Genomic signatures of ecological divergence between savanna and forest populations of a Neotropical tree

André Carneiro Muniz, Renata Santiago de Oliveira Buzatti, José Pires de Lemos-Filho, Myriam Heuertz, Alison Gonçalves Nazareno, Maria Bernadete Lovato

Annals of Botany **132**: 523–540, 2023
https://doi.org/10.1093/aob/mcad120, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

Genomic signatures of ecological divergence between savanna and forest populations of a Neotropical tree

André Carneiro Muniz[1,*], Renata Santiago de Oliveira Buzatti[1], José Pires de Lemos-Filho[2], Myriam Heuertz[3], Alison Gonçalves Nazareno[1] and Maria Bernadete Lovato[1,*]

[1]Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, CP 486, Belo Horizonte, MG 31270-901, Brazil, [2]Departamento de Botânica, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil and [3]Biogeco, INRAE, Univ. Bordeaux, 69 route d'Arcachon, 33610 Cestas, France
[*]For correspondence. E-mail lovatomb@icb.ufmg.br or andrecarneiromuniz@yahoo.com.br

Received: 1 May 2023 Returned for revision: 3 August 2023 Editorial decision: 18 August 2023 Accepted: 25 August 2023

• **Background and Aims** In eastern Neotropical South America, the Cerrado, a large savanna vegetation, and the Atlantic Forest harbour high biodiversity levels, and their habitats are rather different from each other. The biomes have intrinsic evolutionary relationships, with high lineage exchange that can be attributed, in part, to a large contact zone between them. The genomic study of ecotypes, i.e. populations adapted to divergent habitats, can be a model to study the genomic signatures of ecological divergence. Here, we investigated two ecotypes of the tree *Plathymenia reticulata*, one from the Cerrado and the other from the Atlantic Forest, which have a hybrid zone in the ecotonal zone of Atlantic Forest–Cerrado.
• **Methods** The ecotypes were sampled in the two biomes and their ecotone. The evolutionary history of the divergence of the species was analysed with double-digest restriction site-associated DNA sequencing. The genetic structure and the genotypic composition of the hybrid zone were determined. Genotype-association analyses were performed, and the loci under putative selection and their functions were investigated.
• **Key Results** High divergence between the two ecotypes was found, and only early-generation hybrids were found in the hybrid zone, suggesting a partial reproductive barrier. Ancient introgression between the Cerrado and Atlantic Forest was not detected. The soil and climate were associated with genetic divergence in *Plathymenia* ecotypes and outlier loci were found to be associated with the stress response, with stomatal and root development and with reproduction.
• **Conclusions** The high genomic, ecological and morphophysiological divergence between ecotypes, coupled with partial reproductive isolation, indicate that the ecotypes represent two species and should be managed as different evolutionary lineages. We advise that the forest species should be re-evaluated and re-stated as vulnerable. Our results provide insights into the genomic mechanisms underlying the diversification of species across savanna and forest habitats and the evolutionary forces acting in the species diversification in the Neotropics.

Key words: Adaptive loci, Atlantic Forest, Cerrado, ecological divergence, ecotypes, forest–savanna boundaries, genotype–environment associations, hybrid zone, Neotropics, local adaptation, *Plathymenia foliolosa*, *Plathymenia reticulata*.

INTRODUCTION

In eastern South America, the Cerrado, a large savanna vegetation, and the Atlantic Forest harbour high floristic diversity, and their habitats are rather different. The genetic mechanisms underlying the diversification of species along environmental gradients and across habitat types remain poorly understood in Neotropical areas. The Cerrado and Atlantic Forest had intrinsic evolutionary relationships, with high lineage exchange between them (Simon *et al.*, 2009; Hughes *et al.*, 2013; Antonelli *et al.*, 2018). The evolution of several Cerrado species from forest ancestors and the colonization of forest by Cerrado lineages indicate that both biomes contributed to the biodiversity of each other (Simon *et al.*, 2009; Hughes *et al.*, 2013; Antonelli *et al.*, 2018). As a savanna vegetation, the Cerrado is characterized by high light availability, regular fires, a harsh dry season and acidic soils, whereas in the Atlantic Forest the dry season is shorter and less severe, and the soils are richer than those in the Cerrado, with the light availability being a limiting factor (Furley, 1992; Pinto *et al.*, 2005; Durigan and Ratter, 2006). These habitat characteristics lead to adaptive responses of trees. For example, Cerrado trees have thick bark that protects against fire (Hoffmann *et al.*, 2012; Dantas *et al.*, 2013), low leaf specific area, thick leaves with trichomes, and high leaf carbon content in response to high solar irradiance and longer and more severe drought periods (Scholz *et al.*, 2007; Hao *et al.*, 2008; Dantas *et al.*, 2013). Moreover, the leaves of Atlantic Forest tree species show differences in traits associated with light competition, showing higher leaf area and leaf specific area (Hoffmann *et al.*, 2005). Although the morphofunctional characteristics associated with the two habitats are well established, little is known about the genetic basis of these adaptations.

The high species exchange between the Cerrado and Atlantic Forest can be attributed, in part, to the large contact zone they share in southeastern Brazil (Simon *et al.*, 2009; Hughes *et al.*, 2013; Antonelli *et al.*, 2018). In these ecotonal areas, patches of savannas and forests occur side by side. Closely related species occur in savanna–forest boundaries, with poor capacity to colonize the alternative habitat owing to their functional traits, which specialize them to either savanna or forest patches (Maracahipes *et al.*, 2018). Hybridization between tree species or ecotypes, i.e. populations genetically adapted to ecologically divergent habitats, such as *Casearia silvestris* (Cavallari *et al.*, 2010), *Dimorphandra* species (Muniz *et al.*, 2020) and *Hymenaea* species (Resende-Moreira *et al.*, 2017), has been detected in the ecotone, where hybrid zones with ancient and recent hybridization were demonstrated. Hybridization is a well-known phenomenon in plants and is well studied among plant species of temperate zones of the Northern Hemisphere (Eaton *et al.*, 2015; Kremer and Hipp, 2019; Leroy *et al.*, 2019; Arteaga *et al.*, 2020; Benowicz *et al.*, 2020) and *Eucalyptus* species in Australia (Foster *et al.*, 2007; Field *et al.*, 2010; Rutherford *et al.*, 2018; von Takach Dukai *et al.*, 2019; Pfeilsticker *et al.*, 2022, 2023), for instance. However, the ecotonal areas of tropical South America are not well studied, and the extent of hybridization between ecologically divergent lineages and its driving processes are poorly studied across tree species.

The analysis of genomic data can be a powerful tool to evaluate the processes underlying the divergence between lineages and to evaluate their association with the environment, especially in closely related species, such as some of the species occurring in savanna–forest boundaries. By studying the patterns of genomic differentiation of lineages, we can investigate the loci associated with the divergence, in addition to the processes leading to reproductive isolation between them (Feng *et al.*, 2020: p. 202; Tonzo *et al.*, 2020; Wang *et al.*, 2022; Hu *et al.*, 2023). For example, a strong association of allelic differences with differences in environmental variables and the lack of effective immigration might suggest a putative reproductive barrier for nascent species, such as the inviability of the hybrid in the alternative habitat (Lowry, 2012). In hybrid zones, where divergent lineages can interbreed, such conditions may be found and can be used as a natural experiment to characterize the divergence of lineages (Payseur and Rieseberg, 2016). Knowledge of the genotypic composition of a hybrid zone and the process related to the ecotypic divergence can provide insights into the evolution of reproductive barriers and speciation.

Plathymenia reticulata (Benth.) is a widespread tree species from South America, occurring mainly in Brazil but also in Bolivia and Paraguay (Warwick and Lewis, 2003). Previously, the *Plathymenia* genus was recognized to contain two species, *P. reticulata* from the Cerrado and *Plathymenia foliolosa* from the Atlantic Forest. However, the continuous distribution of morphological differences based on herbarium data and the finding that a population of *P. foliolosa* in the ecotone had gene flow with *P. reticulata* (Lacerda *et al.*, 2002) led to the recognition of *P. reticulata* and *P. foliolosa* as a single species, named *P. reticulata* (Warwick and Lewis, 2003). Later studies showed differences between the *Plathymenia* populations of the Cerrado and the Atlantic Forest related to adaptive responses in morphological and physiological characters to divergent habitats, characterizing them as two ecotypes (de Lemos-Filho *et al.*, 2008). Putative adaptive differences between the Cerrado and Atlantic Forest based on field observation were found in vegetative phenology (Goulart *et al.*, 2005), fruit and seed and morphology (Goulart *et al.*, 2006) and the stem radial increment (Toledo *et al.*, 2012). Furthermore, a common garden study showed that differences in vegetative traits associated with savanna and forest ecotypes were maintained, indicating a genetic determination of these characteristics (Goulart *et al.*, 2011). These studies also indicated that the ecotonal individuals generally showed intermediate trait values for the characters evaluated (de Lemos-Filho *et al.*, 2008), supporting the hypothesis of hybridization of the ecotypes in ecotonal areas between the Cerrado and Atlantic Forest.

High genetic divergence has been observed between the ecotypes of *P. reticulata* based on molecular data (Lacerda *et al.*, 2002; Muniz *et al.*, 2022), but no clear separation between the savanna and forest lineages was found in a phylogeographical study using chloroplast DNA (Novaes *et al.*, 2010). A recent study using microsatellites (Muniz *et al.*, 2022) identified a hybrid zone in ecotonal areas, with a predominance of later-generation hybrids. In addition, Muniz *et al.* (2022) showed that climatic and edaphic factors drove the genetic divergence between the ecotypes. A genomic study in *P. reticulata* using a wider sampling of the species range would allow for a better characterization of the genetic diversity and structure in the species and could increase the power of the hybrid classification of the hybrid zone. Furthermore, the phylogenetic relationships among populations and putative ancient gene flow between ecotypes can also be evaluated better using a genomic dataset, which could provide insights into the evolutionary history of the species. Studies of genotype–environment association in *P. reticulata* populations can determine the environmental drivers of ecological divergence and identify adaptive loci and their biological functions.

In this study, we used double-digest restriction site-associated DNA sequencing (ddRADseq) to genotype single nucleotide polymorphisms (SNPs) to investigate the evolutionary history of the divergence in *P. reticulata*. We used a wide sampling of the species range, including populations of the Atlantic Forest, Cerrado, Caatinga (a seasonally dry forest) and the ecotonal range between the Atlantic Forest and Cerrado. Considering the high genetic divergence found in the previous studies and the adaptive morphophysiological differences between the ecotypes, we hypothesize a partial reproductive barrier between them owing to ecological divergence. Finding only early-generation hybrids would indicate the inviability of the hybrids in the alternative habitat, supporting the hypothesis of a reproductive barrier, whereas finding later-generation hybrids would disprove it. Owing to high morphophysiological divergence between ecotypes associated with the habitat characteristics, we hypothesized that divergent selection could lead to allelic differences associated with the environment of each ecotype. Our specific aims were to answer the following questions:

1. What is the degree of genetic divergence between the ecotypes?

2. Do forest and Cerrado populations show evidence of ancient gene flow?
3. What is the genotypic composition of the hybrid zone?
4. Is there evidence of reproductive isolation between the two ecotypes?
5. What are the environmental drivers of genetic divergence?
6. How many putative SNPs are under natural selection and to what function they are related?
7. What is the degree of genomic diversity of the Atlantic Forest, Cerrado, Caatinga and the ecotone Atlantic Forest–Cerrado?

We discuss the results in the context of the incipient speciation process of *P. reticulata*, considering the genomic signatures of ecological divergence and the putative presence of reproductive barriers. We also discuss the implications for conservation, considering the divergence of the ecotypes and the genetic diversity of populations. Finally, our study using *P. reticulata* as a biological model can reveal the genomic differences associated with local adaptation, providing insights into the mechanisms responsible for the lineage divergence between the Cerrado and the Atlantic Forest and the high diversity in eastern South America.

MATERIALS AND METHODS

Population sampling

We sampled 96 individuals of *P. reticulata*, comprising savanna and forest ecotypes, in 14 localities from four different habitats: Atlantic Forest, Cerrado, Caatinga and the ecotone between the Atlantic Forest and Cerrado biomes (Fig. 1). We classified the individuals in ecotypes based on phenotypic differences in size, trunk and bark of the trees. The savanna individuals did not exceed 10 m in height and had tortuous, suberous and twisted trunks, whereas forest individuals measured between 15 and 30 m in height and showed straighter and thinner trunks than savanna individuals (Fig. 2). Six to seven individuals were sampled in four localities of the Atlantic Forest (AJF, BPF, CAF and SJF), in five of the Cerrado (DES, PRS, PTS, SPS and VZS), one in the Caatinga biome (a seasonal dry forest, CRF) and four in the ecotonal range (NEE, COE, SUE and FEE) (Fig. 1). In SUE and FEE localities, both ecotypes were found and sampled, whereas in COE and NEE only the savanna or forest ecotype was found, respectively. To distinguish the populations of each ecotype in the ecotone range, we used the letters ‘s’ and ‘f’

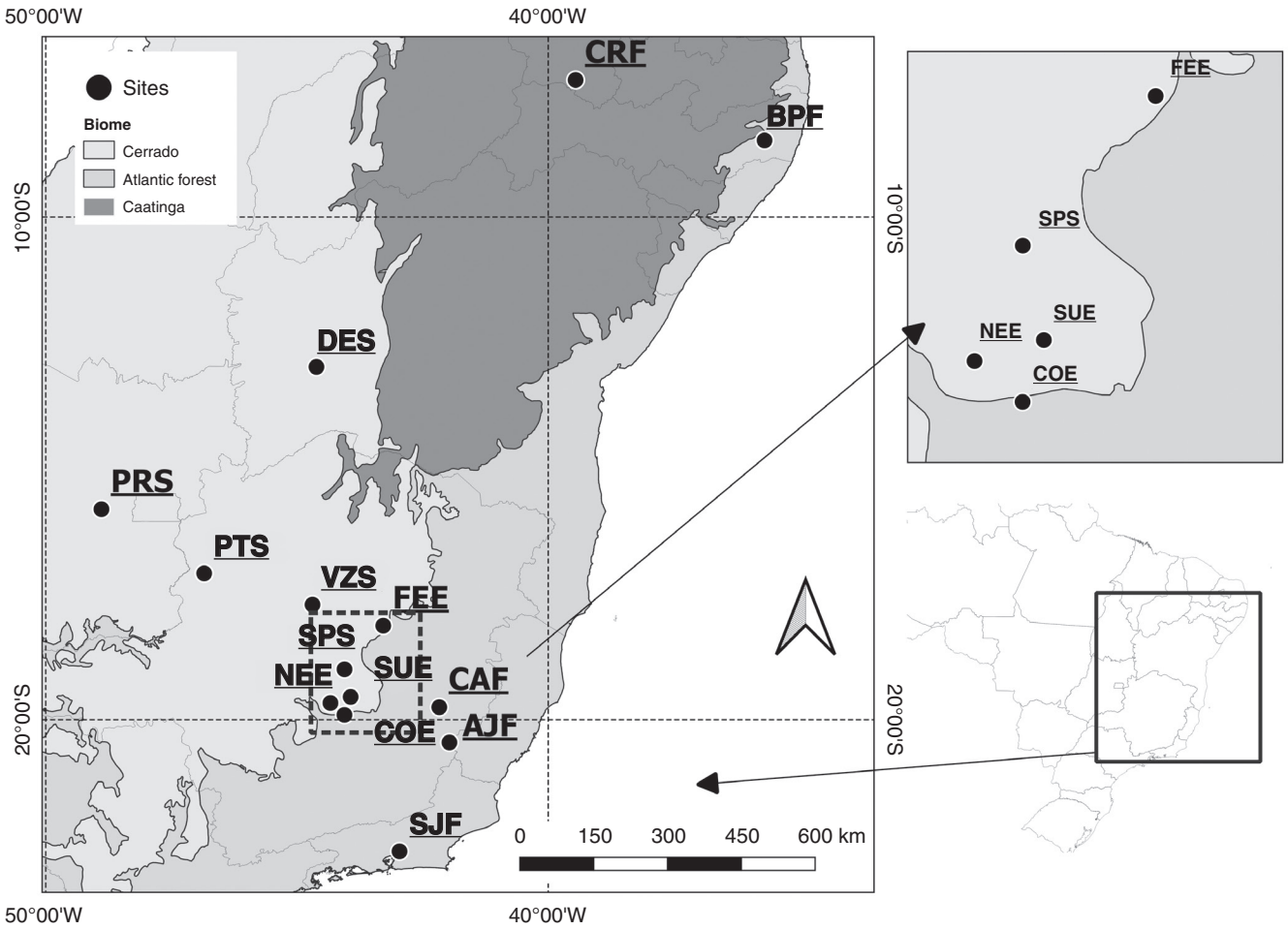

FIG. 1. Map showing the sampling localities of *Plathymenia reticulata* in the Brazilian Atlantic Forest, Cerrado, Caatinga and the ecotone between the Cerrado and Atlantic Forest.

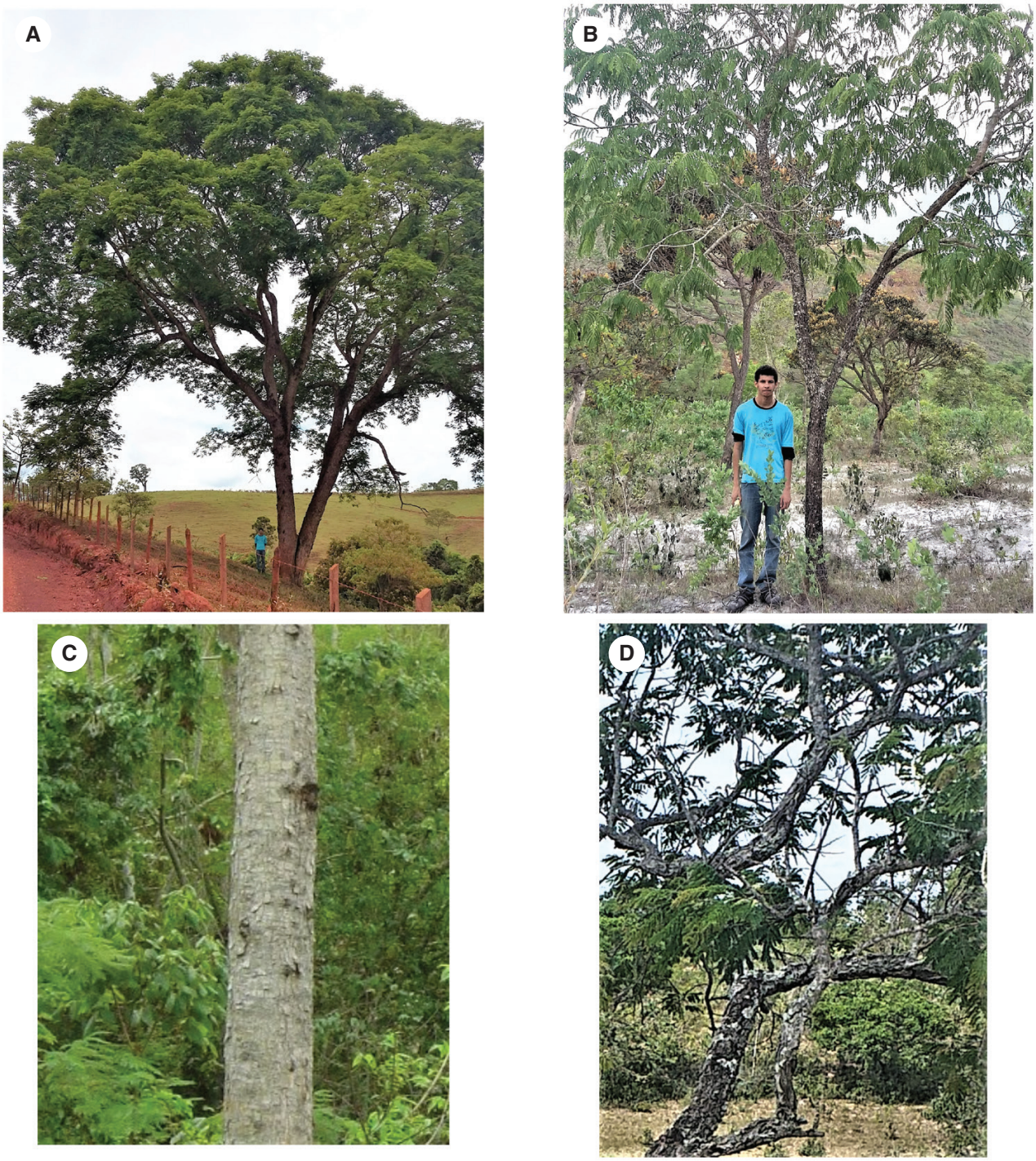

FIG. 2. Photographs of *Plathymenia reticulata* trees, showing a tree of forest ecotype (A), a tree of savanna ecotype (B), details of the forest ecotype trunk (C) and details of the savanna ecotype suberous bark (D).

before the locality code to refer to savanna and forest individuals, respectively. Therefore, the locality SUE was split into f-SUE and s-SUE samples, and FEE was split into f-FEE and s-FEE samples. Hereafter, the sampled individuals of each ecotype in each polymorphic locality are treated as separate populations.

DNA isolation and ddRADseq

Cambium tissue or leaves dried in silica gel were used for genomic DNA isolation. The sorbitol buffer-based method (Souza *et al.*, 2012) and a modified CTAB method (Novaes *et al.*, 2009) were used for DNA extraction of leaves and cambium tissue, respectively. The DNA was quantified using the Qubit

dsDNA Assay Kit (Invitrogen), and 0.30 µg of each sample was used to build the genomic libraries for 96 individuals using ddRADseq adapted from the protocol of Peterson *et al.* (2012). The genomic DNA was digested with the restriction enzymes EcoRI and MseI. Each sample was ligated with adaptors with unique barcodes. The libraries were pooled, then size selected to 350–450 bp using the Pippin Prep (Sage Science, Beverly, MA, USA). More details on ddRADseq preparation are provided in the Supplementary Data (Text S1). The library was sequenced (100 bp single-end reads) on a lane of an Illumina HiSeq 2000 flow cell (Illumina Inc., San Diego, CA, USA) at the Centre for Applied Genomics in Toronto (ON, Canada).

Identification of SNPs and genotyping

We used the SAGARANA high perfomance computing cluster inCento de processamento de alto desempenho/Instituto de Ciencias Biológicas- Universidade Federal de Minas Gerais (CEPAD-ICB-UFMG) to perform all genomic assembly and variant call analyses. The raw sequence reads were processed in Stacks 2.45 using a *de novo* assembly pipeline (Catchen *et al.*, 2013; Rochette *et al.*, 2019). The *process_radtags* program was used to demultiplex and filter poor-quality and missing reads in the expected EcoRI cut site (options -barcode_dist-1 2 -c -q -r -t85 -E phred33 -e ecoRI). The software FastQC (Andrews, 2010) was used to evaluate the quality of reads per individual after demultiplexing. A total of ~114 million reads for the 96 individuals were retained after running the *process_radtags* program, with an average of 1 192 901.2 ± 1 023 773.6 reads per individual. After the demultiplexing and quality control, 81 individuals were maintained with an average of 82 % of reads per individual and a mean coverage per locus of 71.3 ± 14.5 after processing and assembly. Fifteen individuals were excluded from further analysis owing to the low number of read retention and read quality and the high number of missing data (< 60 %) after variant SNP calling. We used the module *ustacks* to find exactly matching stacks (putative alleles) using a maximum likelihood framework. A catalogue of consensus loci was built using *cstacks* module, and genotypes were identified using *sstacks*. The minimum depth of coverage (*m*) was set to three, the maximum distance allowed between stacks (*M*) to two and the maximum number of stacks per *de novo* locus (*n*) to three. The SNPs were called using the software *tsv2bam* and *gstacks* using the default settings (Rochette *et al.*, 2019). The processed genomic data resulted in 102 250 genotyped loci for *P. reticulata*. The module *populations* was used to retain only one SNP per locus genotyped in ≥80 % of individuals across populations. The SNPs were filtered using two minor allele frequency (MAF) thresholds, 0.05 and 0.01, in addition to no filtering, and 10 % thresholds of missing data for each locus. Linkage disequilibrium within populations was tested using the *poppr* package through the estimation of the r_d index (Agapow and Burt, 2001). To avoid false positives attributable to multiple comparisons, the Bonferroni correction was applied.

Genetic diversity, population structure and ecotype divergence

The mean observed heterozygosity (H_O) and unbiased mean expected heterozygosity (uH_E) were estimated using *dartR* v.1.1.1 (Gruber *et al.*, 2017). The allelic richness (A_R) was calculated using rarefaction, with a standardized sample size of three diploid individuals, using the *Popgenreport* R package (Adamack and Gruber, 2014). We combined the total population of Atlantic Forest, Cerrado, forest ecotype in the ecotone and savanna ecotype in the ecotone to test differences in the A_R, H_O and H_E among these groups using Student's *t*-test implemented in R software with the Bonferroni correction for multiple comparisons using the package *rstatix* v.0.7.2 (Kassambara, 2023).

We calculated the genetic divergence among populations (F_{ST}) and the genetic divergence between ecotypes (F_{CT}) to investigate the hierarchical distribution of genetic variation with analysis of molecular variance (AMOVA) in the *poppr* package (Kamvar *et al.*, 2014). We set the ecotypic level of *P. reticulata* as the highest hierarchical level, followed by among-population and within-population levels. The pairwise fixation indexes (F_{ST}) were estimated in the package *Stampp* v.1.6.3 (Pembleton *et al.*, 2013).

The genetic structure and the individual admixture coefficients were evaluated with the maximum-likelihood algorithm implemented in Admixture v.1.3 software (Alexander *et al.*, 2009). The best number of panmictic populations (*K*) was determined based on the lowest error rate across 100-fold cross-validation iterations, with *K* ranging from 1 to 16. The package *ggplot2* v.3.3.1 (Wickam, 2016) was used to produce bar plots of estimated ancestry proportions given by the Q-value estimates of Admixture.

The proportion of fixed differences in different *P. reticulata* groups was calculated using *dartR* v.1.1.1 (Gruber *et al.*, 2017). The individuals were pooled into five sub-datasets: (1) all individuals of the Atlantic Forest; (2) all individuals of the Cerrado; (3) CRF locality; (4) individuals of the forest ecotype in the ecotone; and (5) individuals of the savanna ecotype in the ecotone. Two analyses were performed with ecotonal individuals, one with hybrid individuals and the other excluding the hybrids.

A principal component analysis (PCA) was performed using the function *gl.pcoa* implemented in *dartR* v.1.9.9.1 (Gruber *et al.*, 2017). Phylogenetic trees were built using SVDquartets (Chifman and Kubatko, 2014) in PAUP* v.4.0a152 (Wilgenbusch and Swofford, 2003) for coalescent-based analysis to investigate the differentiation between *P. reticulata* populations. As the root group, we used two individuals from PRS populations because there is a hypothesis of emergence of the species in the central Cerrado (Novaes *et al.*, 2010). To evaluate the support of the tree nodes, a non-parametric bootstrapping using 100 replicates was used. For both PCA and SVDquartets analyses, we used the following three datasets: (1) all individuals of *P. reticulata*; (2) only the pure individuals identified by NewHybrids and hybrid index analysis (see Results); and (3) the Atlantic Forest, Cerrado and Caatinga, excluding individuals from ecotonal areas.

Genotypic composition of ecotonal zone

The individuals were assigned to genotypic classes using the NewHybrids software v.1.1 (Anderson and Thompson, 2002). First, the individuals were classified into two purebred parentals, F1 and F2 hybrids, and two backcrosses of F1 with each parent, totalling six genotypic classes. In further

analysis, the NewHybrids was run using 12 genotypic classes to test for later-generation backcrosses in the dataset. These later-generation backcrosses comprise one generation of backcrosses between the F2 and the two parental classes, and second and third generations of backcrosses of F1 with the two parental classes (Supplementary Data Table S1). We ran NewHybrids using Jeffreys-like priors for 500 000 iterations of Markov chain Monte Carlo (MCMC) with a burn-in period of 100 000 chains and with no a priori information about the genotypic classes of individuals. Given the limitation of the number of SNPs in NewHybrids, the R package *dartR* v.1.1.1 (Gruber *et al.*, 2017) was used to select 200 SNPs for the analysis based on their polymorphic information content. To test the power of this analysis in the discrimination of hybrid classes, we simulated 100 individuals of each of the 12 hybrid classes using the pure individuals (posterior probability > 95 %) of Atlantic Forest and Cerrado as parental individuals in the simulation. The function *hybridize* from the *adegenet* v.2.1.10 package (Jombart, 2008) was used to produce the genotypic data of the individuals. The hybrid index was estimated using the *gghybrid* package v.0.0.0.9 in R (Bailey, 2022), with the total Atlantic Forest and Cerrado populations being used as parental individuals of the ecotonal populations. The posterior probability for the individual hybrid index was estimated after 5000 MCMC iterations of burn-in and 10 000 MCMC iterations.

Historical introgression in Plathymenia reticulata

To investigate the topological relationships and ancient introgression among the *P. reticulata* populations, the TreeMix v.1.13 method was used (Pickrell and Pritchard, 2012). The method estimates the maximum likelihood tree for a set of populations based on the covariance of allele frequencies, which are produced by shared drift or gene flow. The ecotonal populations were excluded because they showed recent hybridization (see Results). The best tree and the best number of migration edges (i.e. introgression) were estimated using 100 iterations for each tree for 0–16 migration edges. The threshold of 99.8 % for the proportion of variance explained was used to investigate the number of migration edges that best explained the trees using optM v.0.1.6 (Fitak, 2021). *Dsuit*e was used to estimate the ABBA-ABBA statistics (Peterson's *D*) and the related statistic, the f_4-ratio, to assess evidence of historical gene flow between *Plathymenia* populations using the *Dtrios* function of the package (Malinsky *et al.*, 2021). We corrected the *P*-value of Peterson's *D* statistics based on the false discovery rate, as suggested by Malinsky *et al.* (2021). The same dataset as TreeMix was used to estimate these statistics.

Genomic signatures of adaptation and environmental influence on genetic divergence

To identify SNPs potentially under balancing and divergent selection, we used the software BayeScan v.2.1 (Foll and Gaggiotti, 2008). The individuals without recent admixture between ecotypes were excluded, in order to determine the putative alleles associated with the genetic divergence of the ecotypes. The software was run for five pilot runs of 10 000 iterations and burn-in of 100 000 iterations, with a final run of 1 000 000 iterations. We set the prior odds of the neutral model as 100 to minimize false positives (Foll and Gaggiotti, 2008). The SNPs with $\log_{10}$ Posterior Odds (PO) > 1.0 were selected as putative outliers because this is considered strong evidence for selection according to Foll and Gaggiotti (2008).

To calculate the correlation between genetic distance among populations with the environmental and spatial distances, we used the partial Mantel test implemented in the *ncf* package in R software (Bjornstad, 2022). The genetic distance was obtained using the pairwise linearized F_{ST} among populations with the same individuals analysed in BayeScan using the *genet.dist* function implemented in *hierfstat* (Goudet, 2005). The pairwise spatial distance was calculated in the *geodist* package v.0.0.7 in R using the population geographical coordinates (Padgham, 2021). Nineteen bioclimatic variables downloaded from the WorldClim database in raster format with a spatial resolution of 2.5 arc-seconds were used as the climatic dataset (www.worldclim.org; Hijmans *et al.*, 2005). In addition, the nitrogen content, the bulk density of the fine earth fraction, soil pH, soil organic carbon and the cation exchange capacity of the soil were used as soil variables, downloaded from https://files.isric.org/soilgrids/latest/data/ with a spatial resolution of 2.5 arc-seconds and based on samples from 15–30 cm below the soil surface. A PCA for climatic variables and another for soil variables were performed to create synthetic components for the Mantel test and to perform the genotype–environment association analyses. Five synthetic climatic variables (95 % of explanatory power) and three synthetic edaphic variables (90 % of explanatory power) were retained, and the score of each population was extracted using the function *extract* from the *terra* package v.1.7-39 (Hijmans *et al.*, 2022). To perform the partial Mantel test, the *Mahalanobis* distance was calculated for the following synthetic variables: (1) climatic variables; (2) soil variables; and (3) soil and climatic variables together.

The genotype–environment association in *P. reticulata* was evaluated using the distance-based redundancy analysis (RDA) (Legendre and Andersson, 1999). The genetic dataset used to perform RDA was the same as in BayeScan. The SNP matrix was imputed with the neighbour method using the function *gl.impute* from the package *dartR* v.1.1.1 (Gruber *et al.*, 2017), because the RDA does not accept missing data. The geographical coordinates of individuals were applied in the principal coordinates of the neighbourhood matrix (PCNM) using the *pcnm* function of the *vegan* package to obtain axes of relevant spatial structure among *P. reticulata* individuals (Oksanen *et al.*, 2022). We performed a stepwise variable selection based on the Akaike information criterion using the *ordiR2step* function in vegan (Oksanen *et al.*, 2022). The variance inflation factor was used to assess the collinearity among the predictors. The significance of the predictors and the canonical axis of the RDA were estimated using *anova.cca* in R software (R Core Team, 2022). To evaluate the proportion of variance explained by each variable (partial R^2), the function *varpart* in the *vegan* v.2.6-4 package was used (Oksanen *et al.*, 2022). The SNPs with locus scores >2.0 standard deviations were considered putative loci under natural selection based on Forester *et al.* (2018) and Rellstab *et al.* (2015).

Outlier SNP annotation and putative gene functions

We searched for similarities of the consensus sequences of the putative outliers with proteins in the database of the National Center for Biotechnology Information (NCBI) nucleotide database (nr) using BLASTx (Basic Local Alignment Search Tool). To increase the power of the analysis, we considered only the SNPs that were identified in both BayeScan and RDA. The best alignments were chosen based on the best hits values and the E-values, i. e. is the number of expected hits of similar quality (score) that could be found just by chance, of the protein alignments. The information about the protein functions was searched in the https://www.uniprot.org/ database (The UniProt Consortium, 2023).

RESULTS

Sequencing results

The resulting dataset for no filtering and filtering with a MAF of 0.01 and 0.05 yielded 590, 546 and 457 SNPs, respectively, with ~4 % of missing data. Given that the MAF of 0.05 showed the best performance for the classification of individuals in hybridization and admixture analysis, we used the dataset with MAF = 0.05 for further analyses. Only the analyses to detect SNP outliers and genotype–environment association were performed with a dataset filtered by a MAF = 0.01, which removed loci present in only one individual, and ~20 % of missing data in order to have a higher number of SNPs available (1500 SNPs). The linkage disequilibrium (r_d) was significant ($P < 0.003$ after Bonferroni correction) for s-COE, f-FEE, f-NEE, PRS, f-SUE and CAF showing the highest values in the ecotonal populations (Supplementary Data Table S2).

Genetic diversity, population structure and ecotype divergence

The allelic richness (A_R) values ranged from 1.13 to 1.59, with higher values in the ecotonal areas in both forest and savanna ecotypes (Table 1). The mean observed heterozygosity (H_O) values ranged from 0.077 in AJF to 0.474 in f-SUE, and the mean expected heterozygosity (H_E) ranged from 0.080 in AJF to 0.338 in f-SUE (Table 1). The total forest ecotype in the ecotone showed the highest genetic diversity indexes of all groups, Cerrado, Atlantic Forest, and savanna ecotype in the ecotone (Supplementary Data Fig. S1). In contrast, the total Atlantic Forest population showed less diversity than all other groups (Supplementary Data Fig. S1).

The hierarchical AMOVA showed that 65.8 % of the total variation was explained by the divergence between ecotypes (Table 2). The F_{ST} and F_{CT} were 0.800 and 0.657, respectively, indicating that most of the divergence in *P. reticulata* occurred between the two ecotypes. All the pairwise F_{ST} values among populations were significantly different from zero ($P < 0.05$), showing values ranging from 0.034 to 0.855. The pairwise F_{ST} was high between Atlantic Forest and Cerrado populations (F_{ST} = 0.784–0.834), while the populations of the same biome showed the lowest genetic divergence values (F_{ST} = 0.074–0.135) (Supplementary Data Table S3). In the ecotonal areas, the divergence between ecotypes was high in general (F_{ST} = 0.059–0.603), although lower in comparison to the Atlantic Forest vs. Cerrado (Supplementary Data Table S3). The exception to this pattern was f-SUE population, which showed lower divergence values with all ecotonal populations.

The Admixture software estimated the lowest cross-validation values for $K = 3$, followed by $K = 2$ and $K = 4$, with very small differences in the cross-validation values between them indicating that they are similarly probable (Supplementary Data Fig. S2). In all models of Admixture, the Atlantic Forest populations were separated from the Cerrado and CRF populations, with ecotonal populations showing admixture between the ecotypes (Fig. 3). The forest populations showed higher admixture with the savanna group than savanna ecotype populations with the forest group. The $K = 3$ and $K = 4$ models provided the separation of the savanna ecotype in two and three lineages, respectively. The PTS and PRS populations were the pure populations of one lineage; the CRF population, which occurs in Caatinga, was a pure lineage of another group; and the remaining savanna populations showed admixture between the other savanna groups (Fig. 3B, C).

We found 26 and 32 % of fixed differences in alleles in the comparison between the Atlantic Forest and Cerrado and the Atlantic Forest and CRF population, respectively (Supplementary Data Table S4). The number of fixed differences between 'pure' individuals (<99 % probability of being pure in NewHybrids and hybrid index <0.05 and >0.95) of savanna and forest ecotypes was slightly greater in the ecotone than in allopatry. The savanna individuals from the ecotone showed 34 and 28% of fixed differences the forest individuals in the ecotone and in the Atlantic Forest, respectively (Supplementary Data Table S4). Finally, the fixed differences between Cerrado and CRF with the

Table 1. *Genomic diversity parameters of* Plathymenia reticulata *populations. Abbreviations:* A_R, *allele richness;* H_O, *mean observed heterozygosity;* N, *mean number of samples;* uH_E, *unbiased mean expected heterozygosity. The last letter of the locality code indicates savanna (S), forest (F) and ecotone (E) populations.*

Population	*N*	A_R (s.d.)	H_O (s.d.)	uH_E (s.d.)
AJF	4.2	1.23 (0.31)	0.077 (0.175)	0.080 (0.166)
BPF	2.6	1.15 (0.31)	0.081 (0.205)	0.085 (0.191)
CAF	4.8	1.15 (0.30)	0.082 (0.189)	0.080 (0.172)
CRF	2.6	1.14 (0.30)	0.093 (0.222)	0.083 (0.181)
DES	4.8	1.29 (0.32)	0.129 (0.215)	0.143 (0.203)
f-FEE	5.3	1.43 (0.32)	0.264 (0.223)	0.233 (0.184)
f-NEE	4.6	1.44 (0.32)	0.275 (0.232)	0.243 (0.186)
f-SUE	3.8	1.59 (0.35)	0.474 (0.349)	0.338 (0.201)
s-COE	4.0	1.49 (0.30)	0.257 (0.209)	0.280 (0.191)
s-FEE	5.5	1.29 (0.33)	0.150 (0.219)	0.143 (0.193)
s-SUE	5.2	1.40 (0.30)	0.198 (0.202)	0.205 (0.183)
PRS	3.7	1.21 (0.32)	0.116 (0.200)	0.120 (0.188)
PTS	3.8	1.21 (0.32)	0.111 (0.205)	0.113 (0.191)
SPS	3.6	1.28 (0.34)	0.151 (0.251)	0.137 (0.204)
VZS	6.0	1.23 (0.32)	0.131 (0.209)	0.127 (0.183)

TABLE 2. *Analysis of molecular variance, showing the distribution of genetic variability between ecotypes, populations and within populations of* Plathymenia reticulata.

Variability	d.f.	Sum of squares	Mean square	Sigma	Percentage of variance
Between ecotypes	1	4674.0	4674.0	112.7	65.8
Between population within ecotypes	14	2180.6	155.8	24.3	14.2
Within populations	65	2230.2	34.3	34.3	20.0
Total	80	9084.8	113.6	171.3	100

F_{ST} = 0.800; F_{SC} = 0.414; F_{CT} = 0.658. Abbreviations: F_{CT}, divergence between ecotypes; F_{SC}, divergence among populations within ecotypes; F_{ST}, divergence among populations.

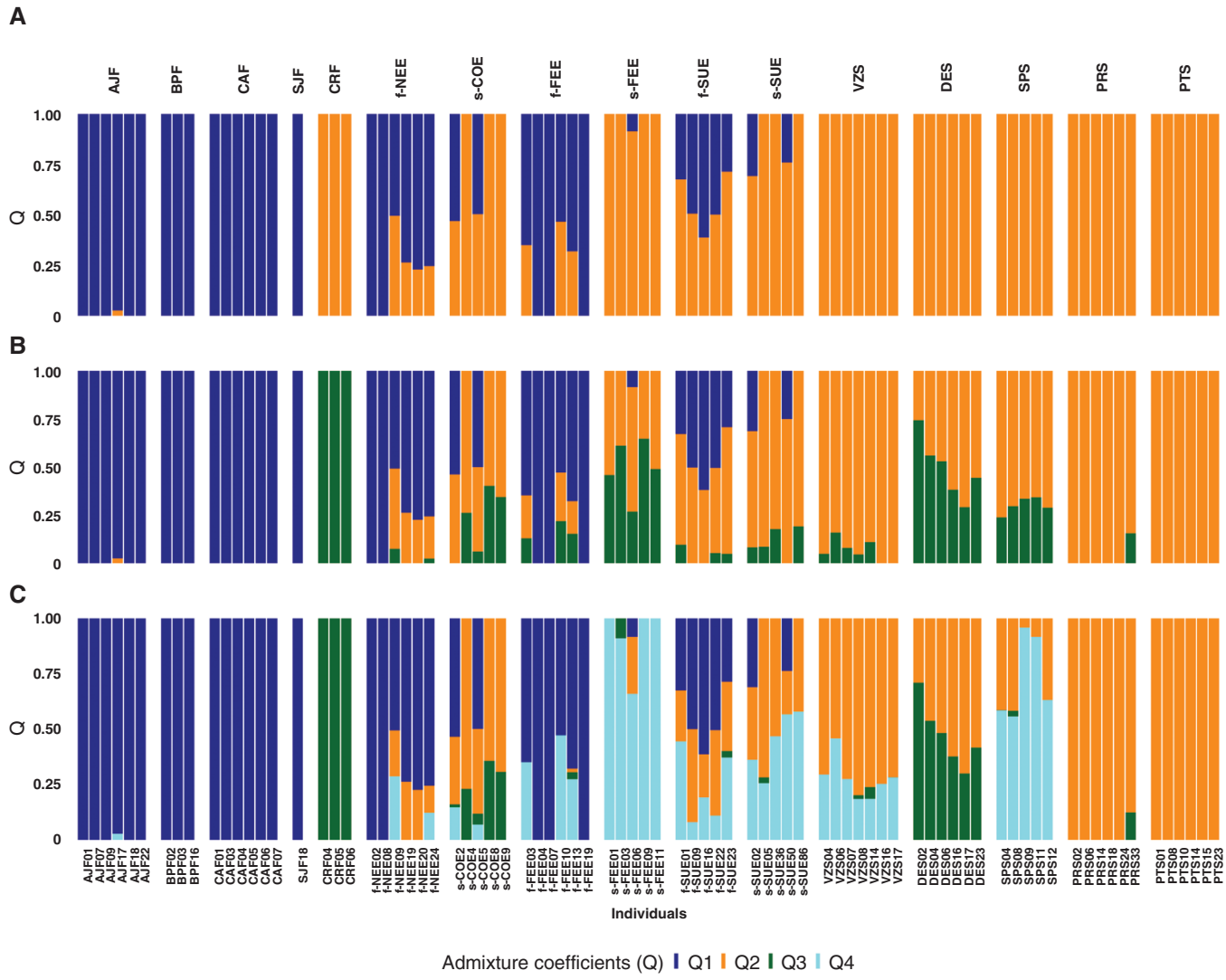

FIG. 3. Population structure of *Plathymenia reticulata*. Bar plot, showing the admixture coefficients of individuals based on Admixture with K = 2 (A), K = 3 (B) and K = 4 (C).

forest population in the ecotone were 31 and 38 %, respectively (Supplementary Data Table S4).

Plathymenia reticulata savanna and forest ecotypes were separated on the first PCA axis of the three PCAs realized, which represented 68–72 % of the genetic variability between the individuals (Fig. 4). The analysis removing the hybrid individuals or the ecotonal populations reinforced the high divergence levels between the ecotypes because no overlap between forest and savanna individuals was observed (Fig. 4B, C).

The SDVquartets phylogenetic tree also clearly showed the separation between forest and savanna ecotypes in all datasets tested (Supplementary Data Fig. S3). When using all

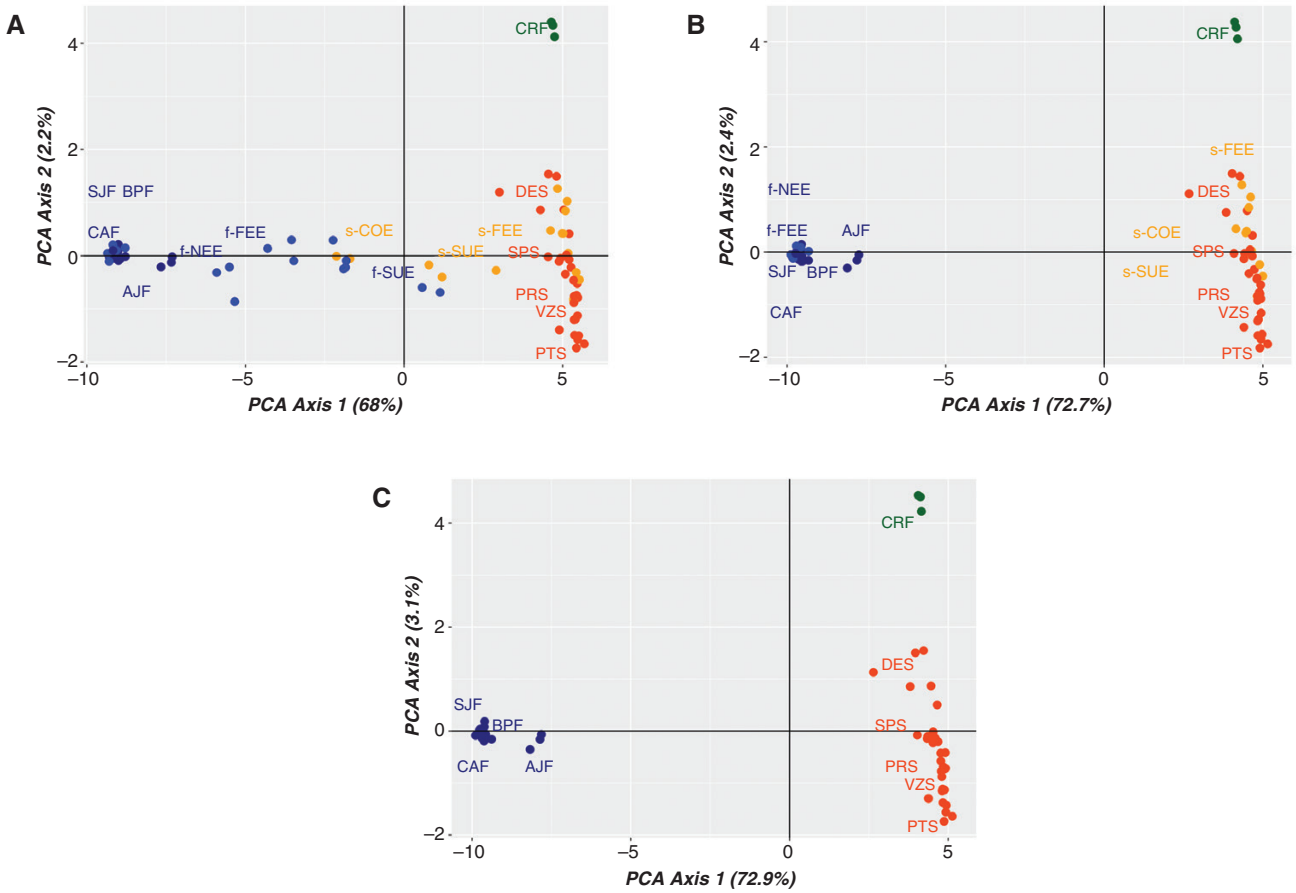

FIG. 4. Principal component analysis (PCA) of *Plathymenia reticulata* individuals, showing the distribution of genetic variability in the taxa. (A) PCA using all individuals. (B) PCA without hybrids. (C) PCA without ecotonal populations. The individuals that showed a posterior probability of >95 % for a given hybrid class in NewHybrids and a hybrid index between 0.05 and 0.95 were removed from the PCA in B.

individuals sampled, even the hybrids, the bootstrap values indicated high support for a forest ecotype group that included ecotone and the Atlantic Forest (Supplementary Data Fig. S3). The three phylogenetic trees presented high concordance, with the discordances only in the topology of DES and CRF (Supplementary Data Fig. S3).

Genotypic composition of the ecotonal populations

The classification of *P. reticulata* in the NewHybrids analysis supported a pure parental composition for the Atlantic Forest and Cerrado individuals, with high posterior probability (>99 %; Fig. 5). In the analysis with six genotypic classes, all ecotonal localities showed hybrid individuals between forest and savanna ecotypes: five F1s, one F2, four backcrosses with a savanna parent and six backcrosses with a forest parent (Fig. 5A). In the analysis with 12 genotypic classes including later generation hybrids, only one individual showed a different classification relative to the analysis with six genotypic classes. This individual was classified as a third-generation backcross {B8 = savanna × [savanna × (F1 × savanna)]} instead of pure parental as in the analysis with six genotypic classes. These results indicated that most hybrids represented early generations of hybridization, i.e. no more than two generations of hybrid classes (Fig. 5B). In accordance with Admixture analyses, the ecotonal populations of the forest ecotype showed the highest proportion of hybrid individuals, with f-SUE, f-NEE and f-FEE showing 100, 66 and 50 % of hybrid individuals, respectively, whereas the savanna populations s-SUE and s-COE showed 40 %, and s-FEE showed 20 % of hybrids. The simulation showed a high capacity to classify parental individuals and early-generation hybrids, although the classification procedure mistakenly classified the third-generation backcrosses as the second-generation backcrosses, indicating a lower capacity to classify these individuals based on the present data (Supplementary Data Fig. S4). The mean hybrid index was consistent with the NewHybrids, showing values of hybrids per population ranging from 20 to 100 %, with a higher proportion of hybrids being found in f-SUE, f-NEE and f-FEE (Fig. 5C). Therefore, both analyses supported that the ecotonal area between the Cerrado and Atlantic Forest is a hybrid zone between forest and savanna *P. reticulata* ecotypes.

Historical gene flow in P. reticulata

TreeMix (without the ecotonal populations) indicated high genetic divergence between the populations of *P. reticulata* from the Cerrado and Atlantic Forest (Fig. 6). The model

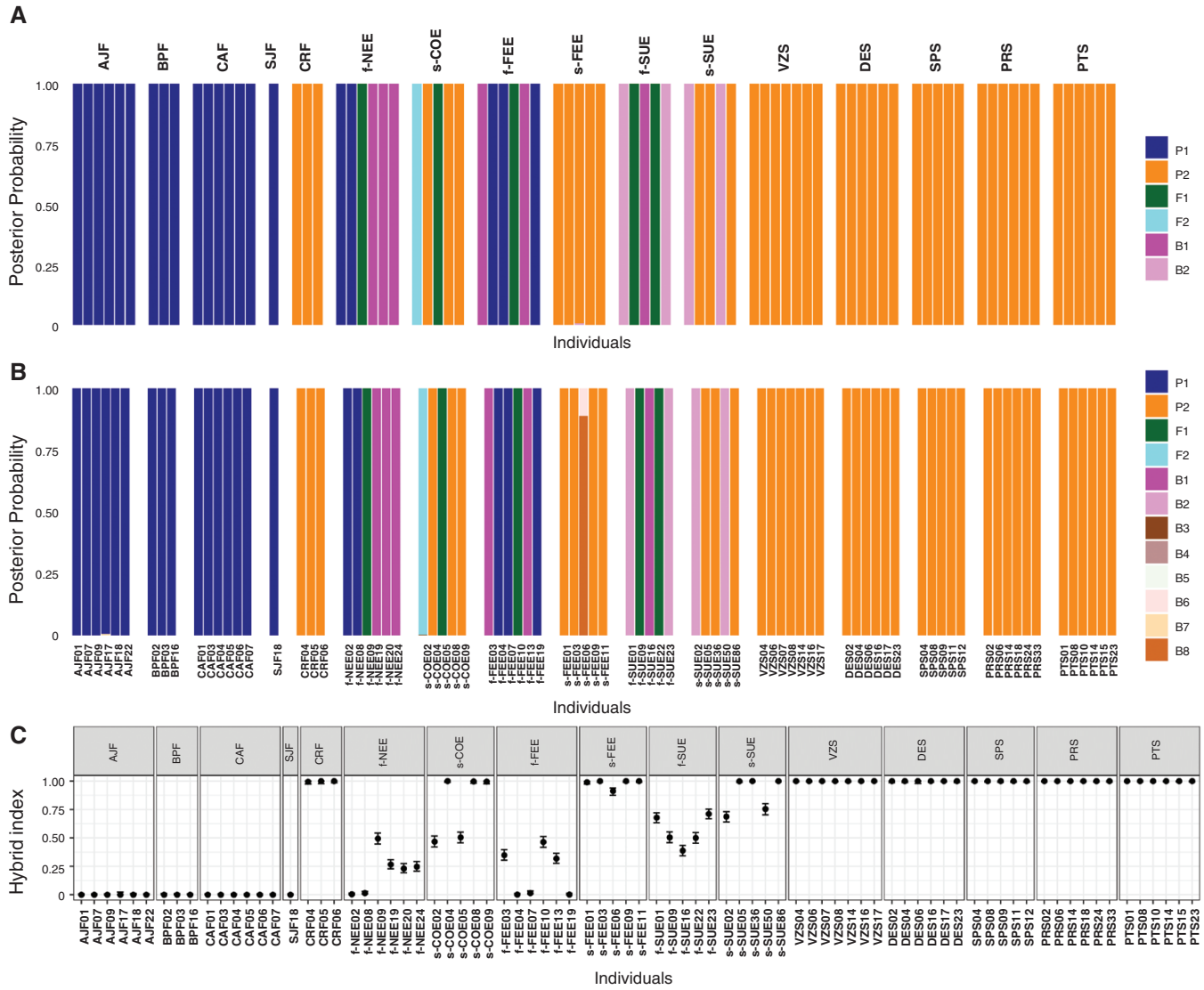

FIG. 5. Genotypic composition of *Plathymenia reticulata* in parental genotypic classes and hybrids based on NewHybrids method. (A) Assignment in six genotypic classes: P1 = forest ecotype, P2 = savanna ecotype, F1 = savanna × forest, F2 = F1 × F1, B1 = F1 × forest and B2 = F1 × savanna. (B) Assignment in 12 genotypic classes: the previous six classes plus B3 = F2 × forest, B4 = F2 × savanna, B5 = forest × (F1 × forest), B6 = savanna × (F1 × savanna), B7 = forest × [forest × (F1 × forest)] and B8 = savanna × [savanna × (F1 × savanna)]. (C) Estimated hybrid index for *Plathymenia reticulata* individuals.

without historical introgression events explained >99.99 % of the variance, indicating that there was no need to add an introgression event to explain the best tree topology. In the phylogenetic tree, the Atlantic Forest populations were very divergent from the Cerrado and CRF populations (Fig. 6). The *Dsuite* analysis estimated the *D*-statistics ranging from 0.005 to 0.392, and the f_4 statistics ranged from 0 to 0.601 (Supplementary Data Table S5). Based on *D*-statistics (ABBA-BABA statistics), no evidence of ancient gene flow between Cerrado and Atlantic Forest populations was found (Supplementary Data Table S5).

Genomic signatures of adaptation and environmental influence on genetic divergence

The BayeScan method identified 144 loci with the $\log_{10}$ of the probability >1.0 supporting their characterization as F_{ST} outlier loci (Supplementary Data Table S6). The method identified 136 SNPs with positive alphas, suggesting divergent selection, while only eight SNPs showed negative alpha values, suggesting purifying or balancing selection. These results show that most outliers might be involved with the divergence between the *P. reticulata* ecotypes.

The partial Mantel test showed a significant association ($P < 0.01$) between genetic distance and soil ($r = 0.430$), climatic ($r = 0.349$) and environmental (soil and climatic; $r = 0.343$) distances when controlling for spatial distance, indicating that the environmental gradient is associated with the genetic divergence in *P. reticulata* (Table 3). The association between genetic distance and spatial distance was not significant ($P > 0.05$) when controlling for environmental factors.

The RDA showed a significant association of the differences in allelic variation of *P. reticulata* individuals with the first soil component ($P = 0.0025$), the fourth climatic component ($P = 0.0433$) and the spatial structure (PCNM1, $P = 0.0001$;

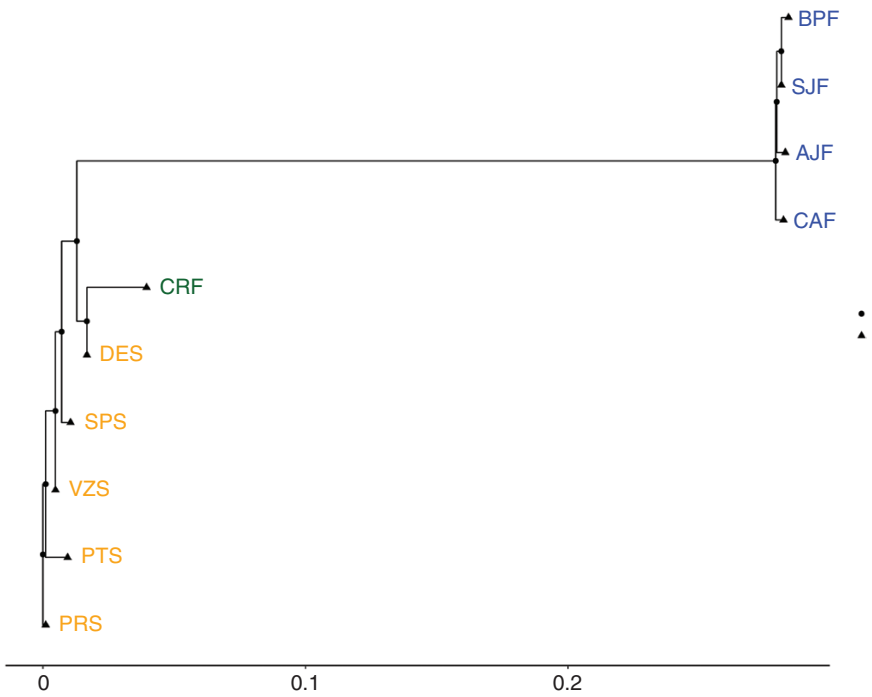

FIG. 6. Consensus tree topology for the divergence of *Plathymenia reticulata* populations based on the TreeMix method. Here, we excluded the individuals of ecotonal populations.

TABLE 3. *Partial Mantel test, showing the correlation of soil, climate and environmental distance (climate and soil) with the genetic distance, while accounting for the spatial distance of* Plathymenia reticulata *populations.*

Soil					
	Genetic × soil	Genetic × spatial	Soil × spatial	Genetic × soil (spatial)	Genetic × spatial (soil)
Correlation	0.505	0.320	0.433	0.430	0.130
P-value	0.000	0.017	0.017	0.001	0.178
Climate					
	Genetic × climate	Genetic × spatial	Climate × spatial	Genetic × climate (spatial)	Genetic × spatial (climate)
Correlation	0.451	0.320	0.538	0.349	0.102
P-value	0.000	0.018	0.000	0.003	0.226
Environment (climate and soil)					
	Genetic × environment	Genetic × spatial	Environment × spatial	Genetic × environment (spatial)	Genetic × spatial (environment)
Correlation	0.449	0.320	0.572	0.343	0.085
P-value	0.000	0.017	0.002	0.005	0.274

PCNM2, $P = 0.0375$). The first soil component, representing 53 % of the variation of the soil dataset, showed a correlation coefficient >0.300 for soil pH and bulk density of the fine earth fraction and <–0.300 for nitrogen content and soil organic carbon (Supplementary Data Table S7). The fourth climatic component, which represented 6 % of the variation in the climatic dataset, showed a correlation >0.300 for temperature seasonality (bio4), mean temperature of wettest quarter (bio8) and mean temperature of wettest quarter (bio7) and <–0.300 with isothermality (bio3) (Supplementary Data Table S7). In addition, four precipitation variables were mildly correlated with the fourth climatic components, with correlation values >0.200 and <0.300 or <–0.200 and >–0.300, which indicates a minor influence of these factors on the divergence of *P. reticulata* (Supplementary Data Table S7). The first soil component and the fourth climatic component explained 7 and 2.5 % of the variation in allele frequencies of *P. reticulata* based on partial RDA, respectively, whereas the spatial distance represented by the PCNMs explained 4.4 %. The first canonical axis of the RDA was the only statistically significant

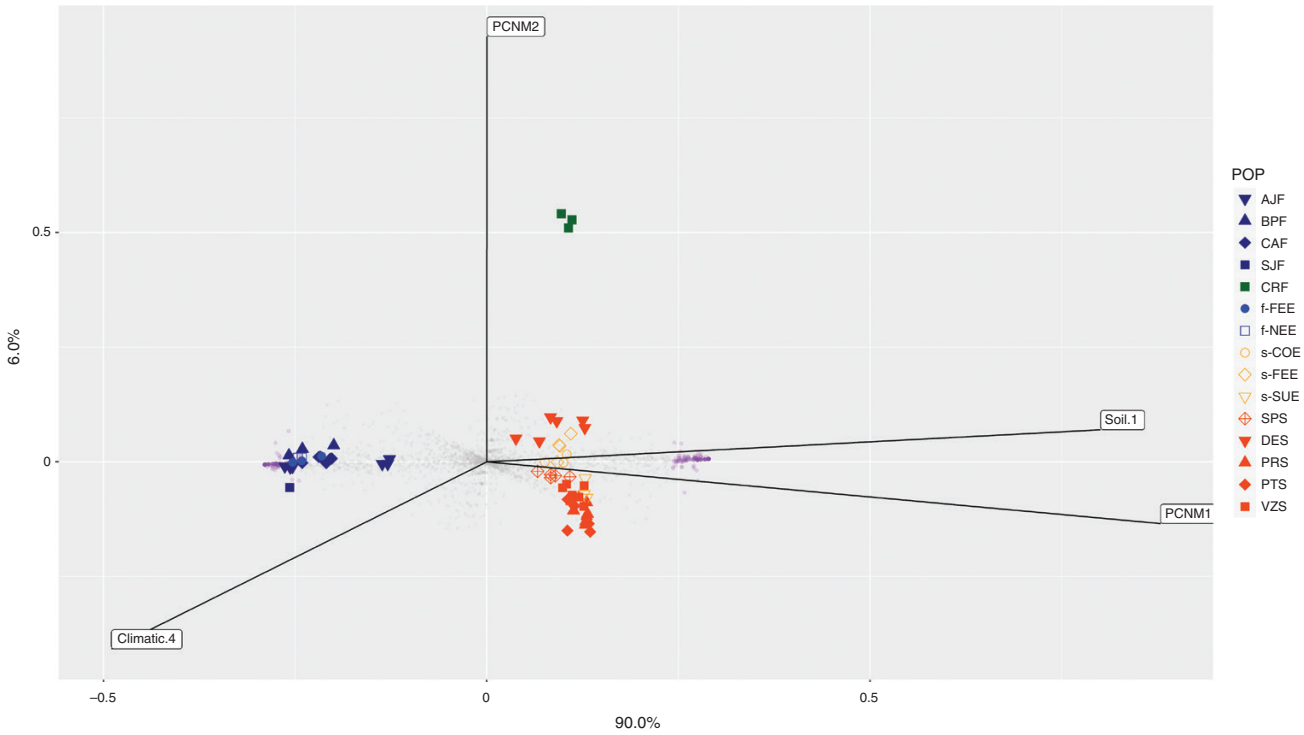

FIG. 7. Plots showing the association between genetic divergence of *Plathymenia reticulata* with the soil, climate and spatial distance based on the RDA. The purple dots represent the outlier SNPs found using the RDA scores, and the arrows represent the correlation coefficient of the predictors with each RDA axis.

axis ($P = 0.0001$), indicating that the environmental gradient coupled with spatial distance separated the forest and savanna individuals (Fig. 7). A high proportion of loci showed scores >0.300 and <–0.300, indicating a high genetic divergence between the two *Plathymenia* ecotypes (Supplementary Data Fig. S5). Based on the locus score of each SNP with the first canonical axis, 138 putative outliers were found (Supplementary Data Table S8), with most SNPs being nearly fixed for alternative alleles when comparing the forest and savanna populations.

Annotation and putative function of SNPs

The number of alleles identified as outliers by both BayeScan and RDA was 81, with all of them being under divergent selection. The annotation of these SNPs resulted in 43 protein alignments, with 28 loci showing putative functions based on the UniProt database (Supplementary Data Table S9). The biological functions of these loci were associated mainly with responses to biotic and abiotic stress, but also with constitutive functions and with plant reproduction and architecture (Fig. 8). Proteins associated with the response to abiotic stress were related to the abscisic acid metabolic path, in addition to auxin and jasmonic acid, which are involved in the response to cold, heat, water and salt stress (Supplementary Data Table S9). In addition, several of these proteins were also related to the immune response to fungi, bacteria and viruses (Fig. 8; Supplementary Data Table S9). The proteins associated with plant development and architecture were mainly related to root and stomatal development, especially the guard cell. Finally, the functions of proteins associated with reproduction were related to gametophyte development (Fig. 8; Supplementary Data Table S9).

DISCUSSION

In this work, we analysed *P. reticulata* populations from the forest and savanna ecotypes sampled in sympatric and allopatric localities with SNPs obtained with the ddRADseq method. A high genetic divergence between the Atlantic Forest and Cerrado was found, and a hybrid zone of early-generation hybrids in the ecotonal localities was detected. The lack of later-generation hybrids supports our hypothesis of partial reproductive isolation between the ecotypes. The soil and the climate were associated with genomic divergence in *P. reticulata* populations, indicating an effect of natural selection in the ecotypic divergence and local adaptations of the ecotypes. A total of 30 SNPs classified as outliers by BayeScan and RDA could be associated with biological processes such as the response to biotic and abiotic stress and plant development and architecture, in agreement with the hypothesis of ecological divergence of the ecotypes. We discuss the results in light of the speciation process of the *P. reticulata* lineages. Finally, we consider the implications of the levels of genetic diversity of populations, the hybrid zone structure and the speciation process of the lineages for the conservation of *P. reticulata* ecotypes.

Divergence, hybrid zone and partial reproductive isolation between ecotypes

The AMOVA showing 65.8 % of the total variation between ecotypes, the PCA, the phylogenetic trees and the high percentage of fixed differences in alleles indicated a high genetic divergence between forest and savanna ecotypes of *P. reticulata*, corroborating the results found in other studies

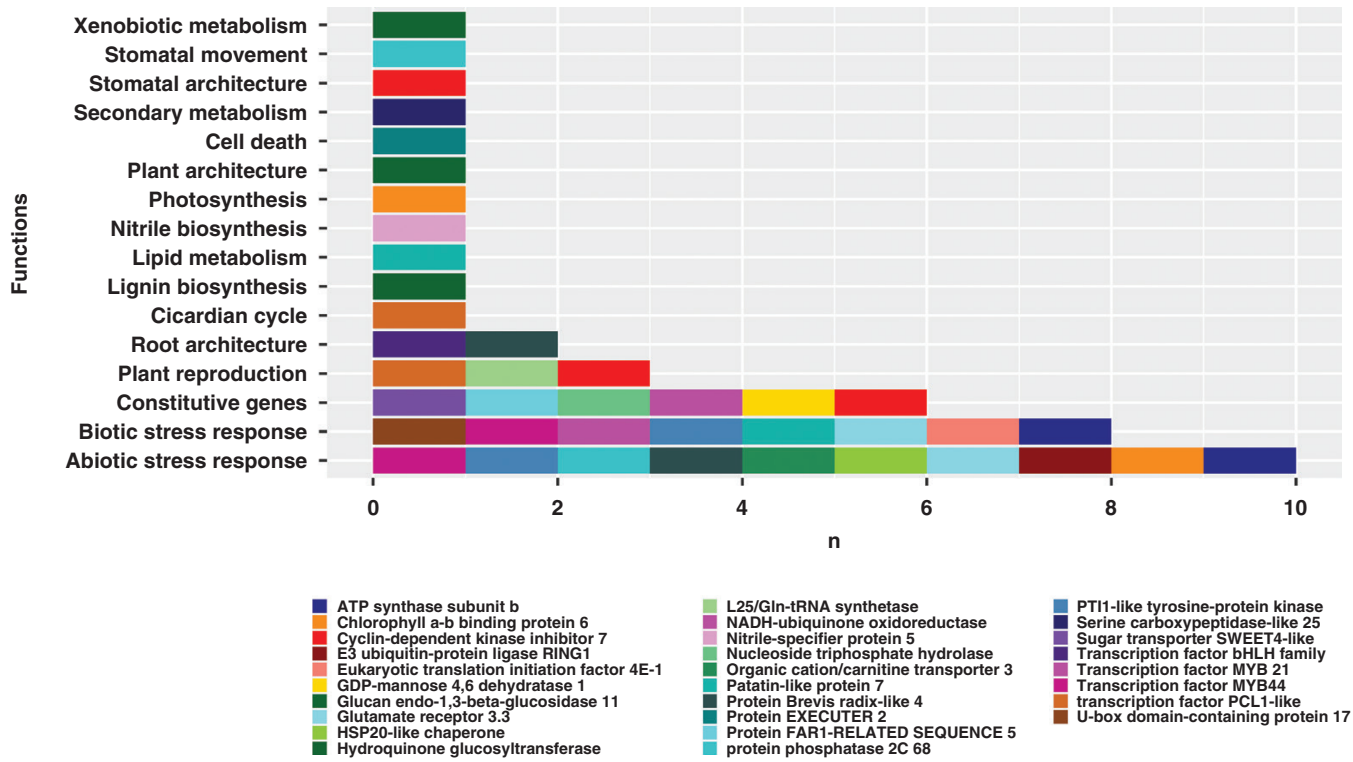

FIG. 8. Plot showing the proteins associated with putative functions based on the UniProt database for *Plathymenia reticulata* outlier loci.

(Lacerda *et al.*, 2002; Muniz *et al.*, 2022). Historical isolation between Cerrado and Atlantic Forest populations contributed to the pattern of high genetic divergence found here, because no evidence of past introgression was found by the TreeMix analysis and the ABBA-BABA statistics (Peterson's D and f_4-ratio). A posterior contact between populations of the Cerrado and Atlantic Forest could lead to the current hybrid zone formation. In fact, phylogeographical and ecological niche modelling studies have shown an eastern and southern retraction of the Cerrado during the Last Glacial Maximum (Souza *et al.*, 2017; Buzatti *et al.*, 2018). The Atlantic Forest was also retracted into two large refuges and a third smaller refuge (Carnaval and Moritz, 2008; Ribeiro *et al.*, 2011). After climate amelioration during the Holocene, the Cerrado and Atlantic Forest expanded and came into contact (Carnaval and Moritz, 2008; Ribeiro *et al.*, 2011; Souza *et al.*, 2017; Buzatti *et al.*, 2018). Thus, it is probable that *P. reticulata* savanna and forest lineages could be in contact at least after the Last Glacial Maximum and that the hybrid zone could be established at that time. An investigation with coalescent analysis to date the time since the divergence of ecotypes and the posterior contact could determine the hybrid zone formation better.

The genomic analysis and a larger population sampling than a previous study (Muniz *et al.*, 2022) allowed the Admixture analysis to reveal two to four genetic groups, with one representing the forest and three representing the savanna and CRF. The CRF population, which occurs in the Caatinga, was grouped with savanna individuals, although it was considered morphologically similar to forest individuals. This population was probably identified mistakenly as from the forest ecotype. The Caatinga is a seasonal dry forest, and the rainfall has an extreme seasonality, with the annual precipitation limited to ≤1000 mm in this biome (Silva and Souza, 2018). As revealed by the phylogenetic trees, the CRF population was more closely related to DES, SPS and VZS than to the geographically closest population of the forest ecotype, BPF, which is from the north of the Atlantic Forest. These results corroborate the results of a phylogeographical study showing that *P. reticulata* colonized the Caatinga from the central Cerrado (Novaes *et al.*, 2010).

The hybridization analyses showed that the ecotonal populations represent a hybrid zone between the *P. reticulata* ecotypes. The individuals from the hybrid zone are mostly first-generation backcrosses (F1 × parent) and F1s, with a higher proportion of hybrids in the population classified morphologically as the forest ecotype, indicating ongoing asymmetric hybridization between the ecotypes. The introgression into the forest lineage could lead to the invasion of a site previously occupied by the savanna ecotype. The invasion of forest vegetation in Cerrado areas has been documented (Hoffmann *et al.*, 2012; Rodrigues-Souza *et al.*, 2015). Using microsatellites, Muniz *et al.* (2022) found a higher proportion of F2s than this study. However, a few simple sequence repeats (SSR) that can show homoplasy (Gramlich *et al.*, 2018) could have a lower power to detect the true genotypic class of individuals than a high number of biallelic SNPs with a high number of fixed allelic differences. In the Cerrado and Atlantic Forest, outside ecotonal areas, no admixture between the savanna and forest ecotype was detected. The absence of later-generation hybrids suggests the presence of partial reproductive barriers between the *P. reticulata* ecotypes (Rieseberg and Blackman, 2010; Baack *et al.*, 2015).

In ecotonal zones, *P. reticulata* savanna and forest ecotypes can be found side by side, but they occupy divergent habitats with sharp habitat discontinuities owing to environmental conditions (Hoffmann *et al.*, 2003; Hao *et al.*, 2008; Barros *et al.*, 2018; Maracahipes *et al.*, 2018). Ecological factors can be strong reproductive barriers, especially in sympatric populations if the differences in habitat are sharp and the establishment in the alternative habitat is low (Feder and Nosil, 2010; Feder *et al.*, 2012; Lowry, 2012; Lowry *et al.*, 2015; Bridle *et al.*, 2019), and might facilitate the speciation process in lineages that show some degree of ecological divergence. The local adaptation owing to morphological and physiological traits of the *P. reticulata* ecotypes has been demonstrated in several field studies (Goulart *et al.*, 2005, 2006, 2011; Toledo *et al.*, 2012) and in laboratory (Lacerda *et al.*, 2004) and common garden (Goulart *et al.*, 2011) studies. The association of SNPs with genes related to differences in functional aspects, such as the stress response to abiotic factors and plant structural development and reproduction, also indicates that the ecotype differences are genetically determined. Altogether, these findings suggest that ecological barriers have an important role in the maintenance of the genetic integrity of the ecotypes, although it is not possible to discount some intrinsic reproductive barrier.

The *Plathymenia* genus was considered to contain two species until 2003, *P. foliolosa* from the forest and *P. reticulata* from the savanna. These species were synonymized posteriorly as *P. reticulata* (Warwick and Lewis, 2003). The high genomic divergence associated with ecological divergence and morphophysiological differences between ecotypes, coupled with partial reproductive isolation, indicate that the ecotypes represent different evolutionary lineages and that the species status of *P. foliolosa* and *P. reticulata* should be restated.

Genomic signatures of adaptation and environmental influence on genetic divergence

The association of the genetic divergence of *P. reticulata* with soil, climate and spatial distances suggests that both selection and genetic drift are driving forces of the divergence between savanna and forest ecotypes. The genomic divergence was associated with differences in the nitrogen content, soil pH and organic carbon of soil, characteristics that are very different between savanna and forest habitat (Hoffmann *et al.*, 2012). In addition, the SNPs were also associated with the bulk density of the fine earth fraction, which is positively correlated with the capacity of the soil to retain water and is very important in environments where there is water shortage throughout the year (Indoria *et al.*, 2020). Furthermore, in several areas of the Cerrado, the climate is similar to the climate of the forest, but the low nutrient and water availability and regular fires of Cerrado sites filter the forest species, and colonization of these areas does not occur (Lloyd *et al.*, 2008). The comparison of several species in savanna–forest boundaries indicates that the establishment of species in the alternative habitat might be difficult (Maracahipes *et al.*, 2018). In *P. reticulata*, the genomic differences reflect the morphophysiological adaptations to forest and savanna present in the two ecotypes and their response to limiting factors occurring in their habitats (de Lemos-Filho *et al.*, 2008).

The annotation of proteins showed that several outliers were associated with constitutive functions of the cell, including DNA duplication and transcription, the transition of the mitotic cell cycle, cell–cell signalling and cell differentiation. The response to biotic and abiotic stress and the root and stomatal architecture were other biological functions found for the annotated proteins. For example, ten proteins, including ATP synthase subunit b, HSP20-like chaperone, Transcription factor MYB44, E3 ubiquitin-protein ligase RING1, protein phosphatase 2C 68 and protein Brevis radix-like 4, were involved in the path of regulation of the stress response by activating the abscisic acid path but also involved in the salicylic acid-mediated defences and auxin activation (Luo *et al.*, 2006; Jung *et al.*, 2008; Ambrosone *et al.*, 2015; Singh *et al.*, 2015). The Cerrado is more prone than the Atlantic Forest to drought and heat stress throughout the year (Maracahipes *et al.*, 2018), hence the differences in these proteins could be related to the response to these stresses. The protein phosphatase 2C 68 is also involved with stomatal movement, which has the function to regulate water loss, indicating its function in response to stress.

The putative transcription factor family and the protein Brevis radix-like 4 are involved in root development (Yang *et al.*, 2015; Breda *et al.*, 2017). The savanna species have a higher root-to-shoot ratio than forest species, which allocate their biomass preferentially to the shoot structures (Hoffmann *et al.*, 2003, 2004; Silva *et al.*, 2019). The high investment in root structures in savannas maximizes nutrient and water uptake, whereas the production of high stems owing to the investment in shoot structures in the forest highlights the importance of light competition in this habitat (Shipley and Meziane, 2002). Thus, the divergence in the tree architecture associated with different forms of proteins involved in root development indicates that divergent selection is acting on the *P. reticulata* ecotypes.

In summary, the results of this study showed that the genes were related to the ecological divergence in a tree occurring in savanna habitat (the Cerrado) and forest (the Atlantic Forest). The evolutionary relationships with lineage interchange between the Cerrado and Atlantic Forest have been pointed out as one of the factors that explain the high biological diversity in eastern South America (Simon *et al.*, 2009; Hughes *et al.*, 2013; Antonelli *et al.*, 2018). However, the genes responsible for local adaptations to each biome are not known. Our study showed that proteins in the path of response to stresses, such as water and heat stress, and plant architecture are important factors shaping the differences between the lineages from these divergent habitats. A few studies comparing the genomic components of species adaptation across habitat types have been performed among the ecosystems of South America. Recent studies comparing species of *Handroanthus* and *Tabebuia* genera also found a similar pattern of response to stress for these trees, but contrary to our study they did not find larger differences across ecosystems when comparing the populations sampled (Collevatti *et al.*, 2019; Vieira *et al.*, 2022). Our results highlight the complexity of the relationship between the Cerrado and Atlantic Forest and help to provide insights into the genomic mechanisms underlying the diversification of species across savanna and forest habitats and the evolutionary forces acting in the lineage diversification in Neotropical areas.

Genomic analysis and the conservation of Plathymenia

The genomic analysis of populations provides important insights into the evolution and conservation of *P. reticulata.* Initially, the genomic analysis of hybrid zone and allopatric populations with no historical admixture suggested a partial reproductive barrier between the two ecotypes owing to ecological factors. The high genomic divergence coupled with the presence of local adaptations indicates that *P. reticulata* ecotypes constitute independent evolutionary lineages. Altogether, the currently termed ecotypes should be managed as different species or at least as different evolutionarily significant units (Crandall *et al.*, 2000).

The *P. reticulata* forest ecotype, previously known as *P. foliolosa*, was considered vulnerable to extinction by the IUCN (World Conservation Monitoring Centre, 1998). Owing to the low levels of genetic diversity of *P. reticulata* populations of the Atlantic Forest, we consider that this status should be maintained and updated from the data generated by our studies. The species should be assessed and listed in the Brazilian Red List, with its extinction risk determined. The lower genetic diversity of these populations might also increase the risk for the populations of the Atlantic Forest, because the isolation of patches is high owing to habitat loss and deforestation in the Atlantic Forest (Ribeiro *et al.*, 2009).

The ecotonal area showed the highest genetic diversity and alleles of all lineages, which indicates that selected individuals could be used as source material for the restoration process. The ongoing admixture occurring in hybridizing populations augments genetic diversity, and the *Plathymenia* individuals of forest ecotype of ecotonal areas could be considered to recover Atlantic Forest populations with low genetic diversity. However, this should be done with care to not disrupt the local adaptations of Atlantic Forest populations. Therefore, progeny tests should be done to test whether differences in local adaptation might be disrupted. The current climate changes can lead to the maladaptation of species that will have to migrate or adapt to new climatic conditions (Aitken *et al.*, 2008). Besides genetic diversity, phenotypic plasticity can also have a role in the response of trees in the face of environmental changes (Souza *et al.*, 2019, 2021; De Kort *et al.*, 2020). The populations of the Atlantic Forest and ecotonal areas of the *Plathymenia* showed high phenotypic plasticity in traits related to the response to light (Goulart *et al.*, 2011), a limiting factor in forest habitats. The pace of environmental changes relative to the generation time of species might create an adaptational lag (Aitken *et al.*, 2008), which is very important in long-lived tree species. Phenotypic plasticity in long-lived trees could alleviate the risk of extinction, countering the adaptational lag. Therefore, the phenotypic plasticity and the risk of mismatch between local adaptation and the future environment in the populations of *Plathymenia* of the Atlantic Forest should be considered when undertaking restoration and reintroduction programmes.

SUPPLEMENTARY DATA

Supplementary data are available at *Annals of Botany* online and consist of the following.

Table S1: tested genotypic categories in NewHybrids analyses showing the expected ancestry proportions of each cross-type. Table S2: tests for overall linkage disequilibrium in *Plathymenia reticulata* populations. Table S3: pairwise F_{ST} between *Plathymenia reticulata* populations. Table S4: percentage of fixed differences among *Plathymenia reticulata* groups. Table S5: ABBA-BABA statistics for *Plathymenia reticulata* populations (excluding the ecotonal populations) to test for ancient introgression. Table S6: outlier loci found by BayeScan method for *Plathymenia reticulata*. Table S7: loadings of the soil and climatic variables with the five first PCA axes of the environmental dataset. Table S8: correlation between allele frequencies of outlier SNPs based on RDA and the selected variables by the model selection for *Plathymenia reticulata*. Table S9: putative function of outlier loci of *Plathymenia* reticulata found by BayeScan and redundancy analysis. Fig. S1: boxplot showing the diversity indexes for total populations and results of Student's *t*-tests across regions with a MAF = 0.05 for allelic richness (A), observed heterozygosity (B) and expected heterozygosity (C) indexes. Fig. S2: plot showing the cross-validation values for sequential *K* of the Admixture software to choose the best *K* for *Plathymenia reticulata* populations. Fig. S3: SVDquartets phylogenetic tree to evaluate the genetic divergence process of *Plathymenia* populations. Fig. S4: classification of simulated hybrid individuals using 12 genotypic classes according with the following crosses: P1 = forest ecotype, P2 = savanna ecotype, F1 = savanna × forest, F2 = F1 × F1, B1 = F1 × forest, B2 = F1 × savanna, B3 = F2 × forest, B4 = F2 × savanna, B5 = forest × (F1 × forest), B6 = savanna × (F1 × savanna), B7 = forest × [forest × (F1 × forest)]) and B8 = savanna × [savanna × (F1 × savanna)]. Fig. S5: distribution of outlier SNP scores of the first axis based on redundancy analysis (RDA) of *Plathymenia reticulata*. Text S1: ddRADseq protocol modified from Peterson *et al.* (2012).

FUNDING

This work was supported by the Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG). In addition, we thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their support of the first author's PhD and Post Doctoral scholarship and research fellowships to J.P.L.-F. and M.B.L., respectively.

ACKNOWLEDGEMENTS

We thank the Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) and the Sistema de Autorização e Informação em Biodiversidade (SISBIO) for providing the licence to collect genetic samples. We thank Caio Cruz for help with DNA isolation. We thank Renan Milagres, Maíra F. Goulart and Mariana Vargas Cruz for help with the fieldwork.

AUTHOR CONTRIBUTIONS

J.P.L.-F., M.B.L. and A.G.N conceived and designed the research. J.P.L-F., A.C.M. and M.B.L. performed field trips to collect samples. A.G.N. and R.S.O.B. conducted the laboratory work. A.C.M.,

M.H. and M.B.L. analysed the data. M.B.L., J.P.L.-F. and A.G.N. provided financial resources for the research. A.C.M. and M.B.L. wrote the manuscript. All authors reviewed the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ACCESSIBILITY

Genotype data and population map of *Plathymenia* individuals is available in https://doi.org/10.5061/dryad.vx0k6djzf and the raw sequences data of *Plathymenia* individuals is available in GenBank in the following link https://www.ncbi.nlm.nih.gov/sra/PRJNA1015739.

LITERATURE CITED

Adamack AT, **Gruber B. 2014**. PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* **5**: 384–387. doi:10.1111/2041-210x.12158.

AgapowP-M,**BurtA.2001**.Indicesofmultilocuslinkagedisequilibrium.*Molecular Ecology Notes* **1**: 101–102. doi:10.1046/j.1471-8278.2000.00014.x.

Aitken SN, **Yeaman S**, **Holliday JA**, **Wang T**, **Curtis-McLane S. 2008**. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111. doi:10.1111/j.1752-4571.2007.00013.x.

Alexander DH, **Novembre J**, **Lange K. 2009**. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**: 1655–1664. doi:10.1101/gr.094052.109.

Ambrosone A, **Batelli G**, **Nurcato R**, ***et al***. **2015**. The Arabidopsis RNA-binding protein AtRGGA regulates tolerance to salt and drought stress. *Plant Physiology* **168**: 292–306. doi:10.1104/pp.114.255802.

Anderson EC, **Thompson EA. 2002**. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229. doi:10.1093/genetics/160.3.1217.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antonelli A, **Zizka A**, **Carvalho FA**, ***et al***. **2018**. Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 6034–6039. doi:10.1073/pnas.1713819115.

Arteaga MC, **Bello-Bedoy R**, **Gasca-Pineda J. 2020**. Hybridization between yuccas from Baja California: genomic and environmental patterns. *Frontiers in Plant Science* **11**: 685. doi:10.3389/fpls.2020.00685.

Baack E, **Melo MC**, **Rieseberg LH**, **Ortiz-Barrientos D. 2015**. The origins of reproductive isolation in plants. *New Phytologist* **207**: 968–984. doi:10.1111/nph.13424.

Bailey RI. 2022. ibailey/gghybrid: gghybrid R package for Bayesian hybrid index and genomic cline estimation (v2.0.0). *Zenodo*. doi:10.5281/zenodo.3676498.

Barros MF, **Pinho BX**, **Leão T**, **Tabarelli M. 2018**. Soil attributes structure plant assemblages across an Atlantic forest mosaic. *Journal of Plant Ecology* **11**: 613–622. doi:10.1093/jpe/rtx037.

Benowicz A, **Stoehr M**, **Hamann A**, **Yanchuk AD. 2020**. Estimation of the F2 generation segregation variance and relationships among growth, frost damage, and bud break in coastal Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) wide-crosses. *Annals of Forest Science* **77**: 1–13. doi:10.1007/s13595-020-0925-9.

Bjornstad ON. 2022. *ncf: Spatial Covariance Functions*. R package version 1.3-2. https://CRAN.R-project.org/package=ncf.

Breda AS, **Hazak O**, **Hardtke CS. 2017**. Phosphosite charge rather than shootward localization determines OCTOPUS activity in root protophloem. *Proceedings of the National Academy of Sciences of the United States of America* **114**: E5721–E5730. doi:10.1073/pnas.1703258114.

Bridle JR, **Kawata M**, **Butlin RK. 2019**. Local adaptation stops where ecological gradients steepen or are interrupted. *Evolutionary Applications* **12**: 1449–1462. doi:10.1111/eva.12789.

Buzatti RSdO, **Pfeilsticker TR**, **de Magalhães RF**, **Bueno ML**, **Lemos-Filho JP**, **Lovato MB. 2018**. Genetic and historical colonization analyses of an endemic savanna tree, *Qualea grandiflora*, reveal ancient connections between Amazonian savannas and Cerrado core. *Frontiers in Plant Science* **9**: 981. doi:10.3389/fpls.2018.00981.

Carnaval AC, **Moritz C. 2008**. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* **35**: 1187–1201. doi:10.1111/j.1365-2699.2007.01870.x.

Catchen JM, **Hohenlohe PA**, **Bassham S**, **Amores A**, **Cresko WA. 2013**. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140. doi:10.1111/mec.12354.

Cavallari MM, **Gimenes MA**, **Billot C**, ***et al***. **2010**. Population genetic relationships between *Casearia sylvestris* (Salicaceae) varieties occurring sympatrically and allopatrically in different ecosystems in south-east Brazil. *Annals of Botany* **106**: 627–636. doi:10.1093/aob/mcq151.

Chifman J, **Kubatko L. 2014**. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**: 3317–3324. doi:10.1093/bioinformatics/btu530.

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. doi:10.1038/s41437-019-0188-0.

Crandall KA, **Bininda-Emonds ORP**, **Mace GM**, **Wayne RK. 2000**. Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* **15**: 290–295. doi:10.1016/s0169-5347(00)01876-0.

Dantas VDL, **Batalha MA**, **Pausas JG. 2013**. Fire drives functional thresholds on the savanna–forest transition. *Ecology* **94**: 2454–2463. doi:10.1890/12-1629.1.

De Kort H, **Panis B**, **Helsen K**, **Douzet R**, **Janssens SB**, **Honnay O. 2020**. Preadaptation to climate change through topography-driven phenotypic plasticity. *Journal of Ecology* **108**: 1465–1474. doi:10.1111/1365-2745.13365.

de Lemos-Filho JPD, **Goulart MF**, **Lovato MB. 2008**. Populational approach in ecophysiological studies: the case of *Plathymenia reticulata*, a tree from Cerrado and Atlantic Forest. *Brazilian Journal of Plant Physiology* **20**: 205–216. doi:10.1590/S1677-04202008000300005.

Durigan G, **Ratter JA. 2006**. Successional changes in cerrado and cerrado/forest ecotonal vegetation in western São Paulo State, Brazil, 1962–2000. *Edinburgh Journal of Botany* **63**: 119–130. doi:10.1017/s0960428606000357.

Eaton DAR, **Hipp AL**, **González-Rodríguez A**, **Cavender-Bares J. 2015**. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* **69**: 2587–2601. doi:10.1111/evo.12758.

Feder JL, **Nosil P. 2010**. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**: 1729–1747. doi:10.1111/j.1558-5646.2010.00943.x.

Feder JL, **Egan SP**, **Nosil P. 2012**. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342–350. doi:10.1016/j.tig.2012.03.009.

Feng X, **Xu S**, **Li J**, ***et al***. **2020**. Molecular adaptation to salinity fluctuation in tropical intertidal environments of a mangrove tree *Sonneratia alba*. *BMC Plant Biology* **20**: 178. doi:10.1186/s12870-020-02395-3.

Field DL, **Ayre DJ**, **Whelan RJ**, **Young AG. 2010**. Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. *Heredity* **106**: 841–853. doi:10.1038/hdy.2010.127.

Fitak RR. 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* **6**: bpab017. doi:10.1093/biomethods/bpab017.

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. doi:10.1534/genetics.108.092221.

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. doi:10.1111/mec.14584.

Foster SA, **McKinnon GE**, **Steane DA**, **Potts BM**, **Vaillancourt RE. 2007**. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist* **175**: 370–380. doi:10.1111/j.1469-8137.2007.02077.x.

Furley PA. 1992. Edaphic changes at the forest-savanna boundary with particular reference to the neotropics. In **Furley PA**, **Proct J**, **Ratter JA**. eds.

Nature and dynamics of the forest savanna boundaries. London: Chapman & Hall, 91–117.

Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical Fstatistics. *Molecular Ecology Notes* **5**: 184–186. doi:10.1111/j.1471-8286.2004.00828.x.

Goulart MF, **Lemos-Filho JP**, **Lovato MB. 2005**. Phenological variation within and among populations of *Plathymenia reticulata* in Brazilian Cerrado, the Atlantic Forest and transitional sites. *Annals of Botany* **96**: 445–455. doi:10.1093/aob/mci193.

Goulart MF, **Lemos-Filho JP**, **Lovato MB. 2006**. Variability in fruit and seed morphology among and within populations of *Plathymenia* (Leguminosae - Mimosoideae) in areas of the Cerrado, the Atlantic Forest, and transitional sites. *Plant Biology* **8**: 112–119. doi:10.1055/s-2005-865964.

Goulart MF, **Lovato MB**, **de Vasconcellos Barros F**, **Valladares F**, **Lemos-Filho JP. 2011**. Which extent is plasticity to light involved in the ecotypic differentiation of a tree species from savanna and forest? *Biotropica* **43**: 695–703. doi:10.1111/j.1744-7429.2011.00760.x.

Gramlich S, **Wagner ND**, **Hörandl E. 2018**. RAD-seq reveals genetic structure of the F_2-generation of natural willow hybrids (*Salix* L.) and a great potential for interspecific introgression. *BMC Plant Biology* **18**: 317. doi:10.1186/s12870-018-1552-6.

Gruber B, **Georges A**, **Berry O**, **Unmack P. 2017**. dartR: importing and analysing SNP and silicodart data generated by genome-wide restriction fragment analysis. R package version 0.93. https://CRAN.R-project.org/package=dartR (10 november 2022, date last accessed).

Hao GY, **Hoffmann WA**, **Scholz FG**, ***et al***. **2008**. Stem and leaf hydraulics of congeneric tree species from adjacent tropical savanna and forest ecosystems. *Oecologia* **155**: 405–415. doi:10.1007/s00442-007-0918-5.

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. doi:10.1002/joc.1276.

Hijmans RJ, **Bivand R**, **Forner K**, **Ooms J**, **Pebesma E**, **Sumner MD. 2022**. *Package 'terra'*. Vienna: Maintainer.

Hoffmann WA, **Orthen B**, **Vargas Do Nascimento PK. 2003**. Comparative fire ecology of tropical savanna and forest trees. *Functional Ecology* **17**: 720–726.

Hoffmann WA, **Orthen B**, **Franco AC. 2004**. Constraints to seedling success of savanna and forest trees across the savanna-forest boundary. *Oecologia* **140**: 252–260. doi:10.1007/s00442-004-1595-2.

Hoffmann WA, **Franco AC**, **Moreira MZ**, **Haridasan M. 2005**. Specific leaf area explains differences in leaf traits between congeneric savanna and forest trees. *Functional Ecology* **19**: 932–940. doi:10.1111/j.1365-2435.2005.01045.x.

Hoffmann WA, **Geiger EL**, **Gotsch SG**, ***et al***. **2012**. Ecological thresholds at the savanna-forest boundary: how plant traits, resources and fire govern the distribution of tropical biomes. *Ecology Letters* **15**: 759–768. doi:10.1111/j.1461-0248.2012.01789.x.

Hu Y, **Wang X**, **Xu Y**, ***et al***. **2023**. Molecular mechanisms of adaptive evolution in wild animals and plants. *Science China Life Sciences* **66**: 453–495. doi:10.1007/s11427-022-2233-x.

Hughes CE, **Pennington RT**, **Antonelli A. 2013**. Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* **171**: 1–18. doi:10.1111/boj.12006.

Indoria AK, **Sharma KL**, **Reddy KS. 2020**. Hydraulic properties of soil under warming climate In: *Climate change and soil interactions*. United Kingdom: Elsevier, 473–508.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405. doi:10.1093/bioinformatics/btn129.

Jung C, **Seo JS**, **Han SW**, ***et al***. **2008**. Overexpression of *AtMYB44* enhances stomatal closure to confer abiotic stress tolerance in transgenic Arabidopsis. *Plant Physiology* **146**: 323–324. doi:10.1104/pp.107.110981.

Kamvar ZN, **Tabima JF**, **Grünwald NJ. 2014**. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**: e281. doi:10.7717/peerj.281.

Kassambara A. **2023**. *rstatix*: *Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.7.2. https://rpkgs.datanovia.com/rstatix/ (15 may 2023, date last accessed).

Kremer A, **Hipp AL. 2019**. Oaks: an evolutionary success story. *New Phytologist* **226**: 987–1011. doi:10.1111/nph.16274.

Lacerda DR, **de Lemos-Filho JP**, **Acedo MDP**, **Lovato MB. 2002**. Molecular differentiation of two vicariant neotropical tree species, *Plathymenia foliolosa* and *P. reticulata* (Mimosoideae), inferred using RAPD markers. *Plant Systematics and Evolution* **235**: 67–77.

Lacerda DR, **Lemos-Filho JP**, **Goulart MF**, **Ribeiro RA**, **Lovato MB. 2004**. Seed-dormancy variation in natural populations of two tropical leguminous tree species: *Senna multijuga* (Caesalpinoideae) and *Plathymenia reticulata* (Mimosoideae). *Seed Science Research* **14**: 127–135.

Legendre P, **Andersson MJ. 1999**. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**: 1–24.

Leroy T, **Rougemont Q**, **Dupouey J**, ***et al***. **2019**. Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist* **226**: 1183–1197. doi:10.1111/nph.16039.

Lloyd J, **Bird MI**, **Vellen L**, ***et al***. **2008**. Contributions of woody and herbaceous vegetation to tropical savanna ecosystem productivity: a quasi-global estimate. *Tree Physiology* **28**: 451–468. doi:10.1093/treephys/28.3.451.

Lowry DB. 2012. Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society* **106**: 241–257. doi:10.1111/j.1095-8312.2012.01867.x.

Lowry DB, **Hernandez K**, **Taylor SH**, ***et al***. **2015**. The genetics of divergence and reproductive isolation between ecotypes of *Panicum hallii*. *New Phytologist* **205**: 402–414. doi:10.1111/nph.13027.

Luo J, **Shen G**, **Yan J**, **He C**, **Zhang H. 2006**. AtCHIP functions as an E3 ubiquitin ligase of protein phosphatase 2A subunits and alters plant response to abscisic acid treatment. *The Plant Journal* **46**: 649–657. doi:10.1111/j.1365-313X.2006.02730.x.

Malinsky M, **Matschiner M**, **Svardal H. 2021**. Dsuite-fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* **21**: 584–595. doi:10.1111/1755-0998.13265.

Maracahipes L, **Carlucci MB**, **Lenza E**, ***et al***. **2018**. How to live in contrasting habitats? Acquisitive and conservative strategies emerge at inter- and intraspecific levels in savanna and forest woody plants. *Perspectives in Plant Ecology, Evolution and Systematics* **34**: 17–25. doi:10.1016/j.ppees.2018.07.006.

Muniz AC, **Lemos-Filho JP**, **Souza HA**, ***et al***. **2020**. The protected tree *Dimorphandra wilsonii* (Fabaceae) is a population of inter-specific hybrids: recommendations for conservation in the Brazilian Cerrado/Atlantic Forest ecotone. *Annals of Botany* **126**: 191–203. doi:10.1093/aob/mcaa066.

Muniz AC, **Pimenta RJG**, **Cruz MV**, ***et al***. **2022**. Hybrid zone of a tree in a Cerrado/Atlantic Forest ecotone as a hotspot of genetic diversity and conservation. *Ecology and Evolution* **12**: e8540. doi:10.1002/ece3.8540.

Novaes RML, **Rodrigues JG**, **Lovato MB. 2009**. An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees. *Genetics and Molecular Research* **8**: 86–96. doi:10.4238/vol8-1gmr542.

Novaes RML, **De Lemos-Filho JP**, **Ribeiro RA**, **Lovato MB. 2010**. Phylogeography of *Plathymenia reticulata* (Leguminosae) reveals patterns of recent range expansion towards northeastern Brazil and southern Cerrados in Eastern Tropical South America. *Molecular Ecology* **19**: 985–998. doi:10.1111/j.1365-294X.2010.04530.x.

Oksanen J, **Blanchet FG**, **Friendly M**, ***et al.*** **2022**. vegan: community ecology package. R package version 2.6-4. https://CRAN.R-project.org/package=vegan (20 march 2023, date last accessed).

Padgham M. **2021**. Geodist: Fast, Dependency-Free Geodesic Distance Calculations. R package version 0.0.7. https://github.com/hypertidy/geodist (11 september 2023 date last accessed).

Payseur BA, **Rieseberg LH. 2016**. A genomic perspective on hybridization and speciation. *Molecular Ecology* **25**: 2337–2360. doi:10.1111/mec.13557.

Pembleton LW, **Cogan NOI**, **Forster JW. 2013**. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* **13**: 946–952. doi:10.1111/1755-0998.12129.

Peterson BK, **Weber JN**, **Kay EH**, **Fisher HS**, **Hoekstra HE. 2012**. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* **7**: e37135. doi:10.1371/journal.pone.0037135.

Pfeilsticker TR, **Jones RC**, **Steane DA**, **Harrison PA**, **Vaillancourt RE**, **Potts BM. 2022**. Expansion of the rare *Eucalyptus risdonii* under climate change through hybridization with a closely related species despite hybrid inferiority. *Annals of Botany* **129**: 1–14. doi:10.1093/aob/mcab103.

Pfeilsticker TR, **Jones RC**, **Steane DA**, **Vaillancourt RE**, **Potts BM. 2023**. Molecular insights into the dynamics of species invasion by hybridisation in Tasmanian eucalypts. *Molecular Ecology* **32**: 2913–2929. doi:10.1111/mec.16892.

Pickrell JK, **Pritchard JK. 2012**. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**: e1002967. doi:10.1371/journal.pgen.1002967.

Pinto JRR, **de Oliveira-Filho AT**, **Hay JDV. 2005**. Influence of soil and topography on the composition of a tree community in a central Brazilian valley forest. *Edinburgh Journal of Botany* **62**: 69–90.

R Core Team. **2022**. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

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. doi:10.1111/mec.13322.

Resende-Moreira LC, **Ramos ACS**, **Scliar MO**, ***et al***. **2017**. Gene flow between vicariant tree species: insights into savanna-forest evolutionary relationships. *Tree Genetics and Genomes* **13**: 36. doi:10.1007/s11295-017-1117-0.

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. doi:10.1016/j.biocon.2009.02.021.

Ribeiro RA, **Lemos-Filho JP**, **Ramos ACS**, **Lovato MB. 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. doi:10.1038/hdy.2010.64.

Rieseberg LH, **Blackman BK. 2010**. Speciation genes in plants. *Annals of Botany* **106**: 439–455. doi:10.1093/aob/mcq126.

Rochette NC, **Rivera-Colón AG**, **Catchen JM. 2019**. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28**: 4737–4754. doi:10.1111/mec.15253.

Rodrigues-Souza J, **Do Prado Júnior JA**, **Do Vale VS**, **Schiavini I**, **De Oliveira AP**, **De Silvério Arantes C. 2015**. Secondary forest expansion over a savanna domain at an ecological reserve in the Southeastern Brazil after 15 years of monitoring. *Brazilian Journal of Botany* **38**: 311–322. doi:10.1007/s40415-015-0146-x.

Rutherford S, **Rossetto M**, **Bragg JG**, ***et al***. **2018**. Speciation in the presence of gene flow: population genomics of closely related and diverging *Eucalyptus species*. *Heredity* **121**: 126–141. doi:10.1038/s41437-018-0073-2.

Scholz FG, **Bucci SJ**, **Goldstein G**, **Meinzer FC**, **Franco AC**, **Miralles-Wilhelm F. 2007**. Biophysical properties and functional significance of stem water storage tissues in Neotropical savanna trees. *Plant, Cell and Environment* **30**: 236–248. doi:10.1111/j.1365-3040.2006.01623.x.

Shipley B, **Meziane D. 2002**. The balanced-growth hypothesis and the allometry of leaf and root biomass allocation. *Functional Ecology* **16**: 326–331. doi:10.1046/j.1365-2435.2002.00626.x.

Silva AC, **Souza AF. 2018**. Aridity drives plant biogeographical sub regions in the Caatinga, the largest tropical dry forest and woodland block in South America. *PLoS One* **13**: e0196130. doi:10.1371/journal.pone.0196130.

Silva MC, **Teodoro GS**, **Bragion EFA**, **van den Berg E. 2019**. The role of intraspecific trait variation in the occupation of sharp forest-savanna ecotones. *Flora* **253**: 35–42. doi:10.1016/j.flora.2019.03.003.

Simon MF, **Grether R**, **de Queiroz LP**, **Skema C**, **Pennington RT**, **Hughes CE. 2009**. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 20359–20364. doi:10.1073/pnas.0903410106.

Singh A, **Jha SK**, **Bagri J**, **Pandey GK. 2015**. ABA inducible rice protein phosphatase 2C confers ABA insensitivity and abiotic stress tolerance in Arabidopsis. *PLoS One* **10**: e0125168. doi:10.1371/journal.pone.0125168.

Souza HAV, **Muller LA**, **Brandão RL**, **Lovato MB. 2012**. Isolation of high quality and polysaccharide-free DNA from leaves of *Dimorphandra mollis* (Leguminosae), a tree from the Brazilian Cerrado. *Genetics and Molecular Research: GMR* **11**: 756–764. doi:10.4238/2012.March.22.6.

Souza HAV, **Collevatti RG**, **Lima-Ribeiro MS**, **de Lemos-Filho JP**, **Lovato MB. 2017**. A large historical refugium explains spatial patterns of genetic diversity in a Neotropical savanna tree species. *Annals of Botany* **119**: 239–252. doi:10.1093/aob/mcw096.

Souza ML, **Lovato MB**, **Fagundes M**, **Valladares F**, **Lemos-Filho JP. 2019**. Soil fertility and rainfall during specific phenological phases affect seed trait variation in a widely distributed Neotropical tree, *Copaifera langsdorffii*. *American Journal of Botany* **106**: 1096–1105. doi:10.1002/ajb2.1333.

Souza ML, **Garcia LE**, **Lovato MB**, **Lemos-Filho JP. 2021**. Leaf trait variation during ontogeny in the endangered Brazilian rosewood tree. *Plant Biology* **23**: 1109–1117. doi:10.1111/plb.13318.

The UniProt Consortium. **2023**. UniProt: the Universal Protein Knowledgebase in 2023 Nucleic Acids Res. **51**:D523–D531. doi:10.1093/nar/gkac1052.

Toledo MM, **Paiva EAS**, **Lovato MB**, **de Lemos-Filho JP. 2012**. Stem radial increment of forest and savanna ecotypes of a Neotropical tree: relationships with climate, phenology, and water potential. *Trees* **26**: 1137–1144. doi:10.1007/s00468-012-0690-y.

Tonzo V, **Papadopoulou A**, **Ortego J. 2020**. Genomic footprints of an old affair: single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. *Molecular Ecology* **29**: 2254–2268. doi:10.1111/mec.15475.

Vieira LD, **Silva-Junior OB**, **Novaes E**, **Collevatti RG. 2022**. Comparative population genomics in *Tabebuia* alliance shows evidence of adaptation in Neotropical tree species. *Heredity* **128**: 141–153. doi:10.1038/s41437-021-00491-0.

von Takach Dukai B, **Jack C**, **Borevitz J**, **Lindenmayer DB**, **Banks SC. 2019**. Pervasive admixture between eucalypt species has consequences for conservation and assisted migration. *Evolutionary Applications* **12**: 845–860. doi:10.1111/eva.12761.

Wang Z, **Guo Z**, **Zhong C**, ***et al***. **2022**. Genomic variation patterns of subspecies defined by phenotypic criteria: analyses of the mangrove species complex, *Avicennia marina*. *Journal of Systematics and Evolution* **60**: 835–847. doi:10.1111/jse.12709.

Warwick MC, **Lewis GP. 2003**. Revision of *Plathymenia* (LEGUMINOSAE – MIMOSOIDEAE). *Edinburgh Journal of Botany* **60**: 111–119. doi:10.1017/s0960428603000106.

Wickam H. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer. https://ggplot2.tidyverse.org (18 july 2023, date last accessed).

Wilgenbusch JC, **Swofford D. 2003**. Inferring evolutionary trees with PAUP*. *Current Protocols in Bioinformatics* **00**: 6.4.1–6.4.28. doi:10.1002/0471250953.bi0604s00.

World Conservation Monitoring Centre. 1998. *Plathymenia foliolosa*. The IUCN Red List of Threatened Species 1998: e.T33970A9817719. https://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T33970A9817719.en (4 August 2023, date last accessed).

Yang K-Z, **Jiang M**, **Wang M**, ***et al***. **2015**. Phosphorylation of serine 186 of bHLH transcription factor SPEECHLESS promotes stomatal development in *Arabidopsis*. *Molecular Plant* **8**: 783–795. doi:10.1016/j.molp.2014.12.014.