



# Selective Sweeps Lead to Evolutionary Success in an Amazonian Hyperdominant Palm

Warita A. Melo<sup>1</sup>, Lucas D. Vieira<sup>1</sup>, Evandro Novaes<sup>2</sup>, Christine D. Bacon<sup>3,4</sup> and Rosane G. Collevatti<sup>1\*</sup>

<sup>1</sup> Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil, <sup>2</sup> Departamento de Biologia, Universidade Federal de Lavras, Lavras, Brazil, <sup>3</sup> Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden, <sup>4</sup> Gothenburg Global Biodiversity Centre, Gothenburg, Sweden

## OPEN ACCESS

### Edited by:

Gyaneshwer Chaubey,  
Banaras Hindu University, India

### Reviewed by:

Eaaswarkhanth Muthukrishnan,  
Dasman Diabetes Institute, Kuwait  
Ivan Scotti,  
Institut National de Recherche pour  
l'agriculture, l'alimentation et  
l'environnement (INRAE), France

### \*Correspondence:

Rosane G. Collevatti  
rosanegc68@icloud.com

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 20 August 2020

**Accepted:** 18 November 2020

**Published:** 23 December 2020

### Citation:

Melo WA, Vieira LD, Novaes E,  
Bacon CD and Collevatti RG (2020)  
Selective Sweeps Lead  
to Evolutionary Success in an  
Amazonian Hyperdominant Palm.  
Front. Genet. 11:596662.  
doi: 10.3389/fgene.2020.596662

Despite the global importance of tropical ecosystems, few studies have identified how natural selection has shaped their megadiversity. Here, we test for the role of adaptation in the evolutionary success of the widespread, highly abundant Neotropical palm *Mauritia flexuosa*. We used a genome scan framework, sampling 16,262 single-nucleotide polymorphisms (SNPs) with target sequence capture in 264 individuals from 22 populations in rainforest and savanna ecosystems. We identified outlier loci as well as signal of adaptation using Bayesian correlations of allele frequency with environmental variables and detected both selective sweeps and genetic hitchhiking events. Functional annotation of SNPs with selection footprints identified loci affecting genes related to adaptation to environmental stress, plant development, and primary metabolic processes. The strong differences in climatic and soil variables between ecosystems matched the high differentiation and low admixture in population Bayesian clustering. Further, we found only small differences in allele frequency distribution in loci putatively under selection among widespread populations from different ecosystems, with fixation of a single allele in most populations. Taken together, our results indicate that adaptive selective sweeps related to environmental stress shaped the spatial pattern of genetic diversity in *M. flexuosa*, leading to high similarity in allele frequency among populations from different ecosystems.

**Keywords:** Population Genomics, neutral evolution, *Mauritia flexuosa*, Arecaceae, adaptation, target sequence capture

## INTRODUCTION

The geographic distribution of genetic diversity results from the interplay between demographic forces (i.e., genetic drift and gene flow), mutation, recombination, and natural selection (Wright, 1931). The study of adaptation to environmental conditions has led to an understanding of how species cope with climate and landscape changes (e.g., Evans et al., 2014; Gratani, 2014; Yeaman et al., 2014; Hornoy et al., 2015). Genome scan approaches have contributed to the understanding of adaptive responses (Savolainen et al., 2013; Tiffin and Ross-Ibarra, 2014), especially because local adaptation may affect genetic variation at specific loci (Smith and Haigh, 1974; Kaplan et al., 1989; Orr, 1998; Kawecki and Ebert, 2004), allowing for the identification of genomic regions that are

tightly linked to or directly under natural selection. However, it is important to note that most adaptive traits are polygenic with large number of alleles with small effects (McKay and Latta, 2002; Pritchard and Di Rienzo, 2010). In these traits, selection may cause only small changes in allele frequencies, and the response to selection may be the outcome of allelic covariances among loci (McKay and Latta, 2002; Pritchard and Di Rienzo, 2010; Le Corre and Kremer, 2012).

Evidence of local adaptation has been identified in plants from temperate climates (Parchman et al., 2012; Yeaman et al., 2014, 2016; Hornoy et al., 2015; Jordan et al., 2017), yet only recently have these approaches been applied to tropical species (e.g., Lanes et al., 2018; Collevatti et al., 2019; Brousseau et al., 2020). Despite recent advances, the detection of selection footprints using genome scans is still limited because of demographic history, which can constrain the detection of other types of selective forces, such as background, balancing, or purifying selection (Otto, 2000; Vitti et al., 2013). The detection of selection footprints may also be hindered by range shifts due to paleoclimatic changes, causing incomplete lineage sorting and high genetic differentiation (Collevatti et al., 2019). Thus, disentangling the effects of demographic history from selection in extant genetic diversity is important to better understand the evolution of species distributions, but remains a challenge.

Target sequence capture has been increasingly used to genotype large numbers of markers because it significantly reduces costs and effort compared with whole-genome sequencing and can be used to study the role of evolution and ecology at the genome-wide scale (Jones and Good, 2016; Silva-Junior et al., 2018; Andermann et al., 2019). In plants, the method has been applied successfully to robustly estimate population genetic parameters in groups such as maize (Fu et al., 2010), pine [*Pinus taeda*, (Neves et al., 2013); *Pinus albicaulis* (Syring et al., 2016)], poplars (Zhou and Holliday, 2012), and tropical trees and palms (e.g., Collevatti et al., 2019; Loiseau et al., 2019).

Palms have high levels of genetic and phenotypic variation and a long and rich fossil history (Baker and Dransfield, 2016). The family has a wide distribution across large geographic areas and latitudinal ranges, but climatic variables cause palms to be spatially restricted along temperature and other geographic and environmental gradients (Eiserhardt et al., 2013; Freitas et al., 2019). *Mauritia flexuosa* L. f. (Arecaceae) is one of the most hyperdominant species in Amazonia and is common in swampy environments (Rull, 1998; Ter Steege et al., 2013; Rull and Montoya, 2014). It has a large geographic range across different ecosystems (**Figure 1**): in lowland tropical forest in the Amazon and Orinoco Basins and the Guiana Shields, and in savannas in the Llanos and the Brazilian Cerrado.

The high abundance and wide geographic distribution of *M. flexuosa* raise the question of how the species reached such level of evolutionary success. The species is dioecious and wind-pollinated (Khorsand and Koptur, 2013). Fossil records and phylogeographic analyses of *M. flexuosa* show a capacity for resilience to climate change, at least through Quaternary range shifts (e.g., de Lima et al., 2014; Melo et al., 2018). Moreover, the species has high genetic differentiation among populations from

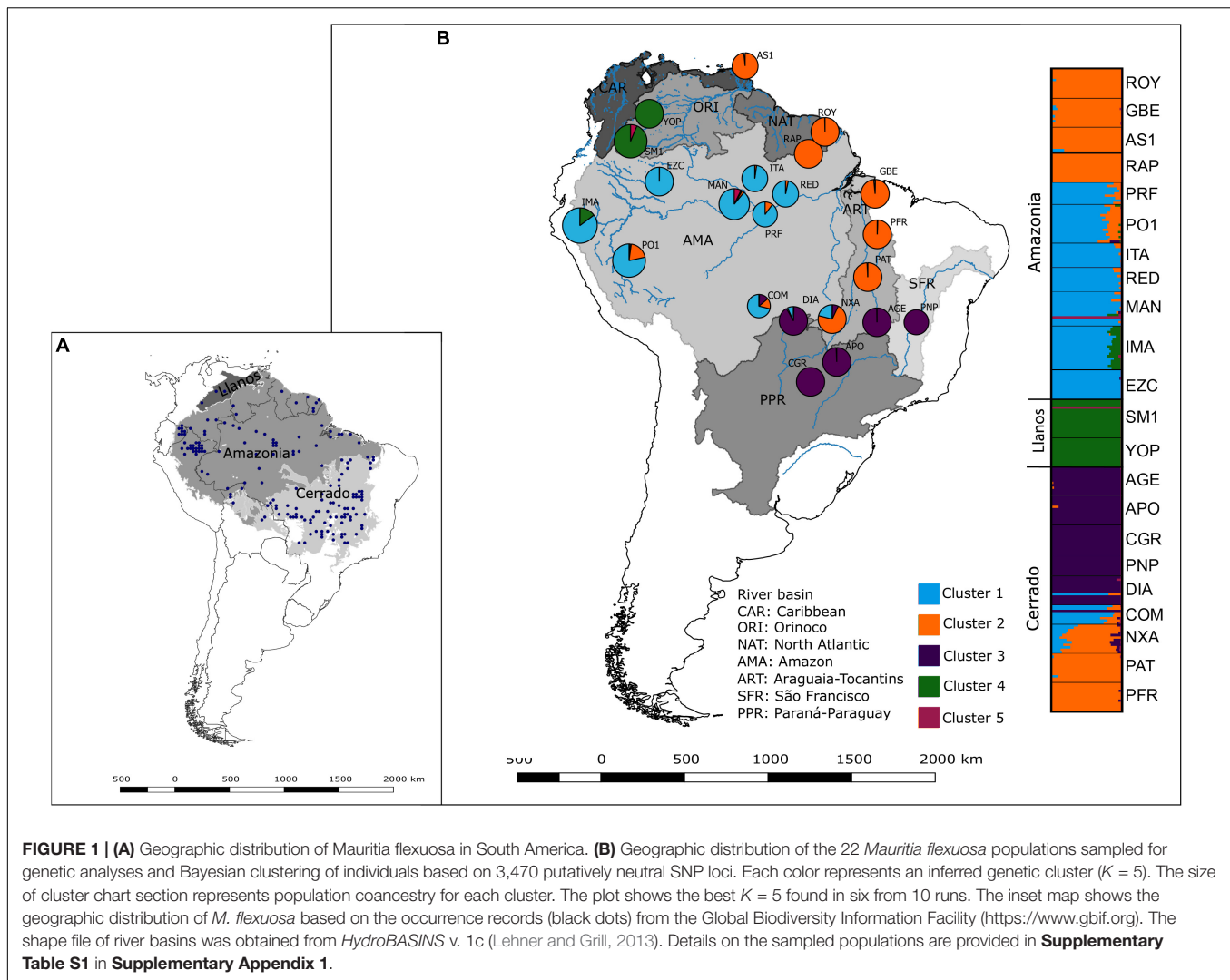
different ecosystems and river basins as shown by chloroplast (de Lima et al., 2014) and microsatellite loci (Melo et al., 2018; Sander et al., 2018), suggesting that genetic drift may shape the current distribution of genetic variation in *M. flexuosa*. A genome-wide scan allows for a comprehensive evaluation of the relative roles of genetic drift and selection in the spatial distribution of genetic diversity and provides an unprecedented window into the evolutionary success of this iconic Neotropical palm species.

Here, we test between local adaptation and genetic drift as a driver of the evolutionary success characterizing the high abundance and broad geographic range of the hyperdominant species *M. flexuosa*. We identify local adaptation through detection of outlier loci and correlation of allele frequencies with environmental variables that indicate ecosystem-based differences among populations. To identify selective sweeps and genetic hitchhiking events, we used likelihood ratio tests. For these tests, we sampled populations in three different ecosystems across its geographic distribution, genotyping 16,262 single-nucleotide polymorphisms (SNPs) with target sequence capture approach. Coalescent simulations were used to reconstruct the demographic history of the species based on putatively neutral SNPs in an attempt to disentangle the effects of demography (genetic drift and gene flow) and natural selection on the distribution of genetic diversity. Because of the contrasting environmental conditions of the different ecosystems, i.e., tropical rain forest and seasonal savanna, we expect to find local adaptation and high differentiation in allele frequencies among ecosystems.

## MATERIALS AND METHODS

### Population Sampling

We sampled 264 individuals from 22 populations of *M. flexuosa* throughout its geographic range (**Figure 1** and **Supplementary Table S1** in **Supplementary Appendix 1**). Populations were sampled in Amazonia (11 populations), Llanos (two), and Cerrado (nine) ecosystems across different river basins: eight populations from the Amazon basin, one in the Caribbean, two in the North Atlantic, two in the Orinoco, three in the Paraná-Paraguay, one in the São Francisco, and five in the Araguaia-Tocantins (**Figure 1**; **Supplementary Table S1** in **Supplementary Appendix 1**). Special authorization in Brazil was not necessary for sampling, because we sampled outside protected reserves. Permits for leaf collection were granted from Trinidad through the University of the West Indies and the Department of Natural Resources and the Environment to S. Federman; from Bolivia by the Ministry of Environment and Water to the Bolivian National Herbarium for carrying out research under the PALMS 2009–2013 project led by H. Balslev; from Colombia by the Ministry of the Environment and Sustainable Development under the Autoridad Nacional de Licencias Ambientales (Resolution 0790) to CES University in collaboration with MJ Sanín Perez; and from Peru through the Instituto Nacional de Recursos Naturales of the Ministry of Agriculture to H. Balslev (690-2007-INRENA-IFFS-DCB). In Brazil, vouchers were collected in each population and compared to exsiccate UB8821 from UNB herbarium



(Universidade de Brasília, Brasília, DF, Brazil). A shape file of river basins was obtained from *HydroBASINS v. 1c* (Lehner and Grill, 2013). Leaves of adults were collected, and geographic coordinates were taken for each individual.

## SNP Genotyping

We used a set of 2,909 120-mer probe sequences targeting single copy loci derived from *Cocos nucifera*, *Elaeis guineensis*, *Nypa fruticans*, *Phoenix dactylifera*, and *Sabal bermudana* covering 169 single-copy nuclear genes for target sequence capture (Heyduk et al., 2016). *M. flexuosa* has  $2n = 30$  and genome size  $1C = 4.7$  pg (Röser et al., 1997). We extracted the DNA using Qiagen DNeasy Plant Mini kit (Qiagen, DK). Targeted DNA enrichment, capture, and sequencing were carried out by RAPiD Genomics LLC (Gainesville, FL, United States) using the SureSelectXT (Agilent Technologies, CA, United States) enrichment system and Illumina DNA sequencing on a HiSeq2000 instrument (Illumina, CA, United States). Genomic DNA libraries were prepared using the Agilent protocols, including DNA shearing followed by ends repair, 3'-end adenylation, adaptor ligation,

and amplification following Neves et al., 2013. Libraries were pooled in equimolar amounts and sequenced in two lanes of an Illumina HiSeq2000 instrument (Illumina, CA, United States),  $2 \times 101$ -bp mode.

Sequencing reads were trimmed for adapters using Trimmomatic v. 0.39 (Bolger et al., 2014). Stretches of low-quality sequences were removed from reads using the option SLIDINGWINDOW: 4:20, and sequences smaller than 32 bp were discarded. Reads were aligned to the probe sequences with BWA v. 0.5.9 (Li and Durbin, 2009) using default settings in the BWA-backtrack algorithm. These steps were performed in a grid computing cluster software system using a pipeline available from the International Cassava Genetic Map Consortium (Alaba et al., 2015). SNP calling was performed using FreeBayes v. 1.1.0 (Garrison and Marth, 2012). FreeBayes genotyped biallelic SNPs only in the target region, using the following options: `-theta 0.01 -ploidy 2 -min-alternate-fraction 0.2 -min-alternate-count 2 -min-coverage 8 -genotype-qualities -min-mapping-quality 10 -min-base-quality 20 -report-genotype-likelihood-max-no-indels-no-mnps-no-complex`. Data were finally filtered using

VCFtools v. 0.1.12b (Danecek et al., 2011) with the following parameters:—max-missing 0.6—mac 2—minGQ 20—minDP 3.

## Genome-Wide Diversity and Genetic Structure

To characterize genome-wide diversity, we calculated the density of SNPs across all probes using a bin size of 100 bp, allele frequencies, and the percentage of missing data. We also inferred the ratio of transitions to transversions substitutions (Ts/Tv) and expected heterozygosity using *VCFtools*.

To characterize population genetic diversity across all SNP loci, we estimated expected heterozygosity under Hardy–Weinberg equilibrium ( $H_e$ ) and inbreeding coefficient ( $f$ ) using Arlequin v. 3.5 (Excoffier and Lischer, 2010). We used a hierarchical analysis of molecular variance (AMOVA) implemented in Arlequin to estimate the genetic differentiation between ecosystems ( $F_{CT}$ ) and among populations within ecosystems ( $F_{SC}$ ). We also analyzed the differentiation among river basins ( $F_{CT}$ ), among populations within river basins ( $F_{SC}$ ), and inbreeding coefficient ( $F_{IS}$ ). Significance levels of 0.05 for each estimate were determined with 10,000 permutations.

Further, we estimated genetic diversity and analyzed genetic structure in a subset of more stringently cleaned, and linkage disequilibrium (LD) pruned loci. We used PLINK 1.9 (Chang et al., 2015) to eliminate loci with more than 15% of missing data (—geno 0.15) and a minor allele frequency (MAF) lower than 1% (—maf 0.01). The LD pruning was also performed with PLINK with sliding windows of 50 SNPs, removing associated loci ( $R^2 > 0.30$ ) using the option —indep-pairwise 50 5 0.3. The list of cleaned and LD-pruned SNPs was checked for SNPs with positive selection signal, and none were identified, based on Outflank and Bayenv2 results. Therefore, this list of potentially neutral and high-quality SNPs was used to estimate genetic diversity and to perform an AMOVA. Finally, we used Bayesian clustering implemented in fastSTRUCTURE software (Raj et al., 2014) to identify the most likely number of genetic clusters among *M. flexuosa* samples. The analysis was performed with number of groups  $K$  ranging from 2 to 22. We assessed the most likely number of clusters supported by the data in Structure Selector (Li and Liu, 2018), with a threshold of 0.5, following (Puechmaille, 2016). To generate output files, we used CLUMPAK v. 1.1 (Kopelman et al., 2015). For this set of potentially neutral loci, we analyzed whether populations are isolated by distance using a linear regression of linearized  $F_{ST}$  on the logarithm of geodesic geographic distance. We also performed an autocorrelation analyses using Moran's  $I$ , implemented in the software SAM (Spatial Analysis in Macroecology; Rangel et al., 2010), to understand the effect of spatial scale in genetic differentiation.

## Simulation of Demographic History

To test whether demographic dynamics in the Quaternary were responsible for the distribution of allele frequencies in extant populations of *M. flexuosa*, we modeled the demographic history and performed simulations for 3,500 potentially neutral loci. We simulated two demographic scenarios derived from ecological

niche modeling (see de Lima et al., 2014; Melo et al., 2018): “Range Expansion” (smaller range size at the LGM compared to present day) and “Range Stability” (similar range size at the LGM and present day). We also simulated the scenario of “Multiple Refugia” (retraction of savanna-like vegetation during glacial periods leading to many refugia of different effective sizes) derived from paleovegetation reconstruction and fossil record. We simulated 22 demes using the software DIYABC v. 2.1 (Cornuet et al., 2014). For model calibration, we used the demographic parameters estimated using coalescent analyses following Melo et al. (2018). Population dynamics were simulated backward (see **Supplementary Figure S1 in Supplementary Appendix 2**) from the present ( $t = 0$ ) to  $t = 525$  generations ago (at 21 ka, using a generation time of 40 years; Melo et al., 2018), with effective size drawn from a uniform distribution with minimum  $N_e = 300$  and maximum  $N_e = 3,000$  (**Supplementary Figure S1 in Supplementary Appendix 2**).

Mean number of alleles, mean expected heterozygosity, and  $F_{ST}$  were inferred from 600,000 simulations and compared with the observed values and the relative fits of the models using approximate Bayesian computation (Cornuet et al., 2014), also implemented in DIYABC. We used the relative proportion of each scenario in the simulated data set closest to the observed data set (hereafter, direct approach) and the logistic regression (hereafter, logistic approach) of each scenario probability on the deviation between simulated and observed summary statistics (Cornuet et al., 2008).

## Genome Scans for Selection Footprints

To test for local adaptation, we used three different genome scan approaches to identify loci under selection. First, OutFLANK (Whitlock and Lotterhos, 2015) was used based on the expected  $\chi^2$  distribution of  $F_{ST}$  in the absence of selection. To generate the distributions, we trimmed the  $F_{ST}$  distribution at 5% and 10% and used a minimum expected heterozygosity of 0.10 and a false discovery rate  $< 0.05$ .

Secondly, we used a Bayesian framework implemented in Bayenv2 (Coop et al., 2010; Günther and Coop, 2013) to identify local adaptation by estimating linear correlations between allele frequencies and environmental variables, while controlling for relationships among populations. Four bioclimatic variables from the WorldClim Global Climate database<sup>1</sup> with a spatial resolution of 30” ( $0.93 \times 0.93 = 0.86 \text{ km}^2$  at the equator) were obtained for the 22 populations sampled. The four bioclimatic variables were identified by low collinearity in factorial analysis with Varimax rotation. These four variables explained 90.4% of the total environmental variation among the 22 populations (**Supplementary Table S2 in Supplementary Appendix 1**). The selected variables were as follows: annual mean temperature, mean diurnal range, precipitation of driest month, and precipitation of the wettest quarter.

We also obtained soil data related to soil fertility from the Harmonized World Soil Database v. 1.2<sup>2</sup>. Varimax factorial

<sup>1</sup><http://www.worldclim.org/bioclim>

<sup>2</sup><http://www.fao.org/docrep/018/aq361e/aq361e.pdf>

analysis selected four soil variables that explained 77.5% of the variation among populations (**Supplementary Table S3 in Supplementary Appendix 1**): topsoil salinity, topsoil organic carbon, topsoil silt fraction, and topsoil reference bulk density. Loci with adaptive selection signal were selected following Günther and Coop, 2013: high Bayes factor ( $> 100.0$ ), i.e., the likelihood ratio of the probability of the linear relationship hypothesis (between allele frequency and environmental variable) and the null hypothesis (no linear relationship) given the data and a Spearman correlation coefficient  $|\rho| > 0.15$ . These values were in the 0.1% quantile of the distribution of the data.

For the loci under selection, we calculated the shift in allele frequency between pairs of populations, i.e., the difference between the allele frequency of the population with highest and lowest value for each climatic and soil variable and for each SNP locus. For each variable, we obtained a vector of differences in allele frequency for the loci, and then, we plotted the density function using the function density implemented in the stat package of R v. 3.6.1 (R Core Team, 2016).

Finally, SweepFinder2 v. 1.0 (Degiorgio et al., 2016) was used to detect selective sweeps and genetic hitchhiking based on deviations of a neutral null hypothesis. We applied the composite likelihood ratio (CLR) to identify possible location of recent selective sweeps in each of the 22 populations, following Nielsen, 2005. We used only contigs containing loci with evidence of adaptive selection identified in Bayenv2. We assumed unknown polarity and considered that the lower frequency allele was the derived state. For each contig, the CLR function was calculated on a grid of 122 positions along the length of the probe sequence. The selective sweep CLR test statistic was the maximum composite likelihood value optimized over all possible positions, compared to the composite likelihood of the neutral null model as calculated by the program. We simulated new neutral data sets with population scaled mutation rate ( $\theta$ ) and population scaled recombination rate ( $\rho$ ) typical for highly heterozygous tropical forest trees estimated from genome-wide data (Silva-Junior and Grattapaglia, 2015). We used  $\theta/bp=0.018 \text{ bp}^{-1}$  and  $\rho/bp=0.0011 \text{ bp}^{-1}$ , assuming homogenous rates among regions. Samples with the same number of individuals in the real populations and sequence length of scaffolds were generated under Wright–Fisher model using the software *ms* (Hudson, 2002) with 100 replicates per scaffold. For each replicate, multiple alignments were performed using MAFFT v. 7.130b (Katoh and Standley, 2013), and variant sites were detected across the samples using SNP sites (Page et al., 2016). The composite likelihood inference was repeated for each replicate in SweepFinder2 using the same grid size as for the real data. Composite likelihood values were collected from all the replicates across all positions along the length of the simulated sequences totaling 10.7 million data points. A cumulative histogram of the values was used to obtain the neutral composite likelihood threshold of 95% (**Supplementary Figure S2 in Supplementary Appendix 2**). The value of 1.22 was used as the significance cutoff. The probable locations of sweeps were taken as the maximum composite values above the significance cutoff calculated for each scaffold.

## Annotation of SNP Loci Under Selection

Putative SNP loci under selection were annotated against the genomes of Embryophytes of the Phytozome v. 12.1 (Goodstein et al., 2012) database<sup>3</sup>. The probe sequences were used in a BlastX, and in the few cases where no Blast hit was obtained, the *E. guineensis* and *M. flexuosa* transcriptomes were used (**Supplementary Table S1 in Supplementary Appendix 3**). Affected genes were analyzed with respect to their functional annotation in terms of Gene Ontology, PFAM, and PANTHER categories obtained from the Blast best hits (Ashburner et al., 2000).

## RESULTS

### SNP Detection and Genotyping

Using target sequence capture, we obtained more than 37 Gbp from the 264 individuals. A total of 366,464,311 reads were sequenced, with a mean of 832,873 reads per individual. The sequencing generated 34,569 SNPs, which after data filtering resulted in 16,262 high-quality polymorphic SNPs, with an average call rate of 93.2%. The average per-sample call rate across all 264 individuals was 93.1%. The transitions-to-transversions ratio was 2.06. The average missing genotype per loci was 6.8% in the sample of 264 individuals, and 6.9% per individual sampled (**Supplementary Figure S3 in Supplementary Appendix 2**). The distribution of MAF showed the expected “L” shape distribution with a larger proportion of low frequency SNPs. The vast majority of the SNPs (82.2%) had MAF  $< 0.15$ , and the median MAF was equal to 0.079. The average read depth at SNPs coordinates across samples was  $25.4x \pm 34.5x$ , and the minimum depth of aligned reads to call a heterozygous genotype for a sample was three with the median equal to 26.

### Genome-Wide Diversity and Genetic Structure

The average SNP density was 1.06 SNP/100 bp (median = 1, min = 0.1, max = 4.20; **Supplementary Figure S3 in Supplementary Appendix 2, Supplementary Table S2 in Supplementary Appendix 3**). Genetic diversity was high in all populations (**Table 1**) and the total genetic diversity considering all the 16,262 SNPs ( $He = 0.338$ ,  $SD = 0.174$ ) was similar to the diversity obtained with only putatively neutral loci ( $He = 0.357$ ,  $SD = 0.023$ ).

To perform a genetic structure analysis for putatively neutral loci, the SNPs were further cleaned to remove loci in LD. The 3,479 loci after LD pruning had no selection signal. Therefore, these neutral LD pruned SNPs were used in a Bayesian clustering method, which supported six genetic clusters ( $K = 6$ ) with low admixture in populations from different ecosystems (**Figure 1 and Supplementary Figures S4, S5 in Supplementary Appendix 2; Supplementary Tables S4–S6 in Supplementary Appendix 1**). However, no individual was assigned to cluster 6 with coancestry coefficient  $Q > 0.001$ , being impossible to display in a map due

<sup>3</sup><https://phytozome.jgi.doe.gov/pz/portal.html>

**TABLE 1** | Genetic diversity in 22 populations of *Mauritia flexuosa*, based on all SNPs (16,262 SNPs) and on 3,479 potentially neutral SNPs.

Ecosystem	River basin	Population	All SNPs					Neutral SNPs			
			<i>n</i>	No. loci	<i>He</i> (SD)	<i>Ho</i> (SD)	<i>f</i>	<i>He</i> (SD)	<i>Ho</i> (SD)	<i>f</i>	
Amazonia	Amazon	COM	8	5,322	0.345 (0.172)	0.502 (0.352)	-0.576	0.361 (0.164)	0.523 (0.344)	-0.509	
		EZC	12	4,457	0.362 (0.169)	0.548 (0.354)	-0.614	0.347 (0.169)	0.515 (0.350)	-0.528	
		IMA	18	5,447	0.311 (0.187)	0.470 (0.368)	-0.593	0.316 (0.178)	0.472 (0.355)	-0.523	
		ITA	10	4,244	0.385 (0.162)	0.593 (0.350)	-0.622	0.363 (0.165)	0.542 (0.350)	-0.557	
		MAN	14	6,264	0.296 (0.180)	0.397 (0.373)	-0.408	0.317 (0.173)	0.436 (0.354)	-0.418	
		PO1	16	3,803	0.353 (0.176)	0.541 (0.367)	-0.658	0.342 (0.174)	0.537 (0.363)	-0.575	
		PRF	9	4,091	0.396 (0.160)	0.608 (0.346)	-0.630	0.374 (0.164)	0.560 (0.348)	-0.569	
		RED	10	4,096	0.392 (0.163)	0.611 (0.356)	-0.636	0.369 (0.166)	0.555 (0.357)	-0.579	
		Mean	12.1	—	0.355	0.534	—	0.354	0.530	—	
	SD	3.6	—	0.171	0.358	—	0.021	0.040	—		
	Total	97	—	0.179 (0.199)	—	-0.599	0.228 (0.190)	—	-0.517		
	Caribbean	AS1	10	3,496	0.392 (0.160)	0.615 (0.353)	-0.655	0.382 (0.161)	0.587 (0.351)	-0.614	
		North Atlantic	RAP	12	3,578	0.372 (0.167)	0.588 (0.363)	-0.666	0.364 (0.166)	0.567 (0.359)	-0.618
			ROY	12	3,593	0.376 (0.167)	0.595 (0.363)	-0.659	0.367 (0.168)	0.569 (0.361)	-0.615
		Mean	12	—	0.374	0.592	—	0.365	0.568	—	
SD		0	—	0.167	0.363	—	0.001	0.569	—		
Total		24	—	0.335 (0.335)	—	-0.661	0.324 (0.181)	—	-0.592		
Overall Amazonia	Mean	12.2	—	0.351	0.525	—	0.363	0.542	—		
	SD	3.0	—	0.167	0.354	—	0.021	0.042	—		
	Total	159	—	0.113 (0.171)	—	-0.509	0.218 (0.191)	—	-0.530		
Lhanos	Orinoco	SM1	16	12,076	0.196 (0.146)	0.161 (0.307)	0.172*	0.286 (0.171)	0.344 (0.355)	-0.281	
		YOP	12	4,123	0.381 (0.165)	0.597 (0.355)	-0.657	0.376 (0.160)	0.577 (0.349)	-0.587	
		Mean	14	—	0.289	0.379	—	0.331	0.344	—	
		SD	2.8	—	0.156	0.331	—	0.045	0.577	—	
Cerrado	Paraná-Paraguay	Total	28	—	0.165 (0.162)	—	-0.119	0.261 (0.180)	—	0.367	
		APO	12	4,194	0.326 (0.182)	0.497 (0.370)	-0.757	0.368 (0.171)	0.591 (0.362)	-0.709	
		CGR	12	4,222	0.314 (0.190)	0.484 (0.383)	-0.630	0.348 (0.179)	0.545 (0.372)	-0.625	
		DIA	12	5,726	0.299 (0.187)	0.448 (0.370)	-0.618	0.346 (0.176)	0.531 (0.367)	-0.570	
		Mean	12	—	0.313	0.476	—	0.348	0.545	—	
		SD	0	—	0.186	0.374	—	0.009	0.025	—	
	São Francisco	Total	36	—	0.209 (0.196)	—	-0.696	0.257 (0.193)	—	-0.649	
		PNP	9	3,436	0.380 (0.169)	0.598 (0.369)	-0.662	0.383 (0.162)	0.593 (0.359)	-0.639	
		Araguaia-Tocantins	AGE	12	5,097	0.303 (0.187)	0.456 (0.366)	-0.649	0.352 (0.177)	0.550 (0.364)	-0.606
			GBE	12	5,613	0.294 (0.186)	0.439 (0.365)	-0.627	0.338 (0.177)	0.520 (0.365)	-0.583
			NXA	12	5,607	0.302 (0.186)	0.454 (0.366)	-0.629	0.341 (0.176)	0.526 (0.359)	-0.588
		PAT	12	4,040	0.337 (0.184)	0.527 (0.380)	-0.656	0.353 (0.175)	0.548 (0.366)	-0.611	
		PFR	12	3,936	0.332 (0.184)	0.511 (0.368)	-0.810	0.384 (0.164)	0.619 (0.346)	-0.731	
		Mean	12	—	0.314	0.477	—	0.352	0.548	—	
		SD	0	—	0.185	0.369	—	0.016	0.035	—	
Total	60	—	0.176 (0.191)	—	-0.696	0.239 (0.191)	—	-0.65			
Overall Cerrado	Mean	11.7	—	0.321	0.49	—	0.352	0.548	—		
	SD	1	—	0.184	0.371	—	0.016	0.032	—		
	Total	105	—	0.154 (0.188)	—	-0.685	0.225 (0.192)	—	-0.638		
Overall ecosystems	Mean	12	—	0.338	0.511	—	0.357	0.546	—		
	SD	2.4	—	0.174	0.361	—	0.023	0.057	—		
	Total	264	—	0.107 (0.168)	—	-0.583	0.206 (0.184)	—	-0.559		

*N*, number of genotyped individuals; *N* loci, number of polymorphic loci; *He*, mean expected heterozygosity under Hardy-Weinberg equilibrium; *Ho*, mean observed heterozygosity; *f*, inbreeding coefficient (all values are significantly different from zero,  $p < 0.05$ ); SD, standard deviation.

to the relative scale (**Supplementary Table S4 in Supplementary Appendix 1**). Thus, we chose the partition of  $K = 5$  (**Figure 1** and **Supplementary Tables S5, S6 in Supplementary Appendix 1**).

The 22 populations showed low and significant genetic differentiation for the 16,262 SNPs ( $F_{ST} = 0.070$ ,  $p < 0.001$ ) and negative and significant inbreeding coefficient ( $F_{IS} = -0.583$ ,  $p < 0.001$ ), showing higher frequency of mating between unrelated individuals than expected at random, as predicted for a dioecious species. Genetic differentiation was higher for the 3,470 putatively neutral loci ( $F_{ST} = 0.104$ ,  $p < 0.001$ ), and the inbreeding coefficient was significant ( $F_{IS} = -0.559$ ,  $p < 0.001$ ). Hierarchical AMOVA showed low, but significant differentiation among ecosystems (Amazonian, Llanos, and Cerrado), for the entire dataset ( $F_{CT} = 0.015$ ,  $p < 0.001$ ) and for the 3,470 putatively neutral, LD-pruned loci ( $F_{CT} = 0.063$ ,  $p < 0.001$ ). Differentiation among populations within ecosystems was low for the entire dataset ( $F_{SC} = 0.064$ ,  $p < 0.001$ ) and for the putatively neutral loci ( $F_{SC} = 0.067$ ,  $p < 0.001$ ). Differentiations among river basins ( $F_{CT} = 0.055$ ,  $p < 0.001$ ) and among populations within river basins ( $F_{SC} = 0.028$ ,  $p < 0.001$ ) were low but significant for the entire dataset, as well as for the 3,470 putatively neutral, LD-pruned loci ( $F_{CT} = 0.080$ ,  $p < 0.001$ ;  $F_{SC} = 0.041$ ,  $p < 0.001$ ). Genetic differentiation among pairs of populations was significantly correlated to geographical distance ( $r^2 = 0.290$ ,  $p < 0.001$ , **Supplementary Figure S6A in Supplementary Appendix 2**). Autocorrelation analysis showed a significant relationship between genetic differentiation and geographical distance for populations up to 1,500 km (**Supplementary Figure S6B in Supplementary Appendix 2**).

## Simulation of Demographic History

The demographic simulations using putatively neutral loci, to analyze the role of genetic drift in shaping the spatial distribution of *M. flexuosa* genetic diversity, supported a scenario of “Multiple Refugia” as the most probable predictor of the observed genetic parameters using the logistic approach (**Supplementary Figure S7 in Supplementary Appendix 2**; **Supplementary Table S7 in Supplementary Appendix 1**). Using the direct approach, we found little variation in likelihood among the different demographic scenarios tested (**Supplementary Figure S6 in Supplementary Appendix 2**).

## Genome Scans for Selection Footprints

The OutFLANK genome scan detected 45 outlier loci (**Supplementary Table S8 in Supplementary Appendix 1**, **Supplementary Figures S8, S9 in Supplementary Appendix 2**, and **Supplementary Table S3 in Supplementary Appendix 3**). These outlier loci had  $F_{ST}$  values ranging from 0.77 to 1.00 (**Supplementary Figures S8, 9 in Supplementary Appendix 2**).

Bayenv2 identified outlier loci with high correlation to climatic variables (17 loci; **Supplementary Table S9 in Supplementary Appendix 1**, **Supplementary Figure S10 in Supplementary Appendix 2**, and **Supplementary Table S4 in Supplementary Appendix 3**) and to soil variables (24 loci; **Supplementary Table S10 in Supplementary Appendix 1**, **Supplementary Figure S11 in Supplementary Appendix 2**, and **Supplementary Table S5 in Appendix 3**). Correlation coefficients

between allele frequency and the environmental variables ranged between  $\pm 0.50$  (**Supplementary Figures S10, S11 in Supplementary Appendix 2**). Climatic and soil variables varied among populations (**Figure 2**, **Supplementary Table S11 in Supplementary Appendix 1**), but did not show a clinal spatial pattern (**Figure 2**). Most loci under selection had small differences in allele frequency between populations with different climatic and soil conditions (**Figure 3**).

We detected hard selective sweeps within 28 of the 41 contigs identified from Bayenv2 (**Supplementary Table S12 in Supplementary Appendix 1**). These 28 contigs contained 134 positions that were likely targets of selective sweeps (**Supplementary Table S12 in Supplementary Appendix 1**). Most populations were fixed for the same allele with selection signal (**Figure 4**, see also **Supplementary Figures S12, S13 in Supplementary Appendix 2**).

## Functional Annotation of SNPs With Selection Signal

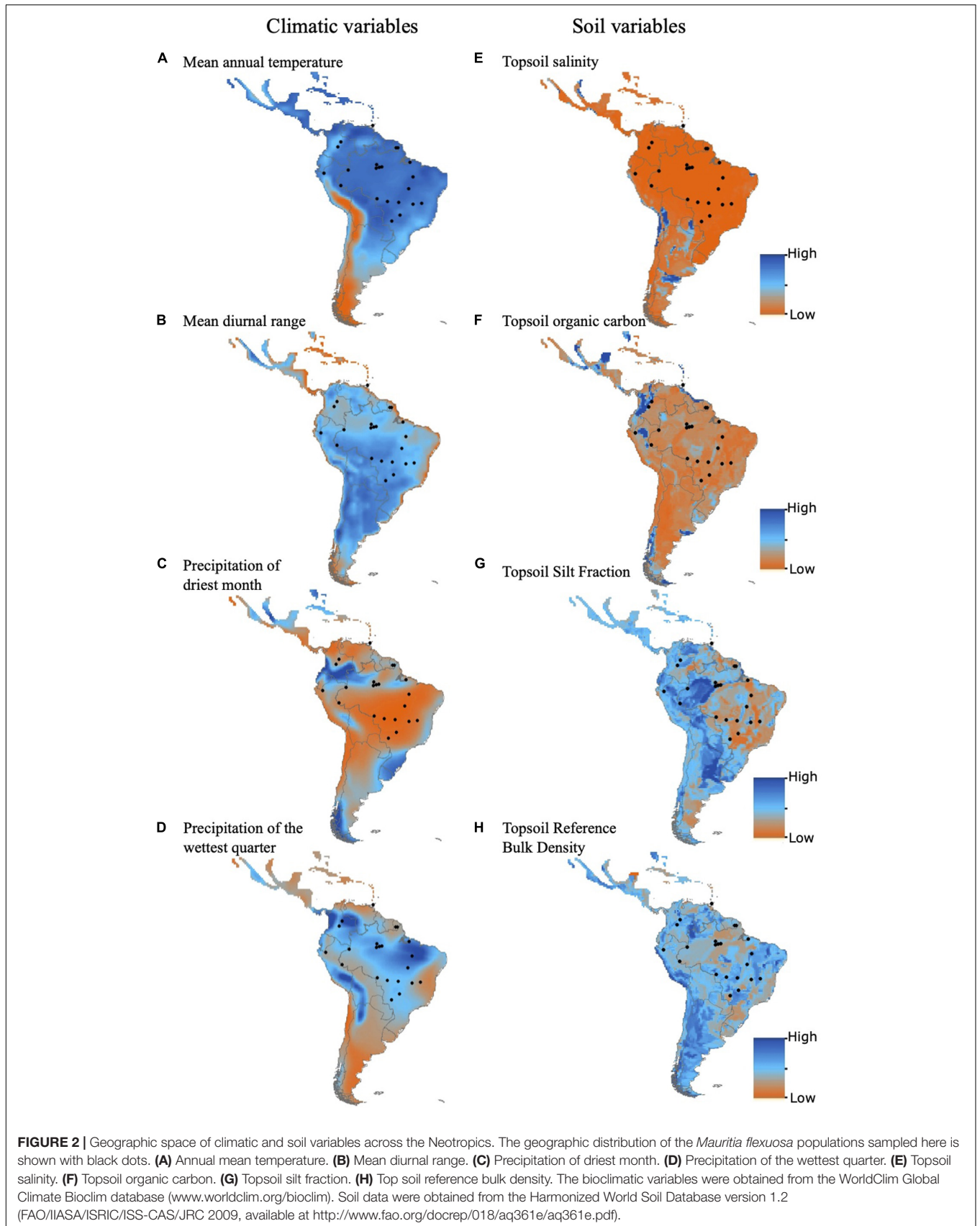
Functional annotation indicated that the 45 outlier SNPs identified by OutFLANK are located in nine genes (**Supplementary Table S8 in Supplementary Appendix 1**). These genes are related to diverse biological processes, including housekeeping genes such as ribonuclease, ATP synthetase, Rab GDP dissociation inhibitors, and the GTPase family.

The 17 loci correlated to climatic variables are located in 16 genes (**Supplementary Table S9 in Supplementary Appendix 1**, **Supplementary Table S6 in Supplementary Appendix 3**), and the 24 loci correlated to soil variables affected 22 genes (**Supplementary Table S10 in Supplementary Appendix 1**, **Supplementary Table S7 in Supplementary Appendix 3**). The 38 genes with putative signals of adaptive selection are involved in a wide spectrum of biological processes such as ATP synthesis, DNA replication, mitotic checkpoint, and transcription factors (elongator acetyltransferase complex-ELP). Additionally, we found SNPs correlated to soil variables potentially affecting magnesium (Mg) transport (Mg transporter *mrs2-4*-related gene) and genes related to NAD(P) activity (**Supplementary Tables S9, S10 in Supplementary Appendix 1**).

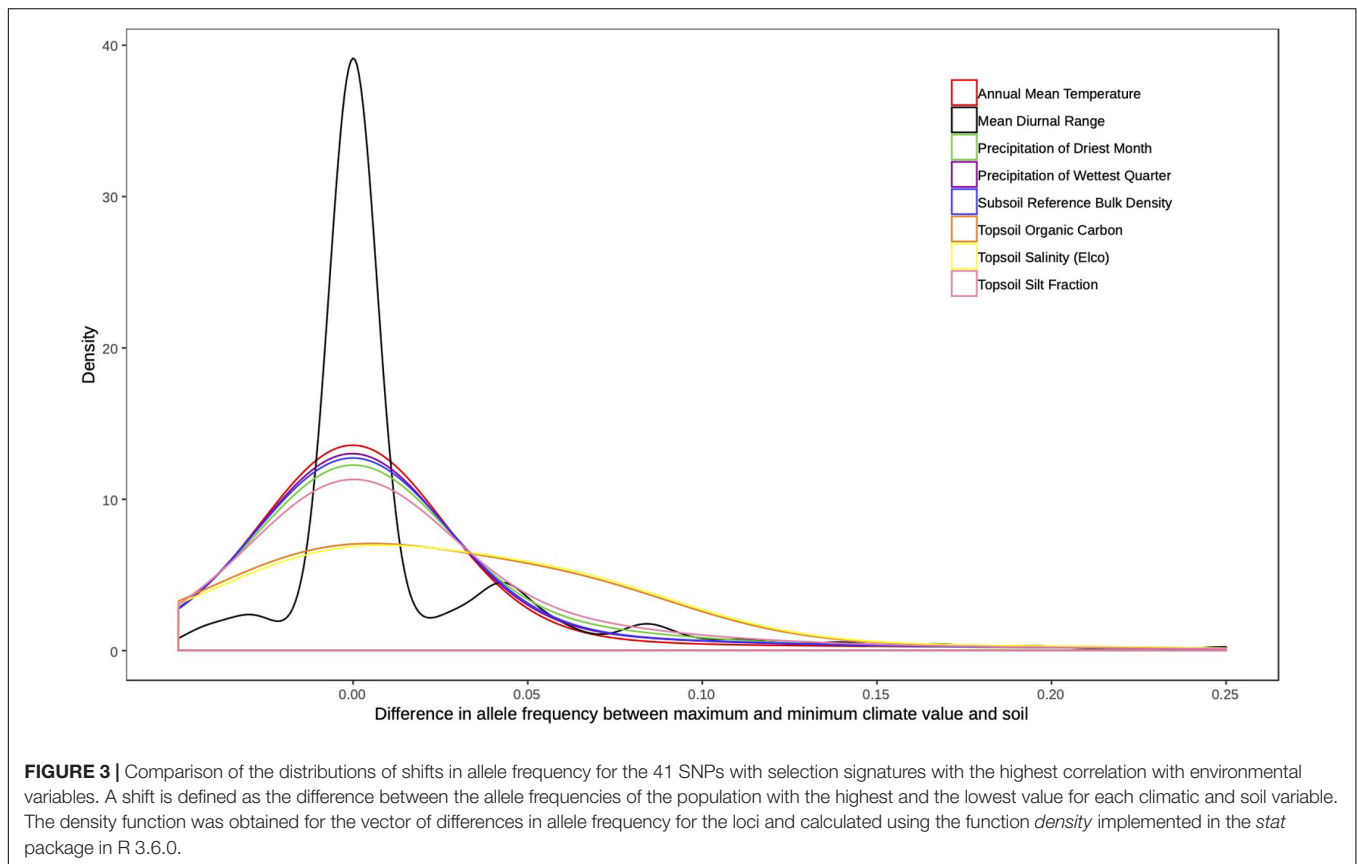
SweepFinder2 analysis detected 28 SNP loci with selective sweeps signal located within 25 genes (**Supplementary Table S12 in Supplementary Appendix 1**). As expected, the list of genes is similar to and nested within those obtained with the Bayenv2 analyses. These include genes involved in cellular metabolism and DNA and RNA synthesis. We also found genes related to UVA radiation resistance, to ion transport through membrane, and Mg transporters.

## DISCUSSION

Sampling populations from across rainforest and savanna sites, our results show evidence of adaptation in *M. flexuosa* in several SNP loci. The OutFLANK approach identified 45 outlier loci with high  $F_{ST}$  values. Correlations of allele frequencies and climatic and soil variables indicated 45 loci putatively affecting adaptation. The test of recent selective sweeps provided evidence of more





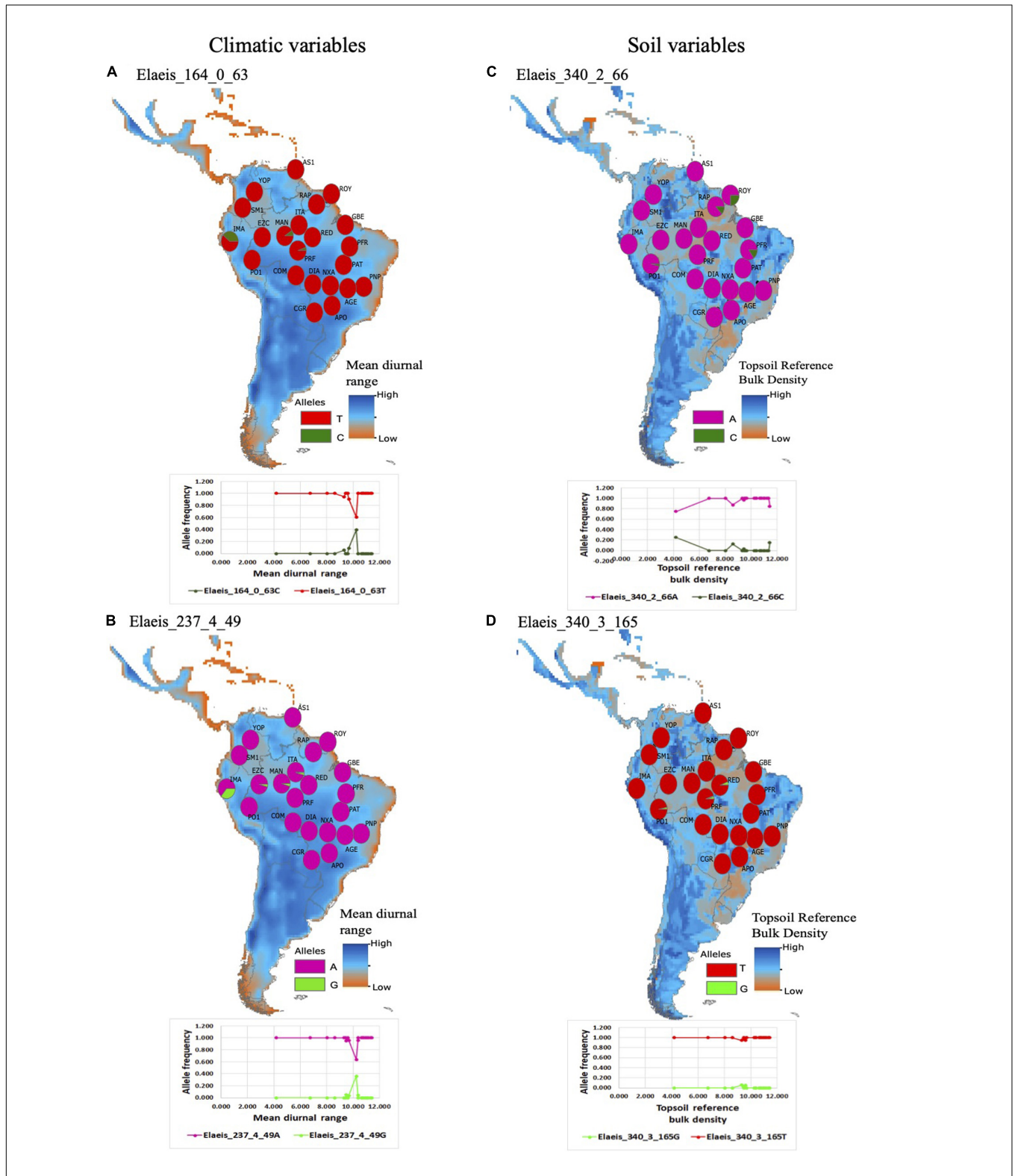


than 28 loci under natural selection, indicating evidence for selection as a driver of the evolutionary success of *M. flexuosa*, but did not support local adaptation.

Despite the climatic and soil differences between Amazonia, the Llanos, and the Cerrado, we found only slight differences in allele frequency at SNPs with selection footprints based on Bayenv 2 analysis (Figure 3), and most populations showed fixation of the same allele with selection signal (Figure 4, Supplementary Figures S12, 13 in Supplementary Appendix 2). The small shifts in allele frequency distributions among contrasting environments for most loci suggest selective sweeps of a single allele favored across most populations. We interpret these results as evidence for adaptation to the relatively stable microhabitat of *M. flexuosa*. Although the species occurs across highly variable macroclimatic conditions, from humid Amazonia rainforest to seasonally dry Cerrado savanna, the species is strictly associated to watercourses and swamps, named “veredas” in Cerrado, “moriches” in the Llanos in Colombia and Venezuela or “aguajales” in the Amazonian areas in Peru. In fact, the species is dependent on water and has physiological and morphological adaptations to germinate in swampy environments (Silva et al., 2014). Taken together, we suggest that the lack of or small differentiation in allele frequencies among populations from different ecosystems is due to the high specialization of *M. flexuosa* to swampy microhabitats, where selective sweeps maintain the same favorable alleles, despite the strong macroclimatic differences.

Bayesian clustering (Figure 1) showed differentiation among populations from different ecosystems in putatively neutral SNPs, but with some admixture between Amazonia and Cerrado, especially in populations from Araguaia/Tocantins, North Atlantic, and Caribbean river basins. In addition, AMOVA evidenced significant, but low differentiation among populations from different ecosystems, and isolation-by-distance is an important driver of genetic differentiation. Differentiation caused by genetic drift not only supports our hypothesis of homogenization among ecosystems due to selective sweeps, but also raises the hypothesis of migration spreading favorable alleles among populations leading to allele fixation (Yeaman and Whitlock, 2011; Savolainen et al., 2013).

Resilience to climate change can ensure evolutionary success (e.g., Whitlock and Barton, 1997). Although highly specialized to swampy microhabitats, *M. flexuosa* is resilient in the face of climatic change ensuring its evolutionary capacity. Suitable habitats for *M. flexuosa* may have been available during the Quaternary glacial cycling, mainly in the Amazon basin, and likely served as sources of migrants to colonize more unstable areas in Central Brazil during interglacial phases (de Lima et al., 2014; Melo et al., 2018). The demographic dynamics may have spread favorable mutations throughout populations leading to allele sharing of adaptive loci among populations. Similar patterns in allele frequency at adaptive loci were found in other widespread Neotropical tree species (Collevatti et al., 2019).



**FIGURE 4 |** Geographic distribution of allele frequencies in climatic and soil geographic space (maps) for SNPs with under adaptive selection evidenced by Bayenv2 analysis and relationship between allele frequency and climatic and soil variables (plots below the maps). **(A)** Allele frequency for the SNP loci affecting the Mon1 trafficking protein gene correlated to mean diurnal range. **(B)** Allele frequency for the SNP loci affecting the Rab protein gene correlated to mean diurnal range. **(C)** Allele frequency for the SNP loci affecting the magnesium transporter MRS2-4 protein gene correlated to topsoil reference bulk density. **(D)** Allele frequency for the SNP loci affecting the magnesium transporter MRS2-4 protein gene correlated to topsoil reference bulk density.

Moreover, despite the macroclimatic differences among populations from different ecosystems, there is a lack of strong latitudinal cline variation in climatic variables in the Neotropics (Figure 2), which may preclude the detection of adaptive signals in many loci, even in widespread species such as *M. flexuosa*. In fact, we identified adaptive selection correlated with mean annual temperature and mean diurnal range in temperature. Despite variation among populations, these variables show no clear geographic trend (Figure 2). The lack of strong latitudinal clines may explain the contrasting results compared to temperate species, for which adaptive selection has been reported for traits with strong latitudinal clines, such as photoperiod response and temperature response-related traits (e.g., Parchman et al., 2012; Yeaman et al., 2014; Hornoy et al., 2015). Top soil organic carbon and salinity have small differences among populations, and although top soil silt fraction and bulk density show larger differences among populations, we identified a patchy distribution among ecosystems, which may further mask the detection of local adaptive selection in traits related to soil adaptation. Nevertheless, all loci with adaptive signal are correlated with these soil variables. In addition, because most adaptive traits have quantitative inheritance (Fisher, 1930; Wright, 1935), the response of *M. flexuosa* to environmental variation may be polygenic, and thereby selection may cause only small changes in allele frequencies, causing weak selection signature on a single locus (Pritchard and Di Rienzo, 2010; Le Corre and Kremer, 2012).

## Selective Sweeps and Adaptation to Environmental Stress

Genes potentially affected by SNPs with selection signal have fundamental roles in cells and are usually highly conserved during evolution. Our analysis was based on functional annotation; thus, the real gene function must be confirmed using experimental assay such as reverse transcriptase-polymerase chain reaction to confirm gene expression. For instance, some genes highlighted in our analysis code for transduction membrane proteins (Wang et al., 2003) or are members of the ATP-binding cassette (ABC) or enzymes with catalytic activity. Adenosine 5' triphosphate (ATP) and NADPH play a central role in metabolism, diffusing throughout the cell to the sites where energy is used for catalytic activities (Bonora et al., 2012). We also found adaptive signal for the proliferating cell nuclear antigen-like complex, which plays important roles in DNA replication and repair, including translesion synthesis, error-free damage bypass, break-induced replication, mismatch repair, and chromatin assembly (Boehm et al., 2016).

We identified genes under adaptive selection related to plant growth and adaptation to stress. The endomembrane trafficking protein MONENSIN SENSITIVITY1 (MonI), which is an important requirement for protein trafficking in cells and plant growth in *Arabidopsis thaliana* (Cui et al., 2014), was significantly correlated to variation in mean temperature diurnal range (Figure 4). The Mg transmembrane transport MRS2-4-related gene, essential for Mg homeostasis and the ability to adapt to a wide range of environmental conditions

in *A. thaliana* (Oda et al., 2016), was correlated to top soil bulk density (Figure 4). This correlation indicates that *M. flexuosa* is adapted to variable and potentially stressful environments, enabling it to persist in adverse conditions.

We also found genes related to drought stress, such as the Rab scort protein geranyltransferase component, which is involved in abscisic acid and auxin signaling in *A. thaliana* (Johnson et al., 2005) and was correlated to mean temperature diurnal range (Figure 4). Interestingly, we found correlations with bioclimatic variables related to temperature, but not with precipitation, most likely because water via swampy habitat preference is required for *M. flexuosa* establishment. Thus, temperature variation among populations may impose higher constraints on the success of *M. flexuosa*.

It is important to note that populations from different ecosystems displayed the same fixed allele for these loci (Figure 4), indicating the role of selective sweeps shaping adaptation and spatial patterns in genetic diversity in *M. flexuosa*. On the other hand, we found weak evidence of local adaptation in different ecosystems, as evidence by few loci with exclusive alleles in just one ecosystem (Supplementary Figures S11, 12 in Supplementary Appendix 2). However, these loci potentially affect genes related to primary metabolism and are likely not directly related to local adaptation to climatic or soil conditions. Instead, they are most likely in LD with genes under selection.

Considering the number of loci scrutinized (16,262) and the transcriptome size of *M. flexuosa* (46,600 genes, Sarah et al., 2017), the number of loci detected with positive selection signal was low. It is important to note that the probe set did not cover all genes in *M. flexuosa* genome; thus, the number of loci that can be potentially detected is much lower than that in the transcriptome. It should be noted that the available methods to detect adaptive selection are based on the assumption that positive selection on a mutation leads to hard selective sweeps and therefore have little power to detect other types of selection such as background and balancing selection (Nielsen, 2005; Przeworski et al., 2005; Messer and Petrov, 2013). Thus, it is possible that our methodology allowed detection of only a fraction of loci under selection.

## Low Neutral Differentiation Among Ecosystems

The putatively neutral SNPs showed low but significant genetic differentiation among populations across ( $F_{CT} = 0.063$ ) and within ecosystems ( $F_{SC} = 0.064$ ). In addition, the Bayesian clustering showed some admixture among populations, especially from Amazonia and Cerrado. These results contrast with the higher differentiation for microsatellite loci ( $F_{CT} = 0.095$ ;  $F_{SC} = 0.137$ ), most likely due to higher mutation rates at microsatellite loci, leading to higher heterozygosity and differentiation. While SNPs are biallelic and evolve under an infinite sites model, microsatellites are multiallelic loci and evolve under stepwise and two-phase mutation models (Valdes et al., 1993; Di Rienzo et al., 1994). However, our findings showed similar patterns of spatial distribution of genetic clusters based on SNPs and microsatellites (Melo et al., 2018), with low admixture among ecosystems, suggesting an important role for genetic

drift in shaping the spatial distribution of genetic diversity in *M. flexuosa*. Further, microsatellites, chloroplast sequences, and putatively neutral SNPs recovered a demographic scenario of multiple refugia present during the last glacial maximum (LGM, 26–19 kya; de Lima et al., 2014; Melo et al., 2018), reinforcing the hypothesis of the role of Quaternary climate change influencing the distribution of genetic diversity in *M. flexuosa*. Given the high fragmentation across current landscapes, it is likely that the genetic structure of *M. flexuosa* between Amazonia and Cerrado will remain in the future because of low gene flow among populations.

We also found higher frequency of mating between unrelated individuals than expected at random (negative inbreeding coefficient). It is important to note that *M. flexuosa* is a dioecious species, which therefore requires outcrossing. However, the high frequency of heterozygotes may also be the outcome of genotyping errors, leading to an overestimation of heterozygous genotypes (Chen et al., 2017). We believe this is unlikely here because we used standard procedures for SNP filtering (see Silva-Junior et al., 2018) with a minimum depth of aligned reads of three (median 26). Coverage depth of 1–2x has been found sufficient for accurate allele calls (Sims et al., 2014).

## CONCLUSION

Using target sequence capture SNP genotyping to generate data for over 16,000 SNPs, we unravel the evolutionary history of a hyperdominant palm species. Our findings suggest that selection together with migration led to the spread of alleles with higher fitness among populations from different ecosystems and promoted the evolutionary success of the widespread, highly abundant Neotropical palm *M. flexuosa* in swampy microhabitats.

## DATA AVAILABILITY STATEMENT

Data and additional supporting information may be found in the online version of this article as supporting information. DNA

## REFERENCES

- Alaba, O. A., Bredeson, J. V., Egesi, C. N., Esuma, W., Ezenwaka, L., Ferguson, M. E., et al. (2015). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* crantz) from 10 populations. *G3 Genes Genomes Genet.* 5, 133–144. doi: 10.1534/g3.114.015008
- Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., et al. (2019). A guide to carrying out a phylogenomic target sequence capture project. *PeerJ* 7:e27968v1. doi: 10.7287/peerj.preprints.27968v1
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Baker, W. J., and Dransfield, J. (2016). Beyond *Genera palmarum*: progress and prospects in palm systematics. *Bot. J. Linn. Soc.* 182, 207–233. doi: 10.1111/boj.12401
- Boehm, E. M., Gildenberg, M. S., and Washington, M. T. (2016). The many roles of PCNA in Eukaryotic DNA replication. *Enzymes* 39, 231–254. doi: 10.1016/b.enz.2016.03.003

datasets are available in GeneBank database (access number: BioProject ID PRJNA588981).

## AUTHOR CONTRIBUTIONS

RC and CB conceived the experiment. RC and CB funded the work. WM conducted the experiment. WM, LV, EN, and RC performed the analysis including assembling, annotation, and curated the data. RC wrote the first draft of the manuscript. All co-authors contributed to the final version of the manuscript.

## FUNDING

Field work and genetic data acquisition were supported by competitive grants from CNPq to RC and CB (PVE/CNPq project no. 407310/2013-4) and Biodiversity and Ecosystem Services in a Changing Climate (BECC) Strategic Research Area at the University of Gothenburg, to RC (Rede Cerrado CNPq/PPBio project no. 457406/2012-7), and CAPES/PROCAD (project no. 88881.068425/2014-01). WM was supported by a Ph.D. FAPEG fellowship and LV received a CAPES fellowship. RC has been supported by productivity grants from CNPq, which we gratefully acknowledge.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.596662/full#supplementary-material>

**Supplementary Appendix 1 | Supplementary Tables S1–S12** with sampling locations, and Bayesian clustering and genome scan results.

**Supplementary Appendix 2 | Supplementary Figures S1–S13** with details on Bayesian clustering and selection analyses.

**Supplementary Appendix 3 | Supplementary Tables S1–S7** with SNP density, nucleotide diversity, OutFLANK result, and gene ontology annotation.

- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bonora, M., Patergnani, S., Rimessi, A., de Marchi, E., Suski, J. M., Bononi, A., et al. (2012). ATP synthesis and storage. *Purinergic Signal.* 8, 343–357. doi: 10.1007/s11302-012-9305-8
- Brousseau, L., Fine, P. V. A., Dreyer, E., Vendramin, G. G., and Scotti, I. (2020). Genomic and phenotypic divergence unveil microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata* Aubl. *Mol. Ecol.* doi: 10.1111/mec.15595 [Epub ahead of print].
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8
- Chen, B., Cole, J. W., and Grond-Ginsbach, C. (2017). Departure from Hardy Weinberg equilibrium and genotyping Error. *Front. Genet.* 8:167. doi: 10.3389/fgene.2017.00167
- Collevatti, R. G., Novaes, E., Silva-Junior, O. B., Vieira, L. D., Lima-Ribeiro, M. S., and 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
- Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185, 1411–1423. doi: 10.1534/genetics.110.114819
- Cornuet, J. M., Pudlo, P., Veysier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30, 1187–1189. doi: 10.1093/bioinformatics/btt763
- Cornuet, J. M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J. M., Balding, D. J., et al. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24, 2713–2719. doi: 10.1093/bioinformatics/btn514
- Cui, Y., Zhao, Q., Gao, C., Ding, Y., Zeng, Y., Ueda, T., et al. (2014). Activation of the Rab7 GTPase by the MON1-CCZ1 complex is essential for PVC-tubule trafficking and plant growth in *Arabidopsis*. *Plant Cell* 26, 2080–2097. doi: 10.1105/tpc.114.123141
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- de Lima, N. E., Lima-Ribeiro, M. S., Tinoco, C. F., Terribile, L. C., and Collevatti, R. G. (2014). Phylogeography and ecological niche modelling, coupled with the fossil pollen record, unravel the demographic history of a Neotropical swamp palm through the quaternary. *J. Biogeogr.* 41, 673–686. doi: 10.1111/jbi.12269
- Degiorio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., and Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics* 32, 1895–1897. doi: 10.1093/bioinformatics/btw051
- Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M., and Freimer, N. B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3166–3170. doi: 10.1073/pnas.91.8.3166
- Eisnerhardt, W. L., Svenning, J. C., Baker, W. J., Couvreur, T. L. P., and Balslev, H. (2013). Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents. *Sci. Rep.* 3:1164. doi: 10.1038/srep01164
- Evans, L. M., Slavov, G. T., Rodgers-Melnick, E., Martin, J., Ranjan, P., Muchero, W., et al. (2014). Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat. Genet.* 46, 1089–1096. doi: 10.1038/ng.3075
- Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Fisher, R. A. S. (1930). *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.
- Freitas, C. G., Bacon, C. D., Souza-Neto, A. C., and Collevatti, R. G. (2019). Adjacency and area explain species bioregional shifts in neotropical palms. *Front. Plant Sci.* 10:55. doi: 10.3389/fpls.2019.00055
- Fu, Y., Springer, N. M., Gerhardt, D. J., Ying, K., Yeh, C. T., Wu, W., et al. (2010). Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.* 62, 898–909. doi: 10.1111/j.1365-313X.2010.04196.x
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1207.3907> (accessed June 23, 2019).
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Gratani, L. (2014). Plant phenotypic plasticity in response to environmental factors. *Adv. Bot.* 2014, 1–17. doi: 10.1155/2014/208747
- Günther, T., and Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics* 195, 205–220. doi: 10.1534/genetics.113.152462
- Heyduk, K., Trapnell, D. W., Barrett, C. F., and Leebens-Mack, J. (2016). Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol. J. Linn. Soc.* 117, 106–120. doi: 10.1111/bij.12551
- Hornoy, B., Pavy, N., Gérardi, S., Beaulieu, J., and Bousquet, J. (2015). Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. *Genome Biol. Evol.* 7, 3269–3285. doi: 10.1093/gbe/evv218
- Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18, 337–338. doi: 10.1093/bioinformatics/18.2.337
- Johnson, C. D., Chary, S. N., Chernoff, E. A., Zeng, Q., Running, M. P., and Crowell, D. N. (2005). Protein geranylgeranyltransferase I is involved in specific aspects of abscisic acid and auxin signaling in *Arabidopsis*. *Plant Physiol.* 139, 722–733. doi: 10.1104/pp.105.065045
- Jones, M. R., and Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* 25, 185–202. doi: 10.1111/mec.13304
- Jordan, R., Hoffmann, A. A., Dillon, S. K., and Prober, S. M. (2017). Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: implications for adaptive potential to projected climate change. *Mol. Ecol.* 26, 6002–6020. doi: 10.1111/mec.14341
- Kaplan, N. L., Hudson, R. R., and Langley, C. H. (1989). The “hitchhiking effect” revisited. *Genetics* 123, 887–889.
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kawecki, T. J., and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecol. Lett.* 7, 1225–1241. doi: 10.1111/j.1461-0248.2004.00684.x
- Khorsand, R. R., and Koptur, S. (2013). New findings on the pollination biology of *Mauritia flexuosa* (Arecaceae) in Roraima, Brazil: linking dioecy, wind, and habitat. *Am. J. Bot.* 100, 613–621. doi: 10.3732/ajb.1200446
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi: 10.1111/1755-0998.12387
- Lanes, É. C., Pope, N. S., Alves, R., Carvalho Filho, N. M., Giannini, T. C., Giulietti, A. M., et al. (2018). Landscape genomic conservation assessment of a narrow-endemic and a widespread morning glory from *Amazonian savannas*. *Front. Plant Sci.* 9:532. doi: 10.3389/fpls.2018.00532
- Le Corre, V., and Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.* 21, 1548–1566. doi: 10.1111/j.1365-294x.2012
- Lehner, B., and Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydro. Process.* 27, 2171–2186. doi: 10.1002/hyp.9740
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, Y. L., and Liu, J. X. (2018). StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* 18, 176–177. doi: 10.1111/1755-0998.12719
- Loiseau, O., Olivares, I., Paris, M., de La Harpe, M., Weigand, A., Koubínová, D., et al. (2019). Targeted capture of hundreds of nuclear genes unravels phylogenetic relationships of the diverse neotropical palm tribe Geonomateae. *Front. Plant Sci.* 10:864. doi: 10.3389/fpls.2019.00864
- McKay, J. K., and Latta, R. G. (2002). Adaptive population divergence: markers, QTL and traits. *Trends Ecol. Evol.* 17, 285–291. doi: 10.1016/s0169-5347(02)02478-3
- Melo, W. A., Freitas, C. G., Bacon, C. D., and Collevatti, R. G. (2018). The road to evolutionary success: insights from the demographic history of an Amazonian palm. *Heredity* 121, 183–195. doi: 10.1038/s41437-018-0074-1
- Messer, P. W., and Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* 28, 659–669. doi: 10.1016/j.tree.2013.08.003
- Neves, L. G., Davis, J. M., Barbazuk, W. B., and Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.* 75, 146–156. doi: 10.1111/tpj.12193
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218. doi: 10.1146/annurev.genet.39.073003.112420
- Oda, K., Kamiya, T., Shikanai, Y., Shigenobu, S., Yamaguchi, K., and Fujiwara, T. (2016). The *Arabidopsis* Mg transporter, MRS2-4, is essential for Mg homeostasis under both low and high Mg conditions. *Plant Cell Physiol.* 57, 754–763. doi: 10.1093/pcp/pcv196
- Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52, 935–949. doi: 10.1111/j.1558-5646.1998.tb01823.x

- Otto, S. P. (2000). Detecting the form of selection from DNA sequence data. *Trends Genet.* 16, 526–529. doi: 10.1016/S0168-9525(00)02141-7
- Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genomics* 2:e000056. doi: 10.1099/mgen.0.000056
- Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., and Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* 21, 2991–3005. doi: 10.1111/j.1365-294X.2012.05513.x
- Pritchard, J. K., and Di Rienzo, A. (2010). Adaptation - Not by sweeps alone. *Nat. Rev. Genet.* 11, 665–667. doi: 10.1038/nrg2880
- Przeworski, M., Coop, G., and Wall, J. D. (2005). The signature of positive selection on standing genetic variation. *Evolution* 59:2312. doi: 10.1554/05-273.1
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* 16, 608–627. doi: 10.1111/1755-0998.12512
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). FastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350
- Rangel, T. F., Diniz-Filho, J. A. F., and Bini, L. M. (2010). SAM: a comprehensive application for spatial analysis in macroecology. *Ecography* 33, 46–50. doi: 10.1111/j.1600-0587.2009.06299.x
- Röser, M., Johnson, M. A. T., and Hanson, L. (1997). Nuclear DNA amounts in palms (Arecaceae). *Bot. Acta* 110, 79–89. doi: 10.1111/j.1438-8677.1997.tb00614.x
- Rull, V. (1998). Biogeographical and evolutionary considerations of *Mauritia* (Arecaceae), based on palynological evidence. *Rev. Palaeobot. Palynol.* 100, 109–122. doi: 10.1016/S0034-6667(97)00060-2
- Rull, V., and Montoya, E. (2014). *Mauritia flexuosa* palm swamp communities: natural or human-made? A palynological study of the Gran Sabana region (northern South America) within a neotropical context. *Quat. Sci. Rev.* 99, 17–33. doi: 10.1016/j.quascirev.2014.06.007
- Sander, N. L., Pérez-Zavala, F., Da Silva, C. J., Arruda, J. C., Pulido, M. T., Barelli, M. A. A., et al. (2018). Rivers shape population genetic structure in *Mauritia flexuosa* (Arecaceae). *Ecol. Evol.* 8, 6589–6598. doi: 10.1002/ece3.4142
- Sarah, G., Homa, F., Pointet, S., Contreras, S., Sabot, F., Nabholz, B., et al. (2017). A large set of 26 new reference transcriptomes dedicated to comparative population genomics in crops and wild relatives. *Mol. Ecol. Resour.* 17, 565–580. doi: 10.1111/1755-0998.12587
- Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14, 807–820. doi: 10.1038/nrg3522
- Silva, R. S., Ribeiro, L. M., Mercadante-Simões, M. O., Nunes, Y. R. F., and Lopes, P. S. N. (2014). Seed structure and germination in buriti (*Mauritia flexuosa*), the Swamp palm. *Flora Morphol. Distrib. Funct. Ecol. Plants* 209, 674–685. doi: 10.1016/j.flora.2014.08.012
- Silva-Junior, O. B., and Grattapaglia, D. (2015). Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytol.* 208, 830–845. doi: 10.1111/nph.13505
- Silva-Junior, O. B., Grattapaglia, D., Novaes, E., and Collevatti, R. G. (2018). Design and evaluation of a sequence capture system for genome-wide SNP genotyping in highly heterozygous plant genomes: a case study with a keystone Neotropical hardwood tree genome. *DNA Res.* 25, 535–545. doi: 10.1093/dnares/dsy023
- Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. doi: 10.1038/nrg3642
- Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35. doi: 10.1017/S0016672300014634
- Syring, J. V., Tennessen, J. A., Jennings, T. N., Wegrzyn, J., Scelfo-Dalbey, C., and Cronn, R. (2016). Targeted capture sequencing in whitebark pine reveals range-wide demographic and adaptive patterns despite challenges of a large, repetitive genome. *Front. Plant Sci.* 7:484. doi: 10.3389/fpls.2016.00484
- Ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., et al. (2013). Hyperdominance in the Amazonian tree flora. *Science* 342, 1243092–1243091. doi: 10.1126/science.1243092
- Tiffin, P., and Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29, 673–680. doi: 10.1016/j.tree.2014.10.004
- Valdes, A. M., Slatkin, M., and Freimer, N. B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* 133, 737–749.
- Vitti, J. J., Grossman, S. R., and Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120. doi: 10.1146/annurev-genet-111212-133526
- Wang, C. W., Stromhaug, P. E., Kauffman, E. J., Weisman, L. S., and Klionsky, D. J. (2003). Yeast homotypic vacuole fusion requires the Ccz1-Mon1 complex during the tethering/docking stage. *J. Cell Biol.* 163, 973–985. doi: 10.1083/jcb.200308071
- Whitlock, M. C., and Barton, N. H. (1997). The effective size of subdivided population. *Genetics* 146, 427–441.
- Whitlock, M. C., and Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. *Am. Nat.* 186, S24–S36. doi: 10.1086/682949
- Wright, S. (1931). Evolution in Mendelian Populations. *Genetics* 16, 97–159.
- Wright, S. (1935). Evolution in populations in approximate equilibrium. *J. Genet.* 30, 257–266. doi: 10.1007/BF02982240
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., et al. (2016). Convergent local adaptation to climate in distantly related conifers. *Science* 353, 1431–1433. doi: 10.1126/science.aaf7812
- Yeaman, S., Hodgins, K. A., Suren, H., Nurkowski, K. A., Rieseberg, L. H., Holliday, J. A., et al. (2014). Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca* × *Picea engelmannii*). *New Phytol.* 203, 578–591. doi: 10.1111/nph.12819
- Yeaman, S., and Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution* 65, 1897–1911. doi: 10.1111/j.1558-5646.2011.01269.x
- Zhou, L., and Holliday, J. A. (2012). Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics* 13:703. doi: 10.1186/1471-2164-13-703

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Melo, Vieira, Novaes, Bacon and Collevatti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.