



# Linking environmental stability with genetic diversity and population structure in two Atlantic Forest palm trees

Kathryn P. Mercier<sup>1,2</sup> | Mariana M. Vasconcellos<sup>1,3</sup> | Euder G. A. Martins<sup>4</sup> |  
 José Rubens Pirani<sup>4</sup> | Fabián A. Michelangeli<sup>2,3</sup> | Ana Carolina Carnaval<sup>1,2</sup>

<sup>1</sup>Department of Biology, City College of New York, New York, New York, USA

<sup>2</sup>PhD Program in Biology, The Graduate Center of the City University of New York, New York, New York, USA

<sup>3</sup>New York Botanical Garden, Institute of Systematic Botany, Bronx, New York, USA

<sup>4</sup>Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

#### Correspondence

Kathryn P. Mercier, Department of Biology at the City College of New York, New York, NY, USA.

Email: [mercier.kate@gmail.com](mailto:mercier.kate@gmail.com)

#### Present address

Mariana M. Vasconcellos, Departamento de Zoologia, Instituto de Biosciências, Universidade de São Paulo, São Paulo, Brazil

#### Funding information

Fundaão de Amparo à Pesquisa do Estado de São Paulo; National Science Foundation

Handling Editor: Anna Papadopoulou

## Abstract

**Aim:** Spatial patterns of biodiversity in the Atlantic Forest are well characterized. However, there is no consensus on the biological processes underlying these patterns, and multiple competing hypotheses have been proposed, several of which centre on climatic stability. Here, we ask if Late Quaternary climatic stability predicts contemporary population structure and genomic-level diversity in two palm species.

**Location:** Atlantic Forest of Eastern Brazil.

**Taxa:** *Syagrus botryophora* and *S. pseudococos* (Arecaceae).

**Methods:** We first model the distribution of suitable environments in 62 time-slice climate projections over the last 120 thousand years, and summarize stability over that period. We then use >25,000 RADseq-generated SNPs to (i) describe the spatial patterns of genomic variation in both species, (ii) test how well genomic variation is explained by isolation by distance and by the environmental resistance imposed by historical instability (isolation by resistance) and (iii) test for a correlation between genetic diversity and historical stability.

**Results:** The contemporary range of *S. botryophora* has been relatively stable over the last 30 thousand years and there are two isolated regions of high stability for *S. pseudococos*. The genomic data recover a clear pattern of isolation by distance in *S. botryophora* and two structured populations in *S. pseudococos*. Consequently, the contribution of isolation by resistance to overall genetic structure is much higher in *S. pseudococos*. Genetic diversity is not significantly correlated with historical stability in either species.

**Main Conclusions:** Based on the concordance between historical stability and genetic structure, Late Quaternary climate stability may have maintained population connectivity within *S. botryophora* and promoted intraspecific divergence in *S. pseudococos*. Conversely, historical stability does not seem to be driving spatial patterns of genetic diversity. This study supports the primary role of climatic stability in determining spatial population structure, but not genetic diversity, in the Atlantic Forest.

#### KEY WORDS

Arecaceae, biodiversity, Brazil, phylogeography, population genetics, RADseq, species distribution modelling



## 1 | INTRODUCTION

The magnitude and extent of climatic stability in a given region has been portrayed as a central determinant of broad patterns of biodiversity (Fjeldså & Lovett, 1997; Hewitt, 2000). At the intraspecific level, climate stability allows populations to persist over longer periods of time compared to unstable regions (Ashcroft, 2010). Long-term persistence is expected to enable the accumulation of genetic diversity (Leffler et al., 2012), and to maintain high levels of gene flow across the region (Fjeldså & Lovett, 1997). Whenever continuous gene flow allows local populations to reach equilibrium, the levels of genetic structuring become low, and genetic isolation by distance (IBD; Wright, 1943) is observed. In contrast, low stability areas are expected to prevent gene flow within the range of a species, leading to genetic divergence (Wright, 1931). As such, populations separated by low stability regions often show high levels of population structure, and a signature of genetic isolation by environmental resistance (IBR; McRae, 2006).

In the Atlantic Forest of Brazil, a climatically diverse biodiversity hotspot (Ribeiro, Martensen, et al., 2011), spatial patterns of biodiversity are well characterized—including the identification of distinct centres of endemism and regions of community turnover (Peres et al., 2020; Werneck et al., 2011). Several hypotheses have been proposed to explain these biogeographical patterns, and many centre on repeated cycles of forest expansion and retraction driven by climatic shifts in the late Quaternary (Damasceno et al., 2014; Vanzolini & Williams, 1981). Still, studies linking historical climatic stability and spatial patterns of genetic diversity within Atlantic Forest species have been largely restricted to animal taxa, particularly vertebrates (Almeida et al., 2020; Brunes et al., 2015; Cabanne et al., 2008; Carnaval et al., 2009; Carnaval & Moritz, 2008; Damasceno et al., 2014; Menezes et al., 2017; Silveira et al., 2019; Thom et al., 2020; Thomé et al., 2010; Valdez & D'Elía, 2013). Little is known about how past climate change influenced the population demography, and therefore patterns of intraspecific genetic diversity, in plants (but see Aguiar-Melo et al., 2019; Neves et al., 2020; Pinheiro et al., 2011; Teixeira et al., 2016).

One geographical area of interest flagged by studies of climate and biodiversity in the Atlantic Forest is the central corridor that extends from southern Bahia to northern Rio de Janeiro (Carnaval et al., 2009; Carnaval & Moritz, 2008; Ledru et al., 2007). Evidence suggests that this area was relatively stable over the climate fluctuations of the Pleistocene, serving as a climate refugium that maintained high levels of genetic diversity in many lineages of amphibians (Carnaval et al., 2009), birds (Cabanne et al., 2008), mammals (de Moraes-Barros et al., 2006) and invertebrates (Batalha-Filho et al., 2010). Yet, while the central corridor has been identified as a centre of endemism for woody plants (Prance, 1982; Werneck et al., 2011), it remains to be seen whether it served as a botanical refuge through climatic stability. We ask if and how the spatial distribution of climatic stability relate to the present-day distribution of

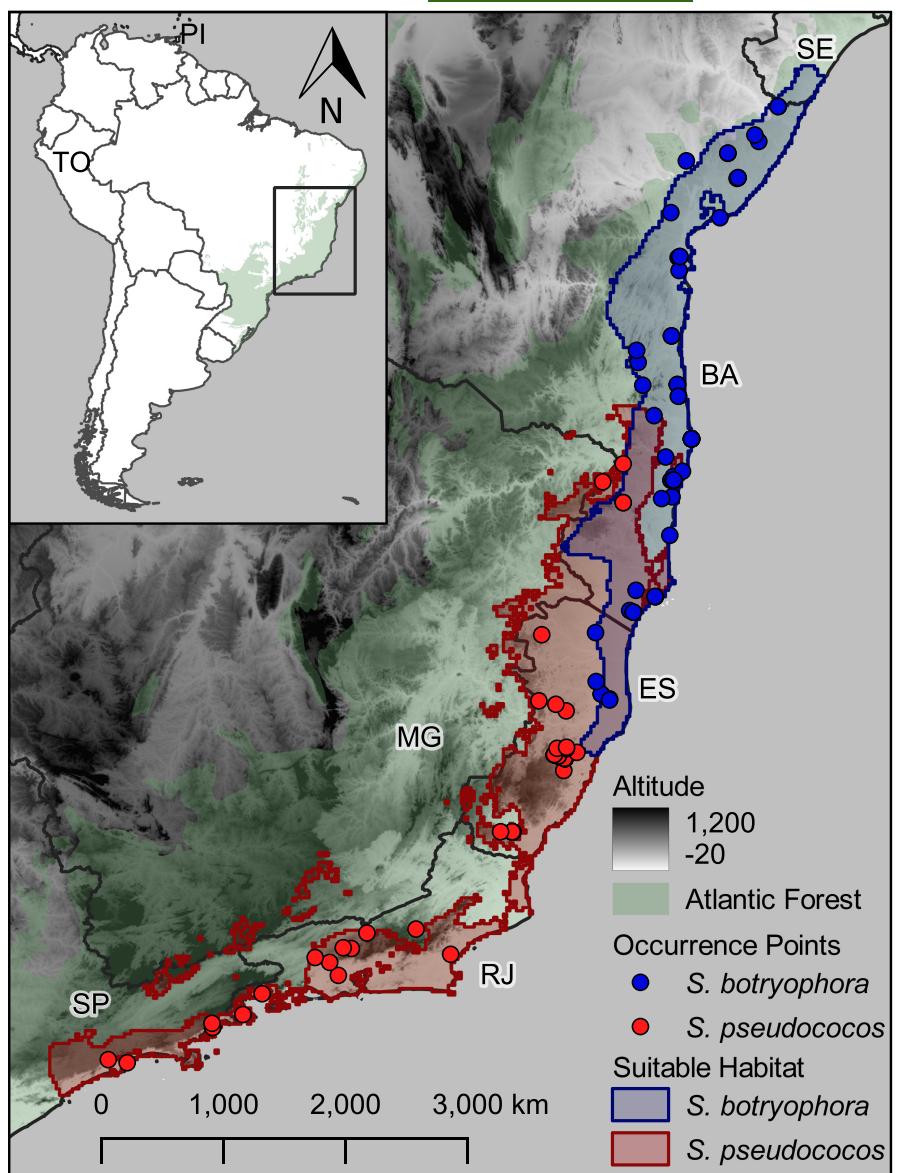
genetic diversity in two closely related species of palm trees endemic to the Atlantic Forest, which occur—although not exclusively—in the central corridor: *Syagrus botryophora* Mart. and *S. pseudococos* (Raddi) Glassman (Figure 1).

*Syagrus* species are generally wind- and insect-pollinated (de Medeiros & Farrell, 2020; Silberbauer-Gottsberger, 1990), with large fruits dispersed by small mammals, and are fast growing (*S. botryophora* is an extreme example; this pioneer species quickly takes advantage of new openings in the canopy; Noblick, 2017). Therefore, while tied and subject to environmental changes, these two species are expected to easily track environmental shifts as they would be able to colonize new open areas but only if within the limited ranges of their dispersers. This makes them a good system to study how environmental change has driven population dynamics over time. While closely related, these are not sister taxa and thus are unlikely to hybridize. Despite having similar dispersal mechanisms and ecologies, *S. botryophora* and *S. pseudococos* appear to have different environmental niches based on their differential occupation of the Atlantic Forest (Noblick, 2017). *Syagrus botryophora* is found in the northern portion of the central corridor of the Atlantic Forest, along the coastal lowlands (less than 400 metres above sea level), primarily in wet and humid evergreen tropical rainforest with clay lateritic soils (Figure 1, blue). *Syagrus pseudococos*, in turn, is found in the southern portion of the central corridor, occupying both lowland and montane systems, primarily in cooler, semi-deciduous forests and the southern evergreen forest with rocky soils (Ledru et al., 2016; Noblick, 2017). According to the IUCN Red List of threatened species *S. botryophora* is classified as near threatened (Noblick, 1998a) and *S. pseudococos* as least concern (Noblick, 1998b). Both species are widely distributed but the habitats that *S. botryophora* occupies are more vulnerable to logging and conversion to pastureland (Noblick, 2017). Additionally, the continuing fragmentation of the Atlantic Forest from human habitat modification as well as anthropogenic climate change, pose potential threats to these species.

Here, we ask if climatic stability during the Late Quaternary explains present day patterns of population structure and genetic diversity in these two palms. To this end, we first examined shifts in suitable habitat associated with Late Quaternary climate change in both species and then used genomic-scale data to characterize patterns of population structure and genetic diversity. Using these data, we tested the following hypotheses and predictions: (1) areas of long term stability within the range of these two species allowed local populations to persist over time and to reach an equilibrium, therefore localities in contiguous regions of high stability show low genetic structuring and a signal of genetic IBD; (2) conversely, areas of low stability prevented population persistence and gene flow through these regions and therefore, localities separated by low stability regions show high levels of population structure and a pattern of IBR and (3) areas of high climatic stability have high genetic diversity. Finally, we examine evidence for the Atlantic Forest central corridor as a climatic refuge for plants.



**FIGURE 1** Distribution of *Syagrus botryophora* (blue) and *Syagrus pseudococos* (red). These two species occur exclusively within the Atlantic Forest of Brazil (green shading shows the pre-Columbian extent of the Atlantic Forest). Points show occurrence records for each species—including our sampling effort and all available geo-referenced museum records. Suitable habitat was inferred with climate-based Species Distribution Modelling, thresholded with a 10% omission rate suitability cutoff. Grey shading shows elevation, where darker colours depict higher altitudes. States are labelled with their abbreviations: Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ) and São Paulo (SP). Map projection is pseudo-Plate Carrée, decimal degree units are treated as if they are linear units.



## 2 | MATERIALS AND METHODS

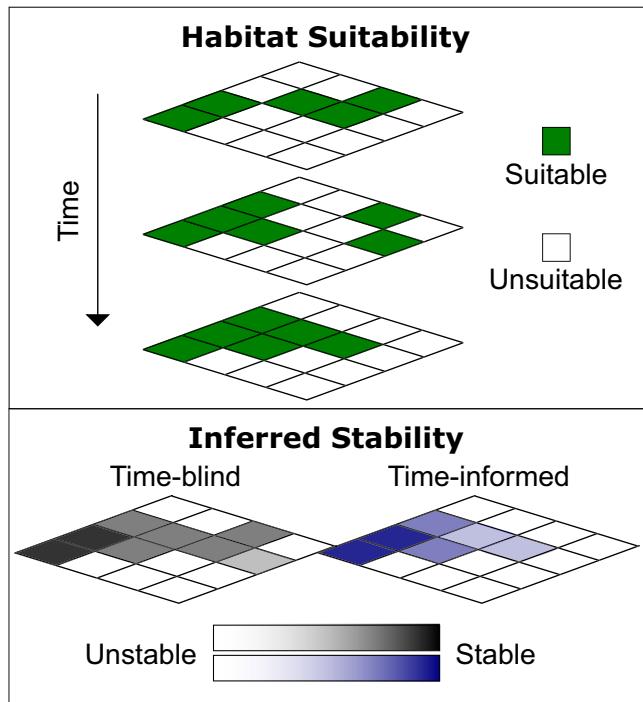
### 2.1 | Overview

Our questions require estimating, over the largest time interval possible, how long the landscape occupied by these palm species has been climatically suitable. Here, this landscape is represented by grid cells in a map. For that, we used current occurrence records of each species and climatic projections over the last 120,000 years (120ky) in correlative Species Distribution Models (SDMs; Townsend Peterson et al., 2011) to infer the geographical extent of suitable climates for *S. botryophora* and *S. pseudococos*. Next, we generated and assembled a double digest restriction-site associated DNA sequencing (ddRADseq) dataset from individuals sampled across the range of each species. We then measured genetic diversity at each sampled locality and characterized population structure within each species. Finally, we examined support for our proposed hypotheses, by testing whether or not historical

stability predicts the empirical genetic structure and diversity patterns for both species.

### 2.2 | Measuring stability through a novel time-informed approach

Traditionally, climatic stability is defined as the persistence of similar climatic conditions in a region over time (Avise, 2000; Owens & Guralnick, 2019). To infer climatically suitable stable areas for specific species, SDMs are built from climate descriptors based on occurrence records of that species. Under this framework, grid cells are categorized as having suitable or unsuitable climates, and stability is quantified by the proportion of time periods when each grid cell presents a suitable climate (e.g. Carstens & Richards, 2007; Hugall et al., 2002; Paz et al., 2019; Vasconcellos et al., 2019; Wieringa et al., 2020). However, quantifying stability solely as the proportion of time periods when climate is suitable for a species is ‘time-blind’;



**FIGURE 2** Graphical representation of time-blind and our novel time-informed metric of habitat suitability over time. Both metrics are based on habitat suitability over time—shown at three time points where green grid cells are suitable and white grid cells are unsuitable. For each grid cell, time-blind stability is measured as a proportion of time periods that were climatically suitable. In contrast, time-informed stability measures how far back into the past a grid cell that is climatically suitable today has been suitable. In other words, for each grid cell that is suitable in the most recent time period, we count backwards until reaching a time point at which the grid cell becomes unsuitable and then ignore any older states of that grid cell.

it fails to identify at which times the climate was suitable or not, and more importantly, how long presently suitable areas have been so. Because the timing of these suitable climate periods matters—especially for scientists interested in demographical responses to environmental change and the resulting patterns of genetic diversity (McDonald-Spicer et al., 2019)—we inferred stability via a novel time-informed metric (Figure 2). Whereas time-blind stability is typically measured as a proportion of time periods that were climatically suitable, time-informed stability measures how far back into the past a grid cell that is climatically suitable today has been suitable. In other words, for each grid cell that is currently suitable, we counted backwards until reaching a time point at which the grid cell becomes unsuitable and ignored any older states of that grid cell.

### 2.3 | Species distribution models

To characterize climate stability for each species over time, we built SDMs by combining occurrence data for the species of interest with spatial environmental data to model how the species occupies

geographical space (Guisan et al., 2017). We used the R package 'walle' v1.0.6.2 (Kass et al., 2018) to create a template of the SDM code, and amended the code as detailed below. Occurrence data from the samples used in our genetic analyses were combined with occurrences listed in the Global Biodiversity Information Facility (GBIF.org), from across the entire species range. To ensure taxonomic and locality accuracy, we used only those GBIF data points based on museum specimens (Table S1).

We used environmental data from the Hadley Climate Center Model (HadCM3) to describe present-day and past climate conditions over the broadest range possible of available times (Singarayer & Valdes, 2010). These data comprise eight bioclimatic variables at a 2.5 min (approximately 5 km) spatial resolution: annual mean temperature (bio1); temperature seasonality (bio4); mean temperature of warmest quarter (bio10); mean temperature of coldest quarter (bio11); annual precipitation (bio12); precipitation seasonality (bio15); precipitation of wettest quarter (bio16); precipitation of driest quarter (bio17). These variables describe the mean and variation in temperature and precipitation in 62 'snap-shots' over the last 120 ky, sampled every 1000 years between the present and 22 ky ago (22 kya), every 2000 years between 24 kya and 80 kya, and every 4000 years from 84 kya until 120 kya.

To sample 10,000 background points needed for the modelling procedure (Hijmans et al., 2013), we defined a background area as the minimum convex polygon encompassing all occurrence points and including a 0.75 degree buffer. Although dispersal distances have not been directly quantified in *S. botryophora* or *S. pseudococos*, ripe fruit carried by land mammals likely represents the farthest dispersal (Souza et al., 2018). Therefore, we chose a 0.75 degree buffer (approximately 80 km) to exclude most inaccessible areas. To ensure that only one occurrence point was sampled from within each grid cell, thus avoiding sampling biases, we thinned the occurrence points to match the spatial resolution of the environmental data, using the R package 'spThin' v0.2.0 (Aiello-Lammens et al., 2015). To evaluate model performance, we split the occurrence data into four geographical sets for training and testing the model. The model was run four times, and each time a different set was withheld for testing the model trained with the remaining three sets. This spatial binning approach is more robust to evaluate how models transfer over environmental spaces not used in the model calibration, which is appropriate when projecting models in time (Muscarella et al., 2014).

We used the R package 'ENMeval' v2.0.0 to build a suite of models, varying the shape of the response curve (linear, or linear and quadratic) and model complexity (regularization multipliers ranging from 0.1 to 4.5 in increments of 0.1; Muscarella et al., 2014). For the modelling procedure, we used the machine learning algorithm MaxEnt as implemented in the R package 'dismo' v1.1–4 (Hijmans et al., 2013), fitting a total of 90 models. To choose the best model from this suite of SDMs, we prioritized those with less model overfitting. First, we selected those models with the lowest 10% omission rates, within 0.05 of the minimum 10% omission rate. From this subset, we chose the most accurate model as measured by the average Area Under the Curve (AUC) of the testing sets (Guisan et al., 2017).



To model the suitable climate of each species through time, we projected the SDMs into the palaeoclimate ‘snap-shots’ from the Hadley Climate Center. To create a binary projection and classify the landscape as suitable or unsuitable at each point in time, we used the 10% omission rate suitability value as a threshold. To overcome limitations inherent to these projections, we implemented two additional methodological steps. First, to account for non-analogous conditions between the contemporary training environment and the palaeoclimate environments, we masked out all non-analogous areas (MESS index <0; Elith et al., 2010). In addition, we limited the geographical extent of the palaeoclimate projections because the actual occupied area of the species is likely a subset of the total suitable area, due to dispersal limitations and possibly other abiotic or biotic interactions (Townsend Peterson et al., 2011). To account for dispersal limitations, and because we are only able to test predictions in areas that are occupied today, we limited the projection area to within 0.75 degrees (approximately 80 km) of the contemporary suitable habitat.

Using these projections, we created maps of time-informed stability by counting the number of time periods each grid cell has been continuously suitable, starting from the present and moving into the past. In order to compare our new stability metric to the traditional metric, we also calculated time-blind stability as the proportion of time periods each pixel was suitable, across all time periods. Grid cells in these maps with higher values are more stable relative to grid cells with lower values.

## 2.4 | Genetic data collection and processing

To test our predictions, we sampled 42 individuals of *S. botryophora* (11 localities) and 51 of *S. pseudococos* (13 localities), including at least three individuals per site across the entire range of each species (Table S2). An individual of *S. oleracea* was also sampled as an outgroup for the phylogenetic analyses (Noblick et al., 2013). Small leaf samples were collected for each individual and dried in individual bags with silica gel for DNA preservation. Total genomic DNA was isolated with the Illustra Nucleon PhytoPure (GE Healthcare Life Sciences) DNA extraction kit, after pulverizing the leaf material in a FastPrep-24™ sample preparation system (MP Biomedicals). We confirmed the quality of extractions by visualizing high molecular weight bands on a 1% agarose gel, and quantified DNA concentrations in a Qubit® 2.0 fluorometer (Life Technologies). All DNA extractions were normalized at 1000 ng of DNA in a total volume of 50 µl of TLE buffer. Genomic library preparation and sequencing was performed at the Texas A&M AgriLife Research Genomics and Bioinformatics Service.

We used a ddRADseq protocol (Peterson et al., 2012) to generate restriction-site associated DNA libraries of thousands of short fragments distributed across the entire genome of the two target species. Each individual genomic DNA extract was double-digested with the enzymes HindIII (5' AAGCTT 3') and BfuCI (5' GATC 3'), then ligated to customized Illumina® adapters containing multiplexed individual barcodes. This enzyme pair was chosen after the

evaluation of several combinations of enzyme digests in a Fragment Analyser™ System, which produced a smooth density curve of fragments around 500 bp. This choice was further validated by an in-silico digestion of two *Elaeis* oil palm genomes (Singh et al., 2013) using the FRAGMATIC software (Chafin et al., 2018), confirming that a sufficient number of fragments would be sequenced at a deep coverage (>10x) considering our sequencing output. Only fragments of 400 bp to 540 bp in size were selected using a PippinPrep (Sage Science) automated gel extraction machine. Fragments were then PCR-amplified and sequenced in a partial lane of the NovaSeq 6000 System Illumina® sequencer, generating paired-end reads of 150 bp each. Sequence cluster identification, including quality pre-filtering, base calling and uncertainty assessment were done in real time using Illumina®’s NCS v1.0.2 and RFV v1.0.2 software with default parameter settings. Base call files were then demultiplexed and formatted into fastq files using the bcl2fastq2 v2.19.0 Illumina® software.

After verifying the sequence quality in FastQC v0.11.3 (Andrews, 2010), we used the ipyrad v0.9.18 pipeline (Eaton & Overcast, 2020) for further quality filtering, de novo assembly and alignment of reads into loci. We proceeded with the pipeline assuming these two species have diploid genomes given the absence of polyploidy in the genus (Gunn et al., 2015; Röser, 2000; Silva-Cardoso et al., 2017). Low-quality sites (PHRED Qscore <20) and reads (>4 low quality bases) were filtered out. The remaining reads were clustered de novo within individuals, using an 85% similarity threshold, in VSEARCH (Rognes et al., 2016). To call genotypes, a consensus sequence was computed for each locus within individuals, considering an estimated sequencing error rate ( $\epsilon$ ) and heterozygosity ( $\pi$ ) across all sites ( $\epsilon = 0.003$  and  $\pi = 0.012$  for this study). Sites with a within-individual coverage larger than 6X were used for base calling, and reads were allowed to have up to 5% undetermined sites. Using the same sequence alignment method and threshold, reads were then clustered across individuals within species. A final alignment step was performed with MUSCLE (Edgar, 2004). Finally, given that the choice of sample coverage can impact not only the number of loci retained but also the type of loci and ultimately the population genetic inferences (Hodel et al., 2017; Huang & Knowles, 2016), we generated five datasets per species, each corresponding to a different level of sample coverage in order to verify the robustness of our results. Specifically, we allowed up to 10%, 30%, 50%, 70% or 90% of individual samples to have missing data at that locus.

## 2.5 | Genetic analyses

We used two approaches to quantify the broad patterns of population structure within each palm species. First, to identify distinct genetic clusters, we estimated ancestry coefficients for each individual using sparse Nonnegative Matrix Factorization (sNMF) as implemented in the R package ‘LEA’ v2.8.0 (Frichot & François, 2015). Using single nucleotide polymorphism (SNP) data, we ran sNMF with the number of genetic clusters,  $K = 1\text{--}12$  and  $1\text{--}14$  (i.e. from one to the number of sampled localities plus one) for *S. botryophora* and *S.*



*pseudococos*, respectively, to ensure we compared the likelihood of each plausible number of populations. Then, for each value of  $K$ , we set the alpha parameter to 1, 10 and 100, to verify the robustness of the results. We ran sNMF in five replicate runs, 2000 iterations, and under default parameters. We determined the most likely number of genetic clusters using the minimum cross-entropy across runs.

Then we used a phylogenetic approach to identify any hierarchical relationships between localities and genetic clusters, using SVDquartets implemented in PAUP\* v4.0a167 (Chifman & Kubatko, 2014; Swofford, 2002). We analysed each species separately, using one individual from the other species, as well as the *S. oleracea* individual, as outgroups. We sampled the quartets exhaustively, using a coalescent species tree approach and assigning individuals to the sampled localities. We performed 1000 bootstrap replicates to infer nodal support.

To determine if and how population genetic structure is influenced by geographical distance and by the environmental resistance to gene flow imposed by historical patterns of stability over time, we tested for patterns of IBD and IBR in each species. For each pair of sampled localities, we first calculated their genetic distance as Nei's distance in the R package 'adegenet' v2.1.3 (Jombart & Ahmed, 2011). Then, we calculated the geographical great-circle distance between each pair as implemented in the R package 'geosphere' v1.5-10 (Hijmans et al., 2016). Lastly, we estimated the resistance between each pair of localities in three ways: (1) based on the contemporary landscape suitability, (2) based on the metric of time-blind stability and finally, (3) using our new time-informed metric of stability. Specifically, using the maps of contemporary suitability or historical stability for each species, we estimated the resistance value between each locality pair, with an eight-neighbour cell scheme based on circuit theory in circuitscape v.4.0 (Shah & McRae, 2008). Localities separated by low suitability or less stable areas have higher resistance compared to localities separated by higher suitability or more stable areas.

Next, we tested how well each of these three resistance maps predicted genetic distance. We performed three distance matrix regressions as implemented in 'ecodist' v2.0.7 (Legendre et al., 1994). Since geographical distance and resistance are correlated, and because we expect geographical distance to be the primary factor driving genetic distance (Guillot & Rousset, 2013), we used the residuals of resistance regressed against geographical distance as the predictor variable representing isolation by resistance in our model. Finally, we compared the explanatory power ( $r^2$ ) of resistance in the three regressions.

To quantify genetic diversity, we calculated the average number of mutations per locus (mut), average number of heterozygous positions at each locus (het pos), average number of haplotypes per locus (hap) and average nucleotide diversity per locus ( $\pi$ ) at each locality in DnaSP v6.12.03 (Rozas et al., 2017). We then tested for a correlation between nucleotide diversity ( $\pi$ ) and each of the stability metrics, as well as contemporary suitability, with a linear model in the R package 'stats' v4.1.0. To account for local variation at sample sites, we averaged the values of grid cells within a 10 km buffer of each sample site.

## 3 | RESULTS

### 3.1 | SDMs and stability metric

The species distribution models showed high accuracy and low overfitting: the best models for *S. botryophora* and *S. pseudococos* had a testAUC of 0.78 and 0.87, and a 10% omission rate of 0.06 and 0.11 respectively (Figure 1; Tables S3 and S4).

The best SDM for *S. botryophora* was based on 32 occurrence points and contained linear and quadratic response curves for the predictor variables (Figure 1, Figure S1). All environmental variables except for annual mean temperature (bio1) and mean temperature of warmest quarter (bio10) were retained in the model (Figure S2). The model suggested that environmental suitability for *S. botryophora* is highest in near-coastal areas of Bahia and northern Espírito Santo, decreasing to the north and south, as well as inland. The binary model (with a suitability threshold of 0.29) predicted suitable habitat in low-elevation near-coastal regions in southern Espírito Santo, expanding into some high elevation areas in northern Espírito Santo, through mostly lower elevation regions of near-coastal Bahia and into southeastern Sergipe (Figure 1).

When projected back in time, the model for *S. botryophora* suggests that the central corridor had relatively high stability over the last 120 ky years (Figure 3). Moreover, coastal southern Bahia had the highest stability, with suitable climates lasting up to 30 kya. The model suggested that the northern portion of the contemporary range of *S. botryophora*, in northeastern Bahia, had moderate stability—with suitable climate only since the beginning of the Holocene (less than 10 ky).

Models of *S. pseudococos* were developed with 28 occurrence points after filtering and thinning (Figure 1, Figure S3). Similar to *S. botryophora*, the best model contained linear and quadratic response curves but included all environmental variables with the exception of annual precipitation (bio12) (Figure S4). Regions with high suitability were less contiguous in this species and included the central corridor as well as near-coastal areas of São Paulo and western Rio de Janeiro. When the model is thresholded (suitability cutoff of 0.39), two major suitable areas are identified, separated by an unsuitable area in northeastern Rio de Janeiro (Figure 1). Areas with both low and high elevation (1000 m above sea level) were predicted as suitable.

In contrast to the pattern seen in *S. botryophora*, the time-informed model of historical climate suitability in *S. pseudococos* show the existence of two main areas of high stability over the last 120 ky: one to the north (in the central corridor in Espírito Santo) and one to the south, between Rio de Janeiro and São Paulo (Figure 3). The region of high stability in the central corridor was, on average, less climatically suitable, but occupied a larger geographical extent compared to the southern area. Here, the model suggested a landscape of small areas with very high stability, with coastal São Paulo and Rio de Janeiro having had continuous suitable climate up to 100 kya (Figure 3). Outside these regions, most of the contemporary range of *S. pseudococos* has only been suitable for the last 10–20 ky.



### 3.2 | ddRADseq data and processing

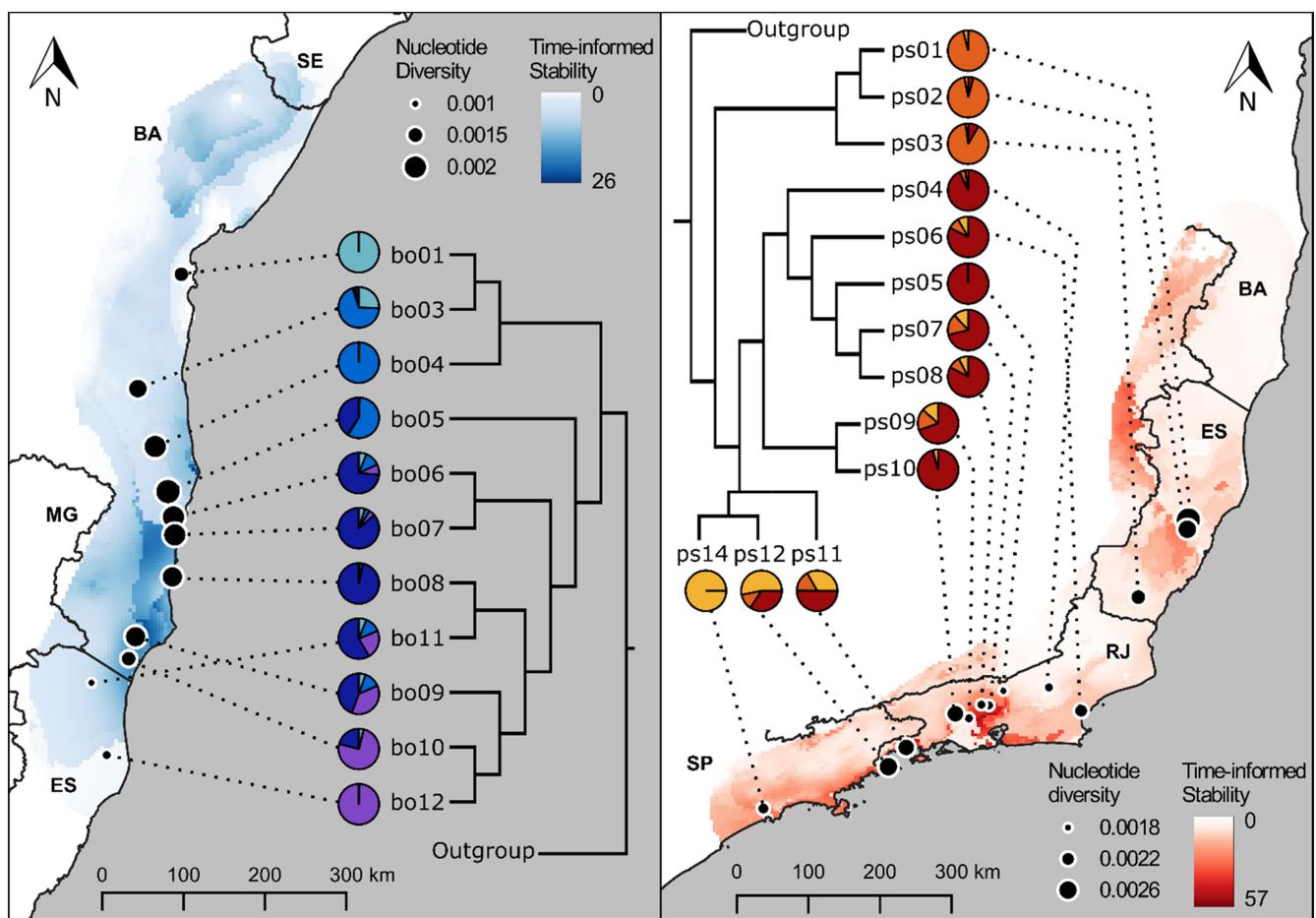
The ddRADseq library prep and sequencing protocol produced an average of 2.7 million reads per individual, ranging from 255,320 to 8,833,066. One individual (ps04\_2) with few reads (34,318) was eliminated. Assembly and filtering of the ddRADseq library for *S. botryophora* resulted in 25,854 loci under the most lenient (90%) missing data filtering scheme, and 92 loci under the strictest (10%). For *S. pseudococos*, we detected 26,908 loci under the most lenient filtering, and 88 loci under the strictest (Table S5).

### 3.3 | Genetic analyses

The population structure analysis in *S. botryophora* inferred between one and six genetic clusters but converged to a single cluster as the

number of loci in the dataset increased (Figure S5). When more than one cluster was inferred, there was high admixture within individuals with closer individuals having similar admixture levels (Figure 3, left). The phylogenetic analyses revealed that the oldest divergence split the three northernmost localities from the remaining ones. Overall, the closer individuals are geographically, the more closely related they were (Figure 3, left). The only exception to this pattern were the individuals from locality bo11, which appeared more similar to populations farther north, but had a higher proportion of missing data than the rest of the samples (Table S6). Very similar tree topologies were inferred from the four largest datasets (Figure S6).

In contrast, *S. pseudococos* exhibited clear population structure, with two clusters on either side of the Espírito Santo–Rio de Janeiro border, the location with the lowest historical stability (Figure 3, right). The sNMF analysis inferred two or three genetic clusters (Figure S7): individuals within the Atlantic Forest central corridor always clustered



**FIGURE 3** Time-informed stability and population structure for *Syagrus botryophora* (on the left, in blue) and *Syagrus pseudococos* (on the right, in red), across their ranges in the Atlantic Forest of Brazil. Time-informed stability is measured as the number of time periods, since the present, that had suitable habitat, and was inferred with Species Distribution Models projected through time. This is overlaid with phylogenetic relationships inferred in SVDQuartets (topology shown reflects results from the majority of missing data treatments, for all topologies recovered see Figures S5 and S6) and population assignment in sNMF (90% dataset, although results from other missing data thresholds show similar patterns). For *S. botryophora*, where the optimal number of clusters was 1, we show population assignment under  $k = 4$  to highlight the pattern of admixture, and similarly we show the population assignment under  $k = 3$  for *S. pseudococos*. The size of the circle at each locality represents the average nucleotide diversity (10% dataset) at that locality. Map projection is pseudo-Plate Carrée, decimal degree units are treated as if they are linear units.



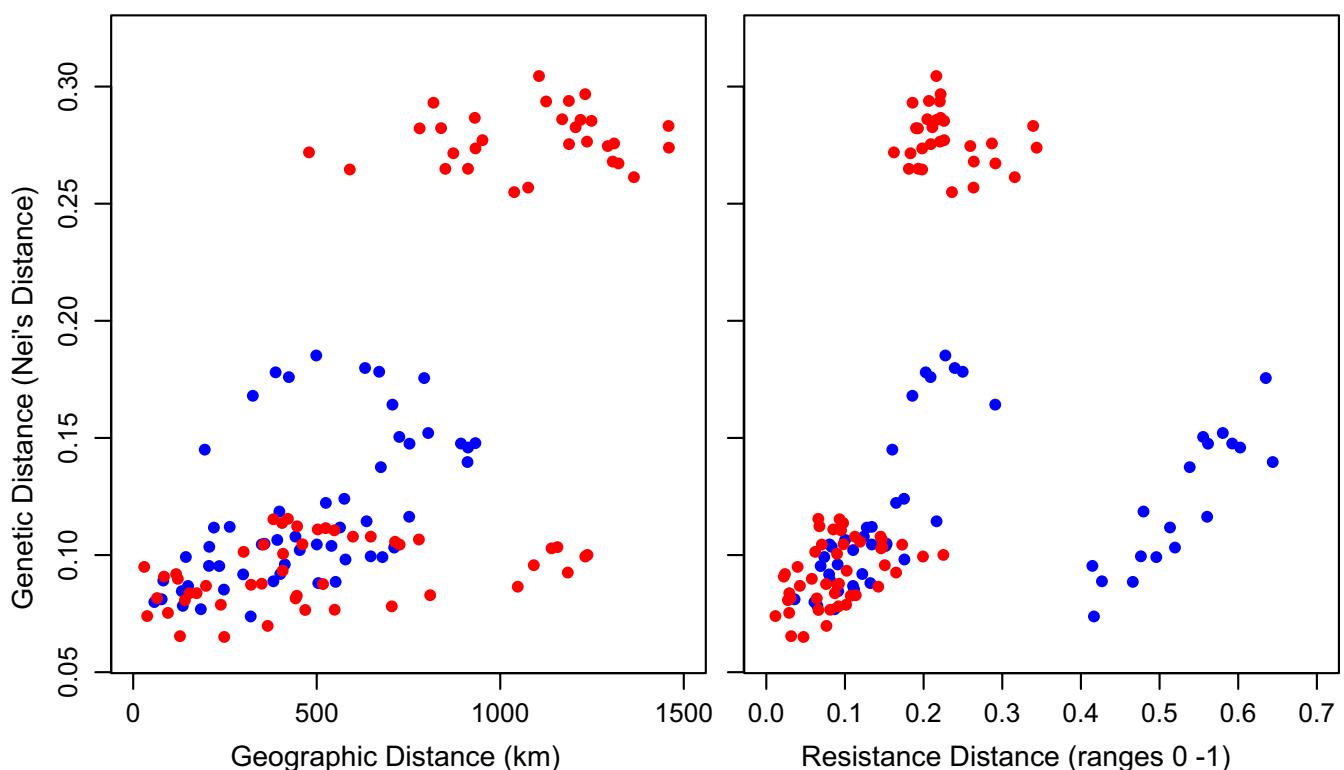
together, and the remaining individuals, south of the central corridor, were recovered either as one cluster or two clusters with admixture. In the phylogenetic reconstruction, the deepest divergence in *S. pseudococos* separated the localities in Espírito Santo from the rest of the localities. This geographical separation overlaps with a region with very low time-informed stability over time (Figure 3, right). Within the large southern clade, there were two monophyletic groups: one comprising the three westernmost localities, and the other comprising the seven remaining localities. As was the case for *S. botryophora*, the tree topology was very similar using the four largest datasets (Figure S8).

Once population structure was characterized, we tested for the effect of IBD and IBR in each species (Figure 4; Tables S7 and S8). In *S. botryophora* geographical distance was a significant predictor of genetic distance in all three models (geographical distance  $p < 0.05$ ). The model including time-informed stability had the highest explanatory power ( $r^2 = 0.36$ ) although the residuals of resistance were not significant (resistance  $p = 0.21$ ). This supports IBD acting as the prevailing pattern in this species with little to no IBR. In *S. pseudococos*, which was split into multiple regions of high stability, geographical distance was also a significant predictor of genetic distance in all models (geographical distance  $p = 0.001$ ). The model which included time-informed stability had the highest explanatory power ( $r^2 = 0.71$ ) and the resistance residuals were significant (resistance  $p = 0.011$ ), suggesting that IBR promoted by differential patterns of climatic stability is acting in this species, in addition to IBD.

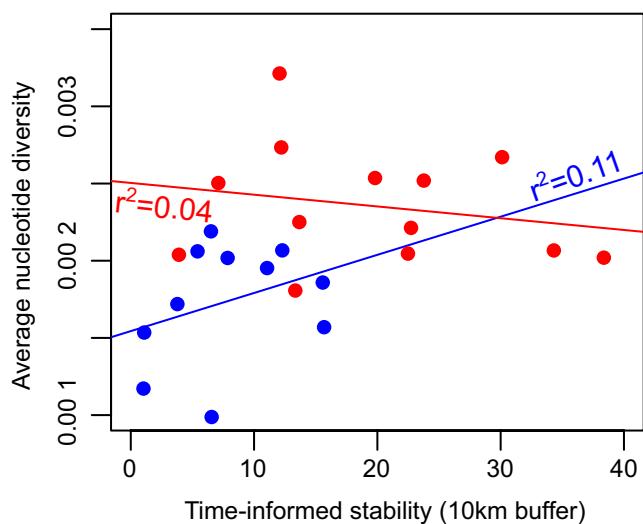
We found variation in genetic diversity levels when comparing localities in both species, but that variation was tied to the number

of loci retained after assembly (Table S6). For that reason, we focus on the results from the most complete dataset (10% missing data threshold), where there was less variation in the number of loci retained across localities. Additionally, we focus on average nucleotide diversity, which was correlated with all the other metrics: average mutations per locus, average heterozygous positions at each loci and average haplotypes per locus.

While stability metrics were not significantly correlated with nucleotide diversity in either species (Table S9), contemporary suitability was a significant predictor of genetic diversity in *S. botryophora* ( $p = 0.046$ ). In *S. botryophora*, there was a positive trend of increased diversity with increased time-informed stability (slope =  $2.47 \times 10^{-5}$ ,  $p = 0.32$ ,  $r^2 = 0.11$ ; Figure 5) and we found the highest nucleotide diversity in the centre of the range (bo5), which decreased towards the edges (Figure 3, left). In *S. pseudococos*, the variation in nucleotide diversity was less centralized; there was no single area with a higher concentration of genetic diversity, but the northernmost individuals, in the central corridor, had the highest average nucleotide diversity overall (Figure 3, right). Neither stability metric was significantly correlated with nucleotide diversity in *S. pseudococos*, but, contrary to *S. botryophora*, a negative trend was detected (slope =  $-7.62 \times 10^{-6}$ ,  $p = 0.5$ ,  $r^2 = 0.04$ ; Figure 5). When we looked at these relationships using different missing data treatments, we saw similar patterns: contemporary suitability is a better predictor of genetic diversity than either stability metric in *S. botryophora* and no metrics significantly predicted genetic diversity in *S. pseudococos*.



**FIGURE 4** Visualization of isolation by distance and isolation by resistance (time-informed stability distance) patterns in *Syagrus botryophora* (blue) and *Syagrus pseudococos* (red).



**FIGURE 5** Relationship between average genetic diversity ( $\pi$ ) at a locality and the average time-informed stability within 10 km of that locality for *Syagrus botryophora* (blue) and *Syagrus pseudococos* (red).

#### 4 | DISCUSSION

We show that Late Quaternary climate change explains present day patterns of population structure in two palm species from the Brazilian Atlantic Forest: *Syagrus botryophora* and *S. pseudococos*. Our models suggest that the Atlantic Forest central corridor served as a centre of stability for both species, supporting the view that it played an important role as a Pleistocene climate refuge for several taxa (Batalha-Filho et al., 2010; Carnaval et al., 2009; Carnaval & Moritz, 2008; de Moraes-Barros et al., 2006; Peres et al., 2020; Werneck et al., 2011). With clear signals of isolation-by-distance and high genetic diversity in this contiguous stable region, we have evidence that populations of *S. pseudococos* and *S. botryophora* were able to persist in different parts of this corridor through the fluctuating climate cycles of the Pleistocene. A break in population structure in *S. pseudococos* along the less stable Espírito Santo–Rio de Janeiro border suggests that persistence and migration through this region has been limited in the past.

These broad phylogeographical patterns are consistent with patterns found in other Atlantic Forest species. Primarily, this study adds to the evidence of the central corridor as a Pleistocene climate refuge in the Atlantic Forest—as found in models of the Atlantic Forest as a whole (Carnaval & Moritz, 2008) as well as taxa specific evidence (Batalha-Filho et al., 2010; Carnaval et al., 2009; de Moraes-Barros et al., 2006). Furthermore, the break in *S. pseudococos* around the Espírito Santo–Rio de Janeiro border has also been found in many other taxa including mammals: *Micoureus* (Costa, 2003), *Akodon montensis* (Valdez & D'Elía, 2013), *Desmodus rotundus* (de Mello Martins, 2011); amphibians: *Dendropsophus elegans* and *Chiromocleis carvalhoi* (Tonini et al., 2013), *Thoropas miliaris* (Fitzpatrick et al., 2009), *Hypsiboas albomarginatus* (de Mello Martins, 2011), *Rhinella crucifer* (Thomé et al., 2010); birds:

*Xiphorhynchus fuscus* (Cabanne et al., 2007) and plants: *Eugenia uniflora* (Turchetto-Zolet et al., 2016), *Dalbergia nigra* (Ribeiro, Lemos-Filho, et al., 2011). This break corresponds to a low elevation region in the Serra do Mar, the Valley of the Paraíba do Sul River. We suggest that the low climatic stability in this region, which affected the long-term persistence and gene flow over time in *S. pseudococos*, may be a main driver behind the genetic break of many more taxa occurring throughout this region.

Large climatically stable regions have the potential to safeguard large population sizes; connecting subpopulations over evolutionary time and favouring gene flow among them, which seem paramount to maintain high levels of genetic diversity within populations (Leffler et al., 2012). This pattern, however, was not evident in our study. Many other factors may have impacted genetic diversity in these species. First, historical climate stability is likely not the only force driving genetic diversity. Other drivers may include the size of the area that was suitable, which was very fragmented in *S. pseudococos* particularly. If effective population sizes are small, from small patches of suitable habitat, genetic drift may reduce standing genetic variation all on its own (Leffler et al., 2012). The southern range of *S. pseudococos* contained areas that were stable for tens of thousands of years, but these were only small patches of habitat. In addition, while we may expect low genetic diversity in a low stability area that was recently colonized (e.g. caused by founder effect), higher genetic diversity may occur if it was colonized by multiple surrounding populations located in pockets of stable climate (Excoffier et al., 2009; Peter & Slatkin, 2015).

Although we have focused here on linking climatic stability with population structure and genetic diversity, we are aware that there likely exist additional factors driving population genetic patterns. Previous studies have shown mixed evidence for the role of geomorphology—such as topographical relief and rivers basins along the Atlantic Forest—as barriers to gene flow (Cabanne et al., 2008; Paz et al., 2019; Souza et al., 2018; Valdez & D'Elía, 2013). Additionally, periodical sea level regression and transgression events during the Quaternary, especially in the region between southern Bahia and northern Espírito Santo (e.g. Leite et al. 2016), may have had an influence on local biotic processes. Moreover, the reproductive biology of these two palms may be a factor in their genetic structure. *Syagrus* species are primarily pollinated by wind, beetles and bees (De Medeiros et al., 2019; Noblick, 2017). Even though detailed data on pollination of *S. botryophora* and *S. pseudococos* are lacking, it is possible that populations with distinct pollinators may be prone to reproductive isolation. Continued study of the factors underlying divergence in *Syagrus*, as well as other Atlantic Forest biota, will bring a more complete understanding of the biogeographical processes underlying the high level of biodiversity in the Atlantic Forest of Brazil.

Previous scholars have stated the need to precisely define stability (McDonald-Spicer et al., 2019), including the interaction between space and time, which is what we aimed to do here. In our implementation, time-informed stability will always be a subset of time-blind (classic) stability, because at the first instance of a grid



cell becoming unsuitable, we stop considering any older states of that grid cell. Admittedly, this is still somewhat simplistic, as is it possible that this scale—one time-slice at one grid cell—is not enough to locally extirpate the species of interest. More complexity could be added to our metric of time-informed stability. For example, one may want to consider the scale at which local extirpation would occur, based on species-specific traits. In spite of this simplicity, the concordance between genetic patterns and time-informed stability, which comes from two independent types of data—DNA sequences and geographical occurrences respectively—suggests that this metric is biologically relevant. We would like to emphasize the utility of our novel time-informed metric of habitat stability and hope that it will be explored and expanded in additional studies.

## ACKNOWLEDGEMENTS

This work was partially funded by a collaborative project supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; 2013/50297-0), the National Science Foundation (NSF; DEB 1343578, DEB-1343612) and NASA. Field collections were done under permit #34946-2 from the Instituto Chico Mendes de Conservação da Biodiversidade—ICMBio. K. Mercier acknowledges funding by the CCNY Levine Fellowship. This research was supported in part by the City University of New York High Performance Computing Center at the College of Staten Island. We would like to thank the Carnaval Lab members for valuable feedback on earlier drafts of the manuscript. We also thank Sereene Kurzum for early laboratory assistance during DNA extractions and Daniela Costa for herbarium assistance curating collected specimens and coordinates of the sampling localities. We thank Christine Bacon and an anonymous reviewer for their valuable comments on an earlier version of this manuscript.

## CONFLICT OF INTEREST

The authors declare no competing interests.

## DATA AVAILABILITY STATEMENT

Occurrence data used for species distribution modelling are available on datadryad.org—<https://datadryad.org/stash/dataset/doi:10.5061/dryad.2z34tmpp?>. R markdown files are available on Github at <https://github.com/kpmmercier/Linking-environmental-stability-with-genetic-diversity-and-population-structure>. Demultiplexed raw fastq files are available on the NCBI Sequence Read Archive (Bioproject PRJNA774804)—<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA774804>.

## ORCID

Kathryn P. Mercier  <https://orcid.org/0000-0001-7898-1353>

Mariana M. Vasconcellos  <https://orcid.org/0000-0002-3175-2649>

José Rubens Pirani  <https://orcid.org/0000-0001-7984-4457>

Fabián A. Michelangeli  <https://orcid.org/0000-0001-7348-143X>

Ana Carolina Carnaval  <https://orcid.org/0000-0002-4399-1313>

## REFERENCES

- Aguiar-Melo, C., Zanella, C. M., Goetze, M., Palma-Silva, C., Hirsch, L. D., Neves, B., da Costa, A. F., & Bered, F. (2019). 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*, 106(7), 971–983.
- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38, 541–545.
- Almeida, J. P. F. A., Thomé, M. T. C., Sturaro, M. J., Pereira, R. J., & Mott, T. (2020). The relative role of glacial refugia and longstanding barriers in the diversification of a fossorial squamate. *Systematics and Biodiversity*, 18, 447–463.
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics Institute.
- Ashcroft, M. B. (2010). Identifying refugia from climate change. *Journal of Biogeography*, 37, 1407–1413.
- Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press.
- Batalha-Filho, H., Waldschmidt, A. M., Campos, L. A. O., Tavares, M. G., & Fernandes-Salomão, T. M. (2010). Phylogeography and historical demography of the neotropical stingless bee *Melipona quadrifasciata* (Hymenoptera, Apidae): incongruence between morphology and mitochondrial DNA. *Apidologie*, 41, 534–547.
- Brunes, T. O., Thomé, M. T. C., Alexandrino, J., Haddad, C. F. B., & Sequeira, F. (2015). Ancient divergence and recent population expansion in a leaf frog endemic to the southern Brazilian Atlantic forest. *Organisms, Diversity and Evolution*, 15, 695–710.
- Cabanne, G. S., d'Horta, F. M., Sari, E. H. R., Santos, F. R., & Miyaki, C. Y. (2008). Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): biogeography and systematics implications. *Molecular Phylogenetics and Evolution*, 49, 760–773.
- Cabanne, G. S., Santos, F. R., & Miyaki, C. Y. (2007). Phylogeography of *Xiphorhynchus fuscus* (Passeriformes, Dendrocolaptidae): vicariance and recent demographic expansion in southern Atlantic forest. *Biological Journal of the Linnean Society*, 91, 73–84.
- Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785–789.
- Carnaval, A. C., & Moritz, C. (2008). Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, 35, 1187–1201.
- Carstens, B. C., & Richards, C. L. (2007). Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, 61, 1439–1454.
- Chafin, T. K., Martin, B. T., Mussmann, S. M., Douglas, M. R., & Douglas, M. E. (2018). FRAGMATIC: In silico locus prediction and its utility in optimizing ddRADseq projects. *Conservation Genetics Resources*, 10, 325–328.
- Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30, 3317–3324.
- Costa, L. P. (2003). The historical bridge between the Amazon and the Atlantic Forest of Brazil: A study of molecular phylogeography with small mammals. *Journal of Biogeography*, 30, 71–86.
- Damasceno, R., Strangas, M. L., Carnaval, A. C., Rodrigues, M. T., & Moritz, C. (2014). Revisiting the vanishing refuge model of diversification. *Frontiers in Genetics*, 5, 353.
- de Medeiros, B. A. S., & Farrell, B. D. (2020). Evaluating insect-host interactions as a driver of species divergence in palm flower weevils. *Communications Biology*, 3(1), 749.
- De Medeiros, B. A. S., Núñez-Avellaneda, L. A., Hernandez, A. M., & Farrell, B. D. (2019). Flower visitors of the licuri palm (*Syagrus coronata*):



- Brood pollinators coexist with a diverse community of antagonists and mutualists. *Biological Journal of the Linnean Society*, 126, 666–687.
- de Mello Martins, F. (2011). Historical biogeography of the Brazilian Atlantic Forest and the Carnaval-Moritz model of Pleistocene refugia: What do phylogeographical studies tell us? *Biological Journal of the Linnean Society*, 104, 499–509.
- de Moraes-Barros, N., de Moraes-Barros, N., Silva, J. A. B., Miyaki, C. Y., & Morgante, J. S. (2006). Comparative phylogeography of the Atlantic Forest endemic sloth (*Bradypus torquatus*) and the widespread three-toed sloth (*Bradypus variegatus*) (Bradypodidae, Xenarthra). *Genetica*, 126, 189–198.
- Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36, 2592–2594.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330–342.
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501.
- Fitzpatrick, S. W., Brasileiro, C. A., Haddad, C. F. B., & Zamudio, K. R. (2009). Geographical variation in genetic structure of an Atlantic Coastal Forest frog reveals regional differences in habitat stability. *Molecular Ecology*, 18, 2877–2896.
- Fjeldså, J., & Lovett, J. C. (1997). Biodiversity and environmental stability. *Biodiversity and Conservation*, 6, 315–323.
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6, 925–929.
- Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, 4, 336–344.
- Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.
- Gunn, B. F., Baudouin, L., Beulé, T., Ilbert, P., Duperray, C., Crisp, M., Issali, A., Konan, J.-L., & Rival, A. (2015). Ploidy and domestication are associated with genome size variation in Palms. *American Journal of Botany*, 102, 1625–1633.
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907–913.
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2013). dismo: Species distribution modeling (1.1-4). <https://cran.r-project.org/web/packages/dismo/index.html>
- Hijmans, R. J., Williams, E., & Vennes, C. (2016). Geosphere: Spherical trigonometry (1.5–10). <https://cran.r-project.org/web/packages/geosphere/index.html>
- Hodel, R. G. J., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: Comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports*, 7, 17598.
- Huang, H., & Knowles, L. L. (2016). Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology*, 65, 357–365.
- Hugall, A., Moritz, C., Moussalli, A., & Stanisic, J. (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences of the United States of America*, 99, 6112–6117.
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071.
- Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C., & Anderson, R. P. (2018). Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution*, 9, 1151–1156.
- Ledru, M.-P., Montade, V., Blanchard, G., & Hély, C. (2016). Long-term spatial changes in the distribution of the Brazilian Atlantic Forest. *Biotropica*, 48, 159–169.
- Ledru, M.-P., Salatino, M. L. F., Ceccantini, G., Salatino, A., Pinheiro, F., & Pintaud, J.-C. (2007). Regional assessment of the impact of climatic change on the distribution of a tropical conifer in the lowlands of South America. *Diversity and Distributions*, 13, 761–771.
- Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology*, 10, e1001388.
- Legendre, P., Lapointe, F.-J., & Casgrain, P. (1994). Modeling brain evolution from behavior: A permutational regression approach. *Evolution*, 48, 1487–1499.
- McDonald-Spicer, C. J., Moritz, C. C., Ferrier, S., & Rosauer, D. F. (2019). The importance of defining measures of stability in macroecology and biogeography. *Frontiers of Biogeography*, 11, 1–13.
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60, 1551–1561.
- Menezes, R. S. T., Brady, S. G., Carvalho, A. F., Del Lama, M. A., & Costa, M. A. (2017). The roles of barriers, refugia, and chromosomal clines underlying diversification in Atlantic Forest social wasps. *Scientific Reports*, 7(7689), 1–16.
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5, 1198–1205.
- Neves, B., Kessous, I. M., Moura, R. L., Couto, D. R., Zanella, C. M., Antonelli, A., Bacon, C. D., Salgueiro, F., & Costa, A. F. (2020). Pollinators drive floral evolution in an Atlantic Forest genus. *AoB Plants*, 12(5), 1–12.
- Noblick, L. R. (1998a). Syagrus botryophora. The IUCN Red List of Threatened Species. <https://doi.org/10.2305/iucn.uk.1998.rlts.t38698a10139959.en>
- Noblick, L. R. (1998b). Syagrus pseudococos. The IUCN Red List of Threatened Species. <https://doi.org/10.2305/iucn.uk.1998.rlts.t38702a10143363.en>
- Noblick, L. R. (2017). A revision of the genus *Syagrus* (Arecaceae). *Phytotaxa*, 294, 1–262.
- Noblick, L. R., Hahn, W. J., & Griffith, M. P. (2013). Structural cladistic study of Cocoseae, subtribe Attaleinae (Arecaceae): Evaluating taxonomic limits in Attaleinae and the neotropical genus *Syagrus*. *Brittonia*, 65, 232–261.
- Owens, H. L., & Guralnick, R. (2019). climateStability: An R package to estimate climate stability from time-slice climatologies. *Biodiversity Informatics*, 14, 8–13.
- Paz, A., Spanos, Z., Brown, J. L., Lyra, M., Haddad, C., Rodrigues, M., & Carnaval, A. (2019). Phylogeography of Atlantic Forest glassfrogs (*Vitreorana*): When geography, climate dynamics and rivers matter. *Heredity*, 122, 545–557.
- Peres, E. A., Pinto-da-Rocha, R., Lohmann, L. G., Michelangeli, F. A., Miyaki, C. Y., & Carnaval, A. C. (2020). Patterns of species and lineage diversity in the Atlantic Rainforest of Brazil. In V. Rull & A. C. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (pp. 415–447). Springer International Publishing.
- Peter, B. M., & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution*, 69, 721–734.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADSeq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135.
- Pinheiro, F., de Barros, F., Palma-Silva, C., Fay, M. F., Lexer, C., & Cozzolino, S. (2011). Phylogeography and genetic differentiation along the distributional range of the orchid *Epidendrum fulgens*: A neotropical coastal species not restricted to glacial refugia. *Journal of Biogeography*, 38(10), 1923–1935.



- Prance, G. T. (1982). Forest refuges: Evidence from woody angiosperms. *Biological Diversification in the Tropics*, 1, 137–158.
- Ribeiro, M. C., Martensen, A. C., Metzger, J. P., Tabarelli, M., Scarano, F., & Fortin, M.-J. (2011). The Brazilian Atlantic Forest: A shrinking biodiversity hotspot. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity hotspots: Distribution and protection of conservation priority areas* (pp. 405–434). Springer Berlin Heidelberg.
- Ribeiro, R. A., Lemos-Filho, J. P., Ramos, A. C. S., & Lovato, M. B. (2011). Phylogeography of the endangered rosewood *Dalbergia nigra* (Fabaceae): Insights into the evolutionary history and conservation of the Brazilian Atlantic Forest. *Heredity*, 106, 46–57.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584.
- Röser, M. (2000). *DNA amounts and qualitative properties of nuclear genomes in palms (Arecaceae)* (pp. 538–544). Systematics and Evolution.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34, 3299–3302.
- Shah, V. B., & McRae, B. H. (2008). Circuitscape: A tool for landscape ecology. In *Proceedings of the 7th Python in Science Conference* (Vol. 7, pp. 62–66). <https://circuitscape.org/>
- Silberbauer-Gottsberger, I. (1990). Pollination and evolution in palms. *Phyton: Annales Rei Botanicae*, 30(2), 213–233.
- Silva-Cardoso, I. M. d. A., de Souza, A. M., & Scherwinski-Pereira, J. E. (2017). The palm tree *Syagrus oleracea* Mart. (Becc.): A review. *Scientia Horticulturae*, 225, 65–73.
- Silveira, M. H. B., Mascarenhas, R., Cardoso, D., & Batalha-Filho, H. (2019). Pleistocene climatic instability drove the historical distribution of forest islands in the northeastern Brazilian Atlantic Forest. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 527, 67–76.
- Singarayer, J. S., & Valdes, P. J. (2010). High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quaternary Science Reviews*, 29, 43–55.
- Singh, R., Ong-Abdullah, M., Low, E.-T. L., Manaf, M. A. A., Rosli, R., Nookiah, R., Ooi, L. C.-L., Ooi, S.-E., Chan, K.-L., Halim, M. A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S. W., He, D., Hogan, M., Budiman, M. A., Lee, E. K., ... Sambanthamurthi, R. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, 500, 335–339.
- Souza, M. C. P., Moura, F., Silva, J. V., & Almeida, C. (2018). Phylogeography of the palm *Syagrus coronata* (Martius) Beccari (Arecaceae): Distribution in the "Caatinga" and Atlantic Forest domains. *Brazilian Journal of Botany*, 41, 849–857.
- Swofford, D. L. (2002). PAUP\*. *Phylogenetic analysis using parsimony (\* and other methods)* (Version 4). Sinauer Associates.
- Teixeira, M. C., Mäder, G., Silva-Arias, G. A., Bonatto, S. L., & Freitas, L. B. (2016). Effects of past climate on *Passiflora actinia* (Passifloraceae) populations and insights into future species management in the Brazilian Atlantic forest. *Botanical Journal of the Linnean Society. Linnean Society of London*, 180(3), 348–364.
- Thom, G., Smith, B. T., Gehara, M., Montesanti, J., Lima-Ribeiro, M. S., Piacentini, V. Q., Miyaki, C. Y., & do Amaral, F. R. (2020). Climatic dynamics and topography control genetic variation in Atlantic Forest montane birds. *Molecular Phylogenetics and Evolution*, 148, 106812.
- Thomé, M. T. C., Zamudio, K. R., Giovanelli, J. G. R., Haddad, C. F. B., Baldissera, F. A., Jr., & Alexandrino, J. (2010). Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution*, 55, 1018–1031.
- Tonini, J. F. R., Costa, L. P., & Carnaval, A. C. (2013). Phylogeographic structure is strong in the Atlantic Forest; predictive power of correlative paleodistribution models, not always. *Journal of Zoological Systematics and Evolutionary Research*, 51, 114–121.
- Townsend Peterson, A., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological Niches and Geographic Distributions* (MPB-49). Princeton University Press.
- Turchetto-Zolet, A. C., Salgueiro, F., Turchetto, C., Cruz, F., Veto, N. M., Barros, M. J. F., Segatto, A. L. A., Freitas, L. B., & Margis, R. (2016). Phylogeography and ecological niche modelling in *Eugenia uniflora* (Myrtaceae) suggest distinct vegetational responses to climate change between the southern and the northern Atlantic Forest. *Botanical Journal of the Linnean Society*, 182, 670–688.
- Valdez, L., & D'Elía, G. (2013). Differentiation in the Atlantic Forest: phylogeography of *Akodon montensis* (Rodentia, Sigmodontinae) and the Carnaval-Moritz model of Pleistocene refugia. *Journal of Mammalogy*, 94, 911–922.
- Vanzolini, P. E., & Williams, E. E. (1981). The vanishing refuge: A mechanism for ecogeographic speciation. *Papeis Avulsos de Zoologia*, 34, 251–255.
- Vasconcellos, M. M., Colli, G. R., Weber, J. N., Ortiz, E. M., Rodrigues, M. T., & Cannatella, D. C. (2019). Isolation by instability: Historical climate change shapes population structure and genomic divergence of treefrogs in the Neotropical Cerrado savanna. *Molecular Ecology*, 28(7), 1748–1764.
- Werneck, M. d. S., Sobral, M. E. G., Rocha, C. T. V., Landau, E. C., & Stehmann, J. R. (2011). Distribution and endemism of angiosperms in the Atlantic Forest. *Natureza & Conservação*, 9, 188–193.
- Wieringa, J. G., Boot, M. R., Dantas-Queiroz, M. V., Duckett, D., Fonseca, E. M., Glon, H., Hamilton, N., Kong, S., Lanna, F. M., Mattingly, K. Z., Parsons, D. J., Smith, M. L., Stone, B. W., Thompson, C., Zuo, L., & Carstens, B. C. (2020). Does habitat stability structure intraspecific genetic diversity? It's complicated. *Frontiers of Biogeography*, 12(2), 1–21.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16, 97–159.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114–138.

## BIOSKETCH

**Kathryn P. Mercier** is a PhD candidate at the City University of New York. She is interested in how environmental change shapes species distributions and dispersal. Her research integrates population genetics, niche modelling and demography, across space and time, with the ultimate goal of conserving biodiversity.

**Author contributions:** Kathryn P. Mercier, Mariana M. Vasconcellos, José Rubens Pirani, Fabián A. Michelangeli and Ana Carolina Carnaval conceived the ideas. Euder G. A. Martins, José Rubens Pirani, Mariana M. Vasconcellos and Fabián A. Michelangeli collected the data. Kathryn P. Mercier analysed the data and led the writing with contributions from all authors.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Mercier, K. P., Vasconcellos, M. M., Martins, E. G. A., Pirani, J. R., Michelangeli, F. A., & Carnaval, A. C. (2023). Linking environmental stability with genetic diversity and population structure in two Atlantic Forest palm trees. *Journal of Biogeography*, 50, 197–208. <https://doi.org/10.1111/jbi.14523>