

Genome-wide SNPs reveal adaptation and population structure in *Syagrus romanzoffiana* (Arecaceae) from southern South America

Kauanne Karolline Moreno Martins

k228341@dac.unicamp.br

State University of Campinas

Matheus Scaketti

Universidade de São Paulo

Ana Flávia Francisconi

Universidade de São Paulo

Igor Araújo Santos de Carvalho

Universidade de São Paulo

Caroline Bertocco Garcia

Universidade de São Paulo

Thiago Deomar Ludwig

Universidade de São Paulo

Enéas Ricardo Konzen

Federal University of Rio Grande do Sul

Maria Imaculada Zucchi

Agência Paulista de Tecnologia dos Agronegócios

Research Article

Keywords: Arecaceae¹, Genetic diversity², Population genomics³, Molecular markers⁴, Atlantic forest⁵, Local adaptation⁶

Posted Date: October 31st, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7862000/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Conservation Genetics on May 14th, 2026.
See the published version at <https://doi.org/10.1007/s10592-026-01788-5>.

Abstract

Palms play crucial ecological and economic roles, and *Syagrus romanzoffiana* is a widely distributed species in South America. Despite its ecological importance and economic potential, population-level genetic studies of this palm remain limited. Here, we assessed the genetic diversity and structure of *S. romanzoffiana* across distinct vegetation types in Brazil and Paraguay using 24,859 single nucleotide polymorphisms (SNPs) derived from genotyping-by-sequencing (GBS). We analyzed 91 individuals from eight populations, revealing high genetic diversity within populations and moderate differentiation among them. Observed heterozygosity was generally high, with some populations showing an excess of heterozygotes, consistent with outcrossing and gene flow. Pairwise F_{ST} and AMOVA confirmed that most variation is maintained within populations, while genetic structure analyses highlighted the influence of vegetation type on differentiation, with *restinga* populations displaying unique signatures. Despite geographic proximity, some populations exhibited greater divergence due to local environmental conditions, supporting isolation by environment. Environmental association analyses (LFMM and RDA) further identified SNPs linked to temperature and precipitation, reinforcing the role of climate as a driver of adaptive differentiation. These findings emphasize the role of ecological factors in shaping genetic variation in *S. romanzoffiana* and illustrate how population genomics of widespread, non-threatened palms can provide critical baselines for conservation, restoration, and sustainable management under ongoing habitat fragmentation and climate change.

Introduction

Palm species (Arecaceae) stand out for their versatility and diversity, with more than 2,600 recognized species worldwide. Over centuries, palms have provided a wide range of services to humankind (Cámara-Leret et al. 2017; Eiserhardt et al. 2011; Levis et al. 2017). Many are ecological keystone species because large numbers of animals depend on their fruit and flower resources to survive (Onstein et al. 2017). Moreover, palms supply non-timber forest products such as food, medicine, fibers, oils, and building materials, supporting both traditional communities and modern industries (Sen and Samanta 2015; Gunjan et al. 2015; Colombo et al. 2018).

Syagrus romanzoffiana (Cham.) Glassman, from the Arecaceae family, commonly known in Brazil as jerivá, is a widespread palm native to southern and eastern South America, occurring in Brazil, Paraguay, northern Argentina and Uruguay (Lorenzi et al. 2010; Falasca et al. 2012). It plays a crucial ecological and socioeconomic role: sustaining frugivorous birds and mammals, is widely cultivated in landscaping and urban forestry, and supporting rural communities through diverse products derived from fruits, leaves, stems, and seeds (Scariot 2015; Zona and Henderson 1989).

Its distribution spans heterogeneous environments, from humid Atlantic Rain Forest to seasonal inland forests and coastal *restinga* systems, the latter characterized by sandy soils, high salinity and wind exposure (Lorenzi et al. 2010; Henderson et al. 1995; Scarano 2009). Such environmental variation

provides a natural framework to test for local adaptation and isolation-by-environment (IBE) (Wang and Bradburd 2014).

Understanding genetic diversity and population structure is fundamental for informed decisions on resource management, sustainable use, and conservation planning (Helyar et al. 2011; Brumfield et al. 2003). For *S. romanzoffiana*, analyzing genetic diversity is particularly relevant for guiding seed sourcing and restoration strategies in degraded Atlantic Forest and restinga habitats, while also providing baseline information that can be used in the future for breeding or domestication efforts according to local adaptation.

To investigate such genetic patterns, molecular markers are essential tools, widely applied in ecological, phylogenetic, and evolutionary studies (Vinson et al. 2018). To investigate such genetic patterns, molecular markers are essential tools, widely applied in ecological, phylogenetic, and evolutionary studies (Helyar et al. 2011; Boutet et al. 2016; Hohenlohe et al. 2010; Morgil et al. 2020; Morin et al. 2004).

In this study, we focused on the southern part of the distribution of *S. romanzoffiana*, sampling populations from southern Brazil and Paraguay to establish a regional genomic baseline. We hypothesized that (i) genetic structure would reflect not only geographic distance but also habitat type (e.g., restinga vs. inland forest), and (ii) climatic variables, particularly temperature and precipitation, would help explain adaptive genomic variation. Specifically, our objectives were to: (i) assess genetic diversity and structure using SNPs obtained through genotyping-by-sequencing (GBS), and (ii) investigate adaptive variation in relation to environmental variables. This is the first SNP-based population study of *S. romanzoffiana*, providing baseline information to support conservation and restoration planning in Atlantic Forest and restinga ecosystems.

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 southern South America: Atlantic Rain Forest, Atlantic Seasonal Forest, and restinga (Henderson et al. 1995; Lorenzi 2010). Sampling sites included Linha Lucena-Cecília, Travessa, and Herval (Venâncio Aires, RS, Brazil; Atlantic Seasonal Forest); San Ramón (Concepción, Paraguay; Seasonal Forest); Pixirica (Morrinhos do Sul, RS, Brazil) and Borrússia (Osório, RS, Brazil; Atlantic Rain Forest); and Condomínio Marítimo, (Cond. Marítimo) (Tramandaí, RS, Brazil) and Gravatá (Laguna, SC, Brazil; restinga) (Fig. 1). Leaf samples were stored in silica gel and transported to the Department of Genetics at Luiz de Queiroz College of Agriculture – University of São Paulo – ESALQ/USP, for DNA extraction.

This study was approved by the Ethics Committee of the University of Campinas and conducted in accordance with Brazilian Ministry of the Environment (MMA) regulations and the National System for

DNA extraction

DNA was extracted from leaf tissue using a modified CTAB protocol (Doyle 1991) optimized for palms. DNA integrity was checked on 1% agarose gels, and concentration was standardized to 30 ng/μL after quantification with a Qubit fluorometer (Thermo Fisher Scientific).

A detailed step-by-step description of the extraction protocol, including buffer composition, centrifugation conditions, and incubation times, is provided in Supplementary Information to facilitate reproducibility.

Genotyping-by-sequencing (GBS) library preparation

Genomic libraries were prepared using a double-digest GBS protocol (Poland et al. 2012) with modifications. Genomic DNA (30 ng/μL) was digested with PstI and MseI, followed by ligation of Illumina-compatible adapters with unique barcodes. Libraries were purified (QIAquick PCR Purification Kit, Qiagen), enriched by PCR with Illumina primers, and quantified by qPCR. Fragment size distribution was evaluated on a 2100 Bioanalyzer (Agilent Technologies), and sequencing was performed on the Illumina NextSeq 1000/2000 platform at Luiz de Queiroz College of Agriculture – University of São Paulo – (ESALQ-USP).

A complete description of adapter design, PCR conditions, and purification steps is provided in Supplementary Information to ensure reproducibility.

SNP discovery and filtering

Raw read quality was first assessed with *FastQC* (Andrews 2010). Demultiplexing and initial filtering were conducted in *Stacks* v2.62 (Catchen et al. 2013) using the *process_radtags* module, to remove low-quality reads (Phred < 10), adapter contamination, uncalled bases (Ns), and sequences lacking restriction sites. *De novo* assembly was performed in *Stacks*: *ustacks* was run with a minimum depth of coverage of 4 (–m 4), allowing up to 5 mismatches between stacks within individuals (–M 5) (Boutet et al. 2016). A catalog of loci was built with *cstacks* (tolerance of 4 mismatches among individuals) (Boutet et al. 2016), followed by *tsv2bam* and genotyping in *gstacks*.

Variant filtering was conducted in *VCFtools* (Danecek et al. 2011). To reduce linkage disequilibrium, one SNP per locus was retained for structure analyses. Filters applied were per-locus missingness ≤ 20%, per-individual missingness ≤ 20%, MAF ≥ 0.05, and Hardy–Weinberg equilibrium within populations ($p < 1 \times 10^{-6}$). Filtering choices followed the guidelines of Hemstrom et al. (2024).

Population genetic diversity

We estimated expected heterozygosity (H_s), observed heterozygosity (H_o), allelic richness (Ar) and fixation index (F_{IS}) using the *hierfstat* package (Goudet 2005). The number of alleles was calculated with *adegenet* package (Jombart and Ahmed 2011), and private alleles were identified using *poppr* package (Kamvar et al. 2014).

Population genetic structure

Wright's F statistics (F_{ST} , F_{IT} , F_{IS}) and pairwise estimates of F_{ST} were estimated using *hierfstat* (Goudet 2005). Pairwise genetic distances (Nei 1987) were visualized as a heatmap with *heatmaply* package (Galili et al. 2018). Analyses of Molecular Variance (Excoffier et al. 1992) were conducted in *poppr* package (Kamvar et al. 2014), and the significance of the AMOVA was tested through 20,000 permutations, using population identity as the hierarchical grouping factor. To minimize bias, loci with more than 5% missing data were excluded.

A Mantel test was performed with *ade4* (Mantel 1967) using 10,000 permutations testing for correlation between genetic and geographic distances (Thioulouse et al. 2018). Population structure was further explored with discriminant analysis of principal components (DAPC) using *adegenet* (Jombart and Ahmed 2011). The optimal number of PCs was determined using α -score optimization, and results were visualized with scatterplots and barplots.

Environmental association and detection of candidate adaptive loci

Environmental association analyses were conducted with Latent Factor Mixed Models (LFMM) implemented in the *LEA* package (Frichot and François 2015). 19 bioclimatic variables from WorldClim v2 (Fick and Hijmans 2017) and elevation were considered. To reduce collinearity among predictors, a Principal Component Analysis (PCA) was performed; the first two PCs explained ~ 85% of the variance (Figure S1, methodological). The most correlated variables were retained: bio06 (minimum temperature of the coldest month) and bio19 (precipitation of the coldest quarter), both commonly reported as drivers of local adaptation in tropical plants (Eckert et al. 2009; Manel et al. 2010; De Kort et al. 2014).

LFMM was run with $K = 5$ latent factors, defined by cross-entropy optimization from preliminary sNMF analyses (10 repetitions, 200,000 iterations; Figure S2, methodological). Each LFMM run used 10,000 iterations (burn-in = 5,000) and 10 replicates per variable. P-values were adjusted using genomic inflation (λ) and the Benjamini–Hochberg false discovery rate ($q < 0.05$) (François et al. 2016; Ahrens et al. 2019).

Candidate SNPs were used for two downstream analyses: (i) a second sNMF restricted to candidate loci, and (ii) Redundancy Analysis (RDA) in *vegan* (Oksanen et al. 2024). In the RDA, SNP genotypes were response variables, with bio06 and bio19 as constraints. Missing data ($\leq 20\%$) were imputed using the *LEA* impute function based on individual assignments (Frichot and François 2015). All analyses were performed in *R* v4.3.0 (R Core Team 2024).

Results

Genetic diversity in southern South 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% (Figure S3A). The average sequencing coverage per individual was 15.25x (Figure S3B), and the mean coverage per locus was 13.68x, with no coverage exceeding 100x or falling below 10x per locus (Figure S4A). The SNP data also revealed a higher number of transitions compared to transversions (Figure S4B).

Genetic diversity varies across sites (Table 1). 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, Gravatá 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). Allelic richness was generally higher in inland forest populations compared to *restinga* sites, with the exception of Gravatá, which also displayed high Ar values (1.473). The highest fixation index (f) showed variation among populations, with the highest value observed in Travessa (0.449), while negative values in Lucena-Cecília (−0.031), San Ramón (−0.028), Gravatá (−0.013), and Cond. Maritmo (−0.004), indicated an excess of heterozygotes.

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 Gravatá – 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
Cond.Maritimo	0.140	0.138	36125	1.387	653	-0.004
Gravata	0.174	0.167	38196	1.473	3291	-0.013
Herval	0.162	0.159	36631	1.452	364	0.259
Lucena-Cecilia	0.169	0.169	37555	1.492	329	-0.031
Pixirica	0.138	0.132	34935	1.372	476	0.013
San Ramón	0.205	0.203	40079	1.567	17426	-0.028
Travessa	0.160	0.163	35518	1.467	98	0.449

Population genetic structure in southern South America

Based on 24,859 SNPs, Wright's F statistics revealed significant differentiation among populations. The estimates across the eight sampled *S. romanzoffiana* populations were: total inbreeding coefficient ($F_{IT} = 0.139$), within-population inbreeding coefficient ($F_{IS} = 0.021$), and genetic differentiation coefficient among populations ($F_{ST} = 0.122$). Pairwise F_{ST} analyses indicated that the San Ramón was consistently the most divergent population, with values reaching 0.371 compared to Pixirica, while the lowest differentiation was observed between Borrússia and Pixirica $F_{ST} = 0.079$, both located in Atlantic Rain Forest (Fig. 2). In contrast, Borrússia and Cond. Maritimo, despite being geographically closer (~ 20 km apart), showed higher differentiation ($F_{ST} = 0.081$), reflecting the influence of habitat type (*restinga* vs. forest) on genetic structure (Fig. 2).

These pairwise comparisons highlight both geographic and ecological influences on population structure. Consistently, AMOVA confirmed significant differentiation among populations, with 29.7% of the genetic variance distributed among populations and 70.3% within populations (Table 2). These findings indicate that most of the genetic diversity is maintained within populations, while moderate structure exists among them. This pattern may result from substantial gene flow within populations and ecological or geographic factors constraining gene flow between populations. However, the Mantel test showed a non-significant correlation between genetic and geographic distances ($r = -0.009$, $p = 0.141$), indicating no evidence of isolation by distance.

Table 2
Analyses 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.

Although no isolation by distance was detected, multivariate approaches further clarified population structure. Discriminant Analysis of Principal Components (DAPC) identified clear population structure. The first discriminant function (LD1) explained 19.22% of the variation and separated San Ramón from all other sites (Fig. 3A). The remaining populations showed partial overlap along LD1, reflecting more subtle differentiation. Membership probabilities (Fig. 3B) further supported these results, with San Ramón forming a distinct cluster, while *restinga* populations (Cond. Maritimo and Gravatá) tended to group together and inland forest populations showed mixed assignment.

Detection of adaptive loci and environmental associations

LFMM identified 3,768 SNPs significantly associated with bio06 (minimum temperature of the coldest month) and bio19 (precipitation of the coldest quarter) (Figures S5–S6). A new sNMF analysis based on these candidate SNPs identified a population structure with $K = 4$ clusters (Fig. 4), indicating that environmental variables contribute to population differentiation beyond neutral processes. San Ramón and the two *restinga* sites formed distinct clusters, while inland forest populations showed partial admixture.

Redundancy Analysis (RDA) based on candidate SNPs confirmed strong genotype–environment associations. The global model constrained by bio19 and bio06 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 constrained axes explained nearly all of the variance in candidate loci (99.9%, Fig. 5).

Individuals from San Ramón were positively associated with bio06, indicating adaptation to colder minimum temperatures, while Cond. Maritimo, Pixirica, Borrússia, and Gravatá were strongly associated with bio19, suggesting adaptive responses to precipitation seasonality in the southern Atlantic Forest and *restinga* environments. These findings confirm that climatic variables related to temperature and precipitation are important drivers of adaptive differentiation in *S. romanzoffiana*.

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, with H_o exceeding H_s in most sites, indicate a potential excess of heterozygotes. This pattern may reflect adaptive responses to local environmental conditions or a mating system that favors outcrossing among genetically distinct individuals. *S. romanzoffiana* is pollinated mainly by nitidulid beetles and other insects and has its fruits dispersed by birds and mammals (Zona and Henderson 1989), which contributes to maintaining gene flow and elevated heterozygosity. Populations with high heterozygosity are generally more resilient to environmental pressures, which may be crucial for persistence in altered habitats (Hoffmann and Sgrò 2011).

In addition to heterozygosity, allelic richness (Ar) was particularly notable in San Ramón and Lucena-Cecília, highlighting their genetic potential for adaptation and long-term evolutionary resilience (Pritchard et al. 2000). The high number of private alleles (Ap) found in Cond. Maritimo and Gravatá emphasizes the genetic uniqueness of *restinga* populations, likely reflecting specific local adaptations to coastal environments characterized by sandy soils, high salinity, and strong wind exposure (Zamith and Scarano 2006). Comparable results have been reported for other neotropical trees, such as *Eugenia uniflora* (Myrtaceae) (Vetö et al. 2023). Similar patterns of habitat-associated divergence have been suggested in palms. For instance, *Euterpe edulis* shows structuring of alleles and signals of selection across contrasting microhabitats (Conte et al. 2008).

Patterns of fixation indices further support the influence of habitat context. High positive (f) values, as observed in Travessa and Herval, suggest increased homozygosity, potentially resulting from inbreeding

in fragmented landscapes. Smaller but positive values in Borrússia and Pixirica also indicate a slight trend toward homozygosity. By contrast, negative values in Lucena-Cecília, San Ramón, Gravatá, and Cond. Maritimo indicate an excess of heterozygotes, which may arise from outcrossing and gene flow in larger or more environmentally heterogeneous populations (Allendorf et al. 2010). Together, these results illustrate how habitat fragmentation and population size shape mating dynamics in palms, with consequences for the maintenance of genetic diversity.

Wright's F-statistics revealed moderate differentiation ($F_{ST} = 0.122$), indicating that while gene flow persists, populations remain sufficiently isolated to maintain genetic differences (Wright 1931, 1949). AMOVA corroborated this finding, showing that most genetic variation occurs within populations (70.3%). This high intrapopulational diversity is consistent with expectations for long-lived, outcrossing tropical trees (Anderson 2001; Griffiths and Tavaré 1997), where large effective population sizes and extensive gene flow maintain diverse alleles within local gene pools.

Interestingly, the lack of a significant correlation between genetic and geographic distances in the Mantel test indicates that geographic separation is not the main driver of differentiation. Instead, ecological factors appear to play a greater role. For instance, Borrússia and Pixirica, although ~ 90 km apart, were genetically more similar to each other than Borrússia and Cond. Maritimo, which are separated by only ~ 20 km but differ in vegetation type (*restinga* vs. Atlantic Rain Forest). This contrast reinforces the importance of isolation by environment (IBE) over isolation by distance, highlighting the role of local ecological conditions and microclimates in shaping genetic divergence (Mantel 1967; Wang and Bradburd 2014).

Consistent with this interpretation, environmental association analyses provided compelling evidence for local adaptation in *S. romanzoffiana*. SNPs associated with climatic variables (bio06 and bio19) highlight the role of temperature and precipitation as selective pressures across vegetation types. Notably, adaptive genetic structure emphasized that selection could act independently of geographic distance or historical demography. Similar patterns of environment-driven divergence have been documented in *Euterpe edulis* (Gaiotto et al. 2003), supporting the idea that neotropical palms frequently undergo local adaptation even across relatively small spatial scales. In this study, 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.

Collectively, these results reinforce the importance of integrating adaptive genetic variation into conservation planning. For species distributed across ecotonal gradients and fragmented landscapes, such as *S. romanzoffiana*, management strategies should explicitly account for both neutral and adaptive processes to preserve evolutionary potential under ongoing environmental change. Although we did not implement a separate Multiple Matrix Regression (MMRR) test, the Redundancy Analysis (RDA) already fulfills this role by formally partitioning genetic variance explained by climatic variables while accounting for neutral structure. Thus, our approach provides a direct test of isolation by environment. We also acknowledge that demographic history (e.g., bottlenecks or drift) was not explicitly tested here, and

future studies with broader sampling or whole-genome data will be necessary to disentangle the relative contributions of drift versus selection to the observed genetic patterns.

Conservation Implications

The results obtained provide important insights into the conservation and management of genetic diversity in *Syagrus romanzoffiana*. Although this palm is widespread and not considered threatened, discussing its genetic patterns in a conservation framework is still relevant. First, genetic variation is not homogeneously distributed across its range: most diversity is maintained within populations, underscoring the importance of protecting local habitats that sustain viable population sizes and gene flow. Second, the distinctiveness of populations from Paraguay and *restinga* habitats demonstrates the value of conserving ecologically unique areas that harbor locally adapted variants. Moreover, because *S. romanzoffiana* is ecologically important and widely used in restoration, landscaping, and local livelihoods, preserving its adaptive genetic variation has practical significance for long-term sustainability. Rather than representing a species-wide conservation concern, the relevance of our findings lies in integrating this genomic baseline into landscape management, restoration initiatives, and sustainable use programs, ensuring that adaptive variation is preserved and can contribute to resilience under ongoing environmental change.

Conclusion

In summary, the analyses of the genetic structure of *S. romanzoffiana* populations revealed moderate differentiation among populations and high diversity within them. The vegetation type appears to be an important factor shaping genetic structure, sometimes overriding geographic distance. These results suggest that local ecological conditions, together with evolutionary history and patterns of gene flow, influence population differentiation. Importantly, our study provides a regional genomic baseline for a widespread but ecologically and socioeconomically relevant palm. Although *S. romanzoffiana* is not threatened, understanding the distribution of its neutral and adaptive genetic variation is valuable for restoration programs, urban forestry, and the management of Atlantic Forest and *restinga* landscapes. Future research should aim to disentangle the relative roles of neutral and adaptive processes in shaping genetic variation, thereby strengthening the integration of genomic data into conservation and sustainable management of neotropical palms. These results illustrate the value of applying population genomics to widespread tropical palms, showing how even non-threatened species provide key baselines for conservation and management in changing environments.

Declarations

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.

Competing interests

The authors declare that they have no conflict of interest.

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.

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>).

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the University of Campinas and conducted in accordance with Brazilian Ministry of the Environment (MMA) regulations and the National System for Genetic Heritage and Associated Traditional Knowledge (SisGen; registration A3A9281).

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

References

1. Ahrens CW, Byrne M, Rymer PD (2019) 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. <https://doi.org/10.1111/mec.15092>

2. Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* 11:697–709. <https://doi.org/10.1038/nrg2844>
3. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
4. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
5. Boutet G, Alves-Carvalho S, Falque M et al (2016) SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics* 17:121. <https://doi.org/10.1186/s12864-016-2447-2>
6. Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* 18:249–256. [https://doi.org/10.1016/S0169-5347\(03\)00018-1](https://doi.org/10.1016/S0169-5347(03)00018-1)
7. Cámara-Leret R, Fortuna MA, Albuixech J et al (2017) Fundamental species traits explain provisioning services of tropical American palms. *Nat Plants* 3:16220. <https://doi.org/10.1038/nplants.2016.220>
8. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140. <https://doi.org/10.1111/mec.12354>
9. Colombo CA, Berton LHC, Diaz BG, Ferrari RA (2018) Macauba: a promising tropical palm for the production of vegetable oil. *OCL* 25:D108. <https://doi.org/10.1051/ocl/2017038>
10. Conte R, Reis MS, Mantovani A, Vencovsky R (2008) Genetic Structure and Mating System of *Euterpe edulis* Mart. Populations: A Comparative Analysis Using Microsatellite and Allozyme Markers. *J Hered* 99:476–482. <https://doi.org/10.1093/jhered/esn055>
11. Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
12. De Kort H, Vandepitte K, Mergeay J et al (2014) 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. <https://doi.org/10.1111/mec.12813>
13. Doyle J (1991) DNA protocols for plants. In: Hewitt GM, Johnston AWB, Young JPW (eds) *Molecular Techniques in Taxonomy*. Springer, Berlin, pp 283–293. https://doi.org/10.1007/978-3-642-83962-7_18
14. Eckert AJ, van Heerwaarden J, Wegrzyn JL et al (2009) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* 182:1289–1302. <https://doi.org/10.1534/genetics.109.102350>
15. Eiserhardt WL, Svenning JC, Kissling WD, Balslev H (2011) Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Ann Bot* 108:1391–1416. <https://doi.org/10.1093/aob/mcr146>
16. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.

- Genetics 131:479–491. <https://doi.org/10.1093/genetics/131.2.479>
17. Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37:4302–4315. <https://doi.org/10.1002/joc.5086>
 18. Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. *Methods Ecol Evol* 6:925–929. <https://doi.org/10.1111/2041-210X.12382>
 19. François O, Martins H, Caye K, Schoville SD (2016) Controlling false discoveries in genome scans for selection. *Mol Ecol* 25:454–469. <https://doi.org/10.1111/mec.13513>
 20. Gaiotto FA, Grattapaglia D, Vencovsky R (2003) Genetic structure, mating system, and long-distance gene flow in heart of palm (*Euterpe edulis Mart*). *J Hered* 94:399–406. <https://doi.org/10.1093/jhered/esg087>
 21. Galili T, O’Callaghan A, Sidi J, Sievert C (2018) heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* 34:1600–1602. <https://doi.org/10.1093/bioinformatics/btx657>
 22. Goudet J (2005) hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* 5:184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
 23. Henderson A, Galeano G, Bernal R (1995) *Field Guide to the Palms of the Americas*. Princeton University Press, Princeton. <https://doi.org/10.2307/j.ctvczzzd>
 24. Hemstrom K, Xuereb A, Prunier J et al (2024) Next-generation data filtering in the genomics era. *Nat Rev Genet* 25:750–767. <https://doi.org/10.1038/s41576-024-00738-6>
 25. Helyar SJ, Hemmer-Hansen J, Bekkevold D et al (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol Ecol Resour* 11:123–136. <https://doi.org/10.1111/j.1755-0998.2010.02943.x>
 26. Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature* 470:479–485. <https://doi.org/10.1038/nature09670>
 27. Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
 28. Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281. <https://doi.org/10.7717/peerj.281>
 29. Levis C, Costa FRC, Bongers F et al (2017) Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* 355:925–931. <https://doi.org/10.1126/science.aal0157>
 30. Lorenzi H (2010) *Flora Brasileira: Arecaceae (Palmeiras)*. Instituto Plantarum, Nova Odessa
 31. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
 32. Manel S, Joost S, Epperson BK et al (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Mol Ecol* 19:3760–3772. <https://doi.org/10.1111/j.1365-294X.2010.04717.x>

33. Onstein RE, Baker WJ, Couvreur TLP et al (2017) Frugivory-related traits promote speciation of tropical palms. *Nat Ecol Evol* 1:1903–1911. <https://doi.org/10.1038/s41559-017-0348-7>
34. Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. <https://doi.org/10.1371/journal.pone.0032253>
35. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959. <https://doi.org/10.1093/genetics/155.2.945>
36. R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.r-project.org/>
37. Scarano FR (2009) Plant communities at the periphery of the Atlantic rain forest: rare-species bias and its risks for conservation. *Biol Conserv* 142:1201–1208. <https://doi.org/10.1016/j.biocon.2009.02.027>
38. Thioulouse J, Dray S, Dufour AB, Siberchicot A, Jombart T, Pavoine S (2018) Multivariate analysis of ecological data with ade4. Springer, New York
39. Vetö NM, Oliveira GCX, Lemos-Filho JP et al (2023) 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. <https://doi.org/10.1093/botlinnean/boac012>
40. Wang IJ, Bradburd GS (2014) Isolation by environment. *Mol Ecol* 23:5649–5662. <https://doi.org/10.1111/mec.12938>
41. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159. <https://doi.org/10.1093/genetics/16.2.97>
42. Wright S (1949) The genetical structure of populations. *Ann Eugen* 15:323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
43. Zamith LR, Scarano FR (2006) Restoration of a *restinga* sandy coastal plain in Brazil: survival and growth of planted woody species. *Restor Ecol* 14:87–94. <https://doi.org/10.1111/j.1526-100X.2006.00108.x>
44. Zona S, Henderson A (1989) A review of animal-mediated seed dispersal of palms. *Selbyana* 11:6–21

Figures

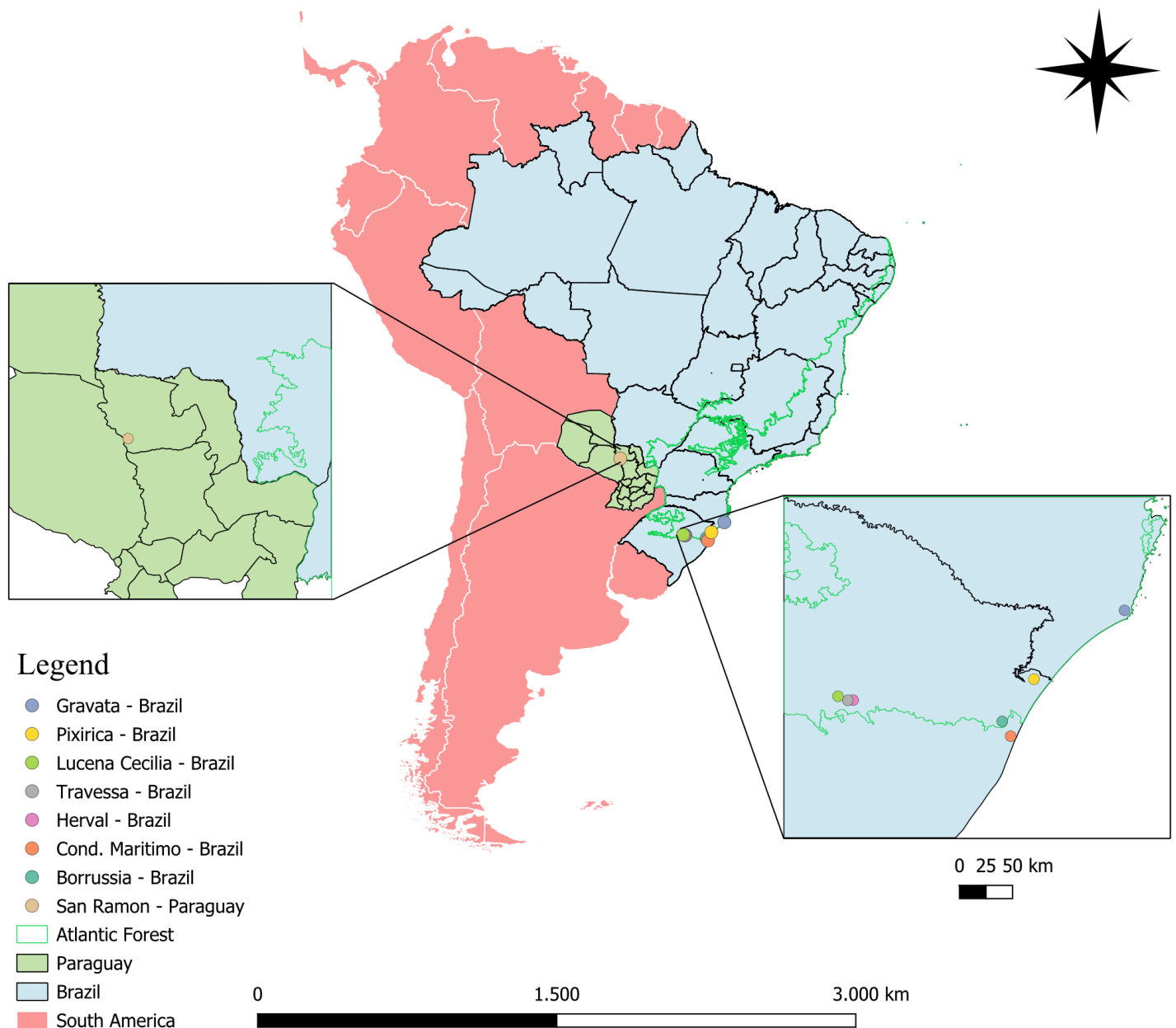


Figure 1

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.

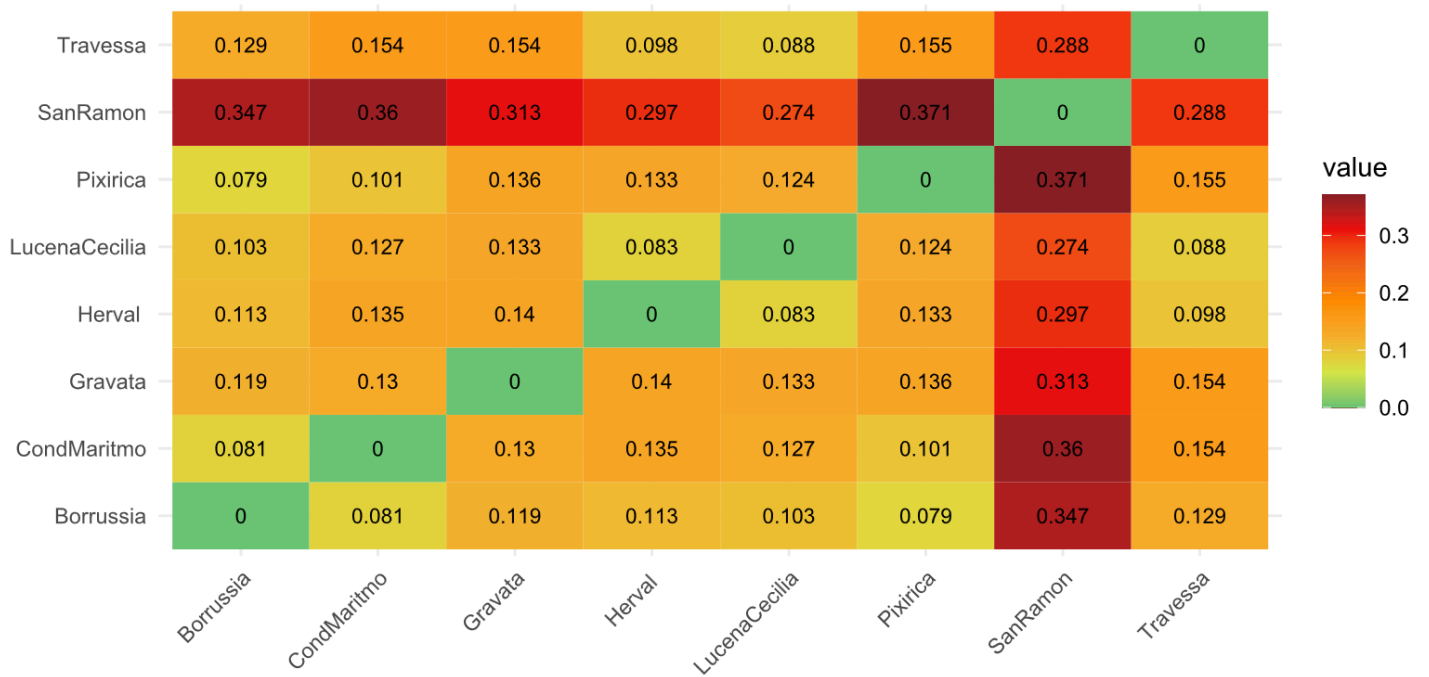


Figure 2

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.

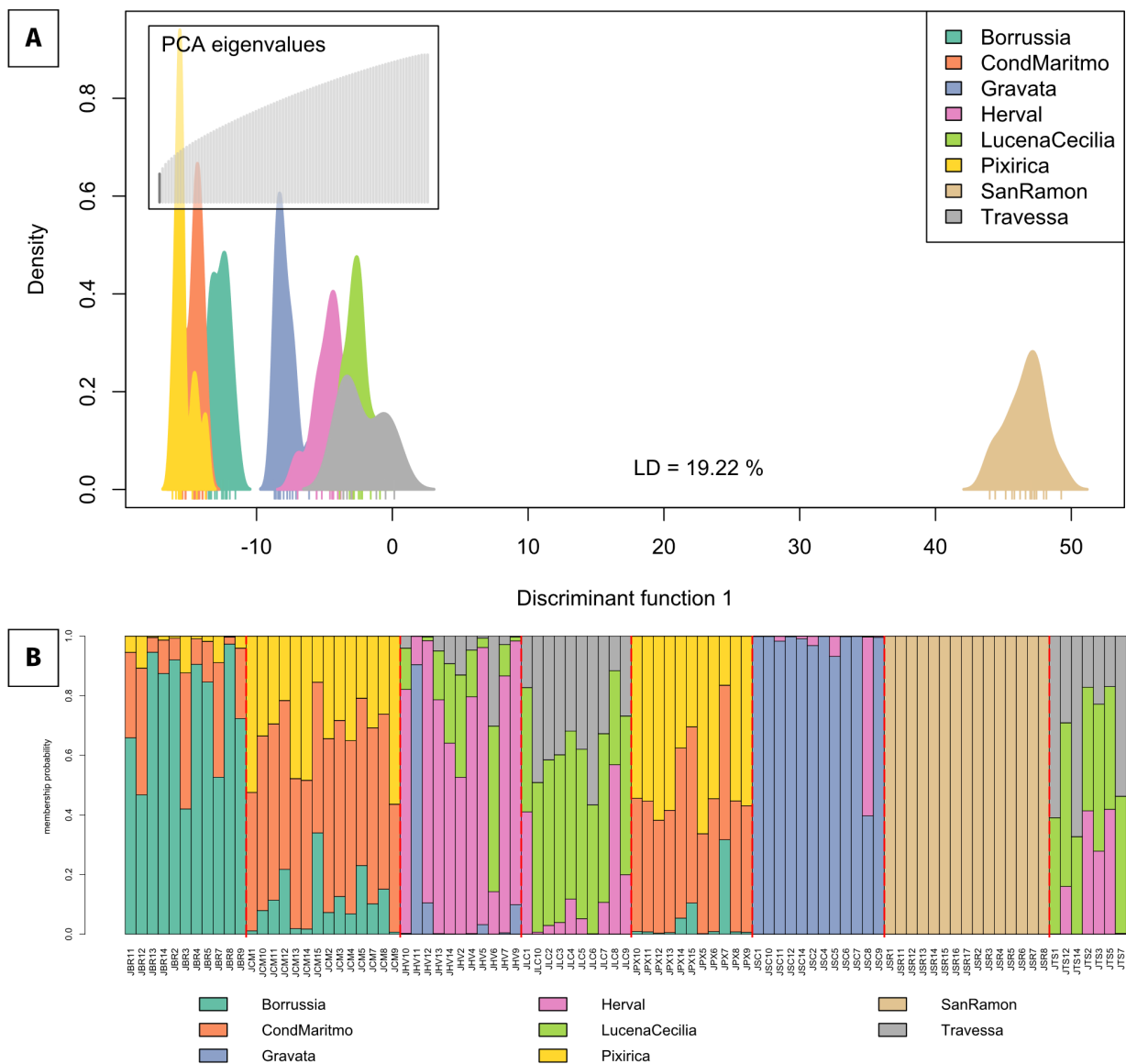


Figure 3

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

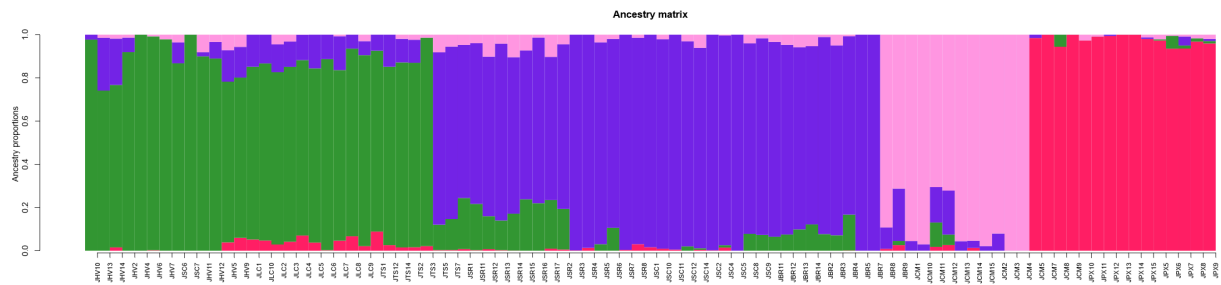


Figure 4

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.

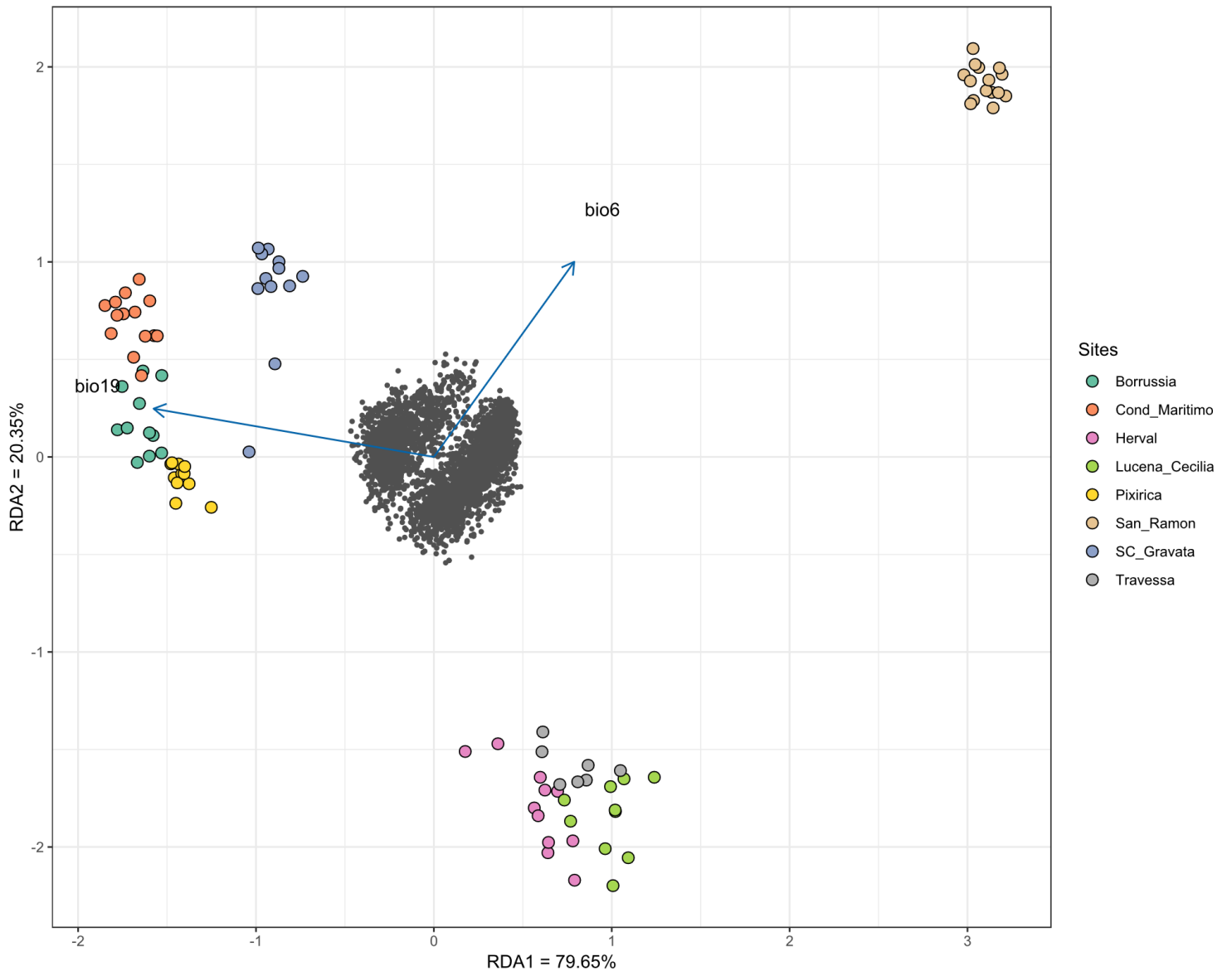


Figure 5

Page 19/20

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 nearly all 99% of the total adaptive variance.

Supplementary Files

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

- [SupplementaryInformationsSyagrus.docx](#)