand the population genetics and the evolution process of this species in the region.

2.3 Molecular markers

Genomic characterization studies involve the use of molecular markers and high-throughput sequencing technologies to generate large amounts of genomic data for a wide range of organisms (OTTENBURGHS et al., 2019). Molecular markers allow researchers to infer population structure, gene flow patterns, and evolutionary relationships among populations (FREELAND; KIRK; PETERSEN, 2011). This information is critical for understanding the mechanisms that drive population divergence, speciation, adaptation, and the factors that shape the distribution and diversity of species (FREELAND; KIRK; PETERSEN, 2011).

Some of the specific ways in which molecular markers are used in population genomics and phylogeography studies include:

- Inferring population structure: Molecular markers can be used to infer the genetic relationships among populations, which can provide insight into the history of population divergence and gene flow (GAN et al., 2021).
- Identifying gene flow patterns: Molecular markers can be used to infer the direction and magnitude of gene flow among populations, which can provide insight into the factors that shape population connectivity and the spread of genetic variation (HUDSON; SLATKIN; MADDISON, 1992; PAPA; GEPTS, 2003).
- Inferring evolutionary relationships: Molecular markers can be used to infer the evolutionary relationships among populations, which can provide insight into the history of speciation and the factors that shape the diversity of species (COUVREUR; FOREST; BAKER, 2011).
- Identifying the genetic basis of adaptation: Molecular markers can be used to identify the genomic regions associated with traits and to infer the genetic basis of adaptation to different environments (ZIMMERMAN; ALDRIDGE; OYLER-MCCANCE, 2020).
- Inferring demographic history: Molecular markers can be used to infer the demographic history of populations, such as population size changes and migration events, which can provide insight into the factors that shape population dynamics and the distribution of genetic variation (VIEIRA et al., 2022).

Each genetic marker has a unique evolutionary rate, that is, the speed with which changes occur in its DNA (MORRIS; SHAW, 2018). Some markers evolve very slowly, while others

change more quickly. This means that different markers can reveal different evolutionary moments in the history of a population or species (ROWE G., SWEET, M., BEEBEE, 2017). The choice of the appropriate genetic marker for a given analysis is essential to ensure the accuracy and reliability of the results obtained. In this study we used three types of molecular markers:

2.3.1 Microsatellites

Microsatellite markers are codominant and are based on PCR amplification of specific regions of the genome using a locus-specific pair of primers. It is generally abundant and polymorphic in unexpressed genomic regions and, consequently, are considered selectively neutral (FREELAND; KIRK; PETERSEN, 2011). Neutral microsatellites evolve rapidly without vital consequences for organisms and have a high mutation rate which favors recent gene flow and divergence studies (GRÜNWALD et al., 2017).

The mutation rate of microsatellites in palm trees has been estimated in a few studies, but it can vary depending on the species and the microsatellite loci used (BHARGAVA; FUENTES, 2010). According to one study the mutation rate for microsatellites was estimated to be around 3.8×10^{-3} . The mutation rate of microsatellites in palm trees is generally high, and it can vary depending on the species. Different mutational models have been proposed to describe the evolution of microsatellites markers (VALDES; SLATKIN; FREIMERT, 1993):

1. The stepwise mutation model (SMM): This model assumes that mutations occur one repeat unit at a time, and that the probability of a repeat unit being added or removed depends on the number of repeat units already present.
2. The multiple-step mutation model (MSM): This model is an extension of the SMM that allows for more than one repeat unit to be added or removed at a time, and the probability of this is dependent on the number of repeat units already present.
3. The infinite allele model (IAM): This model assumes that each microsatellite locus has a finite number of alleles and that new alleles can arise by mutation.
4. The two-phase model (TPM): This model assumes that microsatellites are inherited from a common ancestor, and that a small proportion of new alleles arise de novo.
5. The coalescent model: This model is used to infer the demographic history of a population or a set of populations. They use the gene genealogies of a sample of DNA

sequences to infer the demographic parameters, such as population size, migration, and population splitting.

The most used models for microsatellite markers are the stepwise mutation model (SMM) and the multiple-step mutation model (MSM) because they can provide an estimation of the mutation rate and effective population size and are also simple to implement. But in some cases, more complex models can be used to better understand the population dynamics and evolutionary history of the population (SAINUDIIN et al., 2004).

2.3.2 Silico-DArT

Genotyping by NGS (Next Generation Sequencing), using methods to reduce the complexity of the genome, allows the identification of polymorphisms, the selection of regions of interest in the genome and it is widely used in population studies (BAIRD et al., 2008; CATCHEN et al., 2013; XU et al., 2014). Different NGS genotyping methods have been developed and applied to studies of species with little or no previously available genomic information (BRÅTE et al., 2019; FOYEZ et al., 2019; NARUM et al., 2013; XIAO et al., 2017).

Diversity Array Technology Sequencing (DArTseq) (KILIAN et al., 2012; PETROLI et al., 2019) a method that combines the DArT methodology of complexity reduction using restriction enzymes with the Illumina technology, based on the analysis of short, randomly selected DNA sequences that are representative of the genome which correspond predominantly to active genes (Fig. 3) (FOYEZ et al., 2019; SANSALONI et al., 2020). This technique allows genotyping Silico-DArT markers and SNPs (Single Nucleotide Polymorphism).

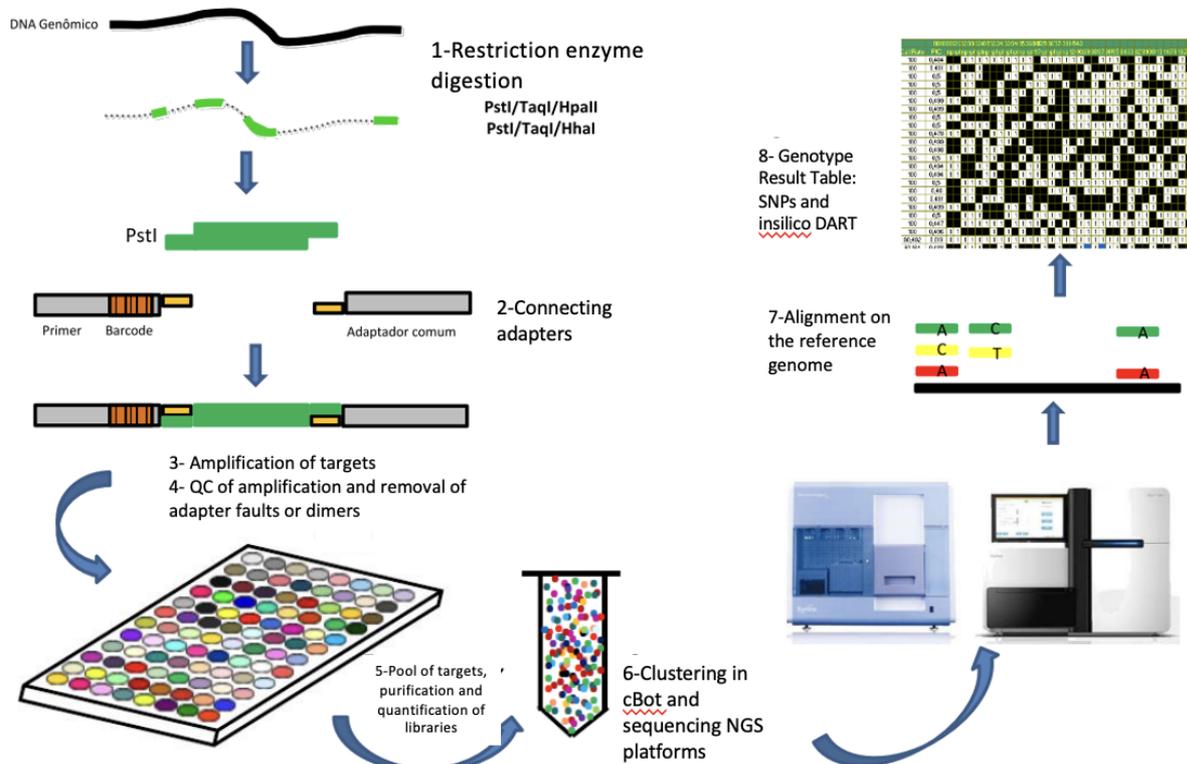


Figure 3: In this image, it is possible to observe steps of the DArTseq methodology. The reduction of genomic complexity and adapter binding as a first measure, followed by target amplification, sample clustering or pooling, cluster amplification, fragment sequencing and finally data analysis. (Sansaloni et al., 2012)

Silico-DArT are dominant markers are based on the presence or absence of restriction fragments in a specific region that are conserved in the genomes of different species, therefore the mutation rate is not based on nucleotide changes as it is in SNP markers (PETROLI; KILIAN, 2019; SANSALONI et al., 2011). The mutation rate of Silico-DArT markers is not well-characterized, as it can vary depending on the specific DNA sequences being analyzed and the species being studied.

In general, the mutation rate of DNA sequences is typically on the order of 10^{-9} to 10^{-10} per base pair per generation. However, the mutation rate of specific DNA sequences can vary depending on a few factors, including the length of the sequence, the specific base pairs involved, and the genetic context in which the sequence is found. Additionally, the mutation rate of Silico-DArT markers may be influenced by the specific techniques and protocols used to generate and analyze the markers. Therefore, it is difficult to provide a general estimate of

the mutation rate of Silico-DArT markers without more specific information about the specific DNA sequences and species being studied.

It's important to note that Silico-DArT markers are a recent development in the field of molecular markers and therefore there is less information available about the most appropriate mutational models to use with them. However, as they combine features of both microsatellites and SNPs, it is likely that models used for these types of markers will also be relevant for Silico-DArT markers (PETROLI; KILIAN, 2019; SANSALONI; GRATTAPAGLIA, 2012). Additionally, due to their high resolution, they can be used to detect fine-scale population structure and infer demographic history of the populations. Also, the ability to identify the genomic regions associated with important agronomic traits can be of great importance for breeding program.

2.3.3 SNP

The methodology DarTseq tracks thousands of SNP markers associated with the restriction site throughout the genome, typically generating 69bp sequences containing two alleles for a given *locus* (CRUZ; KILIAN; DIERIG, 2013). SNPs are a single point mutation of DNA in which one nucleotide at a particular *locus* is replaced by another. Due to their codominance and bi-allelic nature, traditional population genetic statistics can be easily applied to them, but a larger number of sufficiently polymorphic *loci* may be needed to achieve the same power as multi-allelic microsatellites *loci* (GUICHOUX et al., 2011).

The mutation rate of SNPs is typically much lower than the mutation rate of other types of DNA sequences, such as microsatellites. The mutation rate of SNPs is typically on the order of 10^{-8} to 10^{-9} per base pair per generation (ROWE G., SWEET, M., BEEBEE, 2017). The low mutation rate of SNPs makes them useful as markers for studying population genetics, as they can provide a high level of genetic stability over time. However, the low mutation rate also means that SNPs may be less sensitive than other types of molecular markers for detecting subtle changes in the genetic makeup of a population (FREELAND; KIRK; PETERSEN, 2011). Additionally, the low mutation rate of SNPs means that it can take a longer time for new SNPs to arise, which can limit the ability of SNPs to track rapid changes in the genetic makeup of a population.

The proposed mutational models to describe the evolution of SNP markers are (KIMURA; OHTA, 1978):

1. The Infinite Site Model (ISM): This model assumes that each SNP site is a separate locus, and that each allele at a site can mutate to any other allele with a constant rate.
2. The Infinite Allele Model (IAM): This model assumes that each SNP site has a finite number of alleles and that new alleles can arise by mutation.
3. The Stepwise Mutation Model (SMM): This model assumes that mutations occur one base at a time, and that the probability of a particular base mutating to another base depends on the base that it is currently in.
4. The Two-Phase Model (TPM): This model assumes that most SNPs are inherited from a common ancestor, and that a small proportion of SNPs arise de novo.
5. The General Time-Reversible Model (GTR): This model assumes that mutations occur at a constant rate, and that the probability of a particular base mutating to another base depends on the base that it is currently in.

In practice, most of the studies use a combination of these models, or even more complex models, to estimate the demographic history of populations and the parameters of the evolutionary process. The choice of the model depends on the aim of the study and the information that the marker can provide.

3 JUSTIFICATIONS

The studies to be carried out in this work are interdisciplinary with applied and technological knowledge around environment and genetic improvement and will provide relevant knowledge for conservation, management, and socio-economic-environmental development. The *E. edulis* species is considered a key species in this biome, as it has an important ecological role, contributing to the survival of countless species that feed on its fruits. The need for studies with *E. edulis* the fact that it appears as vulnerable to extinction in the Livro Vermelho Flora do Brasil due to the devastation and fragmentation of the Atlantic Forest and its immense exploitation due to the high desire for the palm heart it produces.

Given the situation of reduced vegetation cover in the Atlantic Forest, the understanding of local biodiversity is necessary to develop effective conservation and restoration actions in this biome. The refinement of the knowledge of the evolutionary history and of the phylogeographic factors that determine the diversity of the species given that previous approaches - which demonstrate that the natural populations of *E. edulis* are found in four main phylogeographic groups, with genetic diversity, little gene flow and inbreeding indicate the need for regional conservation strategies for the species natural populations. In addition, given the current climate change scenarios, actions to restore the connectivity of populations through the mobility of pollinators and seed dispersers may be relevant for the maintenance of the species.

Faced with the need to expand the knowledge about genomic resources for the species *Euterpe edulis*, performing a transcriptome of this species will allow exploring information related to the structure and organization of the genes of this species. Knowing the transcriptome of a species allows performing structural and functional analysis through the transcriptome, providing progress in studies of genetic improvement both for the species and for nearby species. It is possible to obtain markers, such as single polymorph nucleotides (SNPs), which can be used for the development of primers for studies of genes of agronomic, phylogenetic and phylogeographic interest. The power of analysis of SNP markers, extensive genotyping, bioinformatics analysis to choose potentially functional *loci* for analysis, may reveal new knowledge about the species and the biome. Thus, in conjunction with other data, it can effectively guide restoration, conservation, and enhancement strategies for the species.

4 AIM

To understand evolutionary and historical processes that led to the genetic diversity and current distribution of the *Euterpe edulis* species in Atlantic Forest of Brazil.

4.1 Specific aims:

- Evaluate the ability of three marker types to estimate genetic diversity and population structure of *Euterpe edulis*.
- To verify the role of natural selection in the pattern of population structuring of *E. edulis*.
- Search for the presence of loci under selection, using F_{ST} distribution and presence of loci correlated with environmental variables.

- Identify which biogeographic variables may be influencing the distribution of populations in Atlantic Forest Brazil.
- Analyze whether populations separated by regions of low stability present high population structure and isolation by resistance.
- Analyze whether populations in contiguous regions of high climate stability will show low population structure and genetic isolation signal by distance.

5 REFERENCE

ADU, B. G. et al. High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz). **PLoS ONE**, v. 16, n. 7 July, p. 1–19, 2021.

AGUIAR-MELO, C. et al. Ecological niche modeling and a lack of phylogeographic structure in *Vriesea incurvata* suggest historically stable areas in the southern Atlantic Forest. **American Journal of Botany**, v. 106, n. 7, p. 971–983, 2019.

AVISE, J. C. Phylogeography. In: **Brenner's Encyclopedia of Genetics: Second Edition**. [s.l: s.n.].

BACCI, L. F. et al. Biogeographic breaks in the Atlantic Forest: evidence for Oligocene/Miocene diversification in *Bertolonia* (Melastomataceae). **Botanical Journal of the Linnean Society**, p. 1–16, 2021.

BAIRD, N. A. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PLoS ONE**, v. 3, n. 10, p. 1–7, 2008.

BARROSO, R. M.; REIS, A. Etnoecologia e etnobotânica da palmeira juçara (*Euterpe edulis* Martius) em comunidades quilombolas do Vale do Ribeira, São Paulo. **Act Bot Bras**, v. 24, n. 2, p. 518–528, 2010.

BHARGAVA, A.; FUENTES, F. F. **Mutational dynamics of microsatellites** **Molecular Biotechnology**, mar. 2010.

BRANCALION, P. H. S. et al. Phenotypic plasticity and local adaptation favor range expansion of a Neotropical palm. **Ecology and Evolution**, v. 8, n. 15, p. 7462–7475, 3 ago. 2018.

BRÂTE, J. et al. Draft genome assembly and transcriptome sequencing of the golden algae *Hydrurus foetidus* (chrysophyceae). **F1000Research**, v. 8, p. 1–12, 2019.

CARDOSO, S. R. S. et al. Genetic differentiation of *Euterpe edulis* Mart. populations

estimated by AFLP analysis. **Molecular Ecology**, v. 9, n. 11, p. 1753–1760, 2000.

CARLA, A. et al. Agroecology and Sustainable Food Systems Market for Amazonian Açai (*Euterpe oleraceae*) Stimulates Pulp Production from Atlantic Forest Juçara Berries (*Euterpe edulis*) Market for Amazonian Açai (*Euterpe oleraceae*) Stimulates Pulp Production from Atla. **Agroecology and Sustainable Food Systems**, v. 39, n. 7, p. 762–781, 2015.

CARNAVAL, A. C. et al. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. **Science**, v. 323, n. 5915, p. 785–789, 2009.

CARNAVAL, A. C. et al. Prediction of phylogeographic endemism in an environmentally complex biome. **Proceedings of the Royal Society B: Biological Sciences**, v. 281, n. 1792, 2014.

CARPIN, D. et al. Valorization of *Euterpe edulis* Mart . agroindustrial residues (pomace and seeds) as sources of unconventional starch and bioactive compounds. v. 85, 2020.

CARVALHO, C. DA S. et al. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. **Heredity**, v. 115, n. 3, p. 216–224, set. 2015.

CARVALHO, C. DA 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.

CARVALHO, C. DA S. et al. Extant fruit-eating birds promote genetically diverse seed rain, but disperse to fewer sites in defaunated tropical forests. **Journal of Ecology**, v. 109, n. 2, p. 1055–1067, 1 fev. 2021.

CARVALHO, C. S. et al. Defaunation leads to microevolutionary changes in a tropical palm. **Scientific Reports**, v. 6, p. 1–9, 2016.

CATCHEN, J. et al. Oregon threespine stickleback determined using RAD-seq. **Molecular Ecology**, v. 22, n. 11, p. 2864–2883, 2013.

CERQUEIRA, A. F. et al. Landscape conservation and maternal environment affect genetic diversity and the physiological responses of *Euterpe edulis* (Arecaceae) progenies to light availability. **Environmental and Experimental Botany**, v. 194, 1 fev. 2022.

COELHO, G. M. et al. Genetic structure among morphotypes of the endangered Brazilian palm *Euterpe edulis* Mart (Arecaceae). **Ecology and Evolution**, v. 10, n. 12, p. 6039–6048, 2020.

CONTE, R. et al. Genetic diversity and recruitment of the tropical palm, *Euterpe edulis* Mart., in a natural population from the Brazilian Atlantic Forest. **Heredity**, v. 91, n. 4, p. 401–406, 2003.

- CONTE, R. et al. Genetic structure and mating system of *Euterpe edulis* Mart. populations: A comparative analysis using microsatellite and allozyme markers. **Journal of Heredity**, v. 99, n. 5, p. 476–482, 2008.
- COSTA, L. P.; LEITE, Y. L. R. Historical Fragmentation Shaping Vertebrate Diversification in the Atlantic Forest Biodiversity Hotspot. In: **Bones, Clones, and Biomes**. [s.l.] University of Chicago Press, 2013. p. 283–306.
- COUVREUR, T. L. P.; FOREST, F.; BAKER, W. J. Origin and global diversification patterns of tropical rain forests: Inferences from a complete genus-level phylogeny of palms. **BMC Biology**, v. 9, 2011.
- DA SILVA CARVALHO, C. et al. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. **Heredity**, v. 115, n. 3, p. 216–224, 2015.
- DA SILVA, J. M. C.; CASTELETI, C. H. M. In **The Atlantic Forest of South America: Biodiversity Status, Threats, and Outlook**, 2003.
- DASILVA, M. B.; PINTO-DA-ROCHA, R.; DESOUZA, A. M. A protocol for the delimitation of areas of endemism and the historical regionalization of the Brazilian Atlantic Rain Forest using harvestmen distribution data. **Cladistics**, v. 31, n. 6, p. 692–705, 1 dez. 2015.
- DE LIMA, R. A. F. et al. The erosion of biodiversity and biomass in the Atlantic Forest biodiversity hotspot. **Nature Communications**, v. 11, n. 1, p. 1–16, 2020.
- DE MORAES, M. C. et al. Genetic diversity in matrices and progenies of *euterpe edulis* mart. In managed area and in natural populations by microsatellites markers. **Ciencia Florestal**, v. 30, n. 2, p. 583–594, 2020.
- EDUARDO SÍCOLI SEOANE, C. et al. **EFEITOS DA FRAGMENTAÇÃO FLORESTAL SOBRE A IMIGRAÇÃO DE SEMENTES E A ESTRUTURA GENÉTICA TEMPORAL DE POPULAÇÕES DE *Euterpe edulis* Mart. 1.** [s.l: s.n.].
- FERREIRA, M. E. .; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3. ed. Brasília: CENARGEM, 1998.
- FOYEZ, S. et al. Application of DArT seq derived SNP tags for comparative genome analysis in fishes ; An alternative pipeline using sequence data from a non-traditional model species , *Macquaria ambigua*. **PLoS ONE**, v. 14, n. 12, p. 1–13, 2019.
- FRANC, M. et al. Spray-dried microcapsules of anthocyanin-rich extracts from *Euterpe edulis* M . as an alternative for maintaining color and bioactive compounds in dairy beverages. v. 56, n. September, p. 4147–4157, 2019.

- FREELAND, J.; KIRK, H.; PETERSEN, S. **Molecular ecology**. [s.l.] Wiley-Blackwell, 2011.
- GAIOTTO, F. A.; GRATTAPAGLIA, D.; VENCOVSKY, R. Genetic Structure, Mating System, and Long-Distance Gene Flow in Heart of Palm (*Euterpe edulis* Mart.). **Journal of Heredity**, v. 94, n. 5, p. 399–406, 2003.
- GALETTI, M. et al. Functional extinction of birds drives rapid evolutionary changes in seed size. **Science**, v. 340, n. 6136, p. 1086–1090, 31 maio 2013.
- GAN, S. T. et al. Assessment of genetic diversity and population structure of oil palm (*Elaeis guineensis* Jacq.) field genebank: A step towards molecular-assisted germplasm conservation. **PLoS ONE**, v. 16, n. 7 July 2021, 1 jul. 2021.
- GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age : ten years of next- generation sequencing technologies. **Nature Publishing Group**, v. 17, n. 6, p. 333–351, 2016.
- HUDSON, R. R.; SLATKIN, M.; MADDISON, W. P. Estimation of levels of gene flow from DNA sequence data. **Genetics**, v. 132, n. 2, p. 583–589, 1992.
- KANG, J. et al. Bioactivities of açai (*Euterpe precatoria* Mart .) fruit pulp , superior antioxidant and anti-inflammatory properties to *Euterpe oleracea* Mart . q. **Food Chemistry**, v. 133, n. 3, p. 671–677, 2012.
- KIMURA, M.; OHTA, T. Stepwise mutation model and distribution of allelic frequencies in a finite population. **Proceedings of the National Academy of Sciences of the United States of America**, v. 75, n. 6, p. 2868–2872, 1978.
- KONZEN, E. R.; MARTINS, M. P. CONTRASTING LEVELS OF GENETIC DIVERSITY AMONG POPULATIONS OF THE ENDANGERED TROPICAL PALM *EUTERPE EDULIS MARTIUS*. **CERNE**, v. 23, n. 1, p. 31–42, mar. 2017.
- LIMA, C. P. . et al. Conteúdo polifenólico e atividade antioxidante dos frutos da palmeira Juçara (*Euterpe edulis* Martius). **Revista Brasileira de Plantas Mediciniais**, v. 14, n. 2, p. 321–326, 2012.
- MARTINELLI, G.; MORAES, M. A. **Livro Vermelho Flora do Brasil**. 1. ed. Rio de Janeiro: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, 2013.
- MATA ATLÂNTICA. **Novos dados do Atlas da Mata Atlântica**.
- MATTHIES, I. E. et al. Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. **Molecular Breeding**, v. 30, n. 2, p. 951–966, 2012.
- MCCORMACK, J. E. et al. Applications of next-generation sequencing to phylogeography

and phylogenetics. **Molecular Phylogenetics and Evolution**, v. 66, n. 2, p. 526–538, 2013.

MERCIER, K. P. et al. Linking environmental stability with genetic diversity and population structure in two Atlantic Forest palm trees. **Journal of Biogeography**, v. 50, n. 1, p. 197–208, 1 jan. 2023.

MORRIS, A. B.; SHAW, J. Markers in time and space : A review of the last decade of plant phylogeographic approaches. n. December 2017, 2018.

MYERS, N. et al. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853–858, fev. 2000.

NARUM, S. R. et al. Genotyping-by-sequencing in ecological and conservation genomics. **Molecular Ecology**, v. 22, n. 11, p. 2841–2847, 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, 1 jan. 2018.

OLIVEIRA, L. C.; PADILHA, S. Cytogenetics Karyotype and genome size in *Euterpe Mart.* (Arecaceae) species. **Comparative Cytogenetics**, v. 10, n. 1, p. 17–25, 2016.

OLIVEIRA, M.; BICUDO, P. Anthocyanins, Phenolic Acids and Antioxidant Properties of Juçara Fruits (*Euterpe edulis M.*) Along the On-tree Ripening Process. p. 142–147, 2014.

OTTENBURGH, J. et al. Population Genomics and Phylogeography. In: **Avian Genomics in Ecology and Evolution**. Cham: Springer International Publishing, 2019. p. 237–265.

PAPA, R.; GEPTS, P. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris L.*) from Mesoamerica. **Theoretical and Applied Genetics**, v. 106, n. 2, p. 239–250, 2003.

PEREIRA, A. G. et al. Patterns of genetic diversity and structure of a threatened palm species (*Euterpe edulis Arecaceae*) from the Brazilian Atlantic Forest. **Heredity**, 2022.

PETROLI, C.; KILIAN, A. Introducción al método de genotipificación DArTseq y los datos obtenidos. **CIMMYT Research Data & Software Repository Network**, V1, 2019.

PICHARDO-MARCANO, F. J. et al. Phylogeny, historical biogeography and diversification rates in an economically important group of Neotropical palms: Tribe Euterpeae. **Molecular Phylogenetics and Evolution**, v. 133, p. 67–81, 1 abr. 2019.

REIS, M. S. DOS et al. Management and Conservation of Natural Populations in Atlantic Rain Forest : The Case Study Management and Conservation of Natural Populations in Atlantic Rain Forest : The Case Study of Palm Heart (*Euterpe edulis Martius*) 1. **Biotropica**, v. 32, n. 4, p. 894–902, 2000.

RIBEIRO, M. C. et al. The Brazilian Atlantic Forest: A Shrinking Biodiversity Hotspot. In:

Biodiversity Hotspots. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 405–434.

ROWE G., SWEET, M., BEEBEE, T. **An Introduction to Molecular Ecology.** 3^a ed. [s.l: s.n.].

RULL, V.; CARNAVAL, A. C. **Neotropical Diversification: Patterns and Processes.** Cham: Springer International Publishing, 2020.

SAINUDIIN, R. et al. Microsatellite mutation models: Insights from a comparison of humans and chimpanzees. **Genetics**, v. 168, n. 1, p. 383–395, 2004.

SANSALONI, C. et al. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. **BMC Proceedings**, v. 5, n. S7, p. P54, 2011.

SANSALONI, C. et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. **Nature Communications**, v. 11, n. 1, p. 4572, 11 dez. 2020.

SANSALONI, C. P.; GRATTAPAGLIA, O. D. **Desenvolvimento e aplicações de DArT (Diversity Arrays Technology) e genotipagem por sequenciamento (Genotyping-by-Sequencing) para análise genética em Eucalyptus.** [s.l.] Universidade de Brasília, 2012.

SANTOS, A. S. et al. Landscape-scale deforestation decreases gene flow distance of a keystone tropical palm, *Euterpe edulis* Mart (Arecaceae). **Ecology and Evolution**, v. 6, n. 18, p. 6586–6598, 1 set. 2016.

SCHULZ, M. et al. Juçara fruit (*Euterpe edulis* Mart.): Sustainable exploitation of a source of bioactive compounds. **FRIN**, v. 89, p. 14–26, 2016.

SILVA, J. Z. DA; REIS, M. S. DOS. Reproductive phenology and production of fruits in *Euterpe edulis* (Martius). **Cienc Florest**, v. 28, n. 1, p. 295–309, 2018.

SOARES, L. A. S. S. et al. Anthropogenic Disturbances Eroding the Genetic Diversity of a Threatened Palm Tree: A Multiscale Approach. **Frontiers in Genetics**, v. 10, 7 nov. 2019.

TURCHETTO-ZOLET, A. C. et al. **Marcadores Moleculares na Era G enômica: Metodologias e Aplicações.** Ribeirão Preto: Sociedade Brasileira de Gnética, 2017.

VALDES, A. M.; SLATKIN, M.; FREIMERT, N. B. **Allele Frequencies at Microsatellite Loci: The Stepwise Mutation Model Revisited** **Genetics**. [s.l: s.n.].

VALE, M. M. et al. Endemic birds of the Atlantic Forest: traits, conservation status, and patterns of biodiversity. **Journal of Field Ornithology**, v. 89, n. 3, p. 193–206, 1 set. 2018.

VIEIRA, L. D. et al. Comparative population genomics in *Tabebuia* alliance shows evidence of adaptation in Neotropical tree species. **Heredity**, v. 128, n. 3, p. 141–153, 2022.

XIAO, Y. et al. The genome draft of coconut (*Cocos nucifera*) St ephanie. n. September, p.

1–11, 2017.

XU, P. et al. Population genomic analyses from low-coverage RAD-Seq data: A case study on the non-model cucurbit bottle gourd. **Plant Journal**, v. 77, n. 3, p. 430–442, 2014.

ZIMMERMAN, S. J.; ALDRIDGE, C. L.; OYLER-MCCANCE, S. J. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. **BMC Genomics**, v. 21, n. 1, 1 jun. 2020.

ZOLET, A. C. T. et al. **Guia prático para estudos filogeográficos**. Sociedade ed. Ribeirão Preto - SP: Sociedade Brasileira de Gnética, 2013.

6 CHAPTER I

Different molecular markers influence genetic diversity parameters of *Euterpe edulis*

Abstract

Knowledge of the genetic diversity and population structure of species can facilitate conservation and guide management programs. The type of molecular markers used can affect estimates of genetic divergence and clustering of genotypes. Here we evaluated the ability of three marker types to estimate genetic diversity and population structure of *Euterpe edulis*. This palm species is endemic to the Atlantic Forest of Brazil and is economically and ecologically important, but it is threatened with extinction. The objective of this research was to compare genetic diversity and population structure accessed by SNP, Silico-DArT and microsatellites (SSR) markers in 138 *E. edulis* individuals sampled from 15 sites. Cluster analysis showed that samples correctly assigned to their cluster of origin for SNP and Silico-DArT. The SNP and Silico-DArT data showed less divergence between the individuals within each cluster in relation to the results with SSR data. The analysis of diversity, using both SNP and SSR markers, shows that the genetic divergence is high among populations when looking at the SNP markers, but when looking at the microsatellites, the biggest genetic divergence is found within each population. Within each population we found low inbreeding using both markers. In addition, two geographical groups, Northern and Southern, divided the Brazilian Atlantic Forest for both markers. Our results suggest that SNP and Silico-DArT markers are effective for assessing population structure, but SSR are better able to detect diversity among individuals. This work affirms the potential of markers SNP and SSR in genetic diversity study which can to reenforce the conservation and breeding of the *E. edulis*.

Keywords: genetic diversity; Juçara; molecular markers; palm.

6.1 INTRODUCTION

Molecular markers are used to quantify genetic diversity, genetic structure, and gene flow of species (AVISE, 2000; SILVA et al., 2020). Molecular markers are classified into two groups codominant and dominant, depending on their ability to distinguish allelic status of a heterozygote from a homozygote (TURCHETTO-ZOLET et al., 2017). Studies from genetic diversity and structure studies show that the type of markers used can affect the results of the association of genotypes/varieties to their original clusters (SIMKO; EUJAYL; VAN HINTUM, 2012; ZIMMERMAN; ALDRIDGE; OYLER-MCCANCE, 2020).

Many of these studies are performed with microsatellites (GAN et al., 2021; VINSON et al., 2018). The amplification variation of primers and fragment size homoplasy potentially reduce the accuracy of genetic estimates inferred from microsatellite markers (TSYKUN et al., 2017). Frequent forward and backward mutations of SSR loci can produce identical alleles in populations that are unrelated or genetically isolated (TSYKUN et al., 2017). The use of microsatellites markers may thus lead to estimates of genetic diversity and differentiation that can underestimate the divergence between population (VÄLI et al., 2008). Comparative analyzes between microsatellites and biallelic markers show that the high mutation rate of microsatellites leads to overestimation of heterozygosity and inbreeding values (SILVA et al., 2020). Furthermore, differentiation index estimates and a superior ability to resolve population structure on a fine scale are better biallelic markers when compared to microsatellites (HODEL et al., 2017; SINGH et al., 2013; ZIMMERMAN; ALDRIDGE; OYLER-MCCANCE, 2020).

Microsatellites (Simple Sequence Repeats, SSR) are short tandem repeats in DNA (TAUTZ; RENZ, 1984). In plants, microsatellite mutation rates range between 10^{-6} and 10^{-2} per locus and generation (BHARGAVA; FUENTES, 2010). Their high mutation rates allow for an accumulation of variation, which is leveraged for the study of population structure and phylogeography (TSYKUN et al., 2017). Because of their high reproducibility, multiallelism, and codominant inheritance, SSRs are frequently used in a diverse array of populations genetics studies (AHMED et al., 2021; ZANE; BARGELLONI; PATARNELLO, 2002).

Silico Diversity Arrays Technology (Silico-DArT) is another tool used for population genetics. Illumina sequencing applied to the original DArT method, called DArTseq, in addition to providing markers of single nucleotide polymorphisms (SNPs), also provides data on Silico-DarT (KILIAN et al., 2012, PETROLI et al., 2019). This method selects genomic

regions is based on the identification of regions that are conserved in the genomes of different species. Silico-DArT represents the presence or absence of restriction fragments of a particular sequence in genomic representations, thus providing less genetic information for a given locus (PETROLI et al., 2019). The mutation rate of Silico-DArT markers is not well-characterized, as it can vary depending on the specific DNA sequences being analyzed and the species being studied.

Single Nucleotide Polymorphisms are a single-point mutation of DNA in which one nucleotide in a particular locus is substituted with another one. The mutation rate of SNPs is typically on the order of 10^{-8} to 10^{-9} per base pair per generation. For being codominant and bi-allelic, population genetic statistics can be applied to them but a higher number of loci sufficiently polymorphic might be necessary to reach the same discriminatory power of individuals as multi-allelic SSR loci (GUICHOUX et al., 2011; TURCHETTO-ZOLET et al., 2017).

Although the SSR is more informative per locus than the SNP, however the SNPs favor the detection of haplotypes, which can be used to characterize evolutionary processes, and the sequencing technique allows a larger part of the genome to be sampled (VALDES; SLATKIN; FREIMERT, 1993a). Microsatellites are efficient for population genetics studies. However, the challenge of correctly interpreting microsatellite data is often heavily underestimated (SILVA et al., 2020) and the question of whether a limited number of microsatellite markers accurately reflects genome-wide diversity remains an issue. Additionally, during the analysis of the results, it is important to consider that each type of marker represents a specific evolutionary moment in the analysis of the results.

Knowledge of the diversity and genetic structure of tree species is a means of supporting conservation and management programs. *Euterpe edulis* (Juçara) is of particular interest because the Juçara fruit is similar to açai (*E. oleracea* and *E. precatoria*), whose fruit is increasingly consumed on the global market (SILVA; REIS, 2018; LUGON et al., 2021; BRASIL FLORA, 2022; DAMASCO et al., 2022). This species is a key species in the Atlantic Coastal Forests of Brazil, a biome with different biotic and abiotic conditions in its extension. In the work of Pereira et al. (2022), detected two genetic groups for *E. edulis* along the Atlantic Forest, as well as detected in other species, which may be related to the evolutionary history of this biome. The use of markers with greater comprehension detection capacity can help in genomic characterization studies of the species.

Although the existing information on the population genetics of *Euterpe edulis*, marker resources and data gathered thereof are still restricted to 18 microsatellite loci

(COELHO et al., 2020; PEREIRA et al., 2022; DA SILVA CARVALHO et al., 2015). To determine genetic diversity and structure of *E. edulis* and contribute to conservation and breeding programs of the species, we analyzed 138 *E. edulis* genotypes from 15 populations using SSR, Silico-DArT, and SNPs. Our objective was to evaluate the impact of the type of molecular marker on the diversity and structuring analysis of *Euterpe edulis*. This species is widely studied using a limited group of microsatellites, and given its wide distribution in the Atlantic Forest, a biome with many distinct phytophysiognomies, the detection of few genetic groups using SSRs may be underestimated. We question whether the type of marker influences the number of genetic groups detected so far as well as the diversity parameters of this species, which is threatened with extinction and highly promising for environmental, socio-economic, and conservation sustainability efforts

6.2 MATERIAL AND METHODS

6.2.1 Sampling information

We sampled 138 *E. edulis* individuals from 15 populations along the Brazilian Atlantic Forest (Fig.1). Sampling ranged from 7 to 10 adult individuals from conserve vegetation areas. The complete list of samples is provided in Supplementary Data S1.



6.2.3 DNA Extraction

The extraction of genomic DNA was performed using the cetyltrimethylammonium bromide or CTAB method by Doyle and Doyle (1990) with modifications described in Ferreira and Grattapaglia (1998). The protocol was adapted to carry out three steps of protein removal with 24:1 isoamyl alcohol chloroform and DNA precipitation without the use of ammonium acetate. The purified DNA was then quantified in NanoDrop® and samples that presented a 260/280 nm ratio between 1.8 and 2.0 were genotyped.

6.2.4 SSR markers

The genotyping was performed using eight microsatellite markers dye-labeled for *E. edulis* published by Gaiotto et al. (2001) (Supplementary Data S2). PCR was performed with forward primers that were labeled with FAM, VIC, or NED dyes (Applied Biosystems). The PCR reaction mixture (13 µl) consisted of approximately 20 ng template DNA, 1× PCR buffer (200 mM Tris-HCl, pH 8.4, and 500 mM KCl), 1 mM MgCl₂, 0.21 mM of each dNTP, and 0.17 µg/µl of BSA. Taq DNA polymerase and primers varied between markers. For EE5, EE9, EE47, EE54, and EE63 the reactions contained 1.2 U of Taq and 0.3 µM for 0.3 µM of each primer in the pair. For EE23 and the EE43-EE45, the reactions contained 1.25 U of Taq and 0.1 µM of each primer. The thermocycling conditions included an initial denaturing period of 4 min at 94 °C, followed by 30 cycles of 94 °C for 90 s, annealing at *T_a* (Supplementary Data S2) for 60 s, extension at 72 °C for 60 s, and a final extension period for 7 min at 72°C. Alleles were sized on an automated DNA sequencer (3500 Genetic Analyzer, Applied Biosystems) using GeneMarker software v.2.4 (HULCE; LI; SNYDER-LEIDBY, 2011) and LIZ 500 Size Standard (Applied Biosystems).

6.2.5 SNP and Silico-DArT markers

The genomic DNA (50ng/µl) of the 138 samples were sequenced by the Genetic Analysis Service for Agriculture Laboratory (SAGA in Spanish, Texcoco - Mexico). Library preparations were performed using two restriction enzymes, *MseI* and *HpaII*. The restriction site-specific adapter is labeled with 96 different barcodes that allow a 96-well microtiter plate to be multiplexed with equimolar amounts of amplification products to run within a flow cell

and sequenced with the Illumina Novaseq 6000 System with single-end read sequencing (PETROLI et al., 2019).

A proprietary analytical pipeline developed by DArT P/L is used to generate allele calling for SNPs and markers of presence/absence variation (Silico-DArT). The SNPs markers are identified *de novo* by comparing the sequences of fragments present in genomic representations (libraries) of samples processed in the execution of DArTsoft14. Thus, identifies and calls SNP markers completely independently of any reference genome. Silico-DArT markers are determined by the amplification (presence) or not (absence) of restriction fragments of a given sequence during sequencing. Silico-DArTs are extracted from sequence data using our proprietary algorithm in DArTsoft14 software. Markers were scored '1' for presence, and '0' for absence and '-' for calls with non-zero count but too low counts to score confidently as "1" for the Silico-DArT while the DArTSeq SNPs were scored '0' for reference allele homozygote, '1' for SNP allele homozygote and '2' for heterozygote (PETROLI et al., 2019).

6.2.6 Data analysis

Marker quality parameters

All marker systems were evaluated for their call rate (%) and polymorphism information content (PIC). Call rate (%), an expression of reliability of the final scores for each markers, was used to eliminate markers with $\geq 5\%$ missing data. PIC indicates how informative the marker is (BOTSTEIN et al., 1980), and how the relative frequencies of presence versus absence of signal are distributed (Sansaloni et al., 2012). The PIC value for the microsatellite markers was calculated by the formula of Botstein et al.(1980). Formula for SNP and Silico-DArT markers:

$$PIC = 1 - \frac{2maf^2(1-maf)^2}{maf^2 + (1-maf)^2}$$

where *maf* is minor allele frequency (SERROTE et al., 2020). In addition, for the SSR markers, the frequency of null alleles with MicroChecker software, version 2.2.3 (VAN OOSTERHOUT et al., 2004).

Genetic relationship analysis

Genetic dissimilarity matrices were constructed in DARwin v. 6.0.13 (PERRIER; ACQUEMOUD-COLLET, J.P, 2006) to identify the genetic relationships among the

individuals. Weighted neighbour-joining dendrograms were constructed in both marker platforms. Statistical support for relationships was tested by 20,000 bootstrap analyses. The distances used were:

Simple matching to SSR were: d_{ij} : dissimilarity between units i and j ; L : number of *loci*; π : ploidy; m_l : number of matching alleles for locus l .

Roger-Tanimoto to SNP, were: d_{ij} : dissimilarity between units i and j ; u : number of unmatching variables; m : number of matching variables.

Jaccard to Silico-DArT were: d_{ij} : dissimilarity between units i and j x_i, x_j : variable values for units i and j a : number of variables where $x_i = \text{presence}$ and $x_j = \text{presence}$ b : number of variables where $x_i = \text{presence}$ and $x_j = \text{absence}$ c : number of variables where $x_i = \text{absence}$ and $x_j = \text{presence}$.

Interpreting group memberships

For the next analysis outlier tests was performed. The test for deviation from neutral selection was implemented in OutFLANK v.0.2.0 (WHITLOCK; LOTTERHOS, 2015). To discard the potential bias due to the much larger number of SNPs and Silico-DArT to detect population differentiation, we randomly selected 100 SNPs and 445 Silico-DArT. The number chosen came from a general indication that around four to 13 SNPs and 23 to 56 Silico-DArT are expected to provide the equivalent power of a single microsatellite for population structure analyses (ROSENBERG et al., 2003; SIMKO; EUJAYL; VAN HINTUM, 2012).

The Discriminant Analysis of Principal Components (DAPC) analysis was conducted with the R package “adegenet” (JOMBART; DEVILLARD; BALLOUX, 2010). DAPC summarizes genotypes in principal components that are then used to construct linear functions that simultaneously maximize among-cluster variation and minimize within cluster variation (JOMBART; DEVILLARD; BALLOUX, 2010). The *dapc* function was used to evaluate the classification of individuals by the DAPC package in relation to the original clusters. Afterward, the function *find.clusters* was used to determine genetic clusters, based on principal component analysis retaining the maximum number of principal components and

the Bayesian information criterion (BIC) to generate a graph of optimum K. This procedure produced a scatterplot, and a smaller number of principal components were retained then, following recommendations in the package.

Genetic diversity

To estimate the population genetic diversity, we estimated i.e. observed (H_o) and expected (H_e) heterozygosity using with the R package “diversity” (KEENAN et al., 2013). We also estimated the genetic differentiation among populations (F_{st}) and inbreeding coefficient (F_{is}), across SNPs and SSR, using analysis of molecular variance implemented in “poppr” (KAMVAR; TABIMA; GRÜNWALD, 2014) based on Weir & Cockerham (1984). Significance levels of 0.05 for each estimate were determined with 10,000 permutations.

6.3 RESULTS

Marker quality parameters

The identified SNP and Silico-DArT markers showed a call rate in the range of 31–100% and 78–100% respectively (Fig. 2A). SSR markers had a call rate in the range of 86–99%. As such, markers with extremely low one ratio (<0.05) were not considered in the analysis. The total number of alleles per SSR locus ranged from 11 to 21 (Supplementary Data S2). Before passing the quality filters we had, eight microsatellites markers with an average PIC of 0.81, 74543 SNP markers with an average PIC of 0.29, and 248549 Silico-DArT markers with an average PIC of 0.05 (Fig. 2B). Considering all the quality parameters, 7833 SNP and 9885 Silico-DArT markers were used for subsequent analysis. These markers have an average PIC value 0.31 and 0.30 respectively.

A

B

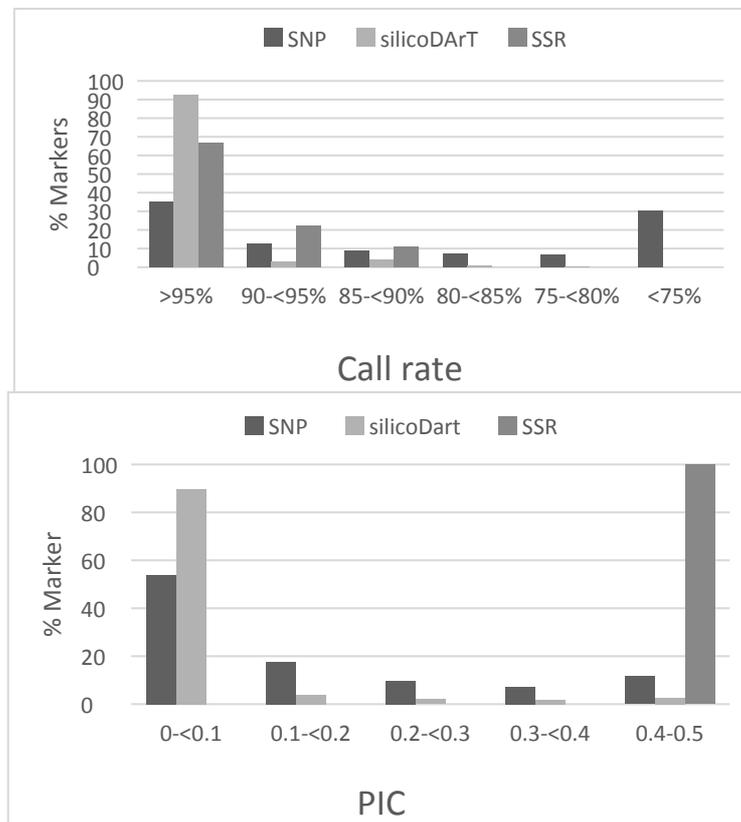
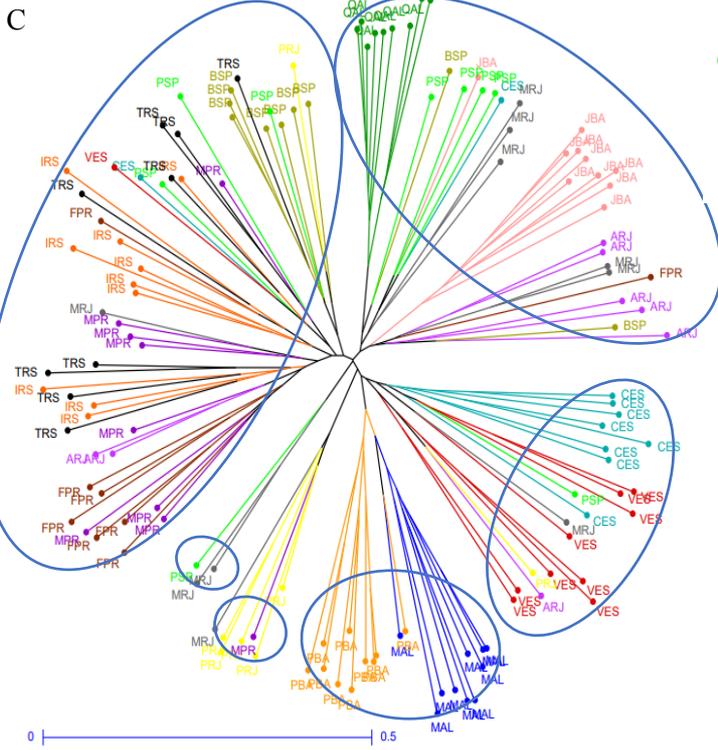
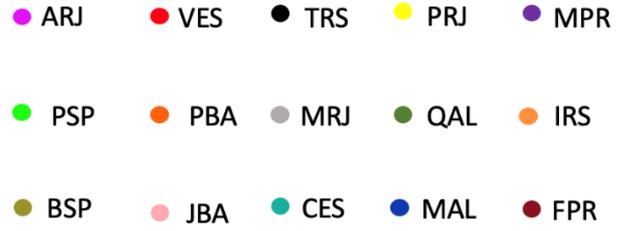
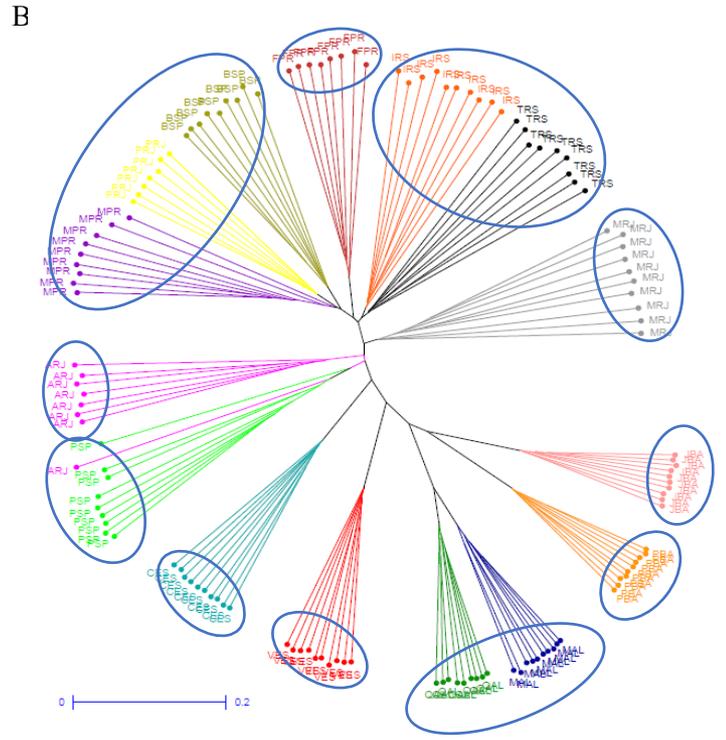
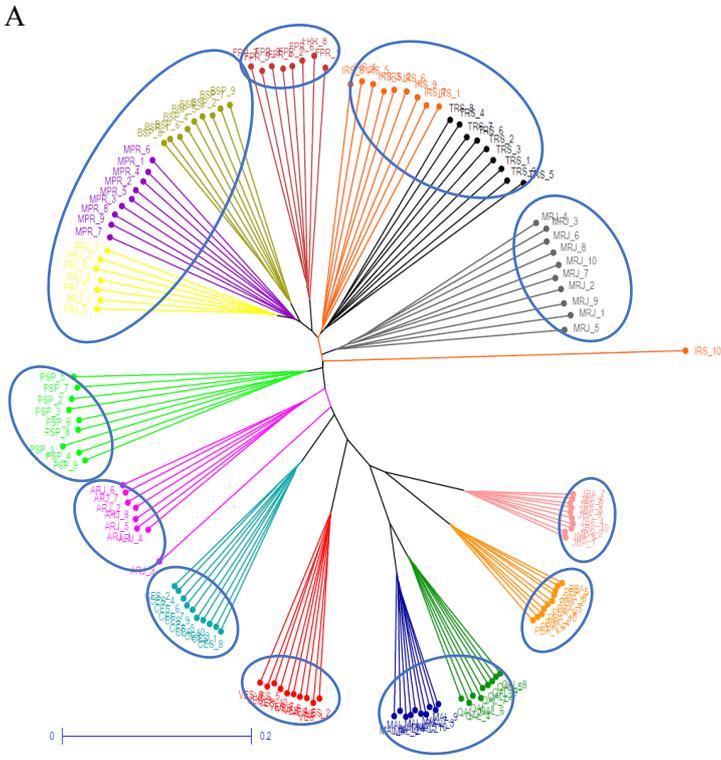


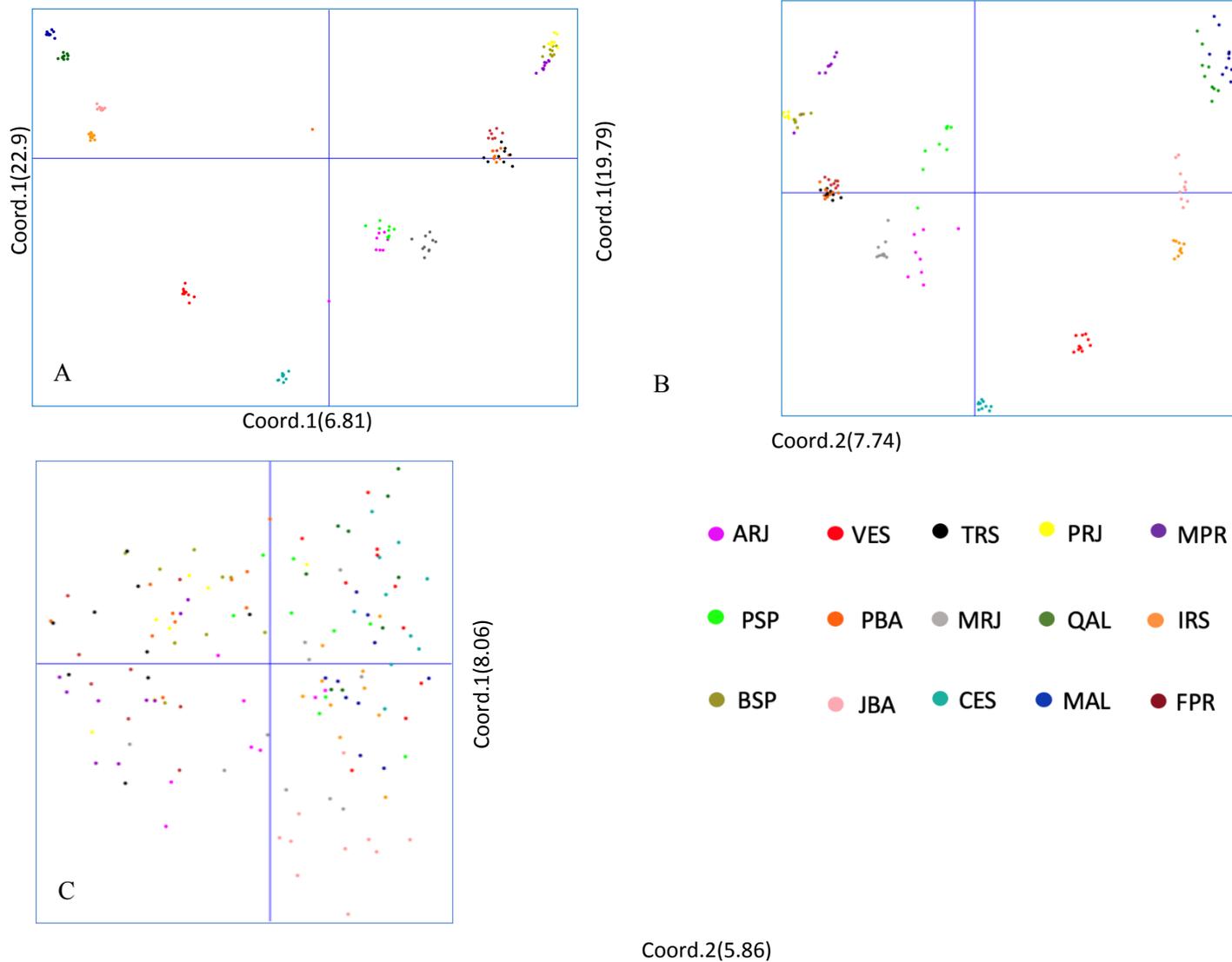
Figure 5: Distribution of SNP, Silico-DArT and SSR marker data for several quality parameters. (A) Percentage distribution of call rate values. (B) Percentage distribution of PIC values.

Genetic relationship analyses

The genetic dissimilarities among the individuals estimated through the Silico-DArT markers ranged from 0.29 to 0.91, SNP were 0.198 to 0.677, and SSR 0.187 to 0.973. In the Neighbour-joining tree (Fig. 3) *E. edulis* genotypes were grouped into eleven clusters for SNP (Fig. 3A) and Silico-DArT (Fig. 3B), and six clusters for SSR (Fig. 3C). Dendrograms obtained with SNP and Silico-DArT markers were similar, both produced several small clusters of genotypes that were collected at the same location. The clustering pattern using SSR markers showed subsets of the samples from the same origin within the larger clusters, but there are samples from these subgroups in more than one of the 6 clusters found.



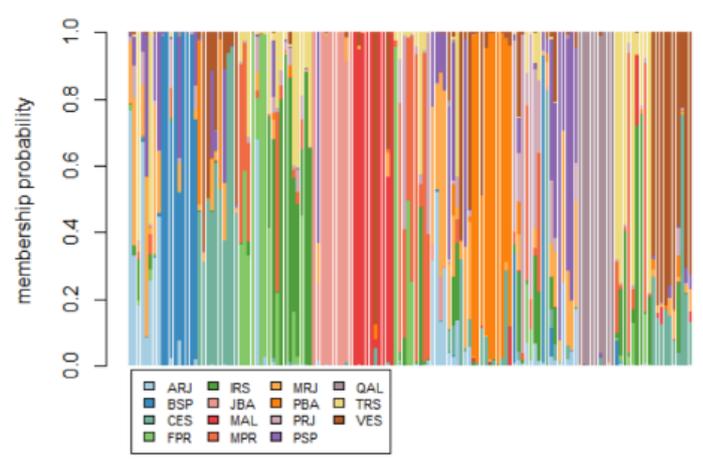
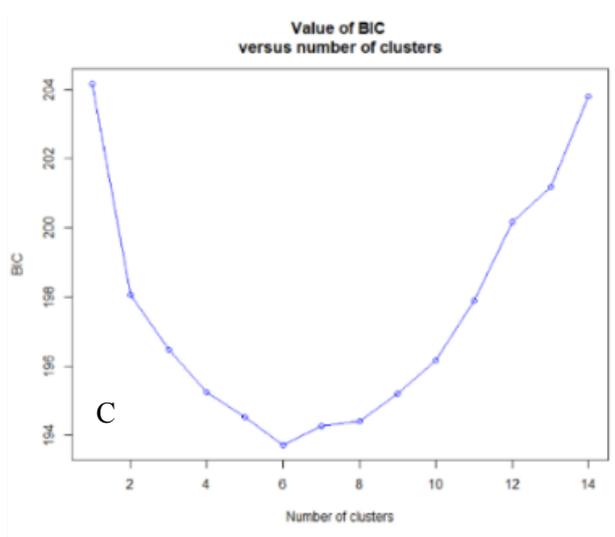
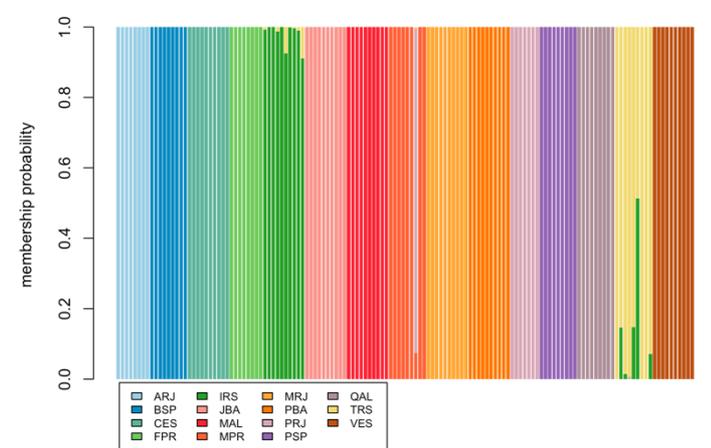
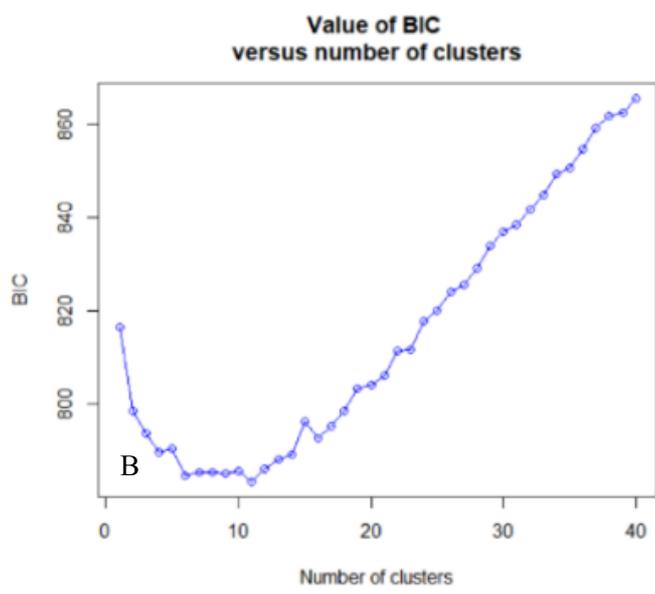
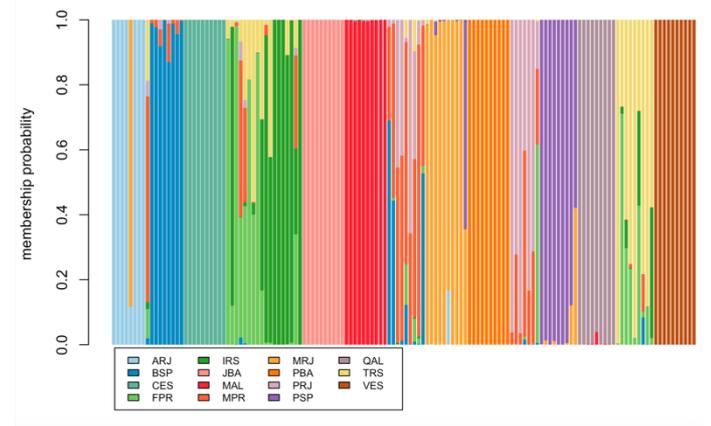
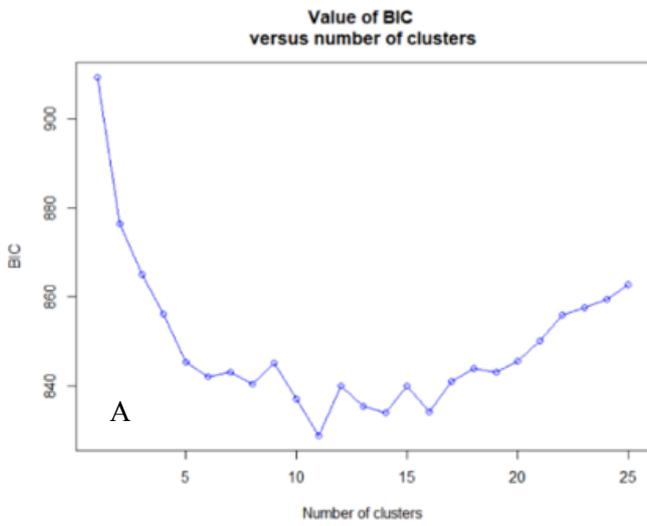
Principal Coordinate Analysis (PCoA) with SSR (Fig. 4C) markers showed that large diversity existed in *E. edulis* genotypes. Genotypes exhibited uniform distributions across the two axes (Fig. 4C). The first two axes explained 13.82% of cumulative variation. In PCoA all genotypes were labeled with different colors based on their different regions to indicate their region specificity (Fig. 4). The intermixing of color across the coordinates further supports the unrooted tree that there is no location-specific grouping of the samples. In the case of SNP and Silico-DArT markers genotypes were grouped eleven clusters across the first two axes (Fig. 4A and 4B). The first two axes of SNP and Silico-DArT explained 29.71% and 27.53% of cumulative variation respectively. In the PCoA, the arrangement of points between coordinates also supports the non-rooted tree of SNP and Silico-DArT markers, which have a specific location grouping.



Interpreting group memberships

We assess the probability of reassignment of the 138 samples to their cluster of origin, the eight SSR markers, a subset of 100 SNPs and a subset of 445 Silico-DArT were used, by DAPC method. Values above 0.5 indicate well-defined clusters, while lower values suggest mixed groups or poorly classified samples (Supplementary Data S3). The average association with the original clusters was 0.92 for SNPs, 0.71 for SSR and 0.99 for Silico-DArT. Of all 138 samples genotyped with SSR, 12 samples were classified in a discrepant way by DAPC in relation to the original cluster, and three populations (ARJ, MRJ and MPR) showed association values lower than 0.5 indicating the presence of mixtures of the other clusters. For the SNPs, the association values were above 0.77, and 1 were identified in clusters different from the original ones. The Silico-DArT marker association values were above 0.88 and all individuals were correctly classified by DAPC in relation to the original clusters. A detailed overview of the association of all studied samples and their origin is given in Supplementary Data S3.

After analyzing the classification of samples by the DAPC method in relation to the original clusters, hypothetical genetic clusters were inferred. The lowest BIC for hypothetical genetic clusters in DAPC corresponded to 6 groups with SSR (BIC = 168,799), 11 groups with the SNPs (BIC = 247,881) and Silico-DArT (BIC = 342,860) (Fig. 5). The optimal number of PCs to include in the DAPC analysis as determined by the a-score method was 7 for SNPs, 5 for SSR and 10 for Silico-DArT.



The SNPs dataset showed two separate clusters in quadrant four that are composed of QAL and MAL samples, and two others slightly overlapping clusters (PBA and JBA) in quadrant three, the other clusters are overlapping at the intersection of quadrants one and two (Supplementary Data S4A). Except for a sample from the ARJ site that was grouped in a different cluster from the others collected at the same site, all the other samples were grouped according to the collection site. The result for the Silico DAPC data (Supplementary Data S4B) were like the SNPs clusters composed of samples from the QAL, MAL, JBA and PBA locations were far from the others, and there is still a cluster with samples from the VES location that diverged from the other clusters that are found overlapping. The grouping of samples in DAPC with SSR identified one cluster separate from the other clusters along the discriminant function, in addition to the inferred clusters have samples from several different locations (Supplementary Data S4C).

Genetic diversity

As there is no way to identify heterozygotes using the Silico-DArT marker, this analysis was only performed for the SNP and microsatellites markers. The overall heterozygosity values for each population in the eight SSR ($H_e=0.69$; $H_o=0.59$) marker were almost four times higher than in the SNPs (data 7833 SNP, $H_e=0.15$; $H_o=0.13$; data 100 SNP, $H_e=0.14$; $H_o=0.12$) (Table 1). However, that did not result in significant higher inbreeding estimates for SSR. The similar direction of inbreeding was observed to SNP data 7833 only for seven out of the 15 populations (BSP, CES, IRS, JBA, MRJ, TRS and VES) and to SNP data 100 was observed nine (BSP, CES, FPR, IRS, JBA, MAL, MRJ, QAL, and TRS).

Overall, wider confidence intervals were observed around the estimates of inbreeding with SNP data 100 and SSR. Moreover, a greater genetic divergence within the population was noticed by SNP 100 and SSR data.

Table 1: Comparative summary of genetic diversity parameters (He expected heterozygosity; Ho observed heterozygosity) and inbreeding coefficient (Fis) with its respective 95% confidence interval (C.I) obtained with diversity package in R software for the two different data sets for the 15 *E. edulis* populations.

SNP data (7833)								SNP data (100)								Microsatellites (8)							
Regio n	Spot	He	Ho	Fis	Low r 95% C.I	Upper r 95% C.I	Sign	Regio n	Spot	He	Ho	Fis	Low r 95% C.I	Upper r 95% C.I	Sign	Region	Spot	He	Ho	Fis	Low r 95% C.I	Upper95 % C.I	Sign
ARJ	1	0,16 9	0,13	0,19	0,216	0,244	*	ARJ	1	0,17 5	0,13 6	0,20 8	0,1	0,336	*	ARJ	1	0,79	0,66	0,15	- 0,006	0,345	ns
BSP	2	0,15 6	0,14	0,08 1	0,089	0,117	*	BSP	2	0,16 2	0,13 7	0,11 5	0,04	0,268	*	BSP	2	0,68	0,7	-0,00 7	-0,2	0,122	ns
CES	3	0,14 4	0,14 7	-0,01 6	- 0,037	- 0,006	*	CES	3	0,12 7	0,11 7	0,07 9	- 0,034	0,202	ns	CES	3	0,71	0,6	0,1	- 0,081	0,363	ns
FPR	4	0,18 4	0,16 7	0,07 1	0,082	0,109	*	FPR	4	0,15 9	0,14 9	0,08 9	0,011	0,227	*	FPR	4	0,62	0,44	0,26	0,031	0,541	*
IRS	5	0,20 5	0,18	0,09 9	0,11	0,134	*	IRS	5	0,21 6	0,18 2	0,13 2	0,053	0,255	*	IRS	5	0,69	0,59	0,18	0,032	0,276	*
JBA	6	0,07 7	0,07 4	0,03 4	0,018	0,06	*	JBA	6	0,06 1	0,06 6	-0,04 8	- 0,314	0,156	ns	JBA	6	0,59	0,43	0,33	0,082	0,49	*
MAL	7	0,09 3	0,08 6	0,06 7	0,059	0,098	*	MAL	7	0,07 6	0,07 3	0,04 2	- 0,162	0,222	ns	MAL	7	0,72	0,66	0,05	- 0,175	0,38	ns
MPR	8	0,17 1	0,14 9	0,09 6	0,114	0,141	*	MPR	8	0,17 5	0,13 9	0,14 6	0,074	0,318	*	MPR	8	0,61	0,5	0,24	- 0,115	0,526	ns
MRJ	9	0,20 2	0,17 7	0,10 2	0,112	0,136	*	MRJ	9	0,18 9	0,15 3	0,15 8	0,088	0,281	*	MRJ	9	0,8	0,81	-0,01	- 0,169	0,133	ns
PBA	10	0,09 2	0,09 4	-0,00 6	- 0,035	0,003	ns	PBA	10	0,08 5	0,09 2	-0,05	- 0,257	0,09	ns	PBA	10	0,65	0,62	0,09	- 0,092	0,244	ns
PRJ	11	0,15 4	0,13 6	0,07 6	0,087	0,12	*	PRJ	11	0,14 4	0,11 2	0,18 4	0,073	0,375	*	PRJ	11	0,75	0,67	0,09	- 0,119	0,32	ns

PSP	12	0,20	0,16	0,14	0,163	0,189	*	PSP	12	0,20	0,13	0,33	0,252	0,458	*	PSP	12	0,79	0,66	0,13	-	0,396	ns
		4	8	4						6	2	4								0,048			
QAL	13	0,10	0,09	0,06	0,064	0,104	*	QAL	13	0,06	0,06	0,01	-	0,245	ns	QAL	13	0,61	0,59	0,03	-	0,173	ns
		5	6	9						3		4	0,189							0,108			
TRS	14	0,20	0,18	0,08	0,096	0,121	*	TRS	14	0,19	0,16	0,13	0,056	0,251	*	TRS	14	0,71	0,58	0,19	0,031	0,343	*
		5	3	7						3	3	5											
VES	15	0,13	0,12	0,05	0,058	0,087	*	VES	15	0,11	0,10	0,04	-	0,183	ns	VES	15	0,71	0,43	0,39	0,146	0,63	*
		5	5	4						6	7		0,047										
	Overall	0,15	0,13	0,07	0,08	0,110	*			0,14	0,12	0,10	-	0,258	ns			0,69	0,596	0,14	-	0,352	ns
	1	3	7	6						3	1	5	0,017					5		8	0,053		

Estimates of within population inbreeding based on Weir and cockerham estimator; estimates contained in a 95% confidence interval containing zero are declared not significantly different (ns) from zero; otherwise, inbreeding was declared significantly different from zero (*)

The AMOVA estimated for 7833 SNPs and 100 SNPs, showed that the genetic divergence using SNP markers is due ~58% and ~55% to the difference among the populations and ~41% and ~44% within populations. On the other hand, the eight SSR results showed that the greatest genetic divergence occurs ~83% within each population and only ~16% among populations (Table 2). Within each population was found high diversity for both markers. In addition, it was noticed a high genetic structure was found among the populations for both markers, and two genetic groups throughout the Atlantic Coastal Forest, Northern and Southern, as seen in Pereira et al. (2022) (Fig. 6).

Table 2: Analysis of Molecular Variance (AMOVA) for SNP 7833, SNP 100 and SSR 8 data, using package “ade4” in R software.

Markers	Source of variation	Degrees of freedom	Sum of squares	Sigma	% of variation	phi	Pvalue
7833 SNP	Among population	14	129806.56	937.642	58.999		
	Within populations	123	80147.04	651.602	41		
	Total	137	209953.60	1589.244	100	0.589	0.00005
100 SNP	Among population	14	1690.346	12.081	55.54		
	Within populations	123	1189.153	9.667	44.45		
	Total	137	2879.499	21.749	100	0.555	0.00005
8 SSR	Among population	14	146.152	0.740	16.91		
	Within populations	123	447.237	3.636	83.08		
	Total	137	593.390	4.376	100	0.169	0.0005

Significance tests were performed with 10,100 permutations.

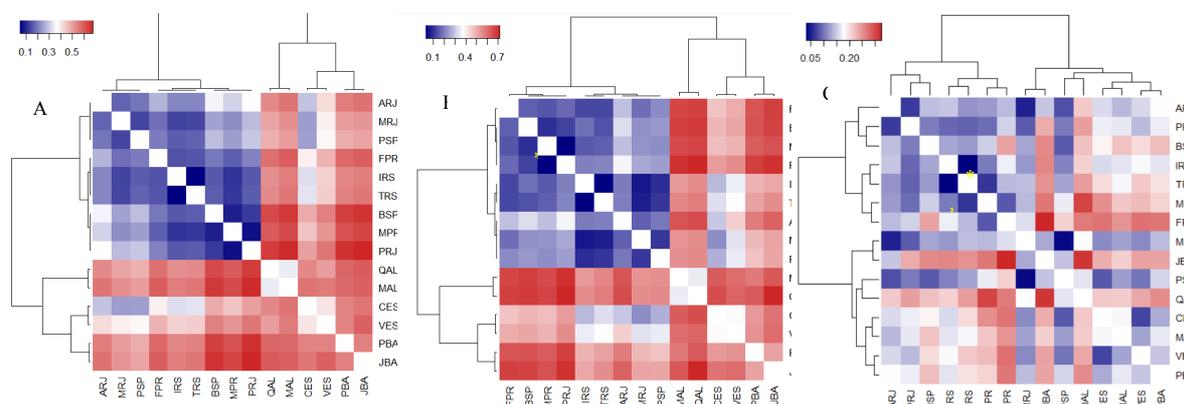


Figure 6: Heat map of pairwise Fst values estimated. (A) 7,833 SNPs data, (B) 100 SNP data and (C) eight microsatellite data across all populations. in blue the Fst estimate means that the

difference between the populations is smaller. in red the Fst estimate shows the greatest difference between populations. *Significant Fst.

6.4 DISCUSSION

In the present study, the analysis allowed accessing information on the structure and genetic diversity of *E. edulis* samples obtained by three types of markers. The type of marker may present differences in the interference of the structure and population diversity, due to different allele frequency spectra, mutation rates and mechanisms and genome-wide distribution patterns (GRÜNWALD et al., 2017; VINSON et al., 2018; YILDIRIM; FORSMAN; SUNDE, 2017). Our results suggest that SNP and Silico-DArT markers are effective for assessing population structure, but SSR are better able to detect diversity among samples. This work affirms the potential of markers SNP and SSR in genetic diversity study of *E. edulis*.

The quality parameter of markers in *E. edulis* were comparable with that of other species (ADU et al., 2021; MATTHIES et al., 2012). There was conservation of Silico-DArT low rate of polymorphism, high frequency of locus with PIC below 0.1, which is expected by restriction sites that are conserved in all species (ADU et al., 2021; MATTHIES et al., 2012). SSR markers (0.81) frequently have higher PIC values than SNP (0.31) or Silico-DArT (0.30) markers, because multiallelic markers are more informative than markers biallelic (SERROTE et al., 2020). Furthermore, due to the biallelic and dominant nature of SNPs and Silico-DArT markers, PIC values can range from 0 to 0.5 (SERROTE et al., 2020). Whereas for SSR markers which are mutli-allelic PIC value goes above 0.5 and can go up to 1.0 (SERROTE et al., 2020). PIC depends on many factors such as breeding behavior of the species, genetic diversity in the samples, size of the samples, sensitivity of the genotyping method and location of primers in the genome used for study (SINGH et al., 2013). The average PIC value found for SNP and Silico-DArT is considered a moderate value, which indicates polymorphism and genetic diversity in the samples studied (GROSSI et al., 2021).

A genetic distance approach based on SNP, SSR and Silico-DArT were successfully used to verify the relationship among the samples. It was already expected to find a greater distance between samples with the use of SSR markers (HURTADO et al., 2008; SINGH et al., 2013a). Due to high recent mutation rates, they differ better the samples than SNP and Silico-DArT (YILDIRIM; FORSMAN; SUNDE, 2017). SNPs are known to have two alleles per locus and follow infinite sites model of evolution, while microsatellites, on the other

hand, possess multiple alleles per locus and are subject to stepwise or two-phase mutation models in their evolution (VALDES; SLATKIN; FREIMERT, 1993). However, there may be underestimation of the divergence between populations, in reason of the homoplasmy by fragment size. The fragments have the same size but there be variation on sequence which is not identify (TSYKUN et al., 2017). That can influence the inference of the populational structure as well (GUICHOUX et al., 2011). A common problem in population genetics studies is a sample to one of cluster based on its genotype and information about distribution of the alleles in the populations (MATTHIES et al., 2012; SIMKO; EUJAYL; VAN HINTUM, 2012) . The clustering analyses showed an increase in precision with SNP and Silico-DArT data that is consistent with previous studies. Hurtado et al. (2008) noticed that 251 DArTs and 36 SSRs generated broadly similar clustering patterns in 436 varieties of cassava; however, greater genetic differentiation was revealed with SSR markers.

Regarding genetic diversity, it was possible to notice low heterozygosity for the population from North using SNPs markers. These populations are more isolated, which limited the gene flow and lead the smaller genetic variation and fixation some alleles. In the work by SILVA et al. (2020) with subtropical conifer tree, it was also possible to observe that the SNPs in relation to the SSRs indicated a greater difference in the average of heterozygosity for a set of populations that were further away from the others studied. Furthermore, higher rates of heterozygosity for microsatellite markers were observed and underestimation of population differentiation, as in our work.

Comparing the results of the two SNPs datasets (7833 SNPs and 100SNPs), it was noted that there was no significant difference in the estimates of genetic diversity. The findings indicate that utilizing a limited number of SNPs can effectively evaluate heterozygosity, inbreeding, and accurately identify and differentiate genetic variation among populations, as well as in the work of Silva et al. (2020). The significant differences between SNPs and SSRs were observed with respect to F_{ST} , but not for F_{IS} . Using SNPs, greater differentiation was noticed between the more distant populations, forming two evident groups (North and South). In contrast, with the SSR data, which the two groups are also observed, but with a smaller difference between them. These two groups and high levels of genotypic diversity were expected in these populations, as observed in Pereira et al. (2022) study with microsatellites. Furthermore, although we found high F_{ST} values for SNPs, indicating low gene flow between populations, a lower value is observed among populations located in the southeast. The lower F_{ST} for SSR markers can be explained because this estimator is related to the number of alleles per locus (JOST, 2008).

Each molecular marker type provides information from different evolutionary moments. The discrepancy in genetic diversity estimates obtained from microsatellites and genome-wide SNP data may be due to recent demographic events. It is important to consider that each marker type follows its own mutational model, which can affect the interpretation of the results. For example, microsatellites have a high mutation rate and are more sensitive to genetic drift, and their variability can be influenced by population size and effective population size. Similarly, SNPs can be affected by natural selection, demographic history, and genetic drift. Silico-DArT markers, being a combination of microsatellites and SNPs, can be affected by both types of mutational models. Microsatellites are in non-coding regions and are mostly selectively neutral, hence of less evolutionary importance. Therefore, genome-wide SNP diversity better reflects functionally important and potentially adaptive genetic variation and should be prioritized, particularly in conservation genetics studies (YILDIRIM; FORSMAN; SUNDE, 2021).

6.5 CONCLUSION

SNPs and Silico-DArT markers are effective for assessing population structure and SSR markers are better to detect diversity between samples. In population studies, SNPs are better able to identify genetic diversity and clusters, mainly if there is a hypothesis of drift or subdivision of populations. But, if the study's focus is on a single population or relates to the reproductive system of the species, the SSRs will better estimate divergence within samples.

Acknowledgments

Financial support to this study was obtained from the Conselho Nacional de Pesquisa (CNPq, Brazil) (Researcher productivity fellowship AF; MFSF #311840/2019-1), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil)—Finance Code 001, and from the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES, Vitória—ES, Brazil) in partnership with VALE and the Swedish Research Council (2017-04980).

6.6 REFERENCES

- ADU, B. G. et al. High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz). **PLoS ONE**, v. 16, n. 7 July, p. 1–19, 2021.
- BACCI, L. F. et al. Biogeographic breaks in the Atlantic Forest: evidence for Oligocene/Miocene diversification in *Bertolonia* (Melastomataceae). **Botanical Journal of the Linnean Society**, p. 1–16, 2021.
- BAKER, W. J.; COUVREUR, T. L. P. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. **Journal of Biogeography**, v. 40, n. 2, p. 274–285, 2013.
- BRANCALION, P. H. S. et al. Phenotypic plasticity and local adaptation favor range expansion of a Neotropical palm. **Ecology and Evolution**, v. 8, n. 15, p. 7462–7475, 3 ago. 2018.
- BROWN, J. L. et al. Paleoclim, high spatial resolution paleoclimate surfaces for global land areas. **Scientific Data**, v. 5, 2018.
- CARNAVAL, A. C. et al. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. **Science**, v. 323, n. 5915, p. 785–789, 2009.
- CARNAVAL, A. C. et al. Prediction of phylogeographic endemism in an environmentally complex biome. **Proceedings of the Royal Society B: Biological Sciences**, v. 281, n. 1792, 2014.
- CARVALHO, C. DA S. et al. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. **Heredity**, v. 115, n. 3, p. 216–224, set. 2015.
- CARVALHO, C. DA 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.
- CARVALHO, C. S. et al. Defaunation leads to microevolutionary changes in a tropical palm. **Scientific Reports**, v. 6, p. 1–9, 2016.
- CINGOLANI, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. **Fly**, v. 6, n. 2, p. 80–92, 2012.

COELHO, G. M. et al. Genetic structure among morphotypes of the endangered Brazilian palm *Euterpe edulis* Mart (Arecaceae). **Ecology and Evolution**, v. 10, n. 12, p. 6039–6048, 2020.

CONABB. **Boletim da Sociobiodiversidade / Companhia nacional de Abastecimento**. Brasilia: [s.n.].

COSTA, L. P.; LEITE, Y. L. R. Historical Fragmentation Shaping Vertebrate Diversification in the Atlantic Forest Biodiversity Hotspot. Em: **Bones, Clones, and Biomes**. [s.l.] University of Chicago Press, 2013. p. 283–306.

CRUZ, V. M. V; KILIAN, A.; DIERIG, D. A. Development of DArT Marker Platforms and Genetic Diversity Assessment of the U . S . Collection of the New Oilseed Crop *Lesquerella* and Related Species. **PLoS ONE**, v. 8, n. 5, p. 1–13, 2013.

DA SILVA, J. M. C.; CASTELETI, C. H. M. **In The Atlantic Forest of South America: Biodiversity Status, Threats, and Outlook**. , 2003.

DANECEK, P. et al. Twelve years of SAMtools and BCFtools. **GigaScience**, v. 10, n. 2, 1 fev. 2021.

DASILVA, M. B.; PINTO-DA-ROCHA, R.; DESOUSA, A. M. A protocol for the delimitation of areas of endemism and the historical regionalization of the Brazilian Atlantic Rain Forest using harvestmen distribution data. **Cladistics**, v. 31, n. 6, p. 692–705, 1 dez. 2015.

DRANSFIELD, J. et al. A new phylogenetic classification of the palm family , Arecaceae. v. 60, n. 4, p. 559–569, 2014.

EXCOFFIER, L.; LISCHER, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, v. 10, n. 3, p. 564–567, 2010.

FISHER, J. A.; FAVREAU, M. B. R. Plasmid purification by phenol extraction from guanidinium thiocyanate solution: Development of an automated protocol. **Analytical Biochemistry**, v. 194, n. 2, p. 309–315, 1991.

FOYEZ, S. et al. Application of DArT seq derived SNP tags for comparative genome analysis in fishes ; An alternative pipeline using sequence data from a non-traditional model species , *Macquaria ambigua*. **PLoS ONE**, v. 14, n. 12, p. 1–13, 2019.

FRICHOT, E.; FRANÇOIS, O. **LEA: An R Package for Landscape and Ecological Association Studies**. [s.l: s.n.].

GALETTI, M. et al. Functional Extinction of Birds Drives Rapid Evolutionary Changes in Seed Size. **Science**, v. 340, n. 6136, p. 1086–1090, maio 2013.

GARRISON, E.; MARTH, G. Haplotype-based variant detection from short-read sequencing. 17 jul. 2012.

GREGORY, T. R. Understanding Natural Selection: Essential Concepts and Common Misconceptions. **Evolution: Education and Outreach**, v. 2, n. 2, p. 156–175, 9 jun. 2009.

GROSSI, L. L. et al. DArTseq-derived SNPs for the genus *Psidium* reveal the high diversity of native species. **Tree Genetics and Genomes**, v. 17, n. 2, 2021.

GRÜNWARD, N. J. et al. Best Practices for Population Genetic Analyses. **Phytopathology®**, v. 107, n. 9, p. 1000–1010, set. 2017.

GUICHOUX, E. et al. Current trends in microsatellite genotyping. **Molecular Ecology Resources**, v. 11, n. 4, p. 591–611, 12 jul. 2011.

HURTADO, P. et al. Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). **Plant Genetic Resources: Characterisation and Utilisation**, v. 6, n. 3, p. 208–214, 2008.

JACCOUD, D. et al. Diversity Arrays : a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, v. 29, n. 4, p. 1–7, 2001.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. **BMC Genetics**, v. 11, n. 1, p. 94, 2010.

JOST, L. G_{ST} and its relatives do not measure differentiation. **Molecular Ecology**, v. 17, n. 18, p. 4015–4026, set. 2008.

KAMVAR, Z. N.; TABIMA, J. F.; GRÜNWARD, N. J. *Poppr* : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. **PeerJ**, v. 2, p. e281, 4 mar. 2014.

KEENAN, K. et al. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. **Methods in Ecology and Evolution**, v. 4, n. 8, p. 782–788, ago. 2013.

KILIAN, A. et al. The fast and the cheap : SNP and DArT-based whole. “**In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution**”, n. May 2003, p. 443–461, 2005.

KILIAN, A. et al. Diversity arrays technology: A generic genome profiling technology on open platforms. **Methods in Molecular Biology**, v. 888, p. 67–89, 2012.

LORENZI, H. J. **Flora brasileira: Arecaceae (palmeiras)**. Instituto ed. Nova Odessa, São Paulo: [s.n.].

- LUGON, M. D. et al. Is Your Açaí Really from Amazon? Using DNA Barcoding to Authenticate Commercial Products. **Food Analytical Methods**, v. 14, n. 8, p. 1559–1566, 11 ago. 2021.
- MARTINELLI, G.; MORAES, M. A. **Livro Vermelho Flora do Brasil**. 1. ed. Rio de Janeiro: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, 2013.
- MATOS-MARAVÍ, P. et al. Mesoamerica is a cradle and the Atlantic Forest is a museum of Neotropical butterfly diversity: Insights from the evolution and biogeography of Brassolini (Lepidoptera: Nymphalidae). **Biological Journal of the Linnean Society**, v. 133, n. 3, p. 704–724, 1 jul. 2021.
- MATTHIES, I. E. et al. Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. **Molecular Breeding**, v. 30, n. 2, p. 951–966, 2012.
- MCRAE, B. H. **ISOLATION BY RESISTANCE** Evolution. [s.l: s.n.].
- MERCIER, K. P. et al. Linking environmental stability with genetic diversity and population structure in two Atlantic Forest palm trees. **Journal of Biogeography**, v. 50, n. 1, p. 197–208, 1 jan. 2023.
- MYERS, N. et al. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853–858, fev. 2000.
- NEELABH. Autosome. Em: **Encyclopedia of Animal Cognition and Behavior**. Cham: Springer International Publishing, 2017. p. 1–3.
- NEI, M. Chapter 8: Genetic Variation Within Species. Em: **Molecular Evolutionary Genetics**. [s.l.] Columbia University Press, 1987. p. 176–207.
- OWENS, H. L.; GURALNICK, R. P. **climateStability: AN R PACKAGE TO ESTIMATE CLIMATE STABILITY FROM TIME-SLICE CLIMATOLOGIES** Biodiversity Informatics. [s.l: s.n.]. Disponível em: <<https://github.com/hannahlowens/climateStability>>.
- PEREIRA, A. G. et al. Patterns of genetic diversity and structure of a threatened palm species (*Euterpe edulis* Arecaceae) from the Brazilian Atlantic Forest. **Heredity**, 2022.
- PICHARDO-MARCANO, F. J. et al. Molecular Phylogenetics and Evolution Phylogeny , historical biogeography and diversification rates in an economically important group of Neotropical palms : Tribe Euterpeae. **Molecular Phylogenetics and Evolution**, v. 133, n. September 2018, p. 67–81, 2019.
- PURCELL, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

- ROSENBERG, N. A. et al. **Informativeness of Genetic Markers for Inference of Ancestry** ***Am. J. Hum. Genet.** [s.l: s.n.].
- RULL, V.; CARNAVAL, A. C. **Neotropical Diversification: Patterns and Processes.** Cham: Springer International Publishing, 2020.
- SANSALONI, C. et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. **Nature Communications**, v. 11, n. 1, p. 4572, 11 dez. 2020.
- SANSALONI, C. P. et al. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. **Plant Methods**, v. 6, n. 16, p. 1–11, 2010.
- SERROTE, C. M. L. et al. **Determining the Polymorphism Information Content of a molecular marker.** GeneElsevier B.V., , 5 fev. 2020.
- SHERMAN, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). **Nucleic Acids Research**, v. 50, n. W1, p. W216–W221, 5 jul. 2022.
- SILVA, P. I. T. et al. A 3K Axiom SNP array from a transcriptome-wide SNP resource sheds new light on the genetic diversity and structure of the iconic subtropical conifer tree *Araucaria angustifolia* (Bert.) Kuntze. **PLoS ONE**, v. 15, n. 8 August, p. 1–25, 2020.
- SIMKO, I.; EUJAYL, I.; VAN HINTUM, T. J. L. Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. **Plant Science**, v. 184, p. 54–62, 2012.
- SINGH, N. et al. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. **PLoS ONE**, v. 8, n. 12, p. 1–14, 2013a.
- SINGH, R. et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. **Nature**, v. 500, n. 7462, p. 335–339, 2013b.
- TSYKUN, T. et al. Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. **Heredity**, v. 119, n. 5, p. 371–380, 1 nov. 2017.
- TURCHETTO-ZOLET, A. C. et al. **Marcadores Moleculares na Era G enômica : Metodologias e Aplicações.** Ribeirão Preto: Sociedade Brasileira de Gnética, 2017.
- VALDES, A. M.; SLATKIN, M.; FREIMERT, N. B. **Allele Frequencies at Microsatellite Loci: The Stepwise Mutation Model Revisited** **Genetics.** [s.l: s.n.].
- VALE, M. M. et al. Endemic birds of the Atlantic Forest: traits, conservation status, and patterns of biodiversity. **Journal of Field Ornithology**, v. 89, n. 3, p. 193–206, 1 set. 2018.

VINSON, C. C. et al. Using molecular markers to investigate genetic diversity, mating system and gene flow of Neotropical trees. **Revista Brasileira de Botânica**, v. 41, n. 2, p. 481–496, 2018.

WEIR, B. S.; COCKERHAM, C. C. Estimating F-statistics for the analysis of population structure. **Evolution**, v. 38, n. 6, p. 1358–1370, 1984.

WHITLOCK, M. C.; LOTTERHOS, K. E. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. **American Naturalist**, v. 186, p. S24–S36, 1 out. 2015.

WIERINGA, J. G.; CARSTENS, B. C.; GIBBS, H. L. Predicting migration routes for three species of migratory bats using species distribution models. **PeerJ**, v. 9, p. e11177, 16 abr. 2021.

WRIGHT, S. ISOLATION BY DISTANCE. **Genetics**, v. 28, n. 2, p. 114–138, 29 mar. 1943.

YILDIRIM, Y.; FORSMAN, A.; SUNDE, J. Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. **BMC Genomics**, v. 18, n. 1, 11 jan. 2017.

6.7 SUPPLEMENTARY DATA

Table S1: Information about local sample.

Spot	Collection number	Local (code)	Local	Locality	n	Latitude	Longitude
1	INEA(056-2017)	ARJ	Reserva Biológica Estadual de Araras	Araras, RJ	8	-22.434.722	-43.257223
2	Private propriety	BSP	Pousada Brejal	Bananal, SP	9	-22.858.592	-44.471370
3	SISBIO-4	CES	Caparaó, Chale dos Faria	Pedra Menina, ES	10	-20.503511	-41.818960
4	SISBIO-4	FPR	Parque Nacional do Iguaçu	Foz do Iguaçu, PR	8	-25.461766	-53.818300
5	SEMA(07/18)	IRS	Reserva Biológica da Mata Paludosa	Itati, RS	10	-29.850833	-50.189184
6	ONG Gambá	JBA	Serra da Jiboia Monte Cruzeiro	Elísio Medrado, BA	10	-12.872034	-39.481399
7	S9344-4	MAL	Estação ecológica de Murici	Murici, AL	10	-9.226930	-35.857060
8	IAP(47.17)	MPR	Parque Estadual Pico do Marumbi	Morretes, PR	9	-25.445412	-48.916382
9	Private propriety	MRJ	Capelinha	Visconde de Mauá, RJ	10	-22.450029	-44.499745
10	SISBIO-4	PBA	Parque Nacional do Descobrimento	Prado, BA	10	-17.093301	-39.310054
11	Private propriety	PRJ	Engenho D'Ouro	Paraty, RJ	7	-23.213713	-44.793159
12	COTEC (080-833-2017)	PSP	Estação Ecológica de Ibicatu	Piracicaba, SP	9	-22.782675	-47.821360
13	Private propriety	QAL	Pedra Talhada	Quebrangulo, AL	9	-9.244635	-36.419087
14	SEMA(07/18)	TRS	Parque Estadual do Itapeva	Torres, RS	9	-29.613333	-50.282784
15	VALE	VES	Reserva Natural da Vale	Linhares, ES	10	-19.151310	-40.070845

Table S2: Description of eight microsatellite loci of *Eutерpe edulis*. The primer sequences are listed with their respective fluorescence markings, annealing temperatures (Ta), number of alleles per locus (A), allele Size (bp), and the mean frequency of null allele.

Locus	Primer Sequence (5'-3')	Array	Fluorescence	Ta	A	Allele Size (bp)	Null Allele Mean Frequency
EE5	F: GAGAACACATAAGCTGC R: GCTTCAGAATTAGGACA	(AG)24	6-FAM	56 °C	19	84 - 146	0.12
EE9	F: TTCTCTCGCATGCCTCG R:GCCACACACACACAGTAGAATC	(AG)19	NED	58 °C	13	80 - 110	0.13
EE23	F: GTTCTGCGATTCATACTCCTG R: TACGAACCAAGATGGAGCAA	(A)14(AG)23	VIC	58 °C	15	84 - 134	0.16
EE43	F: GCGAAAGGCTAACAACGTTAT R: AGCGAACCAACCAAGAAGAC	(AG)16	VIC	56 °C	14	84 - 150	0.19
EE45	F: AAAGAAATTGGCGTGACATC R: AACCAGTCTTCTCCCTCTCG	(AG)28	6-FAM	56 °C	15	90 - 120	0.16
EE47	F: CGAAATCAATGGTTTCAGTG R: AATTATTGTTGTGGGCAGC	(AG)20	6-FAM	56 °C	20	212 - 268	0.18
EE54	F: CATGTATCTAAGGAACAAGG R: CTGTGCTCTCTCATTCTCA	(AG)25	NED	56 °C	21	100 - 152	0.11
EE63	F: CCGATATGCTCAAATCAATG R: ACGAGAGGAATCAAAGAACC	(AG)18	6-FAM	58 °C	11	94 - 130	0.17
Overall sum					12		
					8		

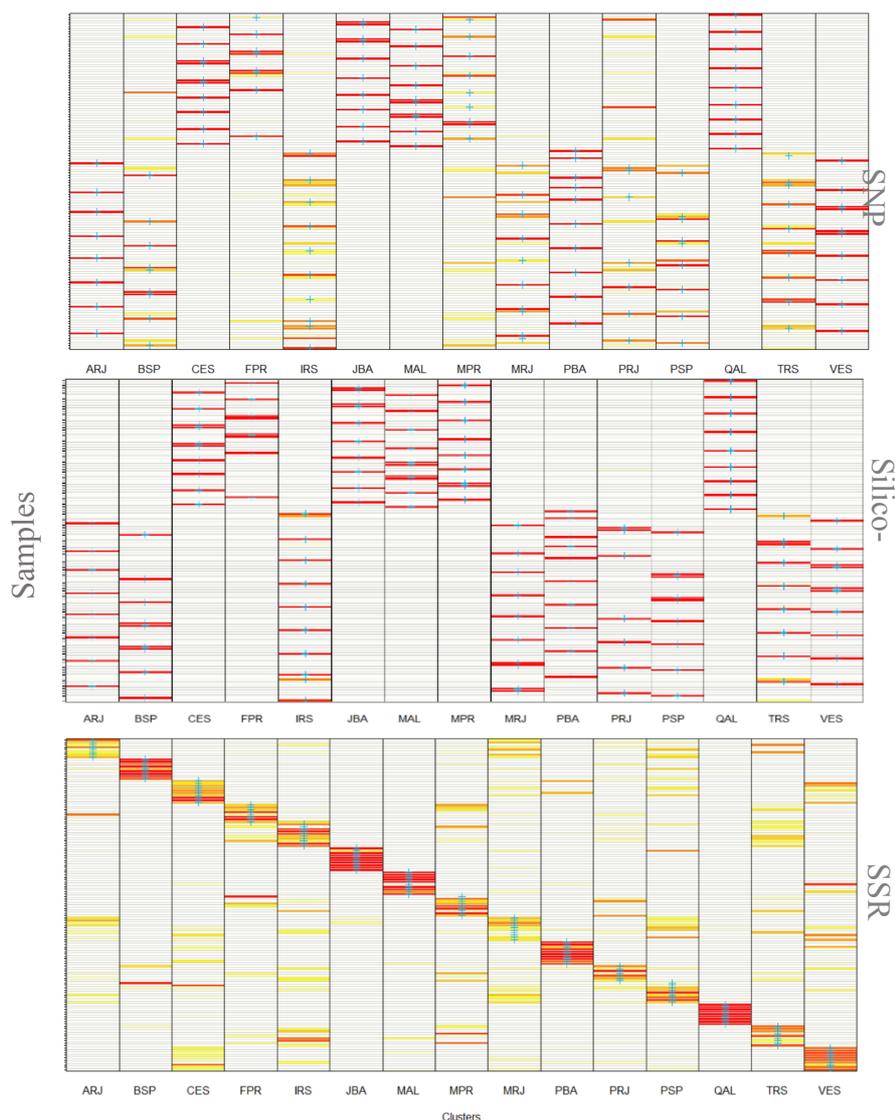


Figure S3: Classification of the 138 samples of *Euterpe edulis* by the DAPC method in relation to the original clusters. SNP data. Silico-DArT data. SSR data. Heat colors represent membership probabilities (red=1, white=0); blue crosses represent the prior cluster provided to DAPC. The DAPC classification is consistent with the original clusters when blue crosses are on red rectangles. Such figure is particularly useful when prior biological groups are used, as one may infer admixed or misclassified samples.

7 CHAPTER II

Population genomics and phylogeography of *Euterpe edulis*: A comprehensive study of genetic diversity and adaptation

Abstract

Euterpe edulis is a species of palm distributed throughout the Atlantic Forest and Cerrado in Brazil. These ecosystems function as corridors between two tropical rainforests that have undergone changes in biogeographic distribution. The phenotypic plasticity of *E. edulis* allows the occurrence in environments with different biogeographical landscapes. This plasticity is believed to be the result of natural selection and gene flow, which allowed *E. edulis* to adapt to diverse environmental conditions, including variations in temperature and precipitation. This study aims to understand evolutionary and historical processes that led to the genetic diversity and current distribution of the *Euterpe edulis*. For this, 261 individuals from 26 populations of this species collected throughout Brazil and genotyped by the DArTseq methodology. We then used 2227 neutral SNPs, 316 outliers SNPs and modeled the distribution of suitable environments over the past 130,000 years to describe the genetic diversity and structure of population of this species, (i) test if the adaptive selection influences this pattern, (ii) test for a correlation between genetic diversity with historical stability and geographic distance. Populations of *E. edulis* showed a high rate of inbreeding and expected heterozygosity. Our results indicate that natural selection related to environmental variables shaped the spatial pattern of genetic diversity in *E. edulis*. The selected SNPs showed the same population structure pattern as the neutral ones, geographically dividing the Atlantic Forest into north and south. Functional annotation of the SNPs under selection identified resulted in association with defense, stress and flowering genes. The results of the ecological modeling showed that the favorable areas for the occurrence of the species would retract in drier periods, with three populations (BDF, FPR and PSP) in a region of low suitability for the occurrence of the species. Furthermore, the contemporary distribution of *E. edulis* has been relatively stable in the southern and central

coastal region of Brazil. The genomic data recover a clear pattern of isolation by distance and resistance. This study contributed to understanding the population genetics and evolutionary history of *E. edulis* and may have important implications for its conservation and sustainable use.

Keywords: conservation; ecological modeling; Juçara; palm; rainforest.

7.1 INTRODUCTION

The Atlantic Forest of Brazil is renowned for its high levels of species richness and endemism, making it one of the regions of high biodiversity (MYERS et al., 2000). Higher species richness is frequently seen in the topographically diverse coastal regions of Rio de Janeiro and São Paulo (DASILVA; PINTO-DA-ROCHA; DESOUZA, 2015; VALE et al., 2018). The Atlantic Forest has been compartmentalized into five main areas of endemism, including Pernambuco, Coastal Bahia, Central Bahia, Serra do Mar, and Paraná/Araucaria (CARNAVAL et al., 2009; RULL; CARNAVAL, 2020). The spatial turnover of lineages within species largely mirrors that of species within communities, with both climate change and geographical barriers acting in combination to produce patterns of diversification (MERCIER et al., 2023).

E. edulis is one of the species that can be studied to explain the diversity pattern in this biome. This palm is distributed throughout the Atlantic Forest and part of the Cerrado in Brazil (MARTINELLI; MORAES, 2013; PEREIRA et al., 2022). These ecosystems function as corridors between two tropical rainforests that have undergone changes in biogeographic distribution, affecting the spatial distribution of the genetic diversity of several species (BACCI et al., 2021; MATOS-MARAVÍ et al., 2021). This species has phenotypic plasticity, which allows it to be present in areas with different biogeographies (BRANCALION et al., 2018). The Atlantic Forest biome has a great diversity of microhabitats, including different soil types, topographies and climates, which may have shaped the local adaptation of this species (DA SILVA; CASTELETI, 2003; MATOS-MARAVÍ et al., 2021).

This phenotypic plasticity is believed to be the result of natural selection and gene flow, which allowed *E. edulis* to adapt to diverse environmental conditions. For example, Brancalion et al. (2018) found that the species exhibited high levels of phenotypic plasticity in response to environmental variation, allowing it to thrive in a range of conditions. Additionally, they found evidence of local adaptation to different environmental conditions,

with populations in the Cerrado exhibiting different phenotypic traits than those in the Atlantic Forest. Furthermore, studies have demonstrated a morphological variation, which is believed to be due to adaptation to local environmental conditions and a result of defaunation (CARVALHO; et al., 2015; CARVALHO et al., 2016; GALETTI et al., 2013). Therefore, it is possible to find individuals with variations in the colors and sizes of the sheaths (COELHO et al., 2020).

In study of Pereira et al. (2022) the geographic division was clear, in north and south mentioned in other studies, in the Atlantic Forest of Brazil. The results showed high levels of genetic diversity within populations, but also significant genetic differentiation between populations. The spatial pattern of genetic differentiation indicates that limited gene flow between the more isolated populations contributed to the genetic structure of the species. However, the genetic structure and differentiation between populations indicate that historical events and current ecological factors influenced the spatial distribution of genetic diversity in this species.

Furthermore, the genetic diversity and structure of plant populations is shaped by many factors, including historical weather events. Studies suggest that historical climate changes and glacial cycles may have affected the expansion and contraction of this species distributions (CARVALHO et al., 2017). Carvalho et al. (2017) showed that areas with more stable historical climates had greater genetic diversity and larger populations of *E. edulis*. Furthermore, they showed that anthropogenic disturbances did not have a significant effect on genetic diversity or population size but did not result in changes in population structure.

Despite these important findings, much remains to be learned about *E. edulis* population genomics. Furthermore, although *E. edulis* is an economically and ecologically important species, much remains to be learned about the genetic basis of important traits (CONAB, 2019; LUGON et al., 2021). In this study, we used population genomics to investigate the genetic diversity and adaptation of this species. Specifically, we tested the hypothesis that adaptation selection and climate stability influenced the genetic diversity and structure of *E. edulis* populations. Selection by adaptation refers to the process by which natural selection acts on the genetic variation of a population, favoring individuals better adapted to their environment (GREGORY, 2009; NEELABH, 2017). Climatic stability refers to the long-term stability of the climate in a given area, which can promote the persistence of populations over time (OWENS; GURALNICK, 2019; WIERINGA; CARSTENS; GIBBS, 2021).

Our hypothesis is that in areas with more stable historical climates populations are less structured and genetic isolation signal by distance and in more unstable areas they present greater structuring and isolation by resistance. Furthermore, we tested whether adaptive selection influences the genetic structure of this species. Overall, our study will contribute to understanding the population genetics and evolutionary history of *E. edulis* and may have important implications for its conservation and sustainable use.

7.2 MATERIAL AND METHODS

7.2.1 Population sampling

We sampled 26 populations of *Euterpe edulis* which 18 populations published in Pereira et al. (2022), throughout the Atlantic Forest and Cerrado of Brazil (Fig. 1). Sampling ranged from five to 20 adult individuals (total 261 individuals). The complete list of material is provided in Supplementary Data S1. Adult individuals were Global Positioning System (GPS) mapped, and small leaf and stipe samples were collected for each individual and dried in individual bags with silica gel for DNA preservation. DNA extraction was carried out using the cetyltrimethylammonium bromide or CTAB method by Doyle and Doyle (1990) with modifications (FERREIRA; GRATTAPAGLIA, 1998). The purified DNA was then quantified in NanoDrop® (considering superior quality samples those that presented a 260/280 nm ratio between 1.8 and 2.0).

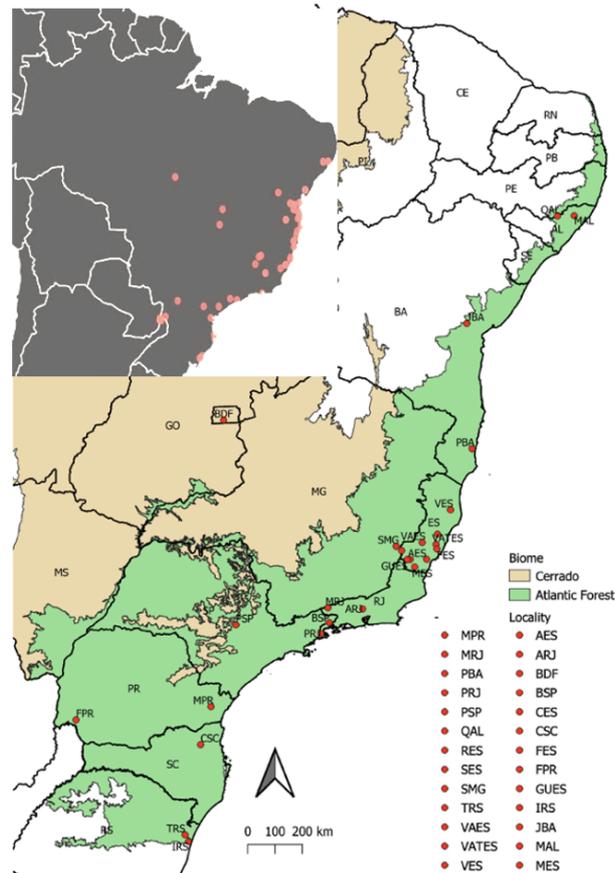


Figure 1: Geographical distribution of the 26 *Euterpe edulis* populations sampled in Brazil. Each location is represented by red dots and each biome by green (Atlantic Forest) and yellow (Cerrado) colors. The smaller graph represents the distribution of *Euterpe edulis* in Brazil.

7.2.2 SNP genotyping

The genomic DNA (50ng/μl) of the 261 samples was sent to the Genetic Analysis Service for Agriculture Laboratory (SAGA in Spanish, Texcoco – Mexico) for checking the quantity and quality. The complexity of the DNA was reduced using two restriction enzymes, *MseI* and *HpaII*. Each sample digestion product was connected to a barcode adapter, allowing for the identification of each sample, and an adapter specific to the Illumina technology. The fragments from each sample were then gathered into a single pool, which was amplified through PCR to enrich the fragments obtained (PETROLINI et al., 2019). After this step, the libraries were validated and quantified before being submitted to the Illumina Novaseq 6000 System sequencer, generating single-end reads of 83 bp each.

All sequencing reads were then evaluated for quality using FastQC software v.0.11.3 (ANDREWS, 2010), followed by trimming for adapters using Trimmomatic software v.0.32,

using the Phred 30, Leading 3, Trailing 3, Slidingwindow 4:30 and Minlen 50 options. (BOLGER; LOHSE; USADEL, 2014). We tested different pipelines for calling variants, and we chose to use the one that we were most familiar with and that would later allow us to associate the variant with a trait. The reads were then aligned to the genome sequence of *Elaeis guinnensis*, GenBank assembly accession GCF_000442705.1 (SINGH et al., 2013b) using BWA v. 0.5.9 (BENTLEY et al., 2008) with `bwa mem`, `-R`, and default settings. Functions from the Picard-Tools package (<http://broadinstitute.github.io/picard>) were used to manipulate the file in a proven way and prepare it for calling variants. The SAM (Sequence Alignment/Map) format files were converted into the BAM (Binary Alignment/Map) format using the `SortSam` function, which also sorts the alignment by coordinate. `MarkDuplicates` tool from Picard was used to mark possible PCR duplicates, which were not used. `Picard AddOrReplaceReadGroups` tool was used to insert identification data into alignment files. SNP calling was conducted using `Samtools v.1.17` (DANECEK et al., 2021), with `mpileup` algorithm, and `Freebayes v.1.3.7` (GARRISON; MARTH, 2012) using default settings.

We characterized genotype quality verifying the percentage of missing data, and the allele frequencies. `VCFtools` with options `--mac 3`, `--minQ 20` and `--max-missing 1` was used for the first filter. Then, a filter was applied using the `PLINK` (PURCELL et al., 2007) to remove variants with minor allele frequency < 0.01 and genotype with missing data > 0.15 . We also obtained the ratio of transition to transversion substitutions (Ts/Tv) and expected heterozygosity (NEI, 1987), using `VCFtools` version 0.1.12b (DANECEK et al., 2021). After separation and filtering, calculations were made genetic analyses.

7.2.3 Genome scans for selection footprints

Two different methods were used to identify candidate loci that are under selection. The first method was the `Fst-Heterozygosity` outlier approach (Flanagan & Jones, 2017) implemented in R software v. 4.2.1 (Core R Team, 2019). The second method was `BayPass v.2.2` (Gautier, 2015a), which was used to detect local selection by estimating linear correlations between allele frequencies and environmental variables while considering the relationships among populations. The environmental variables were selected from the `PaleoClim` (Brown et al., 2018), which is a high-resolution paleoclimate dataset for global land areas with a resolution of 2.5 minutes (~5km). Seven bioclimatic variables were selected using the variance inflation factor (VIF) (Chatterjee & Hadi, 2006) in R software v.4.2.1 (Core R Team, 2019) to avoid collinearity: mean diurnal range (bio2), temperature seasonality (bio4), mean temperature of driest quarter (bio9), annual precipitation (bio12),

precipitation seasonality (bio15), precipitation of warmest quarter (bio18) and precipitation of coldest quarter (bio19). These variables describe the mean and variation in temperature and precipitation over the last 130,000 years, and were sampled at 10 years ago, 0.3 ka, 4.2ka, 11.7ka, 12.9ka, 14.7ka, 21ka and 130ka.

To estimate the covariance matrix of allele frequencies among populations, BayPass v.2.2 was used with 316 SNPs separated by at least 10 kb across all populations for each species, with 100,000 steps through the chain. The program was run independently for each bioclimatic variable with 1,000,000 iterations. Following the instructions of the BayPass v.2.2 manual, SNPs were considered as candidates if they had high approximated Bayesian p values in the log10 scale ($eBF_{mc} > 3$).

After these analyses, we obtained two files, one with the loci under selection obtained by both approaches and one with only the neutral ones. To define the final neutral loci file, a linkage equilibrium analysis was performed. To do this, we utilized the "--indep-pairwise" function in PLINK v1.90b6.16 (PURCELL et al., 2007). This method involved setting a window size of 50 SNPs, shifting the window by 5 SNPs at each step, and using a threshold of 0.3 for the r^2 value. This process generated a subset of SNPs that are approximately in linkage equilibrium with one another, as determined by their pairwise genotypic correlation.

7.2.4 Demographic dynamics

To determine the impact of genetic drift caused by demographic changes, we employed the pairwise sequentially Markovian coalescent (PSMC) model. The PSMC model, developed by Li and Durbin in 2011, was implemented through the R package PSMCR available at <https://github.com/emmanuelparadis/psmcr>. Using the genome assembly of *Elaeis guinnensis* GCF_000442705.1 (SINGH et al., 2013b), as the reference genome, we ran PSMCR on each population with the following parameters: 30 iterations, 100 bootstrap replicates, a θ/ρ ratio of 0.018 based on highly heterozygous tropical forest trees estimated from genome-wide data (SILVA-JUNIOR; GRATTAPAGLIA, 2015), representing the ratio of mutation to recombination rate, a pattern of $4 + 5 \times 3 + 4$ for the number of free atomic time intervals, and a maxt value (the largest possible value for the time to the most recent common ancestor) of 10.n

7.2.5 Genetic diversity and population structure

The genetic diversity was analyzed for 2227 neutral SNPs, to characterize the genetic diversity analyses were performed for this species, implemented in the Arlequin 3.5 program (Excoffier & Lischer, 2010). The number of polymorphic sites expected heterozygosity under Hardy-Weinberg equilibrium (H_e) and observed heterozygosity (H_o) of the average population per locus, inbreeding coefficient (F_{is}), and nucleotide diversity (π) were calculated.

In addition, to analyze population structure, we performed a molecular variance analysis (AMOVA) using the program Arlequin 3.5 (EXCOFFIER; LISCHER, 2010) to estimate the genetic differentiation among populations from regions (North and South) identified by LEA package (F_{ct}) and among populations within regions (F_{sc}). Arlequin estimates genetic structure indices using information on the allelic content of haplotypes and their frequencies. We also performed hierarchical AMOVA using only SNP loci with selection signal (316 SNPs) to compare the relative effects of genetic drift and selection in genetic structure.

To analyze the population structure of the species, we performed simulations based on a Bayesian clustering method, implemented the Nonnegative Matrix Factorization (sNMF) method as implemented in the R package 'LEA' v2.8.0 (FRICHOT; FRANÇOIS, [s.d.]). We performed the analysis using single nucleotide polymorphism (SNP) data and ran sNMF with the number of genetic clusters ranging from 1 to 27, which is one more than the number of sampled localities. We varied the alpha parameter values to test the robustness of the results, setting it to 1, 10, and 100. The sNMF was run in five replicates with 2,000 iterations and default parameters. To determine the most appropriate number of genetic clusters, we used the minimum cross-entropy method across all runs.

7.2.6 Candidate SNPs and annotation analysis

The prediction of the effects of SNPs on the function of *E. edulis* genes, considering changes in amino acids, using was obtained with the SnpEff v.4.2 software (CINGOLANI et al., 2012) using *Elaeis guinnensis* as reference genome. If a single nucleotide polymorphism (SNP) was identified in both the coding region of one gene and upstream of another, only the annotation for its presence in the coding region was retained. The functional characterization of the impacted genes was then conducted by DAVID Bioinformatics Resources (SHERMAN et al., 2022).

7.2.7 Ecological niche models

To determine the length of time the palm species have occupied the landscape represented by grid cells, we needed to estimate over the longest time interval possible. We utilized occurrence records for each population and analyzed climatic projections from the past 130,000 years using correlative Species Distribution Models (SDMs), as described by Townsend Peterson et al. (2011). This approach allowed us to infer the geographic range of suitable climates for *Euterpe edulis*.

To create Species Distribution Models (SDMs) for each species, occurrence data and spatial environmental data were combined. The 'sdm' in R package was used to create a template for the SDM code, which was altered using instance data from samples we collected. Environmental data from the PaleoClim (BROWN et al., 2018) was used to describe past and present climate conditions over the last 130 ky, comprising seven bioclimatic variables at a spatial resolution of approximately 5 km: mean diurnal range (bio2); temperature seasonality (bio4); mean temperature of driest quarter (bio9); annual precipitation (bio12); precipitation seasonality (bio15); precipitation of warmest quarter (bio18); precipitation of coldest quarter (bio19). These variables describe the mean and variation in temperature and precipitation sampled in eight periods over the last 130kay: Anthropocene (1978-2013); Pleistocene: late-Holocene (4.2-0.3ka); Pleistocene: mid-Holocene (8.32-4.2ka); Pleistocene: younger dryas stadial (12.9-11.7ka); Pleistocene: Heinrich stadial (17.0-14.7ka); Pleistocene: Last Glacial Maximum (ca.21ka); Pleistocene: Last Interglacial (ca. 130ka).

To evaluate model performance, the occurrence data was split into four geographical sets, and the model was run four times, with each set withheld once for testing. The machine learning algorithm MaxEnt, as implemented in the R package "dismo" (HIJMANS, 2015) was used to fit a total of 40 models, varying the shape of the response curve and model complexity. Usually, presence-absence data is used for building and validating ecological niche models. However, due to the unavailability of reliable absences, we randomly generated pseudo-absences across the entire area while excluding cells with presences to ensure a complete environmental space sampling and minimal extrapolation when projecting to different historical periods. To optimize model performance, pseudo-absences were weighted to maintain a prevalence of 0.5. We evaluated model stability using random data partitioning (80% training, 20% testing) and repeated each model five times. The accuracy of the models was assessed using the True Skill Statistic–TSS and the area under the receiver curve–AUC. We generated stability maps based on the climatic projections, which indicate

the duration of continuous climatic suitability for each grid cell. The stability maps were created by tracing back from the present to the past, and grid cells with higher values represent higher stability compared to those with lower values.

7.2.8 Predictors variables

To evaluate the impact of geographical distance and environmental factors on population genetic structure, we assessed patterns of Isolation by Distance (IBD) and Isolation by Resistance (IBR) in *Euterpe edulis*. Isolation by distance (IBD) is a population genetics concept that describes the pattern of genetic differentiation among populations due to geographic distance (WRIGHT, 1943). The underlying idea is that as geographic distance increases, gene flow between populations decreases, leading to increased genetic differentiation among populations.

Isolation by resistance (IBR) is a concept used in landscape genetics to describe how landscape features (MCRAE, 2006). IBR models aim to measure the effect of landscape resistance on gene flow between populations, taking into account the spatial distribution of habitats and the physical characteristics of the landscape. IBR is calculated using different approaches, but one common method is the use of circuit theory algorithms, which simulate the movement of individuals across a landscape based on the resistance values assigned to each cell of a raster grid. Resistance values are assigned to each cell based on landscape features that are known or suspected to influence gene flow, such as distance, elevation, land cover type, and barriers to movement. By identifying regions of high resistance and low gene flow, IBR models can also help guide conservation efforts and inform the design of corridors or other landscape features that facilitate gene flow and maintain genetic diversity.

We calculated the genetic distance between each pair of sampled localities using Nei's distance in the 'adegenet' R package (JOMBART & AHMED, 2011). We then determined the geographical distance between each pair using the 'geosphere' R package (HIJMANS et al., 2016).

The resistance values between each pair of localities were estimated using an eight-neighbour cell scheme based on cost-distance matrix based on random paths in the 'raster' and 'gdistance' R software (ETTEN, 2017). This method was based on Mercier et al. (2023). For this we used the commute-time distance which is an alternative to effective resistance from circuit theory. Commute-time is the expected time it takes for a random walk between nodes and has been shown to correlate highly with effective resistance (MARROTTE and

BOWMAN, 2017). We then evaluated the ability of these resistance map to predict genetic distance by performing three distance matrix regressions using ‘ecodist’ (LEGENDRE et al., 1994). Since geographical distance and resistance are correlated, and geographical distance is expected to be the primary driver of genetic distance (GUILLOT & ROUSSET, 2013), we used the residuals of the regression of resistance against geographical distance as the predictor variable in our model. Finally, we compared the explanatory power (r^2) of resistance in the two regressions.

7.3 RESULTS

7.3.1 SNP genotyping

We obtained over 9 Gbp for the 261 individuals. A total of 9,658,427 reads were sequenced, with a mean of 34,994 reads per individual. The variant call identified 37036 SNPs, after the removal of sites with more than 20% missing data and filtering a total of 2899 on-target high-quality polymorphic SNPs with call rate $> 66.8\%$ was detected. The average per-sample call rate across all 261 individuals was 91.7%. *Euterpe edulis*. SNPs revealed a Ts/Tv of 1.38. The average of missing data per individual is 8,2% and per locus 8,3%. Diversity nucleotide, considering values per locus, had an average of $\pi = 0.227$.

7.3.2 Genome scans for selection footprints

We detected 178 loci with selection based on F_{st} -Heterozygosity outlier approach (Fig.2). BayPass detected 147 SNPs correlation with variables, being nine it was detected with previous approach (Supplementary Data S2 – S8). We found high eBF_{mc} values, ranging from 3.0 to 9.97. It was detected 15 outliers for the variable mean diurnal range (bio2); 15 for the temperature seasonality (bio4); 39 for the mean temperature of driest quarter (bio9); 43 for the annual precipitation (bio12); 13 for the precipitation seasonality (bio15); eight for the precipitation of warmest quarter (bio18); 38 for the precipitation of coldest quarter (bio19).

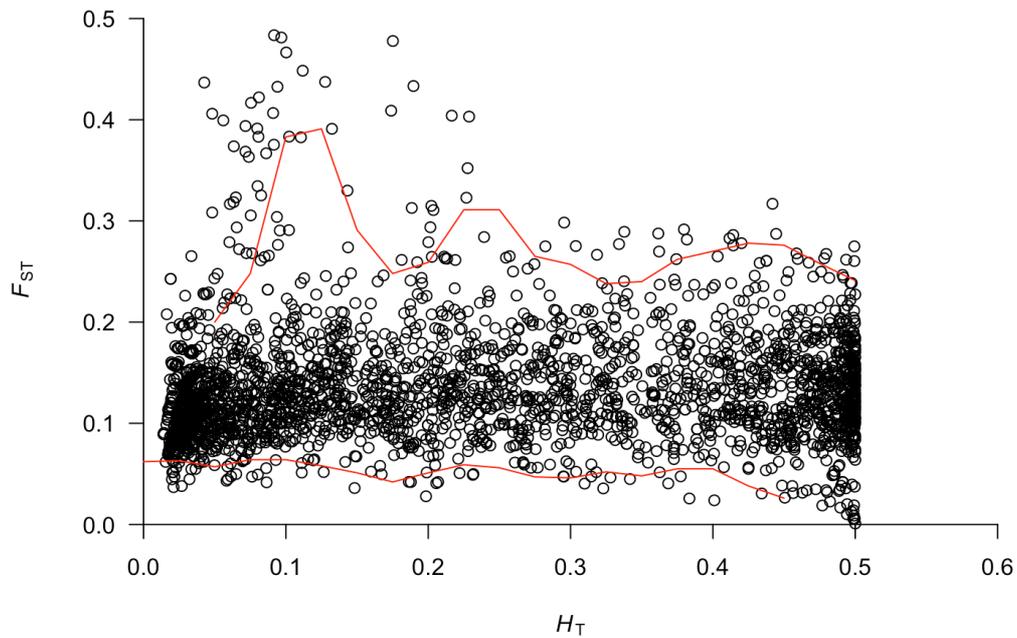


Figure 2: Locus detection under selection by Fst-heterozygosity outlier approach in *Euterpe edulis* based on 2899 SNPs.

7.3.3 Demographic dynamics

For a better visualization of the graphs, we divided the populations into two groups North (Fig.3) and South (Fig. 4). PSMC recovered the demographic dynamics from ~280 to ~2500 years and low value of effective population size (N). We can see that there was a decrease in the effective population size of edulis populations around 500 years.

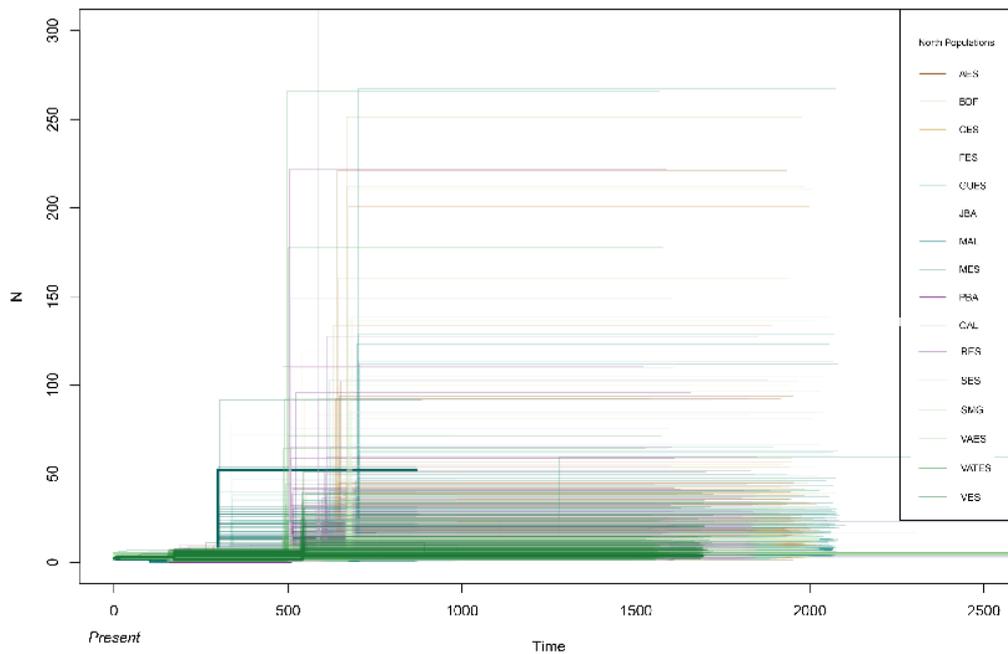


Figure 3: Demographic dynamics for populations of *Euterpe edulis* to the north using a Pairwise Sequentially Markovian Coalescent (PSMC) model and neutral Single Nucleotide Polymorphism (SNP) markers. Each colored line represents a population.

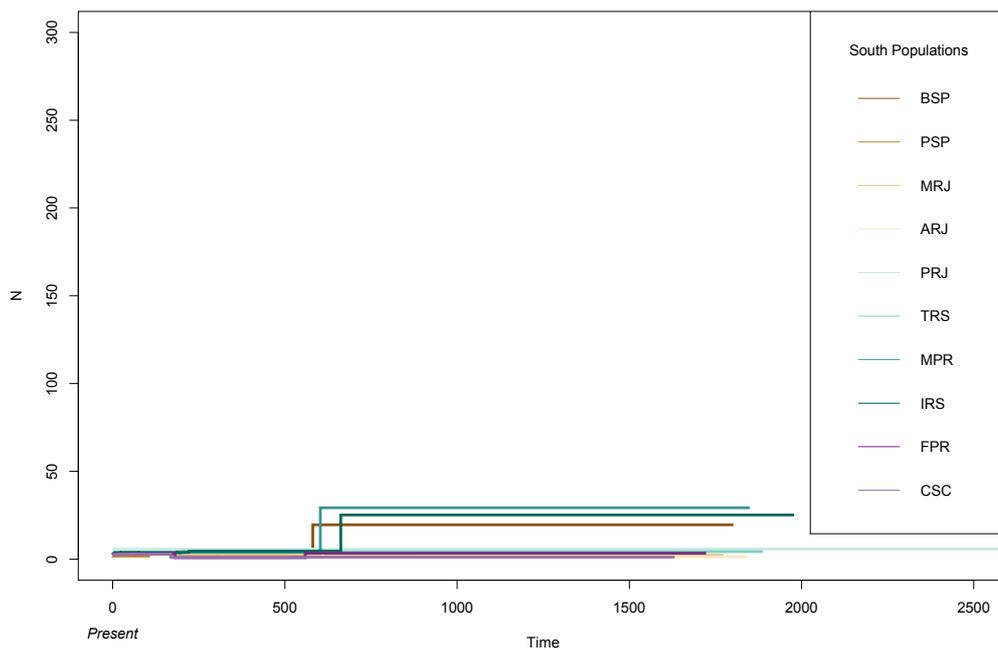


Figure 4: Demographic dynamics for populations of *Euterpe edulis* to the south using a Pairwise Sequentially Markovian Coalescent (PSMC) model and neutral Single Nucleotide Polymorphism (SNP) markers. Each colored line represents a population.

7.3.4 Genetic diversity and population structure analyses

The mean genetic diversity of the populations was high ($He = 0.369$). The QAL, MAL, JBA, PBA and SES populations showed lower nucleotide diversity ($\pi < 0.15$) compared to the others. The populations showed a very high and significant inbreeding rate ($Fis > 0.91$).

Table 1: Genetic diversity of *Euterpe edulis*, based on 2227 neutral loci in 26 populations. N number of individuals genotyped, N loci number of polymorphic loci, He mean expected heterozygosity under Hardy–Weinberg equilibrium, Ho mean observed heterozygosity, Fis inbreeding coefficient (*values significant, $p < 0.05$), SD standard deviation.

Populatio n	N	N loci	Ho	He	Fis	$\pi(SD)$
QAL	1	1346	0,03	0.37	0,916	0.134(0.066
	0		1	0	*)
MAL	1	1408	0,03	0.36	0,910	0.124(0.062
	0		3	8	*)
JBA	1	1576	0,03	0.36	0,896	0.126(0.063
	0		8	5	*)
PBA	1	1325	0,03	0.37	0,901	0.133(0.066
	0		7	2	*)
BDF	1	1107	0,03	0.35	0,905	0.158(0.079
	0		4	8	*)
AES	5	1725	0,03	0.43	0,929	0.184(0.978
			1	9	*)
GUES	5	1703	0,03	0.43	0,911	0.181(0.096
			9	8	*)
VAES	1	1303	0,03	0.36	0,913	0.186(0.092
	0		2	8	*)
VATES	5	1761	0,04	0.43	0,893	0.168(0.089
			7	8	*)
MES	1	1529	0,03	0.36	0,918	0.191(0.095
	0			5	*)
CES	1	1457	0,03	0.36	0,918	0.169(0.084
	0			5	*)
RES	2	731	0,02	0.33	0,922	0.213(0.103
	0		6	4	*)
FES	7	1477	0,03	0.39	0,924	0.169(0.086
				4	*)
SES	1	1282	0,02	0.34	0,933	0.147(0.072
	4		3	5	*)

VES	1	1491	0,03	0.36	0,894	0.165(0.082)
	0		9	8	*)
SMG	1	1325	0,03	0.36	0,911	0.177(0.088)
	0		3	9	*)
BSP	1	162	0,01	0.37	0,952	0.188(0.095)
	0		8	6	*)
PSP	1	1103	0,02	0.36	0,940	0.195(0.097)
	0		2	4	*)
MRJ	1	1369	0,02	0.36	0,925	0.183(0.091)
	0		7	1	*)
Populatio	N	N	Ho	He	Fis	π(SD)
n		loci				
ARJ	1	1273	0,02	0.36	0,924	0.186(0.093)
	0		8	8	*)
PRJ	1	1109	0,03	0.33	0,905	0.178(0.087)
	5		2	6	*)
TRS	1	849	0,01	0.30	0,938	0.246(0.123)
	0		9	8	*)
MPR	1	1356	0,02	0.36	0,920	0.168(0.084)
	0		9	3	*)
IRS	1	1397	0,03	0.35	0,911	0.188(0.093)
	0		2	9	*)
FPR	1	1379	0,03	0.35	0,916	0.178(0.089)
	0			9	*)
CSC	1	1132	0,02	0.36	0,939	0.189(0.094)
	0		2	1	*)

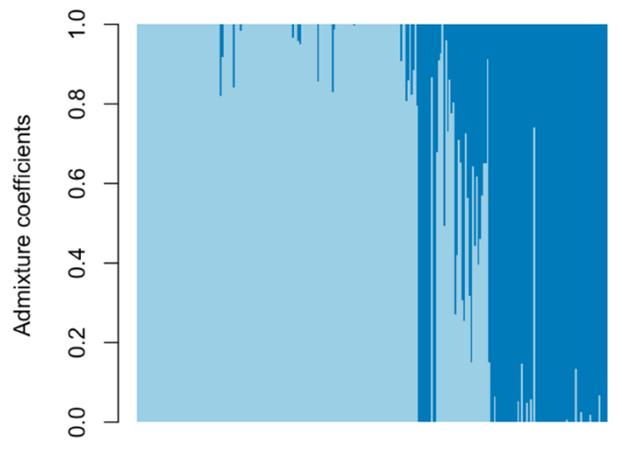
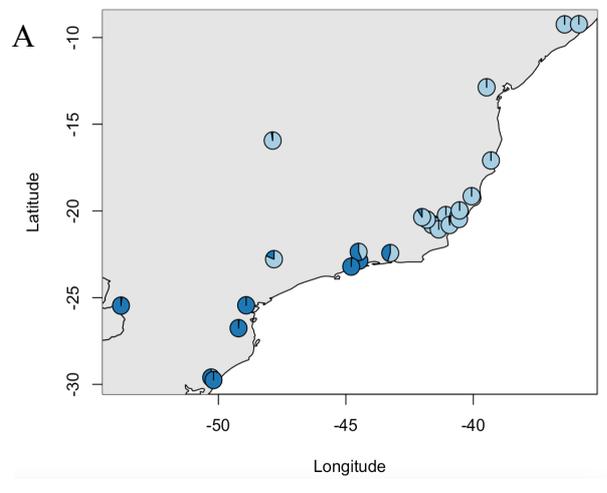
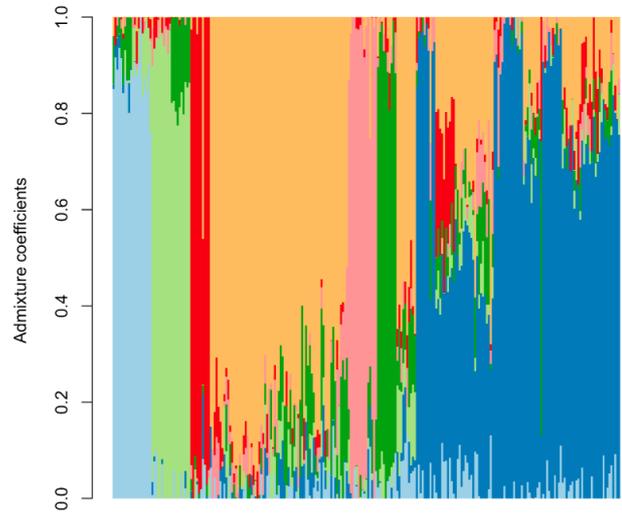
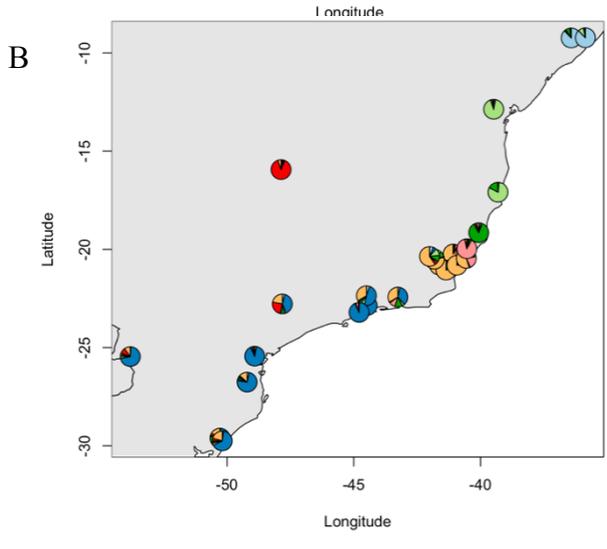
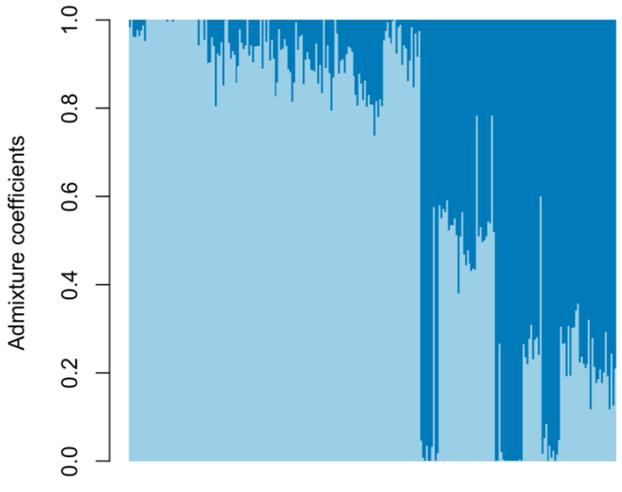
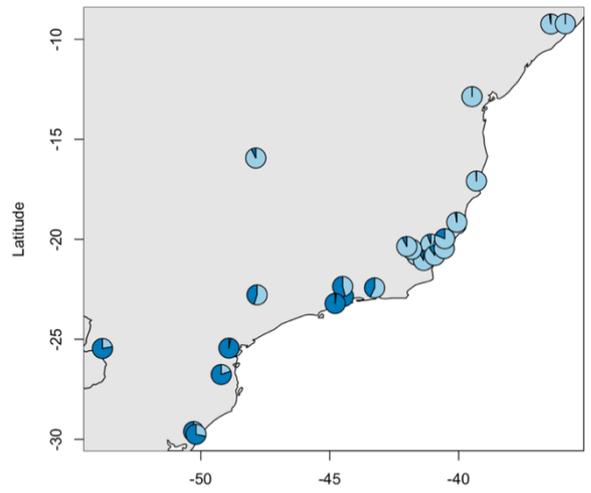
E. edulis showed moderated genetic differentiation among region, North and South ($F_{CT}=0.103$, $p<0.000$) and among populations within regions ($F_{SC} = 0.21$, $p<0.000$). When genetic differentiation was analyzed based on SNPs with selection signal, we found moderated differentiation among regions ($F_{CT} = 0.123$, $p=0.000$) than among populations within regions ($F_{SC} = 0.126$, $p<0.001$). The greatest diversity for both datasets was found among individuals within populations ($F_{ST} = 0,29$ for 2227 SNPs and $F_{ST} = 0.23$). The QAL, MAL, JBA, PBA, BDF and SES populations showed higher F_{ST} .

Table 2: Analysis of Molecular Variance (AMOVA) between regions (North and South) identified by LEA package, for 2227 neutral loci and 316 loci sob selection, using ARLEQUIN software.

Data	Groups	Source of variation	Degrees of freedom	Sum of squares	Variance components	% of variation	Fixation Indices	Pvalue
2227 loci neutral	North and South	Among regions	1	3955.130	14.34565	10.38	0.103	0.00000
		Among populations within regions	24	14828.674	26.08783	18.87	0.210	0.00000
		Among individuals within populations	496	48521.420	97.82544	70.76	0.292	0.00000
316 outliers	North and South	Among regions	1	732.704	2.87281	12.31	0.123	0.00000
		Among populations within regions	24	1662.790	2.57855	11.05	0.126	0.00000
		Among individuals within populations	496	8869.430	17.88191	76.64	0.233	0.00000

The population structure analysis revealed a statistically significant K value of seven for both the neutral and outlier datasets (Supplementary Data S9). Subsequent analysis of the other K values indicated high levels of mixing within individuals, with greater similarity observed among individuals that were geographically closer. The isolated populations exhibited a high frequency of a single genetic cluster (Fig.5 and Fig.6). Outliers exhibited a genetic structure similar to the neutrals. Analysis of K values greater than four revealed a distinct genetic cluster specific to the SES population in Espírito Santo.

A



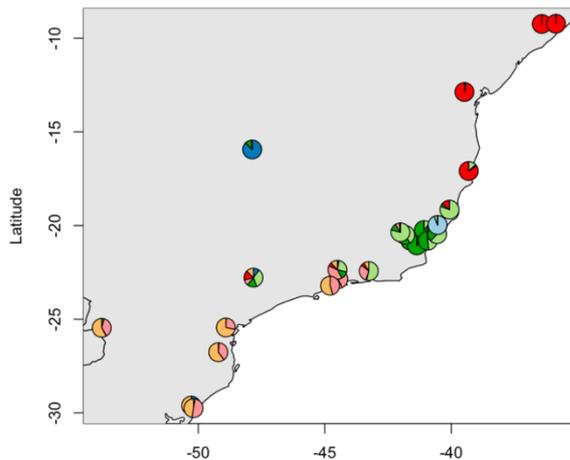


Figure 6: Population structure analysis obtained by the LEA package for *Euterpe edulis* based on 316 outliers SNPs and 26 populations. On the right, a graph showing the contribution of each genetic cluster in each population distributed on the map. On the left is a graph showing this contribution in barplot, with populations from the north, southeast and south of Brazil followed from left to right. (A) K=2. (B) K= 7. The different colors represent the different genetic groups. The divisions of each circle represent the contribution of

7.3.5 Functional annotation of SNPs with selection signal

The average number of SNPs per chromosome was 4, ranging from 1 to 25. Regarding the polymorphism types, transition (Ts) was the most abundant 16,699 events, (58,82%), being most frequently cytosine to thymine, followed by transversions (Tv) with 11,682 events (41,18%). The ratio of Ts/Tv was 1.42. A total of 1655 functional effects for SNP variants were predicted for 316 SNPs, providing information on the location. The predicted effects by impact were of modifier type (83,32%), low impact (14,26%), moderate impact (2.17%), and high impact (0.24%). The percent of effects by region was 15.89% in exon, 16.01% in introns, 1.81% in UTR region, 1.93% in splicing sites and 35,04% in transcript, of which 222 had GO identified (Supplementary Data S10). These genes were related to a variety of mechanisms, such as plant development and multiple stress response pathways.

We found genes potentially affected by 316 SNPs that were classified with the gene ontology (GO) category related to defense response (putative disease resistance protein RGA3), stress factors (AP2-like ethylene-responsive transcription factor At2g41710) and

genes involved in floral development (zinc finger protein STAMENLESS 1 and protein GIGANTEA) (Fig. 7).

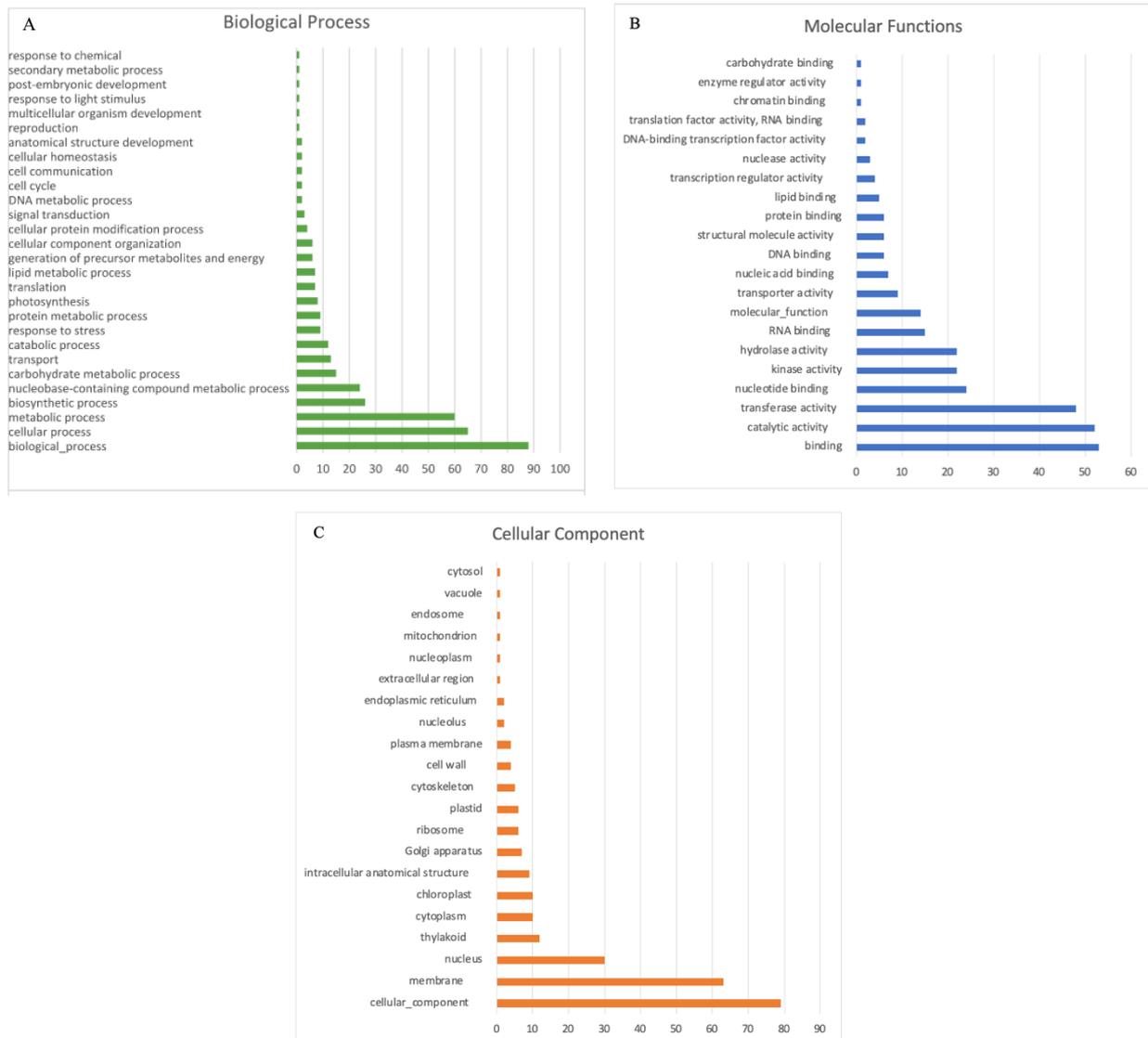


Figure 7: List of gene classes that had their gene ontology (GO) annotated related to the 316 SNPs outliers. (A) biological processes, (B) molecular function and (C) cellular components..

7.3.6 Climatic suitability and resistance

The accuracy and over-fitting of the species distribution models were high, with the best model having a test AUC of 0.89, and a 10% omission rate of 0.06 and 0.11 respectively. The model for *E. edulis* was based on 26 occurrence points. The model contained linear and quadratic response curves, and all environmental variables were retained. The models predicted the suitable habitat areas for both species, and when projected back in time, they

suggested the historical climate suitability of the species and the areas of high stability over the last 130 kay.

The environmental variables that most interfered in the model were mean diurnal range (bio2) and mean temperature of driest quarter (bio9) (Supplementary Data S11). The model showed that the greatest aptitude for *Euterpe edulis* is found in coastal regions of Brazil. This area is not continuous, there are three breaks: one between Alagoas and Bahia, another between Bahia and Espírito Santo and a third between Espírito Santo and Rio de Janeiro (Fig. 8).

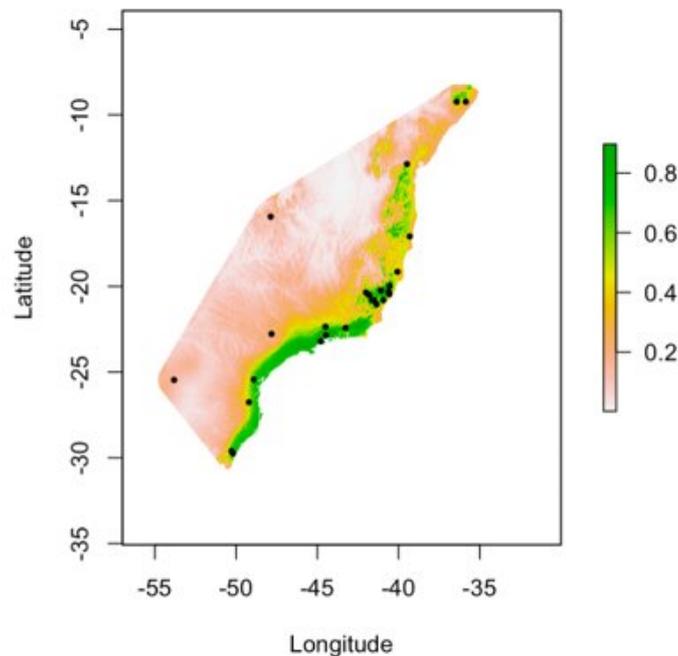


Figure 8: Continuous projection of the best SDM for *Euterpe edulis*. Green indicates the maximum suitability of 1 to yellow moderate suitability 0.5 to white minimum suitability 0 (zero).

The models suggest that over the past 130 Kay there has been relatively high stability extending from Rio Grande do Sul to Rio de Janeiro. At late-Holocene 0.3ka, this area formed a suitable corridor for the occurrence of *E. edulis*, connecting with Espírito Santo and southern Bahia. In addition, in the northern region of Espírito Santo and southern Bahia there was a moderate to high retraction through time. The region from northern Bahia to Alagoas during the Last Glacial maximum and now in the Anthropocene showed low stability for species suitability.

The geographical distance was a significant predictor of genetic distance ($r^2 = 0.34$ and $p < 0.001$). The resistance distance was also a significant predictor of genetic distance (r^2

= 0.24 and $p < 0.001$). However, model that included stability informed by time and geographical distance had explanatory power equal to $r^2 = 0.34$, although the residuals of resistance were not significant ($p = 0.92$). These findings suggest that the predominant pattern in this species is Isolation by Distance (IBD), with minimal to no Isolation by Resistance (IBR) effect.

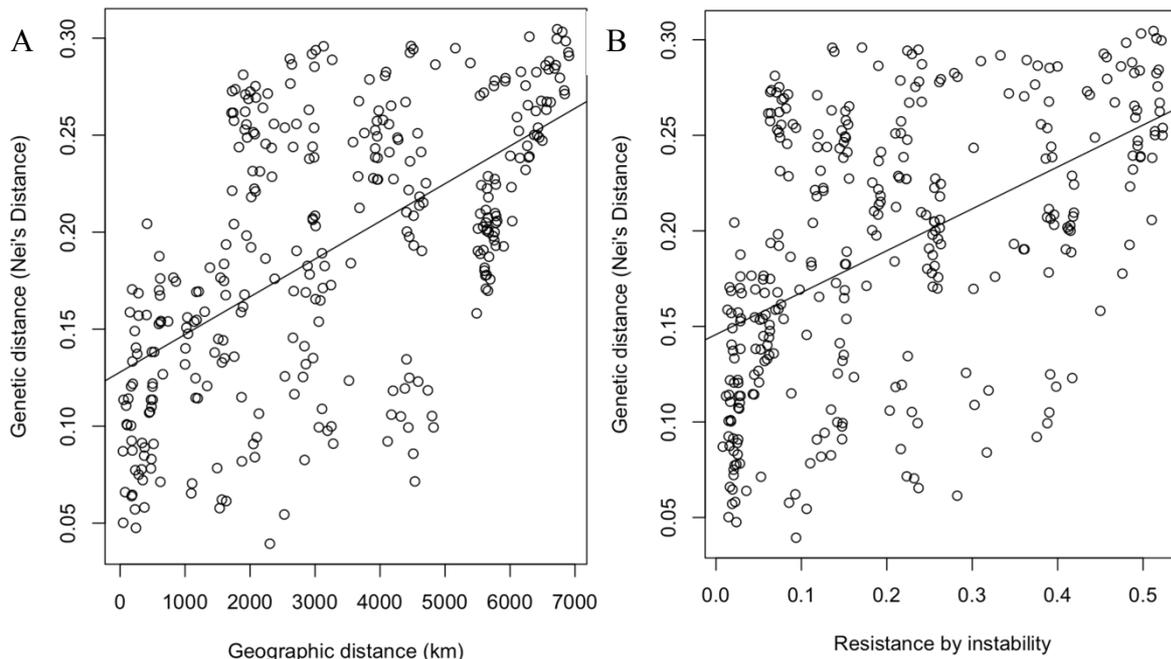


Figure 9: Correlation graph of genetic diversity with geographic distance (A) and climate stability (B).

7.4 DISCUSSION

We showed that in addition to genetic drift, another evolutionary force acting on population structure and genetic diversity in *E. edulis* is natural selection. Reinforced our expectations (ii) by the high presence of SNPs related to environmental variables. We found a significant genetic differentiation among populations and between regions, north and south, that geographically divide the Atlantic Forest both for the neutral loci and for the loci under selection. These groups were defined by analysis of population structure and have been evidenced before for the species under study and others (PEREIRA et al., 2022).

The more isolated populations QAL, MAL, JBA, PBA and BDF showed greater structure and less nucleotide diversity. This can be attributed to reduced gene flow and population connectivity caused by isolation, which limits dispersal between populations (WRIGHT, 1943). Isolation can upset the migration-drift balance by impeding the rescue of

alleles lost by random genetic drift through gene flow (CARVALHO; et al., 2015). In addition to these populations, a specific population in Espírito Santo also showed high structure, presenting an exclusive gene cluster. Individuals in this population have morphological characteristics that are distinct from other individuals of the *E. edulis* species. Its fruits are larger, the sheath yellow, flowering reddish, among others (WENDT et al., 2011). We consider this population as another morphotype that has already been described in the literature with *Euterpe espiritosantensis*, being synonymous with *Euterpe edulis* (COELHO et al., 2020).

Our results contrast with the previous chapter, recovering high values for expected heterozygosity (He) and inbreeding coefficients (Fis). This may have something to do with the variant calling methodology. The use of a phylogenetically distant reference genome can influence the value of heterozygosity. Furthermore, in a reference genome alignment, the reading frames are aligned to a specific region in the genome capturing variation for that location (SCHILBERT; REMPEL; PUCKER, 2020). When calling a *de novo* variant there is no specific region, there is a variation within all reads (LANTICAN et al., 2019).

Despite the high values found, other studies using microsatellites markers report the presence of inbreeding with a high expected heterozygosity value in populations of *Euterpe edulis* (CARVALHO et al., 2017). Pereira et al. (2022) using microsatellites also found high inbreeding ($Fis = 0.12$) in all populations and high expected heterozygosity ($He = 0.76$) in their study. Noting that 18 of our sampled populations come from this work. Assortative mating, possibly due to short-distance pollination, spatial distribution of plants, fine-scale spatial genetic structure, or flowering patterns, could be a reason for the observed high expected heterozygosity (He) despite high inbreeding coefficients (Fis) in some *E. edulis* populations (GHAZOUL, 2005).

Genetic diversity can also be influenced by the demographic history of the species, the dynamics of effective population size over time. Our work shows that there was a decrease in the effective population size around 500 years ago. This may be related to the period of colonization of Brazil, and consequently consumption of Juçara palm heart and deforestation. Juçara was intensely exploited for the extraction of heart of palm for commercialization until 1998 when it became illegal (REIS et al., 2000; SCHULZ et al., 2016). Some studies show that forest fragmentation and defaunation have a significant effect on the genetic diversity of *Euterpe edulis*. Furthermore, they found that fragmented populations exhibited microevolutionary changes, indicating an adaptation to fragmentation

conditions, such as lower reproductive effectiveness and increased self-pollination (CARVALHO et al., 2016; SOARES et al., 2019).

Carvalho et al. (2015) noted that effective population size affects genetic diversity. Expected heterozygosity (H_e) and allelic richness (AR) was best explained by the model that contained historical effective population size alone, with sites with higher historical effective population size showing higher genetic diversity. Overall, these factors interact to shape the genetic diversity and population structure of *E. edulis*, highlighting the importance of considering multiple factors in conservation efforts.

Evidence of selection were detected in genes related to flowering

The study identified selection footprints in *E. edulis*, indicating natural selection may cause distinct signatures across different niches. The allele frequency of most SNPs with selection footprints showed high differentiation between the northern and southern Atlantic Forest regions. Interestingly, the spatial pattern in allele frequency is highly similar to the distribution of genetic clusters of neutral loci, possibly due to historical connectivity between populations and the spread of favorable alleles (VITORINO et al., 2016).

Several SNPs with selection signals were observed in genes associated with both constitutive and adaptive traits. Some of the selected SNPs were identified in housekeeping genes responsible for fundamental cellular processes such as DNA and RNA replication, regulation, and protein processing. Additionally, other SNPs were found to be associated with adaptation to biotic and abiotic stress, including genes such as those that encode proteins that respond to cellular dehydration and signal transduction kinases.

For example, we identified selection signals in the disease resistance protein RGA2-like and RGA3 (GO:0006952) that guard the plant against pathogens that contain an appropriate avirulence protein via a direct or indirect interaction with this a virulence protein (CASTRO; MARTIN, 2018). We also detected the proteins involved in stress factor, AP2-like ethylene-responsive transcription factor At2g41710 (GO:0003677) acts as a transcriptional activator (OHTO et al., 2005). Binds to the GCC-box pathogenesis-related promoter element. May be involved in the regulation of gene expression by stress factors and by components of stress signal transduction pathways. We also detected the proteins involved in floral development (zinc finger protein STAMENLESS 1 and protein GIGANTEA). Zinc finger protein STAMENLESS 1 (GO:0048437) regulates floral organ identity and cell proliferation in the inner floral whorls (XIAO et al., 2009). Probably specifies the identities

of lodicule and stamen through positive regulation of MADS16 expression. May contribute to morphogenesis by suppressing OSH1 expression in the lateral organs. Protein GIGANTEA (GO:2000028) is involved in regulation of circadian rhythm and photoperiodic flowering (BRANDOLI et al., 2020). Is involved in phytochrome B signaling.

There is evidence that floral development genes may be related to seed size in plants. This is because seed production is an important part of the reproductive cycle of plants and involves many developmental processes, including the formation of fruit and seed structures (OHTO et al., 2005). Genes involved in the regulation of floral development may indirectly influence seed size by affecting the amount and type of nutrients that are allocated to the seed during fruit development. Furthermore, studies have also shown that the expression of specific floral development genes is correlated with the production of seeds of different sizes in some plant species (OHTO et al., 2005).

This is an interesting fact to be highlighted, as some studies show that defaunation influences the size of the *E. edulis* seed (CARVALHO et al., 2021, 2016). Galetti et al. (2013a) found that the extinction of large-bodied seed dispersers led to a reduction in seed size, which was driven by natural selection for smaller seeds that were more likely to be dispersed by remaining small birds and mammals. The study also found that this rapid evolutionary change had consequences for the ecological role of the palm, as smaller seeds were associated with reduced seedling survival and recruitment. This leads to the suggestion that a selection of alleles of these genes related to floral development may be taking place.

Influence of climate stability on the genetic structure of *Euterpe edulis* populations

We showed a retraction in areas suitable for *E. edulis* over the last 130ka years and that the current pattern of genetic diversity of *E. edulis* is a result of the geographic distance between populations and little related to resistance isolation (ii). According to our results, the Atlantic Forest central corridor acted as a stable center and played a crucial role as a climate refuge for multiple species during the Pleistocene era (CARNAVAL et al., 2009; MERCIER et al., 2023). Our findings reveal that populations of *E. edulis* were able to survive in various parts of this stable corridor despite the fluctuating climate cycles of the Pleistocene, as indicated by evidence of isolation by distance and low population structure in this continuous stable region. Similar to the findings of Carvalho et al. (2017), our ecological niche models demonstrate that the current distribution of *E. edulis* is narrower than it was during the middle Holocene and Last Glacial Maximum (LGM). Nonetheless, some areas along the Brazilian

coast have remained climatically stable over time, making them potentially refuges for *E. edulis*.

We found three breaks in the population structure of *E. edulis*, which may be related to Pleistocene refuge areas already identified in other works (CARNAVAL et al., 2009; MERCIER et al., 2023). The first one between Espírito Santo and Rio de Janeiro. The authors suggest that this break corresponds to a low altitude region in Serra do Mar, Vale do Rio Paraíba do Sul (BACCI et al., 2021). Two other breaks were also found, one between the north of Espírito Santo and the south of Bahia and the third between the north of Bahia and Alagoas. This may be related to Serra dos Orgãos and the periodic events of sea level regression and transgression during the Quaternary, respectively (LEITE et al., 2016).

The results of insulation by resistance and by distance corresponded to what was found by Mercier et al. 2023. Stable climates promote the persistence of populations for longer periods of time, allowing for the accumulation of genetic diversity and high levels of gene flow (Ashcroft, 2010). Areas of low stability, on the other hand, hinder gene flow within a variety of species, leading to genetic divergence (Leffler et al., 2012). Genetic isolation by distance is observed in areas where continuous gene flow allows local populations to reach equilibrium (WRIGHT, 1943). In contrast, populations separated by regions of low stability generally show high levels of population structure and a signature of genetic isolation by environmental resistance, as noticed for the populations QAL, MAL, JBA, PBA, BDF.

Although we have focused here on linking climate stability and geographic distance to population structure, we recognize that there are likely additional factors that determine population genetic patterns. Previous studies have provided other evidence, such as topographic relief, watersheds, and fragmentations along the Atlantic Forest, as barriers to gene flow (CARDOSO et al., 2000; CARVALHO et al., 2016; EDUARDO SÍCOLI SEOANE et al., 2005).

7.5 CONCLUSION

The selected SNPs showed the same population structure pattern as the neutral ones, geographically dividing the Atlantic Forest into north and south. Functional annotation of the SNPs under selection identified resulted in association with defense, stress and flowering genes. The results of the ecological modeling showed that the environmental variables that most interfered in the model were mean diurnal range (bio2) and mean temperature of driest quarter (bio9). In addition, the genomic data recover a clear pattern of isolation by distance

and resistance. This study contributed to understanding the population genetics and evolutionary history of *E. edulis* and may have important implications for its conservation and sustainable use.

Acknowledge

Financial support to this study was obtained from the Conselho Nacional de Pesquisa (CNPq, Brazil) (Researcher productivity fellowship AF; MFSF #311840/2019-1), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil)—Finance Code 001, and from the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES, Vitória—ES, Brazil) in partnership with VALE and the Swedish Research Council (2017-04980).

7.6 REFERENCES

- ANDREWS, S. **FastQC A Quality Control tool for High Throughput Sequence**. Disponível em: <<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>. Acesso em: 20 maio. 2020.
- BACCI, L. F. et al. Biogeographic breaks in the Atlantic Forest: evidence for Oligocene/Miocene diversification in *Bertolonia* (Melastomataceae). **Botanical Journal of the Linnean Society**, p. 1–16, 2021.
- BENTLEY, D. R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. **Nature**, v. 456, n. 7218, p. 53–59, nov. 2008.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. **Bioinformatics (Oxford, England)**, v. 30, n. 15, p. 2114–2120, 2014.
- BRANCALION, P. H. S. et al. Phenotypic plasticity and local adaptation favor range expansion of a Neotropical palm. **Ecology and Evolution**, v. 8, n. 15, p. 7462–7475, 3 ago. 2018.
- BRANDOLI, C. et al. Gigantea: Uncovering new functions in flower development. **Genes**, v. 11, n. 10, p. 1–15, 2020.
- CARDOSO, S. R. S. et al. Genetic differentiation of *Euterpe edulis* Mart. populations estimated by AFLP analysis. **Molecular Ecology**, v. 9, n. 11, p. 1753–1760, 2000.
- CARNAVAL, A. C. et al. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. **Science**, v. 323, n. 5915, p. 785–789, 2009.
- CARVALHO, C. DA S. et al. Contemporary and historic factors influence differently

genetic differentiation and diversity in a tropical palm. **Heredity**, v. 115, n. 3, p. 216–224, set. 2015.

CARVALHO, C. DA 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.

CARVALHO, C. DA S. et al. Extant fruit-eating birds promote genetically diverse seed rain, but disperse to fewer sites in defaunated tropical forests. **Journal of Ecology**, v. 109, n. 2, p. 1055–1067, 1 fev. 2021.

CARVALHO, C. S. et al. Defaunation leads to microevolutionary changes in a tropical palm. **Scientific Reports**, v. 6, p. 1–9, 2016.

CASTRO, D. G.; MARTIN, S. G. Differential GAP requirement for Cdc42-GTP polarization during proliferation and sexual reproduction. **Journal of Cell Biology**, v. 217, n. 12, p. 4215–4229, 2018.

COELHO, G. M. et al. Genetic structure among morphotypes of the endangered Brazilian palm *Euterpe edulis* Mart (Arecaceae). **Ecology and Evolution**, v. 10, n. 12, p. 6039–6048, 2020.

DANECEK, P. et al. Twelve years of SAMtools and BCFtools. **GigaScience**, v. 10, n. 2, 1 fev. 2021.

DOYLE, J. CTAB Total DNA Isolation. **Molecular Techniques in Taxonomy**, v. 12, n. 27, p. 13–15, 1990.

EDUARDO SÍCOLI SEOANE, C. et al. **EFEITOS DA FRAGMENTAÇÃO FLORESTAL SOBRE A IMIGRAÇÃO DE SEMENTES E A ESTRUTURA GENÉTICA TEMPORAL DE POPULAÇÕES DE *Euterpe edulis* Mart. 1.** [s.l: s.n.].

FERREIRA, M. E. ; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética.** 3. ed. Brasília: CENARGEM, 1998.

GALETTI, M. et al. Functional extinction of birds drives rapid evolutionary changes in seed size. **Science**, v. 340, n. 6136, p. 1086–1090, 2013.

GHAZOUL, J. Pollen and seed dispersal among dispersed plants. **Biological Reviews of the Cambridge Philosophical Society**, v. 80, n. 3, p. 413–443, 2005.

LANTICAN, D. V et al. De Novo Genome Sequence Assembly of Dwarf Coconut (*Cocos nucifera* L . ‘ Catigan Green Dwarf ’) Provides Insights into Genomic Variation Between Coconut Types and Related Palm Species. **Genome Report**, v. 9, n. August, p. 2377–2393, 2019.

LEITE, Y. L. R. et al. Neotropical forest expansion during the last glacial period challenges

refuge hypothesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 113, n. 4, p. 1008–1013, 2016.

MERCIER, K. P. et al. Linking environmental stability with genetic diversity and population structure in two Atlantic Forest palm trees. **Journal of Biogeography**, v. 50, n. 1, p. 197–208, 1 jan. 2023.

OHTO, M. A. et al. Control of seed mass by APETALA2. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 8, p. 3123–3128, 2005.

PEREIRA, A. G. et al. Patterns of genetic diversity and structure of a threatened palm species (*Euterpe edulis* Arecaceae) from the Brazilian Atlantic Forest. **Heredity**, 2022.

REIS, M. S. DOS et al. Management and Conservation of Natural Populations in Atlantic Rain Forest : The Case Study Management and Conservation of Natural Populations in Atlantic Rain Forest : The Case Study of Palm Heart (*Euterpe edulis* Martius) 1. **Biotropica**, v. 32, n. 4, p. 894–902, 2000.

SCHILBERT, H. M.; REMPEL, A.; PUCKER, B. Comparison of read mapping and variant calling tools for the analysis of plant NGS data. **Plants**, v. 9, n. 4, 1 abr. 2020.

SCHULZ, M. et al. Juçara fruit (*Euterpe edulis* Mart.): Sustainable exploitation of a source of bioactive compounds. **FRIN**, v. 89, p. 14–26, 2016.

SILVA-JUNIOR, O. B.; GRATTAPAGLIA, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. **New Phytologist**, v. 208, n. 3, p. 830–845, 1 nov. 2015.

SOARES, L. A. S. S. et al. Anthropogenic Disturbances Eroding the Genetic Diversity of a Threatened Palm Tree: A Multiscale Approach. **Frontiers in Genetics**, v. 10, 7 nov. 2019.

VITORINO, L. C. et al. Demographical history and palaeodistribution modelling show range shift towards Amazon Basin for a Neotropical tree species in the LGM. **BMC Evolutionary Biology**, v. 16, n. 1, p. 1–15, 2016.

WENDT, T. et al. An evaluation of the species boundaries of two putative taxonomic entities of *Euterpe* (Arecaceae) based on reproductive and morphological features. **Flora: Morphology, Distribution, Functional Ecology of Plants**, v. 206, n. 2, p. 144–150, fev. 2011.

WRIGHT, S. ISOLATION BY DISTANCE. **Genetics**, v. 28, n. 2, p. 114–138, 29 mar. 1943.

XIAO, H. et al. STAMENLESS 1, encoding a single C2H2 zinc finger protein, regulates floral organ identity in rice. **Plant Journal**, v. 59, n. 5, p. 789–801, 2009.

7.7 SUPPLEMENTARY DATA

Table S1: Information about local sample.

Spot	Collection number	Local (code)	Local	Locality	n	Latitude	Longitude
1	Private propriety	QAL	Pedra Talhada	Quebrangulo, AL	10	-9.244635	-36.419087
2	S9344-4	MAL	Estação ecológica de Murici	Murici, AL	10	-9.226930	-35.857060
3	ONG Gambá	JBA	Serra da Jiboia Monte Cruzeiro	Elísio Medrado, BA	10	-12.872034	-39.481399
4	SISBIO-4	PBA	Parque Nacional do Descobrimento	Prado, BA	10	-17.093301	-39.310054
5	IBGE	BDF	-	Brasília, DF	10	-15.948.141	-47.878.507
6	Private propriety	AES	-	Alegre, ES	5	-20.807.722	-41.515.805
7	Private propriety	GUES	-	Guaçuí, ES	5	-20.808.555	-41.623.972
8	Private propriety	VAES	-	Vargem Alta, ES	10	-2.024.206	-41.085.19
9	Private propriety	VATES	-	Vargem Alta, ES	5	-20.312	-40.593
10	Private propriety	MES	-	Mimoso, ES	10	-21.062.313	-41.363.282
11	SISBIO-4	CES	Caparaó, Chalé dos Faria	Pedra Menina, ES	10	-20.503511	-41.818960
12	Private propriety	RES	Bonalotti	Rio Novo, ES	20	-20.807.598	-40.934.519
13	INMA	FES	Parque Fundão	Fundão, ES	7	-20.463.05	-40.564.93
14	INMA	SES	Reserva Santa Luzia	Santa Tereza, ES	14	-19.965.455	-40.540.530
15	VALE	VES	Reserva Natural da Vale	Linhares, ES	10	-19.151310	-40.070845
16	Private propriety	SMG	-	Manhumirim, MG	10	-20.356.290	-42.011.789
17	Private propriety	BSP	Pousada Brejal	Bananal, SP	10	-22.858.592	-44.471370
18	COTEC (080-833-2017)	PSP	Estação Ecológica de Ibicatu	Piracicaba, SP	10	-22.782675	-47.821360
19	Private propriety	MRJ	Capelinha	Visconde de Mauá, RJ	10	-22.450029	-44.499745

20	INEA (056-2017)	ARJ	Reserva Biológica Estadual de Araras	Araras, RJ	10	-22.434.722	-43.257223
21	Private propriety	PRJ	Engenho D'Ouro	Paraty, RJ	15	-23.213713	-44.793159
Spot	Collection number	Local (code)	Local	Locality	n	Latitude	Longitude
22	SEMA (07/18)	TRS	Parque Estadual do Itapeva	Torres, RS	10	-29.613333	-50.282784
23	IAP(47.17)	MPR	Parque Estadual Pico do Marumbi	Morretes, PR	10	-25.445412	-48.916382
24	SEMA(07/18)	IRS	Reserva Biológica da Mata Paludosa	Itati, RS	10	-29.850833	-50.189184
25	SISBIO-4	FPR	Parque Nacional do Iguaçu	Foz do Iguaçu, PR	10	-25.461766	-53.818300
26	Private propriety	CSC	-	Rio dos Cedros, SC	10	-26.678.569	-49.406.123

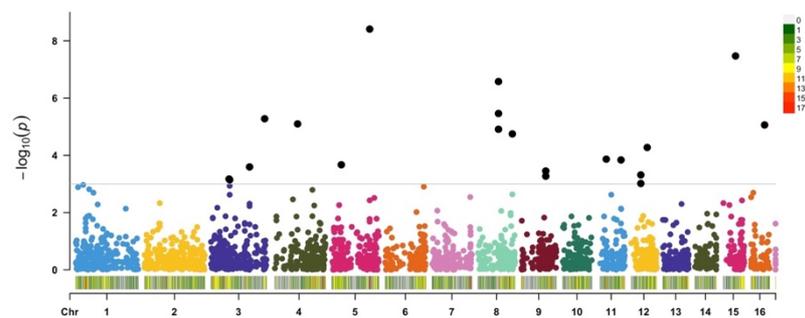


Figure S2: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio2.

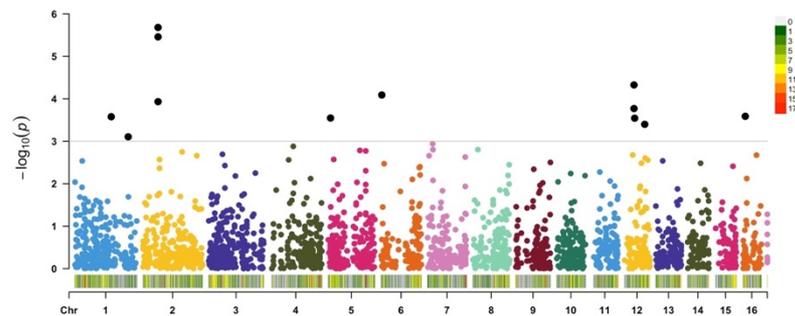


Figure S3: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio4.

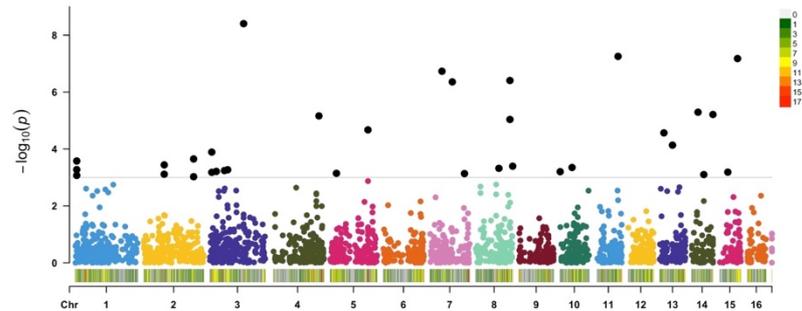


Figure S4: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio9.

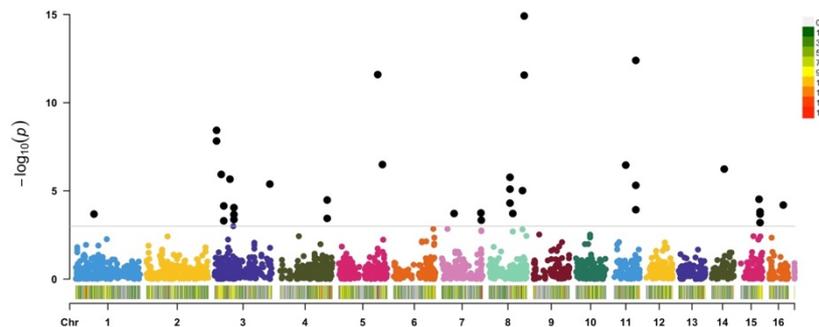


Figure S5: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio12.

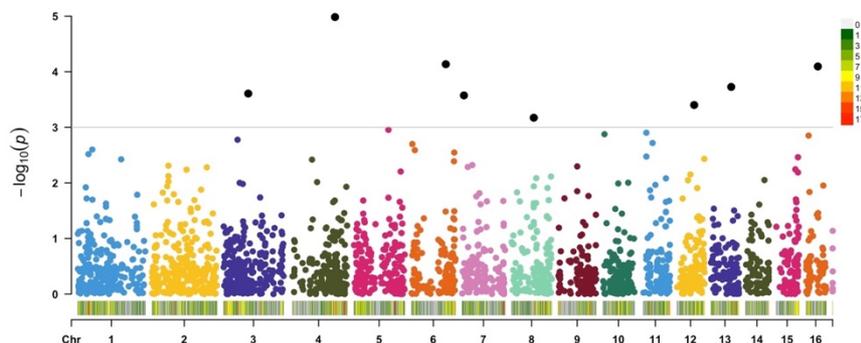


Figure S6: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio15.

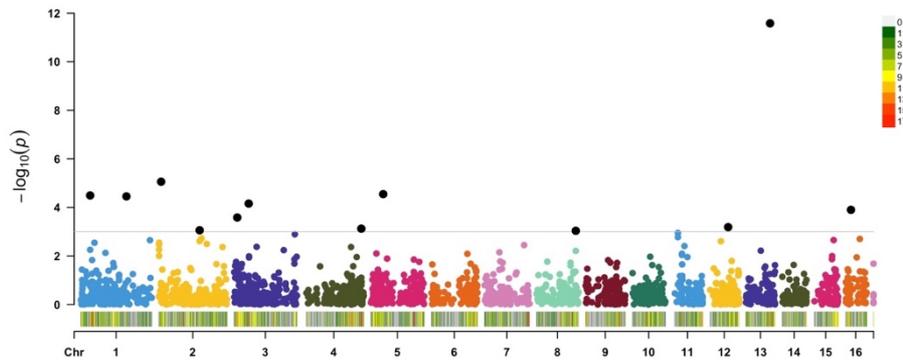


Figure S7: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio18.

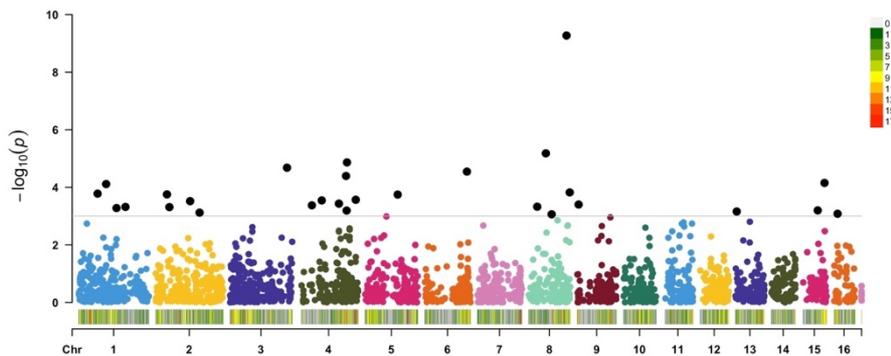


Figure S8: Manhattan plot showing p values from 2899 SNPs aligned by position on chromosomes 1–16. Black dots correspond to 316 SNPs identified as outliers by Baypass that were correlated with environmental parameters bio19.

A

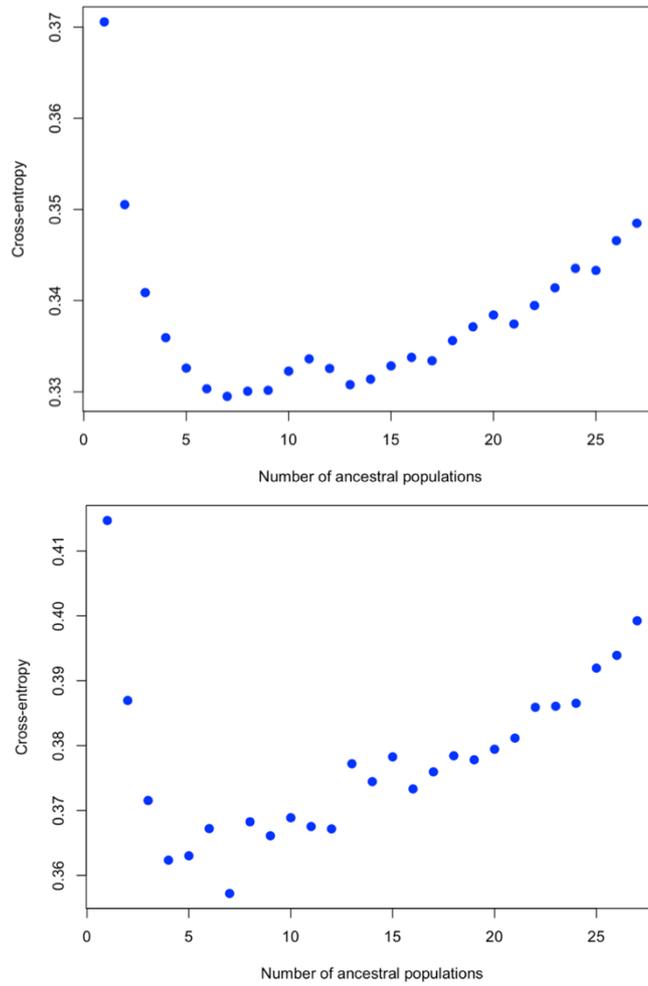


Figure S9: Value of the cross-entropy criterion as a function of the number of populations in snmf. **A** 2227 neutral loci. **B** 316 outlier loci.

Table S10: Table with the class, the identifier of the 222 genes affected by the 316 outlier SNPs found, the genetic ontology identifier and the name of the gene.

Slim_GO_Name	Input_Accession	Input_GOID	Input_GO_Name
cellular_component	105052600	GO:0016592	mediator complex
cellular_component	109504955	GO:0005750	mitochondrial respiratory chain complex III
cellular_component	105050216	GO:0000922	spindle pole
cellular_component	105054983	GO:0000139	Golgi membrane
cellular_component	105053764	GO:0005689	U12-type spliceosomal complex
cellular_component	105058912	GO:0005774	vacuolar membrane
cellular_component	105033408	GO:0009522	photosystem I
cellular_component	105054455	GO:0000145	exocyst
cellular_component	105037975	GO:0032040	small-subunit processome
cellular_component	105058040	GO:0005669	transcription factor TFIID complex
cellular_component	105050125	GO:1990904	ribonucleoprotein complex
cellular_component	105050068	GO:0005874	microtubule
cellular_component	12079378	GO:0009535	chloroplast thylakoid membrane
cellular_component	105051107	GO:0071203	WASH complex
cellular_component	105051390	GO:0035550	urease complex
cellular_component	105057779	GO:0031510	SUMO activating enzyme complex
cellular_component	105060353	GO:0032039	integrator complex
cellular_component	105057822	GO:0016021	integral component of membrane
cellular_component	105057090	GO:0031225	anchored component of membrane
cellular_component	105056822	GO:0000932	P-body
cellular_component	105044306	GO:0030532	small nuclear ribonucleoprotein complex
cellular_component	12079462	GO:0009523	photosystem II
extracellular region	105058444	GO:0005576	extracellular region
cell wall	105050985	GO:0005618	cell wall
nucleus	105057588	GO:0005634	nucleus
nucleolus	105048454	GO:0005730	nucleolus
cytoplasm	105038988	GO:0005737	cytoplasm
endosome	105056134	GO:0005768	endosome
endoplasmic reticulum	105052599	GO:0005783	endoplasmic reticulum
Golgi apparatus	105040229	GO:0005794	Golgi apparatus
cytosol	105044953	GO:0005829	cytosol
ribosome	105038265	GO:0005840	ribosome
cytoskeleton	105049197	GO:0005856	cytoskeleton
plasma membrane	12079379	GO:0005886	plasma membrane
chloroplast	105038672	GO:0009507	chloroplast
thylakoid	105053237	GO:0009579	thylakoid
membrane	105040640	GO:0016020	membrane
nucleotide binding	105059455	GO:0000166	nucleotide binding
nucleotide binding	105054899	GO:0000062	fatty-acyl-CoA binding
nucleotide binding	105059780	GO:0005525	GTP binding

nucleotide binding	105057822	GO:0005524	ATP binding
molecular_function	105040229	GO:0005457	GDP-fucose transmembrane transporter activity
molecular_function	105061577	GO:0015079	potassium ion transmembrane transporter activity
molecular_function	12079378	GO:0046933	proton-transporting ATP synthase activity, rotational mechanism
molecular_function	105045088	GO:0015297	antiporter activity
molecular_function	105056823	GO:0015145	monosaccharide transmembrane transporter activity
molecular_function	105058444	GO:0004601	peroxidase activity
molecular_function	105043678	GO:0004386	helicase activity
molecular_function	105040639	GO:0005249	voltage-gated potassium channel activity
molecular_function	105058661	GO:0022857	transmembrane transporter activity
molecular_function	105044026	GO:0004362	glutathione-disulfide reductase (NADPH) activity
molecular_function	105044718	GO:0005096	GTPase activator activity
nucleic acid binding	105056068	GO:0000977	RNA polymerase II transcription regulatory region sequence-specific DNA binding
nucleic acid binding	105043350	GO:0019843	rRNA binding
nucleic acid binding	105039422	GO:0003676	nucleic acid binding
DNA binding	105057588	GO:0003677	DNA binding
chromatin binding	105050005	GO:0003682	chromatin binding
DNA-binding transcription factor activity	105040317	GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific
DNA-binding transcription factor activity	105038793	GO:0003700	DNA-binding transcription factor activity
RNA binding	105041374	GO:0003723	RNA binding
RNA binding	105059366	GO:0000049	tRNA binding
catalytic activity	105060233	GO:0004298	threonine-type endopeptidase activity
catalytic activity	105049033	GO:0004497	monooxygenase activity
catalytic activity	105034598	GO:0004657	proline dehydrogenase activity
catalytic activity	105042936	GO:0004842	ubiquitin-protein transferase activity
catalytic activity	105038672	GO:0004316	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
catalytic activity	105040545	GO:0004638	phosphoribosylaminoimidazole carboxylase activity
catalytic activity	105061405	GO:0004739	pyruvate dehydrogenase (acetyl-transferring) activity
catalytic activity	105044953	GO:0016920	pyroglutamyl-peptidase activity
catalytic activity	105058350	GO:0097573	glutathione oxidoreductase activity
catalytic activity	105059357	GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
catalytic activity	105059272	GO:0004674	protein serine/threonine kinase activity
catalytic activity	105050891	GO:0003989	acetyl-CoA carboxylase activity
catalytic activity	105039150	GO:0018024	histone-lysine N-methyltransferase activity
catalytic activity	105055466	GO:0004252	serine-type endopeptidase activity

catalytic activity	105049198	GO:0016491	oxidoreductase activity
catalytic activity	105056822	GO:0000175	3'-5'-exoribonuclease activity
catalytic activity	105051861	GO:0061630	ubiquitin protein ligase activity
catalytic activity	105061390	GO:0003755	peptidyl-prolyl cis-trans isomerase activity
catalytic activity	105042399	GO:0004843	thiol-dependent deubiquitinase
catalytic activity	105058225	GO:0003824	catalytic activity
catalytic activity	105055899	GO:0004672	protein kinase activity
catalytic activity	105048454	GO:0008649	rRNA methyltransferase activity
catalytic activity	105045658	GO:0004807	triose-phosphate isomerase activity
nuclease activity	105052040	GO:0004527	exonuclease activity
nuclease activity	105060355	GO:0004518	nuclease activity
structural molecule			
activity	105038265	GO:0003735	structural constituent of ribosome
binding	105033408	GO:0016168	chlorophyll binding
binding	105050985	GO:0030145	manganese ion binding
binding	105048129	GO:0003779	actin binding
binding	12079499	GO:0005506	iron ion binding
binding	105049197	GO:0051015	actin filament binding
binding	105050068	GO:0008017	microtubule binding
binding	105058040	GO:0046982	protein heterodimerization activity
binding	105059883	GO:0005507	copper ion binding
binding	105057764	GO:0000287	magnesium ion binding
binding	105040990	GO:0046983	protein dimerization activity
binding	105035759	GO:0036094	small molecule binding
binding	105035685	GO:0046872	metal ion binding
binding	105055948	GO:0008270	zinc ion binding
translation factor activity,			
RNA binding	105035404	GO:0003746	translation elongation factor activity
translation factor activity,			
RNA binding	105049865	GO:0003743	translation initiation factor activity
lipid binding	105048299	GO:0035091	phosphatidylinositol binding
lipid binding	12079458	GO:0008289	lipid binding
kinase activity	105058743	GO:0004594	pantothenate kinase activity
kinase activity	105060354	GO:0000823	inositol-1,4,5-trisphosphate 6-kinase activity
kinase activity	105050075	GO:0003872	6-phosphofructokinase activity
kinase activity	105033444	GO:0016301	kinase activity
transferase activity	105056662	GO:0008483	transaminase activity
transferase activity	105058717	GO:0004400	histidinol-phosphate transaminase activity
transferase activity	105050124	GO:0016157	sucrose synthase activity
transferase activity	105053237	GO:0004364	glutathione transferase activity
transferase activity	105058113	GO:0008184	glycogen phosphorylase activity
transferase activity	105039087	GO:0008915	lipid-A-disaccharide synthase activity
transferase activity	105053854	GO:0016746	acyltransferase activity
transferase activity	105052944	GO:0004659	prenyltransferase activity

transferase activity	105045373	GO:0008080	N-acetyltransferase activity
transferase activity	105058598	GO:0004315	3-oxoacyl-[acyl-carrier-protein] synthase activity
transferase activity	105052170	GO:0004845	uracil phosphoribosyltransferase activity
transferase activity	105060106	GO:0008168	methyltransferase activity
transferase activity	105043661	GO:0016757	glycosyltransferase activity
transferase activity	105054366	GO:0016758	hexosyltransferase activity
transferase activity	105038463	GO:0016740	transferase activity
hydrolase activity	105051390	GO:0009039	urease activity
hydrolase activity	105052599	GO:0003924	GTPase activity
hydrolase activity	105045830	GO:0016791	phosphatase activity
hydrolase activity	105045829	GO:0004848	ureidoglycolate hydrolase activity
hydrolase activity	105038262	GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds
hydrolase activity	105044951	GO:0004650	polygalacturonase activity
hydrolase activity	105059779	GO:0004575	sucrose alpha-glucosidase activity
hydrolase activity	105047243	GO:0003860	3-hydroxyisobutyryl-CoA hydrolase activity
hydrolase activity	105040640	GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity
hydrolase activity	105050290	GO:0016787	hydrolase activity
carbohydrate binding	105039683	GO:0030246	carbohydrate binding
transcription regulator activity	105043784	GO:0003712	transcription coregulator activity
transcription regulator activity	105058705	GO:0003714	transcription corepressor activity
transcription regulator activity	105053213	GO:0003713	transcription coactivator activity
reproduction	105040294	GO:2000028	regulation of photoperiodism, flowering
carbohydrate metabolic process	105042937	GO:0010411	xyloglucan metabolic process
carbohydrate metabolic process	105043661	GO:0006004	fucose metabolic process
carbohydrate metabolic process	105053237	GO:0033355	ascorbate glutathione cycle
carbohydrate metabolic process	105058113	GO:0005975	carbohydrate metabolic process
carbohydrate metabolic process	105045658	GO:0006096	glycolytic process
carbohydrate metabolic process	105039087	GO:0009245	lipid A biosynthetic process
carbohydrate metabolic process	105059779	GO:0005987	sucrose catabolic process
carbohydrate metabolic process	105050124	GO:0005986	sucrose biosynthetic process
generation of precursor	105048129	GO:0006099	tricarboxylic acid cycle

metabolites and energy generation of precursor			mitochondrial electron transport, ubiquinol to cytochrome c
metabolites and energy generation of precursor	109504955	GO:0006122	
metabolites and energy generation of precursor	105033408	GO:0009765	photosynthesis, light harvesting
metabolites and energy generation of precursor	12079462	GO:0009772	photosynthetic electron transport in photosystem II
nucleobase-containing compound metabolic process	105052170	GO:0006223	uracil salvage
nucleobase-containing compound metabolic process	105058040	GO:0006352	DNA-templated transcription, initiation
nucleobase-containing compound metabolic process	105056068	GO:0045944	positive regulation of transcription by RNA polymerase II
nucleobase-containing compound metabolic process	105056135	GO:0000122	negative regulation of transcription by RNA polymerase II
nucleobase-containing compound metabolic process	105045598	GO:0006139	nucleobase-containing compound metabolic process
nucleobase-containing compound metabolic process	105040545	GO:0006189	'de novo' IMP biosynthetic process
nucleobase-containing compound metabolic process	105056822	GO:0034427	nuclear-transcribed mRNA catabolic process, exonucleolytic, 3'-5'
nucleobase-containing compound metabolic process	105058743	GO:0015937	coenzyme A biosynthetic process
nucleobase-containing compound metabolic process	105037975	GO:0006364	rRNA processing
nucleobase-containing compound metabolic process	105060353	GO:0016180	snRNA processing
nucleobase-containing compound metabolic process	105059366	GO:0006432	phenylalanyl-tRNA aminoacylation
nucleobase-containing compound metabolic process	105053764	GO:0000398	mRNA splicing, via spliceosome
nucleobase-containing compound metabolic process	105035129	GO:0006355	regulation of transcription, DNA-templated

process			
nucleobase-containing compound metabolic process	105061405	GO:0006086	acetyl-CoA biosynthetic process from pyruvate
nucleobase-containing compound metabolic process	12079503	GO:0006397	mRNA processing
nucleobase-containing compound metabolic process	105048454	GO:0000453	enzyme-directed rRNA 2'-O-methylation
nucleobase-containing compound metabolic process	105036417	GO:0006367	transcription initiation from RNA polymerase II promoter
nucleobase-containing compound metabolic process	105054832	GO:0006396	RNA processing
nucleobase-containing compound metabolic process	105061517	GO:0001522	pseudouridine synthesis
nucleobase-containing compound metabolic process	105042936	GO:0036297	interstrand cross-link repair
nucleobase-containing compound metabolic process	105052040	GO:0006281	DNA repair
translation	105038265	GO:0006412	translation
cellular protein modification process	105054366	GO:0006486	protein glycosylation
cellular protein modification process	105057779	GO:0016925	protein sumoylation
cellular protein modification process	105060447	GO:0045116	protein neddylation
lipid metabolic process	105047723	GO:0046856	phosphatidylinositol dephosphorylation
lipid metabolic process	105044200	GO:0008610	lipid biosynthetic process
lipid metabolic process	105038672	GO:0006633	fatty acid biosynthetic process
lipid metabolic process	105044026	GO:0001516	prostaglandin biosynthetic process
transport	105058661	GO:0006857	oligopeptide transport
transport	105054455	GO:0006887	exocytosis
transport	105043172	GO:0006869	lipid transport
transport	105038988	GO:0015031	protein transport
transport	105058261	GO:0048193	Golgi vesicle transport
transport	105050234	GO:0016192	vesicle-mediated transport
transport	105052599	GO:0006886	intracellular protein transport
transport	105051860	GO:0006897	endocytosis

transport	105048299	GO:0043328	protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway
response to stress	12079499	GO:0009635	response to herbicide
response to stress	105036411	GO:0006952	defense response
response to stress	105058444	GO:0006979	response to oxidative stress
cell cycle	105052432	GO:0007049	cell cycle
cell cycle	105049958	GO:0007064	mitotic sister chromatid cohesion
cell communication	105058744	GO:0048015	phosphatidylinositol-mediated signaling
cell communication	105044034	GO:0007166	cell surface receptor signaling pathway
signal transduction	105042423	GO:0007165	signal transduction
biological_process	105058717	GO:0000105	histidine biosynthetic process
biological_process	105052944	GO:0010189	vitamin E biosynthetic process
biological_process	105044532	GO:0007018	microtubule-based movement
biological_process	105033300	GO:0006796	phosphate-containing compound metabolic process
biological_process	105057090	GO:0010215	cellulose microfibril organization
biological_process	105050216	GO:0000902	cell morphogenesis
biological_process	105050290	GO:0009850	auxin metabolic process
biological_process	105048924	GO:0007030	Golgi organization
biological_process	105050075	GO:0006002	fructose 6-phosphate metabolic process
biological_process	105056137	GO:0031047	gene silencing by RNA
biological_process	105056664	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process
biological_process	105060106	GO:0032259	methylation
biological_process	105044221	GO:0010027	thylakoid membrane organization
biological_process	105051717	GO:0061635	regulation of protein complex stability
biological_process	105040520	GO:0001678	cellular glucose homeostasis
biological_process	105042399	GO:0006511	ubiquitin-dependent protein catabolic process
biological_process	105045383	GO:2000762	regulation of phenylpropanoid metabolic process
biological_process	105043350	GO:0000027	ribosomal large subunit assembly
biological_process	105049197	GO:0007010	cytoskeleton organization
biological_process	105034598	GO:0006562	proline catabolic process
biological_process	105060354	GO:0032958	inositol phosphate biosynthetic process
biological_process	105057764	GO:0009082	branched-chain amino acid biosynthetic process
biological_process	105061390	GO:0006457	protein folding
biological_process	105051390	GO:0043419	urea catabolic process
biosynthetic process	105056662	GO:0009058	biosynthetic process
photosynthesis	12079391	GO:0015979	photosynthesis

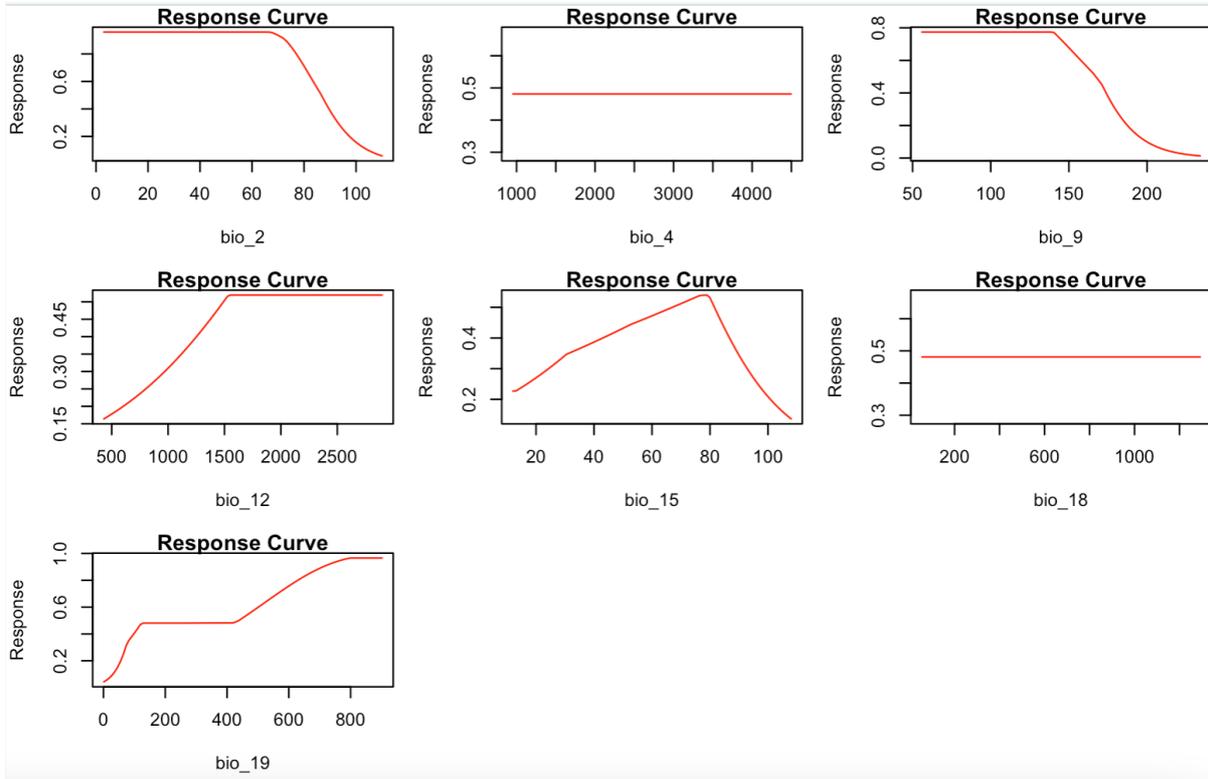


Figure S11: Maxent response curves from the best SDM of *Euterpe edulis*.