Botanical Journal of the Linnean Society, 2023, **201**, 100–113. With 3 figures.

Population structure and signals of local adaptation in *Eugenia uniflora* (Myrtaceae), a widely distributed species in the Atlantic Forest

NICOLE MOREIRA VETÖ[1,+], DRAGOS POSTOLACHE[2,+], FRANK L. GUZMAN ESCUDERO[3], ELIA VAJANA[4], RICARDO BURGO BRAGA[5], FABIANO SALGUEIRO[6], ROGÉRIO MARGIS[1,7], GIOVANNI G. VENDRAMIN[8] and ANDREIA C. TURCHETTO-ZOLET[1,*]

[1]*Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 91501-970, Rio Grande do Sul, Brazil*
[2]*National Institute for Research and Development in Forestry 'Marin Drăcea', Cluj-Napoca Research Station, Strada Horea 65, Cluj-Napoca, 400202, Romania*
[3]*Universidad Científica del Sur, Lima, Peru*
[4]*Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
[5]*INNTI Consultoria Ambiental e Agricultura, Porto Alegre, 90880-410, Rio Grande do Sul, Brazil*
[6]*Laboratório de Biodiversidade e Evolução Molecular, Departamento de Botânica, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 22290-240, Rio de Janeiro, Brazil*
[7]*Departamento de Biofísica, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 91501-970, Rio Grande do Sul, Brazil*
[8]*Istituto di Bioscienze e Biorisorce, Consiglio Nazionale delle Ricerche, Area di Ricerca di Firenze, Sesto Fiorentino, Italy*

Received 4 May 2021; revised 23 December 2021; accepted for publication 8 February 2022

Evolutionary diversity in species can arise in many ways, including local adaptation. Despite the global importance of tropical forest ecosystems, few studies have explored patterns of local adaptation in tropical tree species. We investigated population genetic structure and adaptive genetic diversity in *Eugenia uniflora*, a widely distributed tree species across the Atlantic Forest Domain (AFD). We sampled *E. uniflora* in distinct environments from the AFD and genotyped 523 single-nucleotide polymorphisms (SNPs) using a candidate gene approach. We studied neutral population structure and used two outlier methods based on population differentiation (Bayescan and LEA) and one gene-environment association analysis (LFMM) to detect putative SNPs under divergent selection. We detected spatial population structuring between the northern and southern range of the species distribution and higher genetic diversity in southern populations from the riparian forest. We found 18 candidate loci potentially involved in local adaptation, among which LEA and LFMM both detected seven loci. Overall, precipitation-related variables were more represented in statistically significant genotype-climate associations (*c.* 60%). Our study provides a first insight into the distribution of adaptive genetic variation in *E. uniflora*, highlighting how tropical tree species may adapt over time and across the AFD.

ADDITIONAL KEYWORDS: adaptation – candidate genes – genetic variation – Neotropics – pitanga – SNPs.

*Corresponding author. E-mails: aturchetto@gmail.com; carina.turchetto@ufrgs.br
+These authors contributed equally.

© The Author(s) 2022. Published by Oxford University Press on behalf of The Linnean Society of London. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

INTRODUCTION

Spatial environmental variation is ubiquitous, and populations may display adaptive divergence due to local adaptation to biotic and abiotic conditions as observed across a wide range of species (Savolainen *et al.*, 2013). Local adaptation is driven by spatially divergent selection, favouring genetic variants that confer an adaptive advantage under specific environmental conditions (Potvin & Tousignant, 1996; Kawecki & Ebert, 2004). Consequently, deciphering the genetic architecture of local adaptation is of primary concern to outline the evolutionary potential of species in the face of environmental change (e.g. global warming).

High-throughput sequencing technologies have substantially contributed to the advancement in detecting adaptive genetic variation in non-model organisms (De Wit *et al.*, 2015). In particular, uncovering single-nucleotide polymorphisms (SNPs) from within candidate genes is considered an effective strategy for discovering loci underlying local adaptation in species with large and complex genomes (Wang *et al.*, 2009; Gayral *et al.*, 2013; De Wit *et al.*, 2015). In these cases, the candidate gene approach turns out to be particularly cost-effective as it allows genome sampling efforts to be targeted towards genes of known biological function. Gene annotation in non-model species has been further facilitated by the availability of transcriptome sequence data that allowed the identifying of previously unknown candidate genes underlying local adaptation (Lalagüe *et al.*, 2014; De Wit *et al.*, 2015).

The study of local adaptation is of high pertinence for the Atlantic Forest Domain (AFD) since several environmental factors, such as strong seasonality shifts in climate, sharp environmental gradients and orography-driven rainfall, make this ecoregion a heterogeneous, complex landscape (Joly *et al.*, 1999; de Mello Martins, 2011). The AFD is also one of the most diverse and threatened biomes on the planet at the same time, where only *c.* 7.3% of its primary forest can be currently considered to remain intact. This biome consists of a set of distinct vegetation formations, e.g. tropical and subtropical moist broadleaf forests, tropical and subtropical dry broadleaf forests, wetland and mangrove forests, tropical and subtropical grasslands and open scrub vegetation of the sandy coastal plains (Joly *et al.*, 1999; Morellato & Haddad, 2000; Oliveira-Filho & Fontes, 2000) where the species can occupy one or more phytogeographic regions.

Despite its ecological relevance, there remains a significant lack of plant studies exploring local adaptation in the AFD (but see Brousseau *et al.*, 2018; Collevatti *et al.*, 2019), since most of the research based on landscape genomics was mainly conducted in other biomes such as temperate habitats in Europe and North America. More specifically, a gap exists in the knowledge of both genetic diversity and population structure of Myrtaceae, one of the most ecologically important plant families in the Neotropics, that may undermine our capacity to tackle climate change and preserve their ecological function in the AFD (Margis *et al.*, 2002; Salgueiro *et al.*, 2004; Franceschinelli *et al.*, 2007; Ferreira-Ramos *et al.*, 2014; Brandão *et al.*, 2015; Turchetto-Zolet *et al.*, 2016; Santos & Gaiotto, 2020, Souza-Neto *et al.*, 2022).

Eugenia uniflora L. (Myrtaceae), commonly known as 'pitanga', is endemic to the AFD. It grows in a variety of habitats, from the north-eastern sandy coastal plain environment of Brazil, regionally called restinga, to the southern riparian vegetation, extending up to 400 km inland, in Brazil, Argentina and Uruguay (Fig. 1A; De Almeida *et al.*, 2012). A remarkable phenotypic variation is observed throughout its distribution range (Fig. 1B). The restinga environmental niche comprises a set of interdependent coastal and plain vegetation habitats, encompassing edaphic conditions in the southern Atlantic coastal tropical/subtropical regimes and high environmental heterogeneity in terms of salinity, sandy soils, topography, temperature amplitudes (daily/seasonal), water availability, wind patterns and related sea spray (Scarano & Ceotto, 2015; Table 1). In contrast, the riparian forest is a mosaic environment, defined by several distinct riverine-border plant communities and habitats (Ribeiro, 2000; Ribeiro *et al.*, 2009), that allows variable access to sunlight, water and flood regimes. Thus, *E. uniflora* experiences multiple and complex environments (Fig. 1), occurring as a shrub or small tree in the sandy coastal plain environments (restinga) and as a tree in southern Brazil (riparian forest). Furthermore, leaves on plants from restinga tend to be thicker than those from riparian individuals.

Previous studies based on plastid markers showed population structuring to reflect environmental heterogeneity in *E. uniflora* (Turchetto-Zolet *et al.*, 2016). Distinct patterns of genetic diversity and demographic history of *E. uniflora* were also observed among populations from restinga and riparian forests, overall making this species a potentially good candidate for adaptation studies.

In this study, we test the hypothesis that local adaptation has been consistently shaping adaptive variation in this species and use 523 SNPs located in 12 stress-related candidate genes to detect putative signatures of divergent selection. In particular, we aimed to: (i) outline the genetic structure of *E. uniflora* populations; (ii) unravel geographical patterns of adaptive genetic variation linked with specific environmental conditions across the geographical

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

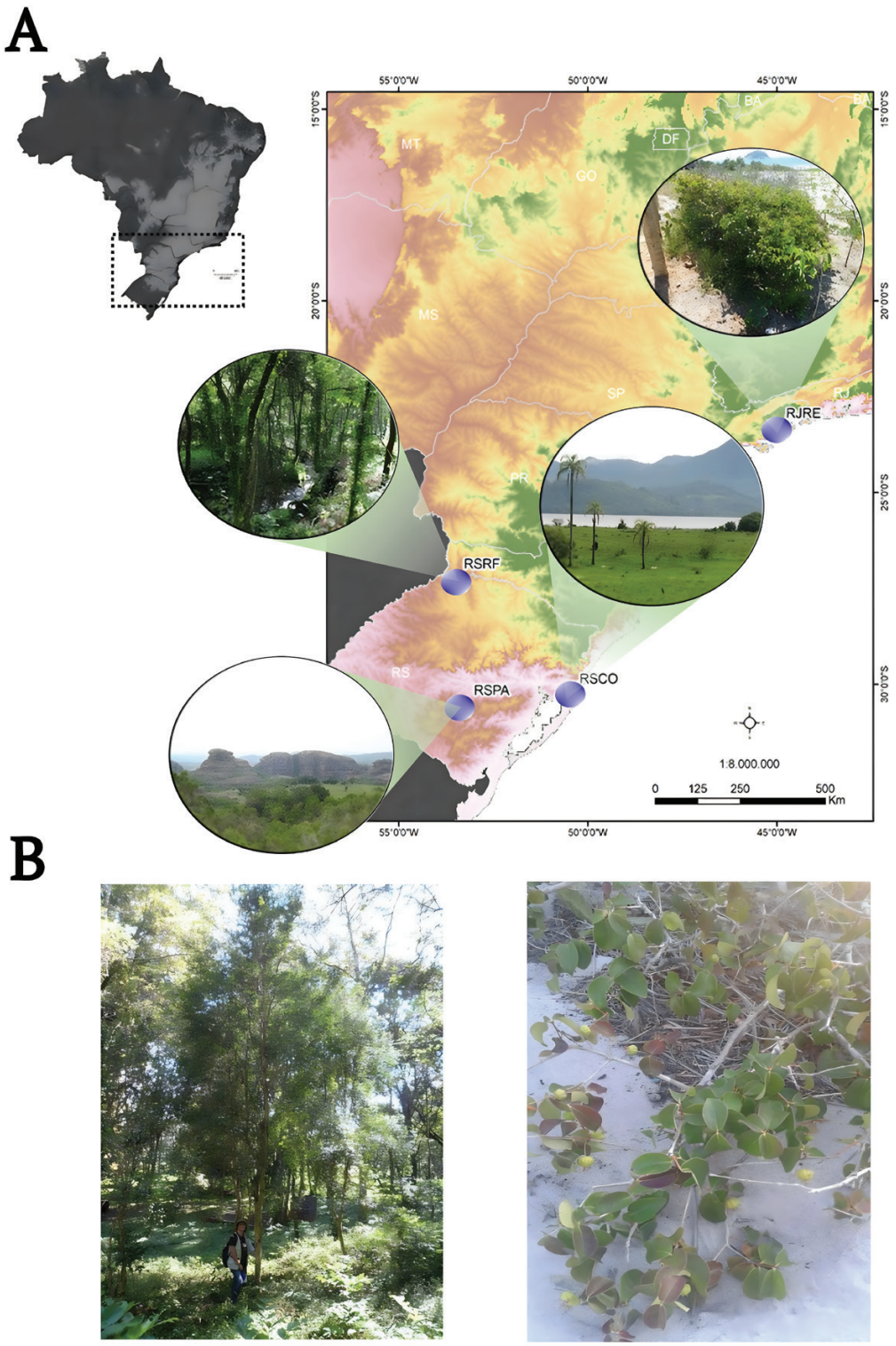


Figure 1. A, map of showing the locations of the four *Eugenia uniflora* populations used in this study and theis ecosystem types. B, the phenotypes of contrasting *E. uniflora* populations, from the riparian forest as a large tree (left) and restinga characterized as a shrub (right).

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

range of *E. uniflora*; and (iii) obtain a set of putative molecular footprints of local adaptation and link them with their biological meaning.

MATERIAL AND METHODS

Sampling collection and candidate gene selection and amplification

We sampled leaves from 96 adult *E. uniflora* trees distributed in four populations of *E. uniflora* growing in distinct environments in the AFD that encompass the natural distribution of the species (Fig. 1; Table 1): one population from the restinga environment in the state of Rio de Janeiro (hereafter RJRE; 32 individuals); three populations in the state of Rio Grande do Sul, including a riparian forest (RSRF; 32 individuals); and two populations from the coastal plain and adjacent grasslands environment, locally referred to as the Pampa biome (RSCO and RSPA, made up of 18 and 14 individuals, respectively; Fig. 1A; Table 1). Representative voucher specimens for two individuals of *E. uniflora* were collected and deposited in the Herbário ICN, Instituto de Biociências (UFRGS, Porto Alegre, Brazil), under the accession numbers ICN167404 and ICN167405.

Environmental stress-related candidate genes were selected based on a literature survey of proteins/genes involved in plant response to abiotic stress. The query was run on the 2014–2015 ISI Web of Knowledge using the keywords 'abiotic and biotic stress', 'drought stress', 'salinity stress' and 'temperature stress'. The sequences of the selected environmental stress-related candidate genes were obtained and used as a query to perform Basic Local Alignment Search Tool (BLAST) searches against the transcriptome of *E. uniflora* (Guzman *et al.*, 2014). The annotation and gene ontology of the genes found were performed using BLAST2GO (Conesa *et al.*, 2005). Finally, the primers for polymerase chain reaction (PCR) amplification were designed in the 5' and 3' untranslated regions (UTRs) using Primer3 (Koressaar & Remm, 2007; Untergasser *et al.*, 2012).

Total genomic DNA was extracted from silica gel dried leaves (Doyle & Doyle, 1990) and quantified using Nanodrop. PCR reactions were performed to amplify the 48 candidate genes found in the 96 DNA samples using an Applied Biosystems 7500 Real-Time PCR system. PCR reactions were carried out in a final volume of 20 µL, containing *c.* 70 ng genomic DNA, 2 mM magnesium chloride, 0.1 mM deoxynucleotide triphosphate mix, 1 × PCR buffer, 0.3 U Platinum Taq DNA polymerase (Invitrogen), 1 × SYBR Green I and 5 µM of each primer. After initial testing, 30 loci were successfully amplified. The thermal cycling conditions were as follows: an initial hot-start step at 95 °C for 5 min, followed by 35 cycles with denaturation at 94 °C for 30 s, an annealing temperature of 60 °C for 30 s, elongation at 72 °C for 3 min, and a final extension of 10 min at 72 °C. The amplification products were checked and quantified based on SYBR Green fluorescence using 7500 Fast System v.2.0.3 software. The respective primer sequences of the 30 genes are described in the Supporting Information (Table S1).

Sequencing, SNP marker discovery and genotyping

PCR products of the 30 genes from the 96 individuals were sequenced using the Illumina Miseq 2000 platform. The libraries were prepared using the Nextera XT DNA Sample Preparation Kit following the manufacturer's instructions. This kit requires only 1 ng template DNA and can be performed with multiple samples being prepared simultaneously. Briefly, the PCR products were pooled for each plant to achieve fragmentation and the addition of unique adapter sequences in the input DNA. Next, DNA was amplified by PCR, and index sequences on both ends of the DNA for each individual were added. Finally, the libraries

Table 1. Information about the four populations sampled: *N*: total number of collected individuals; Sample collection site: Lat.: latitude, Long.: longitude and Elev.: elevation; Soil type - descriptions for each environment where populations are located. The populations were signed as RJRE, RSRF, RSPA and RSCO

Population	*N*	Sample collection site			Soil type
		Lat. (°)	Long. (°)	Elev. (m)	
Restinga (RJRE)	32	-23.06	-43.53	11	Nutrient poor, sandy, saline and acid
Riparian Forest (RSRF)	32	-27.32	-53.33	536	Nutrient rich, organic material and water accumulation
Pampa (RSPA)	14	-30.52	-53.49	378	Neosols, shallow lands, very low permeability
Coastal Plain (RSCO)	18	-30.19	-50.49	7	Sandy, water stressed, low fertility, lagoon and marine environments

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

were quantified using a Qubit Fluorometer and pooled for sequencing using MiSeq v.3 from the Instituto de Ciência e Tecnologia dos Alimentos (ICTA), UFRGS, Porto Alegre, Brazil.

The overall procedure for identifying SNP markers is shown in the Supporting Information (Fig. S1). First, the quality of the raw reads was visualized using FastQC software v.0.11.3 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the remaining adapter sequences and low-quality ends < 30 were trimmed off at the 3' end using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). All cleaned reads were *de novo* assembled separately for each individual using a Trinity assembler (Grabherr *et al.*, 2011) to obtain the corresponding genomic sequences of the candidate genes used for reference in this study. This program is commonly employed to assemble transcriptomes and assembled contigs using low reads coverage. After the assembly process, the complete sequence of the 96 individuals was chosen as the final reference gene, considering the greatest length and the lack of ambiguity of the bases.

Next, the genic structure of the obtained reference sequences was predicted with the FGENESH software (Solovyev *et al.*, 2006) relying on the *Eucalyptus grandis* W.Hill ex Maiden model. Validation of the predicted exons and introns was done by anchoring *E. uniflora* RNAseq reads in the genomic references (Guzman *et al.*, 2014); with this aim, TopHat2 (Kim *et al.*, 2013) and BLASTN (Altschul *et al.*, 1990) were used between the predicted transcript and the unigene used in the primer designs. This information was used to classify the SNPs by location, i.e. in coding or noncoding DNA. The cleaned reads were mapped to the reference using BWA v.0.7.12 (Li & Durbin, 2010), with the parameter -n (number of allowed mismatches) set as default to detect variations between the reads of each individual and the reference. The SAMtools software (Li *et al.*, 2009) was used to convert the sequence alignment/map (SAM) files into sorted binary alignment/map (BAM) files. The SNPs and INDELs were then identified by using the module mpileup of SAMtools and BCFtools (http://samtools.github.io/bcftools/bcftools.html), respectively. The identification was limited to biallelic sites and separated from the generated VCF file using VCFtools (Danecek *et al.*, 2011). SNPs with a genotype quality (GQ) of < 99 and a read depth (DP) of < 20 were filtered out using the VCFlib software (https://github.com/vcflib/vcflib). These stringent criteria were applied to the known variants to eliminate suspicious loci and select only high-confidence SNPs to be used in the subsequent analyses.

Finally, the software plink v.2.0 (Chang *et al.*, 2015; Prive *et al.*, 2018) was used for checking the overall quality of the molecular data set: SNPs with minor allele frequency (MAF) < 1% were eliminated together with markers/individuals with a call rate < 90%, potential genotyping errors were evaluated through the Hardy–Weinberg equilibrium exact test (Wigginton *et al.*, 2005) and eventual replicates identified through the KING-robust kinship coefficient estimator (Manichaikul *et al.*, 2010).

Molecular data sets

An ad hoc R script that relies on the LD function from the R package genetics (Warnes *et al.*, 2021) was used to remove markers showing statistical evidence of linkage disequilibrium (Supplementary material S10). *P*-values from the LD tests were corrected for multiple testing using the p.adjust function in R with the Bonferroni method. Linkage disequilibrium was considered to be statistically significant at a nominal significance threshold equal to 0.01. After removing linked loci, the LD-pruned version of the SNP data set comprised 372 SNPs which were used in Bayescan, LEA and LFMM analyses.

Furthermore, an ad hoc molecular data set for genetic structure analyses was obtained by removing the outliers as defined by LEA from the LD-pruned molecular data set to decrease the possible bias induced by adaptive loci in the study of neutral structure; then, this version of the data set was used as input for STRUCTURE, DAPC, PCA and sNMF analyses.

Genetic diversity and population structure

GenAlEx v.6.5 (Peakall & Smouse, 2012) was used to assess the genetic diversity in the studied populations of *E. uniflora*. We analysed the observed and expected heterozygosity (H_O and H_E, respectively), tested for Hardy–Weinberg equilibrium and examined the molecular variance (AMOVA) and G_{ST} values to obtain pairwise population estimates of genetic differentiation. We also studied genetic differentiation among populations using the pairwise F_{ST} analysis and Hardy–Weinberg equilibrium test as implemented in Arlequin v.3.1.1 (Excoffier *et al.*, 2005).

The nucleotide diversity and substitutions of synonymous and nonsynonymous sites were estimated at the gene level based on the complete sequence of each candidate gene successfully sequenced using Arlequin and DnaSP (Rozas *et al.*, 2017). We performed neutrality tests (Tajima's D and Fu's Fs) at the gene level in Arlequin to test whether the frequency distribution of polymorphic sites deviated from the neutral equilibrium expectation.

Further, the software STRUCTURE v.2.3.4 (Pritchard *et al.*, 2000), the discriminant analysis of principal components (DAPC) and principal components analysis (PCA) were used to analyse the neutral genetic structure in sampled populations of

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

E. uniflora. Posterior probabilities in STRUCTURE were computed using a Markov chain Monte Carlo (MCMC) method with > 2 000 000 iterations of each chain, following a 100 000-iteration burn-in period. Inference of population structure was made by applying the admixture model with correlated allele frequencies (Falush *et al.*, 2003). We performed ten independent runs for each K value, with K values ranging from one to five. The best K was selected based on the log-likelihood value, and populations were assigned to each genetic cluster by considering the membership probabilities of each individual. Then, we used CLUMPAK (Kopelman *et al.*, 2015) to order and visualize the STRUCTURE results.

DAPC (Jombart & Collins, 2015; Jombart *et al.*, 2010) was also used to identify the main genetic clusters within the data set. According to the DAPC procedure, a PCA was first conducted to summarize genetic information into a few but representative synthetic dimensions (here, accounting for 99% of the original molecular variance); then, K-means analysis was performed based on the retained dimensions and assuming up to ten genetic clusters to individuate the most likely K as based on the Bayesian information criterion (BIC). Subsequently, DAPC was run based on both the best K and an intentionally excessive amount of synthetic dimensions to obtain an a-score profile and individuate an optimal number of principal components to avoid overfitting issues. Finally, DAPC was rerun by relying on the identified optimal number of principal components and assuming a number of genetic clusters equal to the best K.

As an alternative to Bayesian clustering algorithms, an individual-based PCA was also used to describe genetic population structure (Lee *et al.*, 2009). The analysis was performed based on the dudi.pca function as implemented in the R package adegenet (Jombart, 2008).

Climate data

The sampled populations were matched with 19 bioclimatic variables derived from the WorldClim database v.1.4 at the highest spatial resolution of 30 arcsec and for the period 1960–1990 (Supporting Information, Table S2) (Hijmans *et al.*, 2005). To avoid testing collinear environmental variables in the gene-environment association tests possibly leading to redundant results while guaranteeing an acceptable degree of interpretability at the same time, a PCA was separately conducted on temperature- and precipitation-related bioclimatic variables to obtain orthogonal and synthetic bioclimatic axes to be processed by LFMM (Rellstab *et al.*, 2015). The analysis was performed with the R package FactoMineR (Lê *et al.*, 2008). Based on the cumulative amount of variance explained by the components, the first three temperature-related PCs (Temp1, Temp2, Temp3) and the first three precipitation-related PCs (Precip1, Precip2, Precip3) were retained for subsequent analyses (Supporting Information, Table S4).

Detection of signatures of divergent selection

Two different population differentiation-based methods (Bayescan and LEA) and one gene-environment association (GEA) method were employed to detect putative signatures of divergent selection. The Bayesian method implemented in BayeScan v.2.1 was used to estimate the posterior probability that a given SNP is under selection by comparing a selection-based model vs. a neutral model (Foll & Gaggiotti, 2008). BayeScan uses a reversible jump MCMC algorithm (Green, 1995) to estimate the posterior probability for each model by relying on a Dirichlet multinomial distribution for gene frequencies (Excoffier *et al.*, 2009; Lotterhos & Whitlock, 2014). Default parameter values were used with a sample size of 5000, two prior odds for the neutral model of 10:1 and 1:1, 20 pilot runs of 5000 iterations each, and an additional burn-in of 50 000 iterations. Outliers in BayeScan were identified applying a false discovery rate cut-off of 0.05. The population differentiation-based method implemented in the R package LEA (Frichot & François, 2015) performs genome scans to detect loci under selection while correcting for the potentially confounding effect due to population structure and calibrating statistical models (Frichot *et al.*, 2015). Two steps were performed using LEA: first, we estimated population genetic structure using the sNMF (Frichot & François, 2015) function within LEA; sNMF was run from one to five ancestral populations with ten replicates per K and recording the cross-entropy value associated with each run. Then, the most likely number of ancestral populations was identified based on the cross-entropy criterion and outlier identification was achieved based on the most likely K previously chosen.

GEA analysis was performed based on the latent factor mixed model (LFMM) approach as implemented in the R package lfmm (Caye & François, 2017). LFMM allows the correction of GEAs for a given number of latent factors representing population stratification as estimated from the molecular data set through PCA (Frichot *et al.*, 2015). In this analysis, such a correction was defined on the basis of ten independent runs of sNMF having as input the molecular data set pruned for LEA outliers (see section 'Molecular data sets'). The analysis tested six focal environmental variables, comprising three temperature-related synthetic variables and three precipitation-related

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

synthetic variables (see section: 'Climate data'). LFMM parameters were estimated with the lfmm_ridge function, which provides estimates for the effect sizes of both the estimated latent variables and the focal environmental variable. Then, the lfmm_test function was run to return the calibrated significance values for each GEA.

The Bonferroni correction with a nominal significance cut-off of 0.01 was applied after both LEA and LFMM analyses to correct for multiple testing (Forester *et al.*, 2018). A Venn representation was produced with the web-based tool InteractiVenn to visualize the overlaps among the three methods (Heberle *et al.*, 2015). Full R scripts, data input and results of LEA and LFMM analyses have been deposited in GitHub: https://github.com/ellios84/Vet-Postolache-et-al.-BOTJLS.

RESULTS

Gene selection, gene sequencing and SNP marker genotyping

The BLAST2go analysis allowed us to retrieve the gene ontology terms (GO terms) for the candidate genes under scrutiny in *E. uniflora*. In particular, these genes resulted in molecular, biological and cellular processes dealing with various mechanisms of response to abiotic stresses like drought, oxidative and osmotic stress, senescence and seed development.

Thirty candidate genes were successfully amplified in the 96 individuals of *E. uniflora* (Supporting Information, Table S1) and then submitted to library construction and sequencing. Finally, 12 out of the 30 amplified candidate genes were successfully sequenced in the studied individuals. Out of the total 20 981 110 candidate gene reads obtained from the 96 samples, 1 471 146 reads (7%) were successfully mapped to the reference transcriptome sequences (Supporting Information, Fig. S1).

SNP data set quality control

From the initial data set of 685 SNPs, we discarded 160 SNPs due to low call rate and two SNPs due to excessive heterozygosity ($H_o > 0.5$). After amplifying, sequencing and filtering for missing data, 523 SNP markers across 12 genes in 85 individuals were successfully genotyped (Table 2; Supporting Information, Table S3). Owing to the removal of 23 linked markers and 128 homozygous SNPs, the LD-pruned data set comprised 372 SNPs from within 12 genes and 80 individuals (Supporting Information, Fig. S2). After removing LEA outliers, the molecular data set used for population structure analysis encompassed 361 SNPs.

Genetic diversity and population structure

Global genetic diversity as expressed by H_o per population ranged from 0.040 (RJRE) to 0.122 (RSRF; Table 3). The northern population (RJRE) showed the lowest percentage of polymorphic loci (16.95%) compared to the southern populations (RSPA, RSCO and RSRF), which showed higher values (51.81%, 45.14% and 73.33%, respectively). In the Arlequin analysis, 33 loci revealed significant deviations from the

Table 2. Information about the 12 successfully sequenced genes, including predicted size, sequenced coverage, reads and SNPs genotyped by gene

Gene	Locus identification	Predicted size	Sequenced coverage	Reads (bp)	SNPs
Dehydrin	*E. Uni 63*	744	1188	147.480	30
Universal stress protein family protein	*E. Uni 7951*	480	1984	105.037	38
Late embryogenesis abundant protein Lea14-A	*E. Uni 15860*	453	701	253.077	22
Dof zinc finger (*dof17*)	*Euni16779_1574*	828	1574	315.877	17
Mitogen-activated protein kinase (*Mapk5*)	*Euni1727_1212*	585	1212	114.325	12
Calcineurin B-like protein 01 (*CBL01*)	*Preuni17960_663*	477	663	28.981	113
Chloroplast chlorophyll A-B binding protein (*CBP3*)	*Preuni31_2107*	426	2107	80.509	38
Histone H4-like protein	*Preuni17184_671*	309	671	214.997	23
Osmotic stress-activated protein kinase (*OSAK*)	*Preuni5552_1340*	699	1340	153.326	68
Phosphoglycerate kinase (chloroplastic isoform 1)	*Preuni328_2200*	873	2200	155.070	88
Plasma intrinsic protein	*Preuni7330_1053*	483	1053	106.777	34
Pentatricopeptide repeat-containing protein At1g15510	*Preuni9300_1515*	1164	1515	431.094	40

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

Hardy–Weinberg equilibrium in the four populations ($P < 0.05$; Supporting Information, Table S5).

Nucleotide diversity was calculated based on the complete sequences of 12 genes in 85 samples (Supporting Information, Table S6). Total nucleotide diversity was calculated for each gene and ranged from 0.002 (mitogen-activated protein kinase; *Mapk5*) to 0.007 (calcineurin B-like protein 01; *CBL01*), and total haplotype diversity ranged from 0.929 (Dof zinc finger; *dof17*) to 0.993 (universal stress protein family protein). When between-population nucleotide and haplotype diversities were compared for each gene, southern populations (RSRF, RSPA and RSCO) always displayed higher estimates than the northern population (RJRE). Notably, the only exception was the pentatricopeptide repeat-containing protein

Table 3. Statistics of genetic diversity in the four populations of *E. uniflora* studied. The indices include: N: number of genotypes; Na: number of different alleles; Ne: number of effective alleles; I: Shannon's information index; H_O: observed heterozygosity; H_E: expected heterozygosity; uH_E: unbiased expected heterozygosity; F: fixation index

Northern		N	Na	Ne	I	H_O	H_E	uH_E	F
RJRE	**Mean**	30.545	1.166	1.062	0.065	0.040	0.041	0.041	0.015
	SE	0.037	0.016	0.008	0.007	0.005	0.005	0.005	0.008
Southern		**N**	**Na**	**Ne**	**I**	H_O	H_E	uH_E	**F**
RSRF	**Mean**	28.885	1.732	1.188	0.214	0.122	0.127	0.129	0.020
	SE	0.018	0.019	0.011	0.009	0.006	0.006	0.007	0.009
RSPA	**Mean**	13.887	1.516	1.171	0.179	0.109	0.109	0.114	-0.010
	SE	0.014	0.022	0.012	0.010	0.007	0.007	0.007	0.011
RSCO	**Mean**	10.880	1.449	1.170	0.175	0.107	0.109	0.114	-0.010
	SE	0.014	0.022	0.012	0.010	0.007	0.007	0.007	0.011

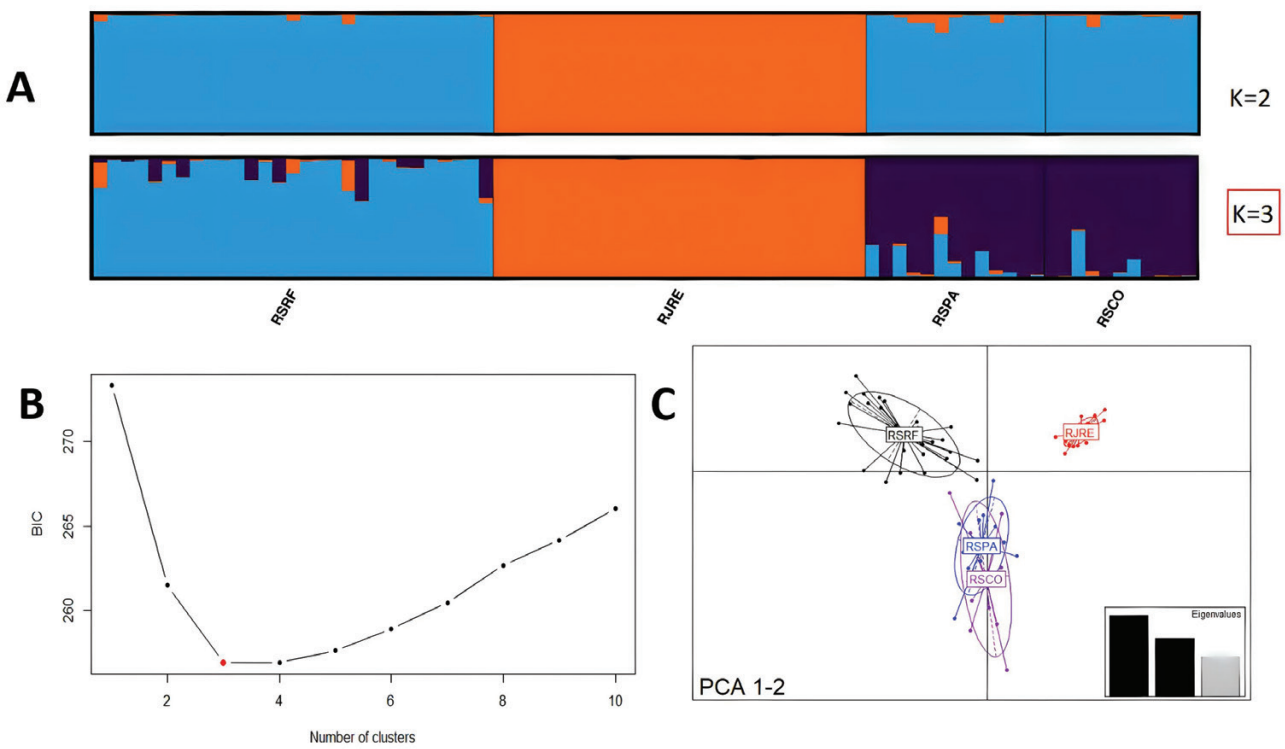


Figure 2. Spatial genetic admixture of 85 *E. uniflora* individuals across four populations, based on genotypes at the 523 SNPs. A, Bayesian analysis of population structure, including all four sampled populations of *E. uniflora*. B, variation in the Bayesian information criterion (BIC) with the possible number of clusters. C, PCA on allelic frequencies across 372 SNPs.

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

gene, for which both indices displayed similar values among populations. Significantly negative Fu's Fs values ($P < 0.05$) were obtained for each of the 12 genes analysed in the four populations, suggesting positive selection or population expansion (Supporting Information, Table S5).

All structure analyses performed (STRUCTURE, DAPC and PCA) pointed to the existence of three main genetic clusters (Fig. 2). The first genetic cluster was represented by the northern population (RJRE) from the restinga environment, the second genetic cluster was composed of individuals from the riparian forest population (RSRF) and the third genetic cluster was composed of the coastal plain (RSPA) and Pampa biome (RSCO) populations (Fig. 2A, C). The first three PCs explained up to 100% of total molecular variance (PC1: 54.46%; PC2: 31.12%). The AMOVA found that 25% of the total molecular variance can be explained by population divergence and 60% to genetic differences between individuals (Supporting Information, Table S7). F_{ST} values ranged between 0.05 and 0.44, with the highest estimate being observed between the Pampa biome (RSCO) and the restinga northern population (RJRE) and the lowest between the coastal plain population (RSPA) and the Pampa biome (RSCO; Supporting Information, Table S8).

Outlier detection tests

No outlier was detected by the BayeScan analysis with prior odds of 10:1 for the neutral model, where the highest posterior probability for the model including selection was 0.33. When using prior odds of 1:1 for both models, the highest posterior probability was 0.77, which corresponds to substantial evidence of divergent selection (Jeffreys, 1961). However, this result was not found to be statistically significant after multiple testing corrections.

The sNMF analysis prior to LEA identified a knee in the cross-entropy decay at K = 3. Then, the population differentiation-based test was conducted by assuming three ancestral populations and identifying 11 SNPs (Supporting Information, Table S9) as potentially under selection.

The sNMF analysis based on the set of putatively neutral SNPs prior to LFMM analysis was unable to identify a clear solution by pointing to K = 3 and K = 4 as the most likely numbers of ancestral populations; therefore, two independent LFMM analyses were performed with three and four latent factors assumed (Supporting Information, Fig. S3) and the resulting 14 outliers shared by these analyses (Supporting Information, Fig. S4) were considered as the final output from the LFMM analysis. Among these putative signatures of selection, seven SNPs (SNP070, SNP086, SNP109, SNP134, SNP139, SNP145, SNP169) were associated with precipitation-related variables, one SNP (SNP514) with temperature-related variables and six SNPs (SNP076, SNP104, SNP151, SNP159, SNP395, SNP418) correlated with both precipitation- and temperature-related variables (Fig. 3; Table 4; Supporting Information, Table S10).

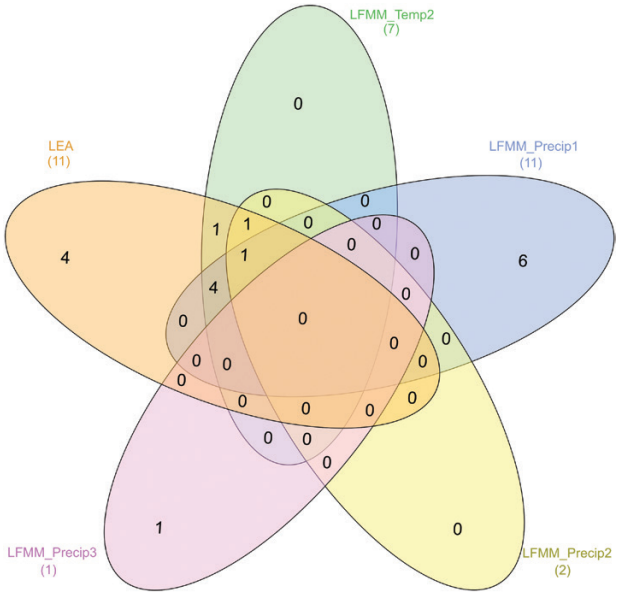


Figure 3. Venn diagram indicating the candidate outlier loci of *E. uniflora* identified among the LEA (11) and LFMM analyses comprising Temp2 (6), Precip1 (11), Precip2 (1) and Precip3 (1).

Table 4. Number of loci outliers found using LEA and LFMM analyses

Method	Total outliers detected
LEA	11
LFMM	14
[Temp2-PC2]	7
[Precip1-PC1]	11
[Precip2-PC2]	2
[Precip3-PC3]	1

Overlapping signatures of selection

Overlapping signatures of selection from LEA and LFMM analyses were detected at seven SNPs, among which SNP104 showed an association with Temp2, Precip1 and Precip2; SNP076, SNP151, SNP159 and SNP395 showed an association with both Temp2 and Precip1; SNP418 with both Temp2 and Precip2 and SNP514 with Temp2 (Fig. 3; Table 5). The number of shared outliers per sequenced gene indicates that SNP076, SNP104, SNP151 and SNP159 are located in the *CBL01* and that the remaining three

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

Table 5. Common loci outliers between LEA and LFMM analyses indicating feature, reference allele (Ref) and alteration (Alt)

Analysis	SNP	Gene	Feature	Ref	Alt
LEA and LFMM (Temp2, Precip1 and Precip2)	SNP104	Calcineurin B-like protein 01 (*CBL01*)	Intron	C	G
LEA and LFMM (Temp2 and Precip1)	SNP076	*CBL01*	3' UTR	A	G
	SNP151	*CBL01*	Intron	A	C
	SNP159	*CBL01*	Intron	C	T
	SNP395	Dehydrin	Intron	A	G
LEA, LFMM (Temp2, Precip2)	SNP418	Plasma intrinsic protein	CDSi	G	C
LEA and LFMM (Temp2)	SNP514	Pentatricopeptide repeat-containing protein At1g15510	CDSo	A	T

SNPs are located in a dehydrin (SNP395), a plasma intrinsic protein (SNP418) and a pentatricopeptide repeat-containing protein At1g15510 (SNP514; Fig. 3; Supporting Information, Table S9).

DISCUSSION

In this research, we identified candidate loci showing evidence of divergent selection in natural populations of *E. uniflora* (Tables 4–5). In particular, the GEA analysis performed made it possible to detect a set of SNPs significantly associated with temperature- and precipitation-related climatic variables, thus suggesting that local adaptation may have played a non-negligible role during adaptation and persistence of *E. uniflora* towards the heterogeneous and diversified habitats that compose the AFD. The AFD includes distinct environments, and thus local adaptations may be an essential process in shaping plant species margins.

Based on a data set of 523 SNPs distributed in 12 candidate genes, higher genetic diversity was detected in *E. uniflora* populations from the riparian forest (southern part of the species range), which substantiates previous findings based on plastid markers (Turchetto-Zolet *et al.*, 2016) and in *E. uniflora* populations distributed along the Brazilian Atlantic Coast, where the highest genetic diversity was observed in the southernmost population (Salgueiro *et al.*, 2004). Similar patterns in spatial genetic distribution have been found in other species in the AFD, including *Epidendrum fulgens* (Pinheiro *et al.*, 2011), which may reflect similar dispersal patterns driven by shared seed dispersers and analogous selective pressures (e.g. soil characteristics, proximity to the sea, potential for drought, temperature and precipitation regimes).

Population genetic structure analyses pointed to the co-existence of three main genetic clusters within the studied metapopulation which correspond to the restinga ecosystem located in the northern AFD (RJRE) and a southern AFD cluster, which included subclusters from the riparian forest (RSRF) and from the Pampa and coastal plain ecosystems (RSPA and RSCO, respectively) in the southern AFD. Such a spatial genetic structure might be possibly driven by the southmost to northmost differentiated environmental variables, such as salinity, temperature, rainfall and soil type (Table 1). Furthermore, this finding (as based on the newly developed SNP data set genotyped in candidate genes) seems to complement a previously published phylogeographic study focused on the species and based on plastid markers that shows the existence of two haplogroups, one from the north and one from the south, which probably diverged *c.* 5 Mya (Turchetto-Zolet *et al.*, 2016). The large proportion of molecular variance explained by genetic differentiation between individuals (here, 75%; Supporting Information, Table S7) is also in line with previous findings reporting that 78.9% of the genetic variation in coastal populations of *E. uniflora* has a within-population origin (Salgueiro *et al.*, 2004).

A recent study focused on *Euterpe edulis*, a widespread tree species in the Brazilian Atlantic Forest, based on SNP markers identified outlier loci possibly under divergent selection in distinct environments in the AFD (Brancalion *et al.*, 2018). The findings from this study were further corroborated by reciprocal transplants that provided evidence of local adaptations in seedlings established in the rainforest and the semi-deciduous forest, but not in the restinga and the submontane rainforest.

The precipitation-related variables resulted in a higher number of significant GEAs (14; Supporting Information, Table S10) compared to temperature-related climatic variables (7); this result can be explained by the known influence of precipitation and elevation in shaping the ecological niche of *E. uniflora* (Turchetto-Zolet *et al.*, 2016). A substantial number of outliers overlapping between LEA and LFMM were in the *CBL01* gene (SNP076, SNP104, SNP151, SNP159) and most of them turned out to be associated with Precip1 and Temp2 (Table 5). Studies have

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

reported *CBL01* to be involved with several stress responses including sensitivity to salinity in corn (*Zea mays*) and drought and cold stresses in *Arabidopsis* (Albrecht *et al.*, 2003; Wang *et al.*, 2007). Several other studies have correlated the dehydrin gene with local adaptations to extreme environmental conditions, such as extreme temperatures and drought conditions (Lopez *et al.*, 2003; Suprunova *et al.*, 2004; Giarola *et al.*, 2015; Blein-Nicolas *et al.*, 2020). Here, *CBL01* and the dehydrin gene were found to correlate with Precip1, which can be interpreted as a precipitation abundancy axis opposing populations with high precipitation during the dry season coupled with low precipitation seasonality (RSRF), to populations with low precipitation regimes during the dry season and high precipitation seasonality (RJRE; Supporting Information, Fig. S5).

LEA analysis pointed to the genes Dof zinc finger (*dof17*), phosphoglycerate kinase (plastid isoform 1) and dehydrin, which are most probably involved in local adaptation towards contrasting edaphoclimatic conditions (Supporting Information, Table S9). Then, possible involvement of these genes in the response against drought stress and competition for light cannot be excluded, especially in the riparian forest, where individuals are usually taller and compete for sunlight with other species, and in the restinga habitat, where most of the environmental heterogeneity rather refers to thermal amplitude and contrasting precipitation regimes. The outliers identified in the intronic regions of these genes might be involved in gene expression and alternative splicing, but further analyses are required to define their function (Blein-Nicolas *et al.*, 2020).

Few studies on South American forest tree species have reported an association between genetic markers and environmental variables. Here, we provide a first insight, on the genetic bases of local adaptation in *E. uniflora* towards highly diversified environmental conditions, and shed light on its evolutionary potential in the face of future environmental changes.

ACKNOWLEDGEMENTS

This work was financially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grant numbers: 306202/2016-6 and 308135/2020-2), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; grant numbers: 88881.064988/2014-01 and 2806/2014), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS; grant numbers: 16/0491-9 and 19/2551-0001677-8), Instituto Nacional de Ciência e Tecnologia (INCT) Plant Stress Biotech grant from Ministério da Ciência e Tecnologia (MCT) of the Brazilian Government and Postgraduate Program in Genetics and Molecular Biology (PPGMB). We acknowledge the contribution of D.B. Anton and F.A. Krause for assistance with protocol tests and L. Pieta and J. Frazzon for assistance with sequencing.

DATA AVAILABILITY

The data that support the findings of this study are openly available in GitHub at https://github.com/ellios84/Vet-Postolache-et-al.-BOTJLS.

REFERENCES

Albrecht V, **Weinl S**, **Blazevic D**, **D'Angelo C**, **Batistic O**, **Kolukisaoglu U**, **Bock R**, **Schulz B**, **Harter K**, **Kudla J. 2003.** The calcium sensor CBL1 integrates plant responses to abiotic stresses. *Plant Journal* **36:** 457–470.

Altschul SF, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ. 1990.** Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403–410.

Blein-Nicolas M, **Negro SS**, **Balliau T**, **Welcker C**, **Cabrera-Bosquet L**, **Nicolas SD**, **Zivy M. 2020.** A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. *Genome Research* **30:** 1593–1604.

Brancalion PHS, **Oliveira GCX**, **Zucchi MI**, **Novello M**, **van Melis J**, **Zocchi SS**, **Chazdon RL**, **Rodrigues RR. 2018.** Phenotypic plasticity and local adaptation favor range expansion of a Neotropical palm. *Ecology and Evolution* **8:** 7462–7475.

Brandão MM, **Vieira F de A**, **Nazareno AG**, **Carvalho D de. 2015.** Genetic diversity of Neotropical tree *Myrcia splendens* (Myrtaceae) in a fragment-corridor system in the Atlantic rainforest. *Flora* **216:** 35–41.

Brousseau L, **Fine PVA**, **Dreyer E**, **Vendramin GG**, **Scotti I. 2018.** Genomics of microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata* Aubl. (Fabaceae). *bioRxiv*, doi: https://doi.org/10.1101/312843, 2 May 2018, preprint: not peer reviewed.

Caye K & **François O. 2017.** Lfmm: Latent Factor Mixed Models. *R package version 0.0*.(Version 0.0).

Chang CC, **Chow CC**, **Tellier LC**, **Vattikuti S**, **Purcell SM**, **Lee JJ. 2015.** Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4:** 7.

Collevatti RG, **Novaes E**, **Silva-Junior OB**, **Vieira LD**, **Lima-Ribeiro MS**, **Grattapaglia D. 2019.** A genome-wide scan shows evidence for local adaptation in a widespread keystone Neotropical forest tree. *Heredity* **123:** 117–137.

Conesa A, **Götz S**, **García-Gómez JM**, **Terol J**, **Talón M**, **Robles M. 2005.** Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21:** 3674–3676.

Cooper B, **Clarke JD**, **Budworth P**, **Kreps J**, **Hutchison D**, **Park S**, **Guimil S**, **Dunn M**, **Luginbuhl P**, **Ellero C**, **Goff SA**, **Glazebrook J. 2003.** A network of rice genes

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

associated with stress response and seed development. *Proceedings of the National Academy of Sciences, USA* **100:** 4945–4950.

Danecek P, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G**, **Durbin R. 2011.** The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

De Almeida DJ, **Faria MV**, **Da Silva PR. 2012.** Biologia experimental em Pitangueira: uma revisão de cinco décadas de publicações científicas. *Revista Ambiência* **8:** 159–175.

De Wit P, **Pespeni MH**, **Palumbi SR. 2015.** SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. *Molecular Ecology* **24:** 2310–2323.

Doyle JJ, **Doyle JL. 1990.** Isolation of plant DNA from fresh tissue. *Focus* **12:** 39–40.

Excoffier L, **Hofer T**, **Foll M. 2009.** Detecting loci under selection in a hierarchically structured population. *Heredity* **103:** 285–298.

Excoffier L, **Laval G**, **Schneider S. 2005.** Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1:** 47–50.

Falush D, **Stephens M**, **Pritchard JK. 2003.** Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164:** 1567–1587.

Ferreira-Ramos R, **Accoroni KAG**, **Rossi A**, **Guidugli MC**, **Mestriner MA**, **Martinez CA**, **Alzate-Marin AL. 2014.** Genetic diversity assessment for *Eugenia uniflora* L., *E. pyriformis* Cambess., *E. brasiliensis* Lam. and *E. francavilleana* O. Berg Neotropical tree species (Myrtaceae) with heterologous SSR markers. *Genetic Resources and Crop Evolution* **61:** 267–272.

Foll M, **Gaggiotti O. 2008.** A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180:** 977–993.

Forester BR, **Lasky JR**, **Wagner HH**, **Urban DL. 2018.** Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology* **27:** 2215–2233.

Franceschinelli EV, **Vasconcelos GMP**, **Landau EC**, **Ono KY**, **Santos FAM. 2007.** The genetic diversity of *Myrciaria floribunda* (Myrtaceae) in Atlantic Forest fragments of different sizes. *Journal of Tropical Ecology* **23:** 361–367.

Frichot E, **Schoville SD**, **De Villemereuil P**, **Gaggiotti OE**, **François O. 2015.** Detecting adaptive evolution based on association with ecological gradients: orientation matters! *Heredity* **115:** 22–28.

Frichot E, **François O. 2015.** LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution* **6:** 925–929.

Gayral P, **Melo-Ferreira J**, **Glémin S**, **Bierne N**, **Carneiro M**, **Nabholz B**, **Lourenco JM**, **Alves PC**, **Ballenghien M**, **Faivre N**, **Belkhir K**, **Cahais V**, **Loire E**, **Bernard A**, **Galtier N. 2013.** Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics* **9:** e1003457.

Giarola V, **Challabathula D**, **Bartels D. 2015.** Quantification of expression of dehydrin isoforms in the desiccation tolerant plant *Craterostigma plantagineum* using specifically designed reference genes. *Plant Science* **236:** 103–115.

GrabherrMG, **HaasBJ**, **YassourM**, **LevinJZ**, **ThompsonDA**, **Amit I**, **Adiconis X**, **Fan L**, **Raychowdhury R**, **Zeng Q**, **Chen Z**, **Mauceli E**, **Hacohen N**, **Gnirke A**, **Rhind N**, **Di Palma F**, **Birren BW**, **Nusbaum C**, **Lindblad-Toh K**, **Friedman N**, **Regev A. 2011.** Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29:** 644.

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82:** 711–732.

Guzman F, **Kulcheski FR**, **Turchetto-Zolet AC**, **Margis R. 2014.** *De novo* assembly of *Eugenia uniflora* L. transcriptome and identification of genes from the terpenoid biosynthesis pathway. *Plant Science* **229:** 238–246.

Heberle H, **Meirelles GV**, **da Silva FR**, **Telles GP**, **Minghim R. 2015.** InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16:** 169.

Hijmans RJ, **Cameron SE**, **Parra JL**, **Jones PG**, **Jarvis A. 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Jeffreys H. 1961. *Theory of probability*, *3rd edn*. Oxford: Oxford University Press.

Joly C, **Aidar M**, **Klink C**, **McGrath DG**, **Moreira AG**, **Moutinho P**, **Nepstad DC**, **Oliveira AA**, **Pott A**, **Rodal MJN**, **Sampaio EVSB. 1999.** Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. *Ciencia e Cultura* **51:** 331–348.

Jombart T, **Collins C. 2015.** *A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0*. Available at: https://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf

Jombart T, **Devillard S**, **Balloux F. 2010.** Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11:** 1–15.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

Kawecki TJ, **Ebert D. 2004.** Conceptual issues in local adaptation. *Ecology Letters* **7:** 1225–1241.

Kim D, **Pertea G**, **Trapnell C**, **Pimentel H**, **Kelley R**, **Salzberg SL. 2013.** TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14:** 1–13.

Kopelman NM, **Mayzel J**, **Jakobsson M**, **Rosenberg NA**, **Mayrose I. 2015.** CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K HHS Public Access. *Molecular Ecology Resources* **15:** 1179–1191.

Koressaar T, **Remm M. 2007.** Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23:** 1289–1291.

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

Lalagüe H, **Csilléry K**, **Oddou-Muratorio S**, **Safrana J**, **de Quattro C**, **Fady B**, **González-Martínez SC**, **Vendramin GG. 2014.** Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. *Tree Genetics & Genomes* **10:** 15–26.

Lê S, **Josse J**, & **Husson F. 2008.** FactoMineR: an R package for multivariate analysis. *Journal of statistical software* **25:** 1–18.

Lee C, **Abdool A**, **Huang CH. 2009.** PCA-based population structure inference with generic clustering algorithms. *BMC Bioinformatics* **10:** S73.

Li H, **Durbin R. 2010.** Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26:** 589–595.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lopez CG, **Banowetz GM**, **Peterson CJ**, **Kronstad WE. 2003.** Dehydrin expression and drought tolerance in seven wheat cultivars. *Crop Science* **43:** 577–582.

Lotterhos KE, **Whitlock MC. 2014.** Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* **23:** 2178–2192.

Manichaikul A, **Mychaleckyj JC**, **Rich SS**, **Daly K**, **Sale M**, **Chen WM. 2010.** Robust relationship inference in genome-wide association studies. *Bioinformatics* **26:** 2867–2873.

Margis R, **Felix D**, **Caldas JF**, **Salgueiro F**, **De Araujo DSD**, **Breyne P**, **Van Montagu M**, **De Oliveira D**, **Margis-Pinheiro M. 2002.** Genetic differentiation among three neighboring Brazil-cherry (*Eugenia uniflora* L.) populations within the Brazilian Atlantic rain forest. *Biodiversity and Conservation* **11:** 149–163.

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.

Morellato LPC, **Haddad CFB. 2000.** Introduction: the Brazilian Atlantic Forest. *Biotropica* **32:** 786–792.

Oliveira-Filho A, **Fontes M. 2000.** Patterns of floristic differentiation among Atlantic forests in southeastern Brazil and the influence of climate. *Biotropica* **32:** 793–810.

Peakall R, **Smouse PE. 2012.** GenALEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28:** 2537–2539.

Pinheiro F, **de Barros F**, **Palma-Silva C**, **Fay MF**, **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:** 1923–1935.

Potvin C, **Tousignant D. 1996.** Evolutionary consequences of simulated global change: genetic adaptation or adaptive phenotypic plasticity. *Oecologia* **108:** 683–693.

Pritchard JK, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Prive F, **Aschard H**, **Ziyatdinov A**, **Blum MGB. 2018.** Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics* **34:** 2781–2787.

Rellstab C, **Gugerli F**, **Eckert AJ**, **Hancock AM**, **Holderegger R. 2015.** A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24:** 4348–4370.

Ribeiro JA. 2000. Flora e vegetação ribeirinha. *Douro-Estudos & Documentos* **5:** 39–45.

Ribeiro MC, **Metzger JP**, **Martensen AC**, **Ponzoni FJ**, **Hirota MM. 2009.** The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation* **142:** 1141–1153.

Rozas J, **Ferrer-Mata A**, **Sánchez-DelBarrio JC**, **Guirao-Rico S**, **Librado P**, **Ramos-Onsins SE**, **Sánchez-Gracia A. 2017.** DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34:** 3299–3302.

Salgueiro F, **Felix D**, **Caldas JF**, **Margis-Pinheiro M**, **Margis R. 2004.** Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain forest. *Diversity and Distributions* **10:** 201–210.

Santos AS, **Gaiotto FA. 2020.** Knowledge status and sampling strategies to maximize cost-benefit ratio of studies in landscape genomics of wild plants. *Scientific Reports* **10:** 1–9.

Savolainen O, **Lascoux M**, **Merilä J. 2013.** Ecological genomics of local adaptation. *Nature Reviews Genetics* **14:** 807–820.

Scarano FR, **Ceotto P. 2015.** Brazilian Atlantic Forest: impact, vulnerability, and adaptation to climate change. *Biodiversity and Conservation* **24:** 2319–2331.

Solovyev V, **Kosarev P**, **Seledsov I**, **Vorobyev D. 2006.** Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology* **7:** 1–12.

Souza-Neto JD, **dos Santos EK**, **Lucas E**, **Vetö NM**, **Barrientos-Diaz O**, **Staggemeier VG**, **Vasconcelos T**, **Turchetto-Zolet AC. 2022.** Advances and perspectives on the evolutionary history and diversification of Neotropical Myrtaceae. *Botanical Journal of the Linnean Society* **199:** 173–195. https://doi.org/10.1093/botlinnean/boab095

Suprunova T, **Krugman T**, **Fahima T**, **Chen G**, **Shams I**, **Korol A**, **Nevo E. 2004.** Differential expression of dehydrin genes in wild barley, *Hordeum spontaneum*, associated with resistance to water deficit. *Plant, Cell and Environment* **27:** 1297–1308.

Turchetto-Zolet AC, **Salgueiro F**, **Turchetto C**, **Cruz F**, **Veto NM**, **Barros MJF**, **Segatto ALA**, **Freitas LB**, **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.

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025

Untergasser A, **Cutcutache I**, **Koressaar T**, **Ye J**, **Faircloth BC**, **Remm M**, **Rozen SG. 2012.** Primer3 - new capabilities and interfaces. *Nucleic Acids Research* **40:** 1–12.

Wang Z, **Gerstein M**, **Snyder M. 2009.** RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10:** 57–63.

Wang M, **Gu D**, **Liu T**, **Wang Z**, **Guo X**, **Hou W**, **Bai Y**, **Chen X**, **Wang G. 2007.** Overexpression of a putative maize calcineurin B-like protein in *Arabidopsis* confers salt tolerance. *Plant Molecular Biology* **65:** 733–746.

Warnes G, **Gorjanc G**, **Leisch F**, **Man M. 2021.** genetics: population genetics. R package version 1.3.8.1.3. Available at: https://rdrr.io/cran/genetics/

Wigginton JE, **Cutler DJ**, **Abecasis GR. 2005.** A note on exact tests of Hardy–Weinberg equilibrium. *American Journal of Human Genetics* **76:** 887–893.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Primer features and conditions used to amplify the 30 sequenced genes, including annealing temperature and forward and reverse sequences.

Table S2. *Eugenia uniflora* population coordinates and climatic characteristics. We extracted the 19 bioclimatic variables from the WorldClim database v.1.4 (30 s resolution) over the period 1950–2000 and computed their average values. MAT: mean annual temperature (°C); TDR: mean diurnal range (°C); isoT: isothermality (MAT/TAR × 100); Tseas: temperature seasonality (CV); MaxTWarm: max. temperature of warmest period (°C); MaxTCold: min temperature of coldest period (°C); TAR: temperature annual range (MaxTWarm-MaxTCold; °C); MeanTWet: mean temperature of wettest quarter (°C); MeanTdry: mean temperature of driest quarter (°C); MeanTwarm: mean temperature of warmest quarter (°C); MeanTcold: mean temperature of coldest quarter (°C); MAP: annual precipitation (mm); Pwet: precipitation of wettest period (mm); Pdry: precipitation of driest period (mm); Pseas: precipitation seasonality (CV); PwetQ: precipitation of wettest quarter (mm); PdryQ: precipitation of driest quarter (mm); PwarmQ: precipitation of warmest quarter (mm); PcoldQ: precipitation of coldest quarter (mm).

Table S3. Reads per sample of the 12 successfully sequenced genes in the 85 individuals.

Table S4. Synthetic environmental variables derived from the PCA of 19 bioclimatic variables processed by LFMM analyses.

Table S5. Loci with significant deviations from the Hardy–Weinberg equilibrium ($P < 0.05$), including population, locus, gene where the SNPs are related, observed and expected heterozygosity (H_O and H_E), *P*-value and standard deviation (SD).

Table S6. Nucleotide diversity of each gene by population.

Table S7. Analysis of molecular variance (AMOVA) in individuals representing the four *E. uniflora* populations.

Table S8. F_{ST} values between pairs of the four populations.

Table S9. Outlier SNPs detected with LEA analyses, including *P*-value: *P*-value of LEA test statistic; pBonf: *P*-value corrected for multiple comparisons using Bonferroni approach; Ref: allele reference; Alt: allele alternative.

Table S10. Outlier SNPs detected with LFMM. CLIM: synthetic climatic variable from PCA on climatic variables; β: LFMM test statistic for the correlation between an environmental variable and allelic frequencies; *P*-value: *P*-value of the regression model; Ref: allele reference; Alt: allele alternative.

Figure S1. Pipeline showing all tools used for processing the output data of NGS (Next Generation Sequencing) in the Illumina platform up to the point of SNP identification and filtering.

Figure S2. Graphic of the 372 SNPs distributed among 12 candidate genes.

Figure S3. Population structure analysis based on sNMF. Ten independent runs were performed from K = 1 to K = 5 (panels from top left to top right). Boxplots are reported representing variability in cross entropy estimation across 20 repetitions per K. The most probable clustering solution oscillates between K = 3 and K = 4 depending on the run examined.

Figure S4. Venn diagram with the number of overlapped SNP outliers detected with LFMM at K = 3 and K = 4.

Figure S5. Allele frequency of SNP outliers shared between LEA with LFMM. Populations are represented by the following colours: RSRF: riparian forest population (blue), RJRE: restinga population (orange), RSPA: Pampa population (grey), RSCO: coastal plain population (yellow).

Downloaded from https://academic.oup.com/botlinnean/article/201/1/100/6553805 by guest on 15 November 2025