

Genome-wide SNP analysis: adaptation and population structure of *Syagrus romanzoffiana* (Cham.) Glassman (Arecaceae) in South America

Kauanne Karolline Moreno Martins

State University of Campinas – UNICAMP

Matheus Scaketti

University of São Paulo – ESALQ/USP

Ana Flávia Francisconi

University of São Paulo – ESALQ/USP

Igor Araújo Santos de Carvalho

University of São Paulo – ESALQ/USP

Caroline Bertocco Garcia

University of São Paulo – ESALQ/USP

Thiago Deomar Ludwig

University of São Paulo – ESALQ/USP

Enéas Ricardo Konzen

Federal University of Rio Grande do Sul – UFRGS

Maria Imaculada Zucchi

mizucchi@sp.gov.br

APTA, UPDR-Piracicaba

Article

Keywords: Arecaceae¹, Genomic diversity², Population genomics³, Molecular markers⁴, Atlantic forest⁵, Adaptive loci⁶

Posted Date: August 31st, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7208849/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Palms play a crucial ecological and economic role, with *Syagrus romanzoffiana* being a prominent species in South America. Despite its widespread distribution and economic potential, genetic studies on this species remain limited. This study assessed the genetic diversity and population structure of *S. romanzoffiana* across distinct vegetation types in Brazil and Paraguay using single nucleotide polymorphisms (SNPs) derived from genotyping-by-sequencing (GBS). A total of 91 individuals from eight populations were analyzed, revealing significant genetic diversity and differentiation. Observed heterozygosity was generally high, with some populations exhibiting excess heterozygosity, indicating potential adaptive significance. Pairwise F_{ST} and AMOVA analysis suggested that most genetic variation is maintained within populations, although moderate differentiation among populations was detected. Genetic structure analysis highlighted the influence of vegetation type on genetic differentiation, with restinga populations displaying unique genetic signatures. Despite geographical proximity, some populations exhibited greater genetic divergence due to local environmental adaptations. These findings underscore the importance of conserving *S. romanzoffiana*'s genetic diversity and highlight the role of vegetation type in shaping population structure. Understanding these patterns is essential for the sustainable management and conservation of this species, particularly in the face of habitat fragmentation and climate change.

Introduction

Palm species (Arecaceae) stand out for its versatility and diversity, with more than 2,600 recognized species worldwide. Over centuries, palms have provided a wide range of services to humankind ^{[1],[2],[3]}. Many palms are considered ecological keystone species because large numbers of animals depend on their fruit and flower resources to survive ^[4]. Moreover, palms have been used for centuries by different communities around the world because their products have different biological properties that are frequently effective against numerous diseases ^[5] including chronic diseases, such as diabetes and Alzheimer's disease ^[6]. In addition, palms are also used in other applications, such as the production of biodiesel and energy ^[7].

Syagrus romanzoffiana (Cham.) Glassman, from the Aracaceae family, commonly known in Brazil as jerivá, is an endemic palm distributed throughout South America ^[8]. *S. romanzoffiana* play a crucial role in the subsistence of many traditional and rural communities, providing a wide range of non-timber forest products. These include food and medicinal uses from fruits, cosmetics, fiber from leaves, building materials from leaves and stems, and handicraft from seeds ^{[9],[10]}. Furthermore, previous studies performed by Moreira et al., ^[11] identified *S. romanzoffiana* as a promising resource for sustainable large-scale production of vegetable oil. The same authors also stated that the oil is a viable alternative to produce biodiesel.

Species delimitation studies are key to understanding genetic diversity and population structure at both intra and interspecific levels. Such insights support informed decisions for resource management, sustainable use, and conservation planning ^{[12],[13]}. Analyzing the genetic diversity of *S. romanzoffiana* is crucial for guiding the selection of the most promising materials for cultivation, maximizing genetic gains, and facilitating the development of commercially viable cultivars, according to local adaptation.

In this context, molecular markers have become essential tools in plant research for examining genetic diversity across ecological, phylogenetic, and evolutionary studies ^[14]. Furthermore, they have been extensively employed for direct management, conservation, and genetic breeding across various species ^[14]. Advances in next-generation sequencing (NGS) technologies have greatly enhanced the discovery of single nucleotide polymorphisms (SNPs). These markers are now among the most widely applied in molecular ecology because they are numerous, cover large portions of the genome, and can be used to assess both neutral and adaptive variation. NGS enables efficient, high-throughput SNP genotyping with low error rates, even in the absence of a reference genome ^{[12],[15],[16]}. SNP markers are diverse, spanning multiple scientific fields, particularly in the development of high-density panels for genomic selection, conservation genetics, and breeding programs ^{[12],[17],[18]}.

The focus of this study is to evaluate and compare the genetic diversity and population structure of the palm *Syagrus romanzoffiana* across different regions in southern Brazil and Paraguay. Specifically, our objectives were to (i) sample individuals from contrasting vegetation types, (ii) assess genetic diversity and structure within and among these populations using SNP markers obtained through genotyping-by-sequencing (GBS), and (iii) investigate patterns of adaptive genetic variation in relation to environmental variables such as temperature and precipitation. We hypothesized that populations occurring in distinct vegetation types would exhibit differentiated genetic structures and that climatic variables would help explain signals of local adaptation. This is the first study to use SNP markers for *S. romanzoffiana*, highlighting the importance of conserving its genetic diversity and understanding its population structure to inform effective management and conservation strategies.

Results

Genetic diversity in Southern America

After data filtering, a total of 24,859 SNP markers were retained for the *S. romanzoffiana*. The average percentage of missing data per individual was 6.11%, with a maximum of 14% (Supplementary Figure 1A). The average sequencing coverage per individual was 15.25x (Supplementary Figure 1B), and the mean coverage per locus, was 13.68x, with no coverage exceeding 100x or falling below 10x per locus (Supplementary Figure 2A). The SNP data also revealed a higher number of transitions compared to transversions (Supplementary Figure 2B).

Observed heterozygosity (H_o) and expected (H_s) heterozygosity were high across populations, with H_o exceeding H_s in six sites. Lucena-Cecília was an exception, showing similar H_o and H_s values (0.169),

while Travessa had a lower H_o than H_s . The highest H_o values were in San Ramón (0.205) and Gravatá (0.174), with the highest H_s also in San Ramón (0.203), followed by Lucena Cecilia (0.169). San Ramón and Lucena-Cecilia had the greatest number of alleles (A , Table 1), with San Ramón and Lucena-Cecilia showing the highest allelic richness (Ar) at 1.567 and 1.492, respectively, followed by Gravatá, which also had a notable number of private alleles (Ap , Table 1). The highest fixation index (f) was computed from Travessa (0.014), Lucena-Cecilia (0.006), and Condomínio-Marítimo (0.004), while it was negative in Gravatá, Herval, and Pixirica, indicating an excess of heterozygotes (Table 1).

Table 1. Population Genomic Diversity of *Syagrus romanzoffiana* based on 24,859 SNPs. H_o = Observed heterozygosity, H_s = Expected heterozygosity, A = Number of alleles, Ar = Allelic richness, Ap = Private alleles, f = Fixation index (bootstraps). Locations in Brazil; Borrussia – RS; CondMaritmo – RS, Herval – RS, Lucena Cecilia – RS, Pixirica – RS, Travessa – RS and Gravata – SC. Location in Paraguay: San Ramón – Concepción.

Location	H_o	H_s	A	Ar	Ap	f	
Borrussia	0.144	0.143	35479	1.415	245	0.010	(0.0012,0.0218)
CondMaritmo	0.140	0.138	36125	1.387	653	-0.004	(-0.0221,-0.0021)
Gravata	0.174	0.167	38196	1.473	3291	-0.013	(-0.0354,-0.0175)
Herval	0.162	0.159	36631	1.452	364	0.259	(0.3303,0.3469)
LucenaCecilia	0.169	0.169	37555	1.492	329	-0.031	(-0.0527,-0.0343)
Pixirica	0.138	0.132	34935	1.372	476	0.013	(0.0231,0.0524)
San Ramón	0.205	0.203	40079	1.567	17426	-0.028	(-0.0534,-0.0341)
Travessa	0.160	0.163	35518	1.467	98	0.449	(0.385,0.4055)

Population genetic structure in Southern America

Based on 24,859 SNPs, Wright's F statistics revealed significant differentiation among populations. The F statistics analysis of the eight sampled *Syagrus romanzoffiana* populations showed the following values: total inbreeding coefficient (F_{IT} = 0.139), within-population inbreeding coefficient (F_{IS} = 0.021), and genetic differentiation coefficient between populations (F_{ST} = 0.122). Pairwise F_{ST} analysis indicated that the San Ramón location was the most divergent, with an F_{ST} of 0.371 compared to the Pixirica. Despite being further apart geographically, Borrússia and Pixirica (~ 90 km apart), both in Atlantic Rain Forest, were more genetically similar, with an F_{ST} of 0.079, than Borrússia and Condomínio Marítimo (~ 20 km apart, F_{ST} of 0.081), the latter in an area of restinga (Figure 1).

The AMOVA results indicated significant differentiation among populations (29.7% of the total variation), but the majority of genetic variation (70.3%) occurred within populations. These findings suggest a

genetic structure in which most of genetic diversity is maintained within populations. This pattern may result from substantial gene flow within populations and moderate barriers to gene flow between them (Table 2). However, the Mantel test showed a very weak and non-significant correlation between genetic and geographic distances ($r = -0.009$, $p = 0.141$), indicating no clear pattern of isolation by distance.

Table 2. Analysis of molecular variance (AMOVA) performed for eight populations of *Syagrus romanzoffiana* collected in the in Brazil and Paraguay.

Source of Variation	df	MS	Est. Var.	PV (%)
Among populations (Among Locations)	7	6617.983	483.8844	29.7%
Within populations (Within Locations)	83	1145.607	1145.6073	70.3%
Total	90	1571.237	1629.4917	100%

p-value = 0.001 (Estimated based on 20000 permutations). Df = Degrees of Freedom; MS = Mean Squares. Est. Var. = Variance component. PV = Percentage of Variation.

For the DAPC with populations, one principal component was retained, followed by one discriminant. The DAPC with predefined populations explained 19.22% of the variation. As observed with the genetic diversity estimates and the pairwise F_{ST} , the San Ramón was the most distinct among the sampled *S. romanzoffiana* locations. The other sites showed some degree of overlap along the LD1 axis, reflecting more subtle genetic differentiation among them (Figure 2).

Detection of adaptive loci and environmental associations

LFMM identified 3,768 SNPs significantly associated with bio06 and bio19 (Supplementary Figures 5 and 6). A new sNMF analysis using only the candidate SNPs resulted in a population structure with $K = 4$ clusters (Figure 3), suggesting that environmental factors are influencing population differentiation beyond neutral processes.

The RDA based on these candidates SNPs confirmed strong genotype–environment associations. The global model constrained by *bio19* (precipitation of the coldest quarter) and *bio6* (minimum temperature of the coldest month) was highly significant ($p < 0.001$), with an adjusted R^2 of 20%. Both variables contributed significantly to the explained genetic variance, and the first two RDA axes were also significant ($p < 0.001$), indicating that climatic factors are important drivers of adaptive genetic variation in *S. romanzoffiana*.

The first two constrained axes explained 99.9% of the total variance (Figure 4). Individuals from San Ramón were positively associated with bio06, indicating adaptation to colder minimum temperatures, while Cond. Marítimo, Pixirica, Borrussia, and SC_Gravatá were strongly associated with bio19, suggesting adaptive responses to precipitation seasonality in the southern Atlantic Forest and restinga environments.

Discussion

The populations of *S. romanzoffiana* analyzed in this study revealed substantial genetic diversity. The high values of observed (H_o) and expected (H_s) heterozygosity found in various sites, with H_o exceeding H_s , except in Lucena Cecília and Travessa, indicates a potential excess of heterozygotes. This phenomenon can be interpreted as an adaptive response to local environmental conditions or as a result of a population structure that favors allogamy among genetically distinct individuals ^[19]. Populations with high heterozygosity are generally more resilient to environmental pressures, which may be crucial for survival in altered habitats ^[20].

Allelic richness (Ar) was also notable, especially when compared to other locations. This richness suggests a high genetic potential, particularly in the populations of San Ramón and Lucena Cecília, which are essential for the adaptation and evolution of populations ^[21]. Additionally, the high number of private alleles (Ap) found in Cond. Marítimo and Gravatá also highlights the genetic uniqueness of these populations, which may represent specific local adaptations or genetic variations not detected in other populations. In fact, these two sites are located in restinga vegetation domains, which are coastal vegetation formations with strong influence of maritime conditions, such as constant wind, sandy and saline soils, alternating with high water supply after heavy rains or even drought episodes ^[22]. Vetö et al., ^[23] found signatures of local adaptation for a Myrtaceae species, *Eugenia uniflora*, where populations located in the coastal plain or in restinga areas showed distinct alleles in comparison to inland populations.

The positive fixation indexes (f) observed in Travessa, Lucena Cecília, and Cond. Marítimo suggest that these populations have a tendency towards homozygosity, which may be concerning in terms of the loss of genetic diversity. In fact, all of these sites are located in a severely fragmented landscape, that may have already led to some crosses among related individuals. In contrast, negative fixation indices in Borrusia, Gravatá, Herval, and Pixirica indicate Hardy-Weinberg equilibrium. These populations are located in larger and more conserved forest fragments than the previous ones. This may reflect a mating dynamic that is favorable for maintaining genetic diversity, essential for the adaptability of populations in constantly changing environments ^{[24],[19]}.

Wright's F-statistics revealed moderate to high differentiation among *S. romanzoffiana* sites. The values of the genetic differentiation coefficient (F_{ST}) indicates that, despite gene flow, populations are sufficiently isolated to maintain genetic differences ^[25].

AMOVA results showed that most genetic variation occurs within populations, reinforce the idea that most of the genetic diversity is maintained at the intrapopulation level. This suggests that populations possess a genetic structure with a considerable variety of alleles ^{[26],[27]}. This pattern may result from barriers to gene flow among populations, combined with a complex population structure.

The Mantel test, which did not find a significant correlation between genetic distance and geographic distance, suggests that the geographic separation of populations is not the sole determining factor for genetic differentiation. This may indicate that other elements, such as ecological barriers (as pointed out), local adaptations, or interactions with other species, are at play and should be considered in future investigations ^{[28],[29]}. In fact, when comparing the genetic differentiation among Pixirica, Borrússia and Cond. Marítimo, Pixirica and Borrússia were more genetically similar to each other despite their higher geographic distance (~90km), in comparison to Borrússia and Cond. Marítimo, that are ~20km apart only but are genetically more distinct than the previous pair. As previously stated, the vegetation type and local climate seem to be important drivers of the genetic differentiation in spite of the geographic proximity. Cond. Marítimo is located in a restinga fragment approximately 5km from the Atlantic Ocean. The area has sandy soils and frequent and strong winds that bring salt particles along with them in the air. Borrússia is located in an elevated area (approximately 200 to 300m in altitude) with soils originated from volcanic depositions, and a vegetation type that belongs with the range of the Atlantic rain forest. Therefore, local adaptations probably explain the genetic differentiation among the studied sites.

The environmental association provided compelling evidence for local adaptation in *Syagrus romanzoffiana*. The identification of SNPs associated with climatic variables (particularly minimum temperature and precipitation) highlighted the selective pressures shaping genetic structure across distinct vegetation types and environmental conditions. Interestingly, the adaptive genetic structure emphasizes that selection may be acting independently of geographic distance or historical demography. This is consistent with findings in other tropical tree species, where environment-driven divergence occurs even at small spatial scales ^[30]. Populations from San Ramón showed strong signals of adaptation to colder winters, while southern Brazilian populations, particularly those in restinga and coastal forest habitats, aligned with wetter and more variable rainfall regimes. These results reinforce the importance of integrating adaptive genetic variation into conservation planning, especially for species distributed across ecotonal gradients or fragmented landscapes.

Material and methods

Plant material

We sampled 91 leaf tissues from adult and reproductive individuals of *Syagrus romanzoffiana* from different locations representing three vegetation types in which they are found in South America: Atlantic Rain Forest, Atlantic Seasonal Forest, and restingas ^{[31],[32]}. The following sites were sampled: Linha Lucena-Cecília, Travessa and Herval, in the municipality of Venâncio Aires, in Rio Grande do Sul state, belonging with the domain of Atlantic Seasonal Forest. The site collected in San Ramón, Paraguay, also represents an area with seasonal forests. Pixirica, in the municipality of Morrinhos do Sul, and Borrússia, in Osório, both in Rio Grande do Sul state belong are part of Atlantic Rain Forest domain. Condomínio Marítimo represents a restinga fragment within the municipality of Tramandaí, also in Rio Grande do Sul. Another restinga site is Gravatá, located in the municipality of Laguna, in Santa Catarina (Figure 5). The samples were collected in Santa Catarina – Brazil, Rio Grande do Sul – Brazil and Concepción –

Paraguay and placed in plastic bags containing silica-gel, duly identified, packed in a cardboard box, transported to the Genetics Department at Luiz de Queiroz College of Agriculture/University of São Paulo. Accordingly, this study received approval from Ethics Committee of University of Campinas and was carried out in accordance with the regulations of Brazilian Ministry of the Environment (MMA) and the National System for Genetic Heritage and Associated Traditional Knowledge (SisGen) under the number A3A9281.

DNA extraction

DNA extraction from *S. romanzoffiana* leaf tissues was performed according to the Doyle and Doyle ^[33] protocol with modifications for palms. Specifically, we applied a modified CTAB method; leaf samples were grounded and placed in a 1.5mL microtube containing 700 µL 2% CTAB plus 2 µL β-mercaptoethanol then incubated at 65°C for 40 min. The microtubes were mixed by inversion every 15 min. Next, 700µL chloroform-isoamylalcohol (24:1) were added and the tubes were gently mixed for 1 min, followed by centrifugation for 15 min at 13,000 rpm. Immediately after centrifuging, 500µL of the supernatant of each tube was transferred to a fresh and identified 1.5 mL tube with 600µL cold isopropanol (–20°C) plus 55µL ammonium acetate 7.5M 1:20. Samples were gently mixed by inversion and conditionate at –20°C for at least 2 hours, followed by centrifugation at 13,000 rpm for 15 min. After centrifugation, the supernatants were removed, and the DNA pellets were washed with 400µL of 70% ethanol and mixed by inversion for 30 seconds followed by centrifugation at 13,000 rpm for 5 min. The ethanol was discarded and 400µL 100% ethanol was added and mixed by inversion for 30 seconds plus centrifuged at 13,000 rpm for 5 min, the ethanol discarded, and the tubes were set to dry for at least 40 minutes. The DNA pellets were suspended in 30µL TE buffer (10 mM Tris–HCl pH 7.6, 1 mM EDTA pH 7.6) plus 2µL ribonuclease (RNAse 20 mg/mL), incubated at 37°C for 30 minutes, and stored at –20°C until quantification. The quantification and quality of the samples were evaluated on a 1% agarose gel. Based on the obtained reading, we standardized the DNA to a concentration of 30ng.µl⁻¹. The final quantification was made using Qubit.

Genotyping by sequencing (GBS) library preparation

A genomic library was developed to obtain SNPs markers by the GBS double digestion technique ^[34] with modifications. We digested 7 µl of the genomic DNA [30ng.µl⁻¹] from each sample with the enzymes Pst1 and Mse1. The first enzyme was used to fragment the DNA into small sites and the second one to fragment the DNA again, but now into several sites. After that, 25µM of specific adapters (barcodes) from Illumina technology were ligated to these digested fragments. Ligation was performed at the ends of the fragments with T4 DNA ligase. After adapter ligation, we purified the samples using QIAquick PCR Purification Kit (Qiagen). Library enrichment was carried out by PCR using primers matching the adapters. Additionally, eight replicates were performed, each PCR reaction contained 10 µL of the purified and amplified ligation product, 12.5 µL of Phusion1 High-Fidelity PCR Master Mix (NEB, New England Biolabs Inc.), and 2 µL of Illumina forward and reverse primers (10 µM), totaling 25 µL. The amplification conditions were: initial denaturation at 95°C for 30 s, followed by 16

cycles at 95°C for 10 s, 62°C for 20 s, 72°C for 30 s, and a final extension at 72°C for 5 min. Libraries were then purified with the QIAquick PCR Purification Kit (Qiagen). Fragment size distribution was assessed using the DNA 12,000 kit on the 2100 Bioanalyzer (Agilent Technologies). The libraries were quantified using quantitative PCR (qPCR) on a CFX 384 real-time thermocycler (Bio-Rad) to ensure accurate assessment of DNA concentration prior to sequencing. PCR sequencing was performed by Illumina NextSeq 1000/2000 platform at Luiz de Queiroz College of Agriculture – University of São Paulo – (ESALQ-USP).

SNP identification

SNP quality control was conducted through *fastqc* (<https://github.com/s-andrews/FastQC>). The DNA sequences (reads) obtained by Illumina sequencing were filtered to obtain SNP markers. Filtering was conducted with Stacks v. 2.62 pipeline^[35]. We then used the *process_radtags* module to demultiplex the samples and remove the low-quality reads (reads with Phred scores lower than 10, or that contained Illumina adapters, uncalled bases - “Ns” - or without restriction sites). Next, the *de novo* assembly was conducted in Stacks, beginning with the *ustacks* module to cluster putatively homologous reads into loci. For each sample, we set a minimum stack depth (-m) of 4, allowed a maximum of 5 mismatches between stacks (-M), and a maximum of 5 mismatches between primary and secondary reads^[15]. After that, using the *cstacks* module, we assembled a catalog of loci, setting the tolerance for sequence variation to a maximum of four mismatches between stacks originating from different individuals^[15]. A conversion of the *tvb2bam* to *gstacks* was performed. Minor allele frequency ($p = t - 1$), and minimum occurrence in 75% of individuals in each location/population ($r = 0.75$). Filtering quality was checked out with *vcftools* (https://vcftools.github.io/man_latest.html).

Population genetic diversity

To characterize the genetic diversity of the collection sites, we estimated expected heterozygosity (H_s), observed heterozygosity (H_o), allelic richness (Ar) and fixation index (F_{IS}), all calculated, through *hierfstat* package^[36]. The number of alleles was estimated using the *adegenet* package^[37], and the number of private alleles was obtained through the *poppr* package^[38]. All analysis were performed in R version 4.3.0^[39].

Population genetic structure

Wright's F statistics (F_{ST} , F_{IT} , F_{IS}) and pairwise estimates of F_{ST} were calculated using the *hierfstat*^[36]. Pairwise estimates of F_{ST} were computed using Nei's distance^[40], and a heatmap was constructed with *heatmaply* package^[41].

To further investigate the genetic structure and inbreeding levels of *S. romanzoffiana*, we conducted an Analysis of Molecular Variance (AMOVA) following Excoffier et al.,^[42]. The hierarchical structure among populations was defined based on a grouping file, and loci with more than 5% missing data were

excluded. The analysis was performed using the *poppr* package^[38], and the significance of the AMOVA was tested through 20,000 permutations. A Mantel test^[28] was conducted to assess the correlation between genetic and geographic distances among populations, to test for isolation by distance. Genetic distances were estimated using Nei's method, and geographic distances were calculated from population coordinates. The analysis was performed with 10,000 permutations using the *ade4* package^[43]. All analysis were conducted in R version 4.3.0^[39].

Finally, we used discriminant analysis of principal components (DAPC) performed by the *adegenet* package^[37], to explore patterns of genetic differentiation and potential admixture among populations. The analysis was conducted using the collection sites as pre-defined genetic clusters. The optimal number of principal components to retain was determined using the α -score optimization procedure. We then generated both scatterplots and barplots to visualize population structure and individual assignment.

Environmental association and detection of candidate adaptive loci

To identify loci potentially under natural selection and associated with environmental gradients, we performed environmental association analysis using LatentFactor Mixed Models (LFMM). A total of 19 bioclimatic variables from the WorldClim v2 dataset were extracted, in addition to elevation data^[44]. To reduce collinearity among predictors, a Principal Component Analysis (PCA) was conducted on the entire set of variables. The first two principal components explained approximately 85% of the total environmental variance (Supplementary Figure 3). Based on this, the two variables most correlated with these components were retained: *bio06* (minimum temperature of the coldest month) and *bio19* (precipitation of the coldest quarter), which are commonly identified as drivers of local adaptation in tropical plants^{[45],[46],[47]}.

LFMM analysis were implemented using the *LEA* package^[48]. The number of latent factors ($K = 5$) was defined using the cross-entropy criterion from a preliminary population structure analysis with a sparse non-negative matrix factorization – sNMF, which was conducted with 10 repetitions and 200,000 iterations (Supplementary Figure 4). LFMM was run for 10,000 iterations with a burn-in of 5,000 and replicated ten times per environmental variable. To control for false positives, p-values were adjusted using the genomic inflation factor (λ) and corrected with the Benjamini–Hochberg procedure for false discovery rate control ($q < 0.05$)^{[49],[50]}. The SNPs significantly associated with the selected bioclimatic variables were retained as candidate loci under selection. These loci were used for two downstream analysis: (i) a second sNMF analysis to investigate adaptive genetic structure, and (ii) a Redundancy Analysis (RDA) to test for genotype–environment associations. Prior to the RDA, missing genotypes (up to 20%) were imputed using the *impute* function from LEA based on individual population assignments^[48]. The RDA was then performed using the *vegan* package^[51], and constrained axes were extracted to visualize the influence of climatic variables on population genetic composition. All analyses were again conducted in R version 4.3.0^[39].

Conservation Implications

The results obtained have significant implications for the conservation of the genetic diversity of *Syagrus romanzoffiana* populations. Recognizing that most genetic diversity is maintained within populations underscores the importance of conservation strategies that prioritize the protection of local habitats and promote practices that favor gene flow. The observed differentiation also suggests that, although populations exhibit genetic variation, conservation should be adapted to the local context, considering the particularities of each population. Moreover, the vegetation type and the terrains that populations are distributed seem to input local adaptations worth being considered in endeavors toward conservation of the species.

Conclusion

In summary, the analysis of the genetic structure of *Syagrus romanzoffiana* populations reveals a complex mosaic of genetic diversity, with considerable differentiation among populations, but also great diversity within them. The vegetation type seems to be an important component for the genetic structure of the populations, in spite of their geographic distance in some cases. The interaction between environmental factors, evolutionary history, and gene flow dynamics remains a rich area for future exploration, especially regarding their implications for the conservation and sustainable management of species in changing environments. Continued research will be vital to better understand these interactions and to implement effective preservation strategies.

Declarations

Data availability

The plant material used in this study was taxonomically identified by Dr. Enéas Ricardo Konzen, an expert in the taxonomy of palms of the genus *Syagrus*. No voucher specimens were deposited in a public herbarium; however, the corresponding genetic data generated from these samples are publicly available in NCBI's GenBank repository (<https://www.ncbi.nlm.nih.gov/>) under the accession number *Syagrus romanzoffiana* (PRJNA1295587; <http://www.ncbi.nlm.nih.gov/bioproject/1295587>).

Acknowledgements

We thank Prof. Guilherme Dubal dos Santos Seger and Ms. Francisco Agustin Vergara for his assistance in guiding and harvesting leaf samples from populations on a few sites of this study.

Authors' contributions

K.K.M.M. First author, conceptualization, performed the experiment, analyzed the results, data curation, original draft writing, review & editing; M.S. Performed the experiment, analyzed the results; A.F.F. Analyzed the results, review & editing; I.A.S.de C. Developing of GBS plates & DNA quantification; C.B.G. Developing of GBS plates; T.D.L. Responsible for submitting the data in the NCBI's GenBank repository

and preparing the data availability information; E.R.K. Collected the samples, responsible for the taxonomic identification of the plant material, data curation, review & editing; M.I.Z. Funding, conceptualization, data curation, review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Funding

We want to thank CAPES (Coordination for the Improvement of Higher Education Personnel) for the promotion granted to those involved. This study was supported by grants, including São Paulo Research Foundation (2011/50296-8), (2021/10319-0) and the National Council for Scientific and Technological Development (CNPq—313417/2023-7). The authors declare that they have no conflict of interest.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Additional information

Supplementary Information The online version contains supplementary material available at:

Correspondence and requests for materials should be addressed to K.K.M.M. and M.I.Z.

References

1. Cámara-Leret, R. *et al.* Fundamental species traits explain provisioning services of tropical American palms. *Nat. Plants* **3**, 16220 (2017).
2. Eiserhardt, W. L., Svenning, J.-C., Kissling, W. D. & Balslev, H. Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Ann. Bot.* **108**, 1391–1416 (2011).
3. Levis, C. *et al.* Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* **355**, 925–931 (2017).
4. Onstein, R. E. *et al.* Frugivory-related traits promote speciation of tropical palms. *Nat. Ecol. Evol.* **1**, 1903–1911 (2017).
5. Sen, T. & Samanta, S. K. Medicinal Plants, Human Health and Biodiversity: A Broad Review. in *Biotechnological Applications of Biodiversity* (ed. Mukherjee, J.) 59–110 (Springer, Berlin, Heidelberg, 2015). doi:10.1007/10_2014_273.

6. Gunjan, M. *et al.* Marketing trends & future prospects of herbal medicine in the treatment of various disease. *World Journal of Pharmaceutical Research* vol. 4 132–155 (2015).
7. Colombo, C. A., Berton, L. H. C., Diaz, B. G. & Ferrari, R. A. Macauba: a promising tropical palm for the production of vegetable oil. *OCL* **25**, D108 (2018).
8. Falasca, S. L., Miranda del Fresno, C. & Ulberich, A. Possibilities for growing queen palm (*Syagrus romanzoffiana*) in Argentina as a biodiesel producer under semi-arid climate conditions. *Int. J. Hydrog. Energy* **37**, 14843–14848 (2012).
9. Scariot, A. O. Palmeiras brasileiras: botânica, ecologia, usos e conservação. in *Palmeiras nativas do Brasil* 15–33 (Embrapa, Brasília, 2015).
10. Zona, S. & Henderson, A. A Review of Animal-Mediated Seed Dispersal of Palms. *Selbyana* **11**, 6–21 (1989).
11. Moreira, M. A. C. *et al.* Characterization of *Syagrus romanzoffiana* oil aiming at biodiesel production. *Ind. Crops Prod.* **48**, 57–60 (2013).
12. Helyar, S. J. *et al.* Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol. Ecol. Resour.* **11**, 123–136 (2011).
13. Brumfield, R. T., Beerli, P., Nickerson, D. A. & Edwards, S. V. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* **18**, 249–256 (2003).
14. Vinson, C. C., Mangaravite, E., Sebbenn, A. M. & Lander, T. A. Using molecular markers to investigate genetic diversity, mating system and gene flow of Neotropical trees. *Braz. J. Bot.* **41**, 481–496 (2018).
15. Boutet, G. *et al.* SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics* **17**, 121 (2016).
16. Hohenlohe, P. A. *et al.* Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLOS Genet.* **6**, e1000862 (2010).
17. Morgil, H. *et al.* Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. in *The Recent Topics in Genetic Polymorphisms* (IntechOpen, 2020). doi:10.5772/intechopen.91886.
18. Morin, P. A., Luikart, G., Wayne, R. K. & Group, the S. workshop. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* **19**, 208–216 (2004).
19. Wright, S. EVOLUTION IN MENDELIAN POPULATIONS. *Genetics* **16**, 97–159 (1931).
20. Hoffmann, A. A. & Sgrò, C. M. Climate change and evolutionary adaptation. *Nature* **470**, 479–485 (2011).
21. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945–959 (2000).
22. Zamith, L. R. & Scarano, F. R. Restoration of a Restinga Sandy Coastal Plain in Brazil: Survival and Growth of Planted Woody Species. *Restor. Ecol.* **14**, 87–94 (2006).
23. Vetö, N. M. *et al.* Population structure and signals of local adaptation in *Eugenia uniflora* (Myrtaceae), a widely distributed species in the Atlantic Forest. *Bot. J. Linn. Soc.* **201**, 100–113

- (2023).
24. Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**, 697–709 (2010).
 25. Wright, S. The Genetical Structure of Populations. *Ann. Eugen.* **15**, 323–354 (1949).
 26. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
 27. Griffiths, R. C. & Tavare, S. Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **344**, 403–410 (1997).
 28. Mantel, N. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.* **27**, 209–220 (1967).
 29. Wang, I. J. & Bradburd, G. S. Isolation by environment. *Mol. Ecol.* **23**, 5649–5662 (2014).
 30. Gaiotto, F. A., Grattapaglia, D. & Vencovsky, R. Genetic Structure, Mating System, and Long-Distance Gene Flow in Heart of Palm (*Euterpe edulis* Mart.). *J. Hered.* **94**, 399–406 (2003).
 31. Henderson, A., Galeano, G. & Bernal, R. *Field Guide to the Palms of the Americas*. (Princeton University Press, 1995).
 32. Lorenzi, H. *Flora Brasileira: Arecaceae (Palmeiras)*. (Instituto Plantarum, Nova Odessa, SP, 2010).
 33. Doyle, J. DNA Protocols for Plants. in *Molecular Techniques in Taxonomy* (eds Hewitt, G. M., Johnston, A. W. B. & Young, J. P. W.) 283–293 (Springer, Berlin, Heidelberg, 1991). doi:10.1007/978-3-642-83962-7_18.
 34. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* **7**, e32253 (2012).
 35. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
 36. Goudet, J. hierfstat, a package for r to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
 37. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
 38. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
 39. R Core Team. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
 40. Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, New York, NY, 1987).
 41. Galili, T., O’Callaghan, A., Sidi, J. & Sievert, C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**, 1600–1602 (2018).
 42. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).

43. Thioulouse, J. *et al.* Multivariate Analysis Graphs. in *Multivariate Analysis of Ecological Data with ade4* (eds Thioulouse, J. *et al.*) 53–76 (Springer, New York, NY, 2018). doi:10.1007/978-1-4939-8850-1_4.

44. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

45. Eckert, A. J. *et al.* Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits. *Genetics* **182**, 1289–1302 (2009).

46. Manel, S. *et al.* Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol. Ecol.* **19**, 3760–3772 (2010).

47. De Kort, H. *et al.* Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Mol. Ecol.* **23**, 4709–4721 (2014).

48. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).

49. François, O., Martins, H., Caye, K. & Schoville, S. D. Controlling false discoveries in genome scans for selection. *Mol. Ecol.* **25**, 454–469 (2016).

50. Ahrens, C. W., Byrne, M. & Rymer, P. D. Standing genomic variation within coding and regulatory regions contributes to the adaptive capacity to climate in a foundation tree species. *Mol. Ecol.* **28**, 2502–2516 (2019).

51. Oksanen, J. *et al.* *vegan: Community Ecology Package.* (2024).

Figures



Figure 1

Heatmap of pairwise F_{ST} values using 91 individuals of *Syagrus romanzoffiana* from different cities. Calculation based on Nei's distance (1987), using 24,859 SNPs.

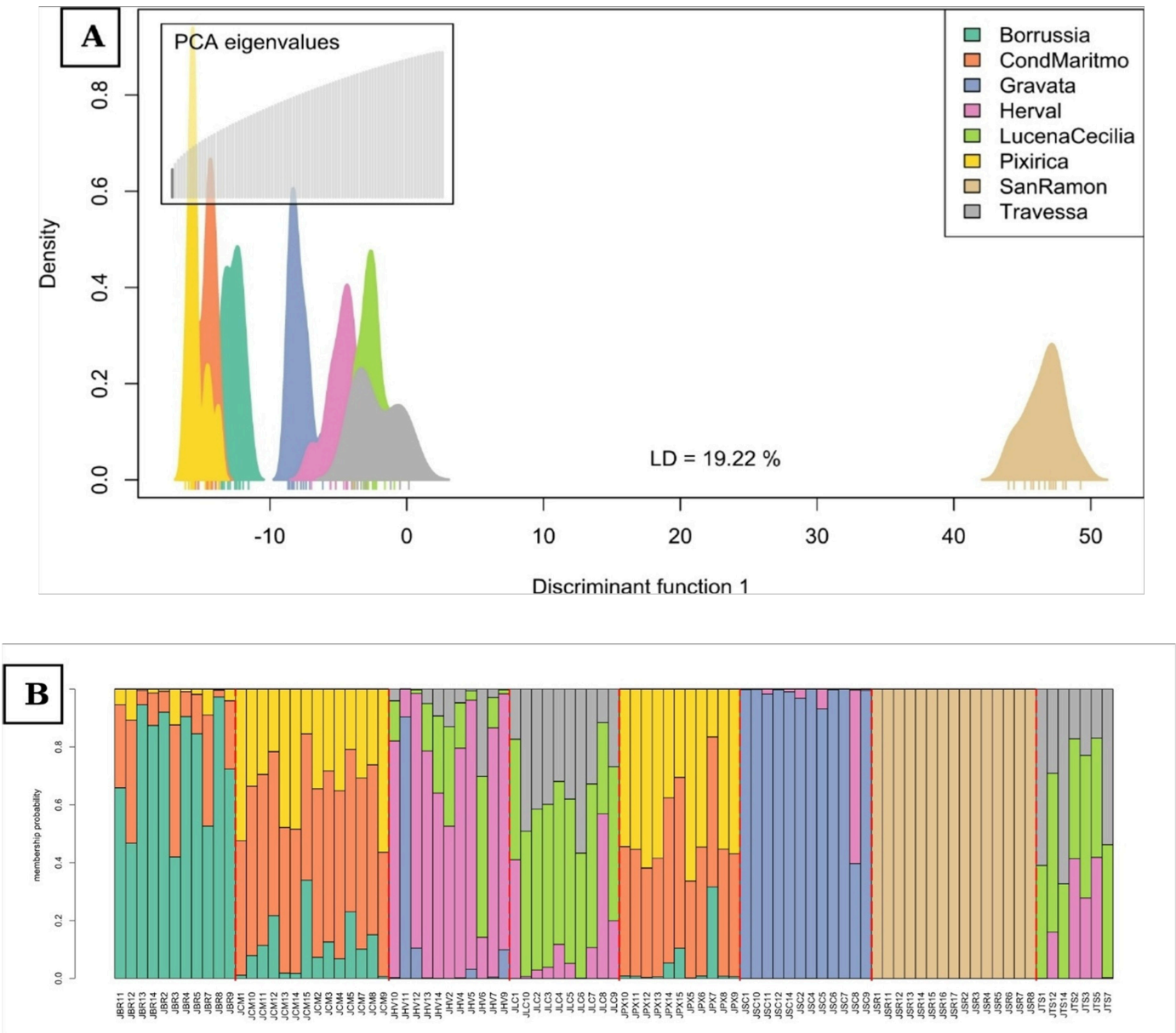


Figure 2

(A) Scatter Plot of Discriminant Analysis of Principal Components for 91 individuals from 8 *Syagrus romanzoffiana* populations. (B) Comomplot with genetic groups.

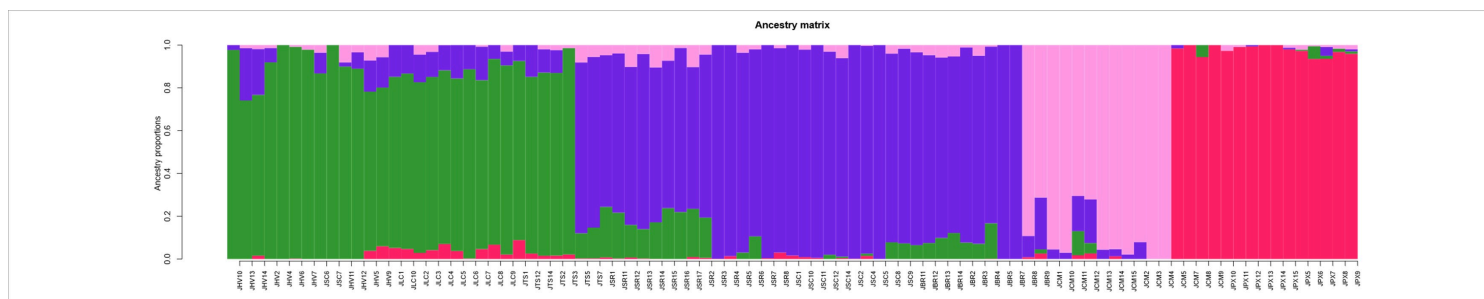


Figure 3

Genetic structure inferred by sNMF using only candidate adaptive SNPs identified by LFMM analysis ($q < 0.05$), with $K = 4$. Each bar represents an individual, and the colors indicate ancestry proportions associated with adaptive genetic clusters, suggesting patterns of local adaptation influenced by environmental factors.

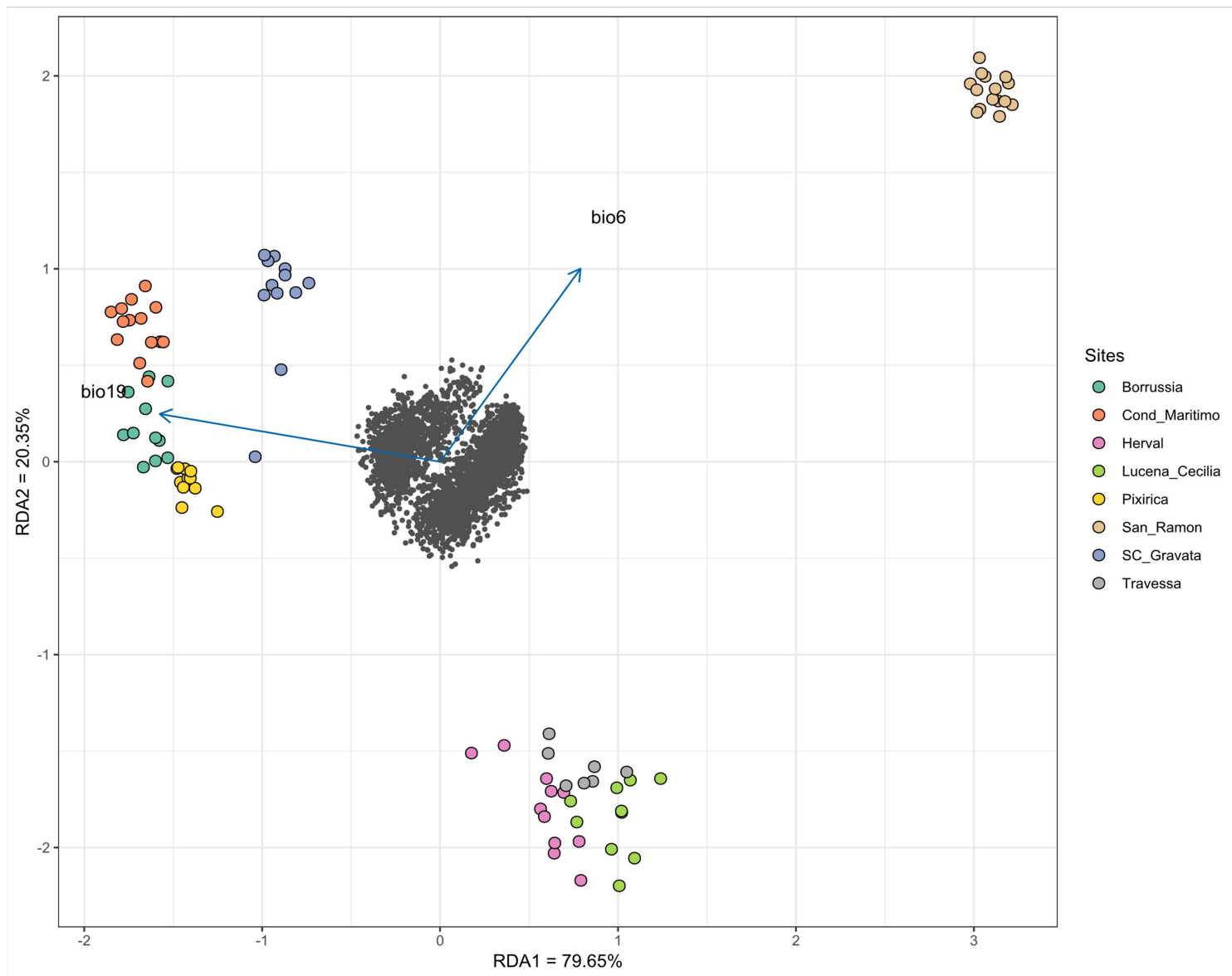


Figure 4

Redundancy analysis (RDA) based on candidate SNPs for selection identified with LFMM. The arrows represent the vectors of the bioclimatic variables bio06 (minimum temperature in the coldest month) and bio19 (rainfall in the coldest quarter), which explain the genotypic variation between *Syagrus romanzoffiana* populations. The colors represent the different sites sampled. The first two restricted axes explain more than 99% of the total adaptive variance.

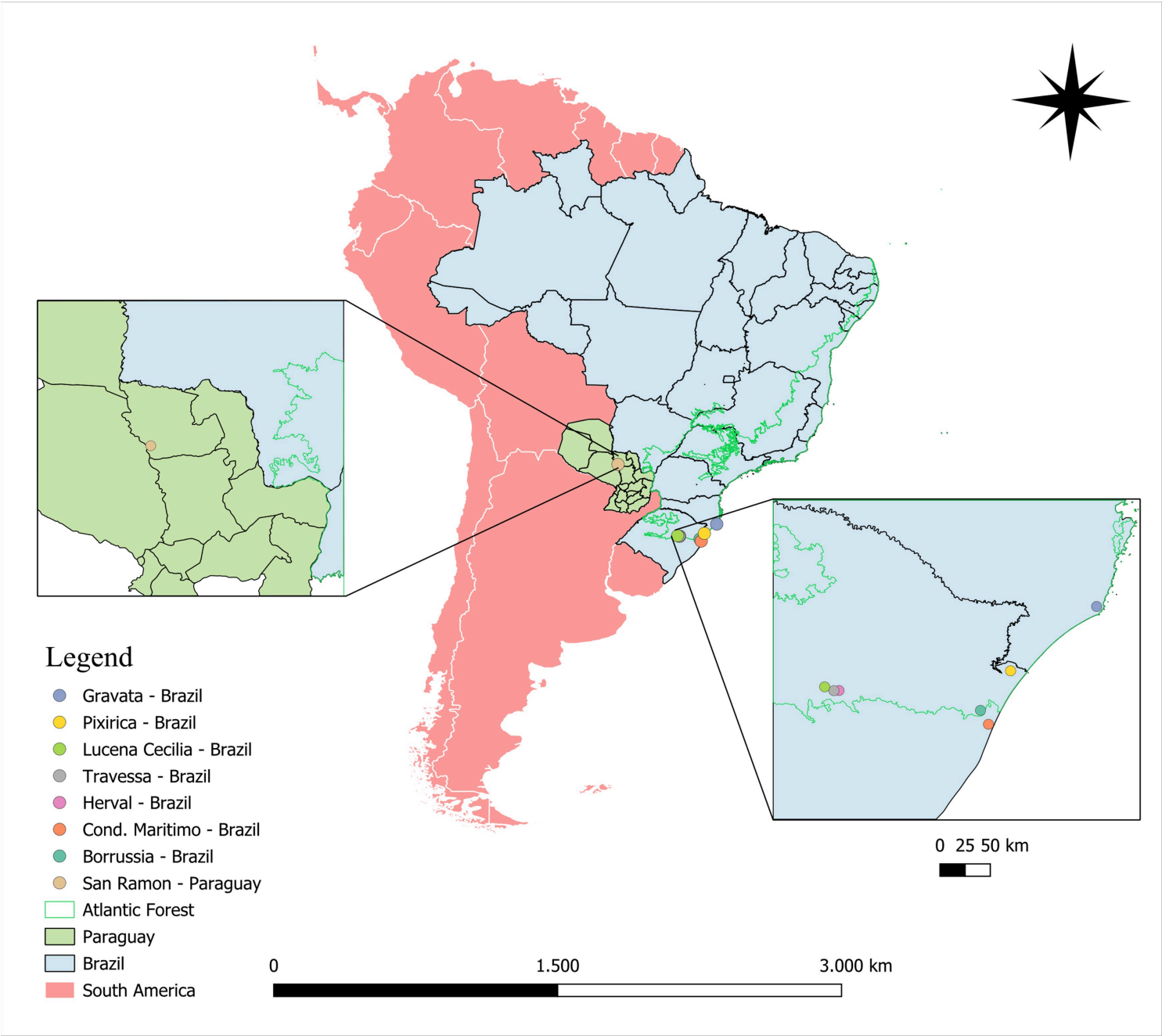


Figure 5

Map of the sampled populations of *Syagrus romanzoffiana* in Brazil and Paraguay. Each marked site corresponds to a sampled population, with dots representing the sampling sites. Populations in Brazil (Light Blue); Borrussia – RS (Dark Green); Cond. Maritimo – RS (Orange), Herval – RS (Pink), Lucena Cecilia – RS (Light Green), Pixirica – RS (Yellow), Travessa – RS (Gray), and Gravata – SC (Pale Blue).

Population in Paraguay (Pale Green): San Ramón – Concepción (Light Brown). Light Pink represents South America and Bright Green, Atlantic Forest biome.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialSyagrus.docx](#)