

Original Article

Sympatry and evolution: relationships between the endemic *Myrcia guarujana* and the widely distributed *M. spectabilis*

Simpatria e evolução: relações entre a endêmica *Myrcia guarujana* e a amplamente distribuída *M. spectabilis*

M. A. G. Magenta^{a*} , L. F. Almeida^a , L. M. Rocha^a , C. E. A. Batista^b  and R. H. G. Priolli^{b,c} 

^aUniversidade Santa Cecília – UNISANTA, Laboratório de Pesquisas em Ecologia Costeira, Santos, SP, Brasil

^bUniversidade de São Paulo – USP, Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ, Departamento de Genética, Piracicaba, SP, Brasil

^cSanta Semente Ltda, Piracicaba, SP, Brasil

Abstract

Sympatry and evolution in *Myrcia* - *Myrcia spectabilis* DC. has a wide distribution in the south and southeast Brazilian regions. *M. guarujana* Sobral, Magenta & Caliari is known from the few collections in submontane ombrophilous forest and a transitional site between restinga and ombrophilous forest in São Paulo state. These species are morphologically similar and occur in sympatry, which prompted us to investigate the morphological aspects, sequence the genomes and consider the effects of such sympatry. In a fragment of Brazilian Atlantic rainforest, we collected leaves from individuals of *M. guarujana* and *M. spectabilis*. Morphological analysis and DNA extraction were performed for sequencing and construction of the GBS library. A total of 3,429 loci containing neutral SNP markers were identified, and 446 represented outlier loci. The fixation index (F_{ST}) had a high value (0.735; $p < 0.05$). By AMOVA, the highest variance percentage found was 73.54% ($P = 0.0079$) between the species. Principal component analysis presented two clusters referring to the species, with the MS1 (*M. spectabilis*- 1) sample isolated from both. On the *neighbour-joining* dendrogram, both species presented strongly sustained branches (100 and 99, respectively), and MS1 emerged as an intermediate. Among the outlier loci, we have LNG1, which promotes foliar elongation, and MYB82, a transcription activator functionally equivalent to GL1, in which high levels of expression result in decreased trichome abundance. Studies with more representative samples, together with the discovery of the SNPs found here, will constitute a powerful tool to understand the evolution, diversification, and conservation of these two species.

Keywords: GBS library, hybridization, Myrtaceae, SNP markers, co-occurrence.

Resumo

Simpatria e evolução em *Myrcia* - *Myrcia spectabilis* DC. tem ampla distribuição nas regiões sul e sudeste do Brasil. *M. guarujana* Sobral, Magenta & Caliari é conhecida das poucas coletas em floresta ombrófila submontana e em um local de transição entre restinga e floresta ombrófila do estado de São Paulo. Essas espécies são morfologicamente semelhantes e ocorrem em simpatria, o que nos levou a investigar os aspectos morfológicos, sequenciar os genomas e considerar os efeitos dessa simpatria. Em um fragmento de floresta atlântica brasileira, coletamos folhas de indivíduos de *M. guarujana* e *M. spectabilis*. Análise morfológica e extração de DNA foram realizadas para sequenciamento e construção da biblioteca GBS. Um total de 3.429 loci contendo marcadores SNP neutros foram identificados e 446 representaram loci outliers. O índice de fixação (F_{ST}) teve um valor alto (0,735; $P < 0,05$). Por AMOVA, o maior percentual de variância entre as espécies encontrado foi de 73,54% ($P = 0,0079$). A análise de PCA apresentou dois clusters referentes às espécies, com a amostra MS1 (*M. spectabilis* 1) isolada de ambos. No dendrograma *neighbor-joining*, ambas as espécies apresentaram ramificações fortemente sustentadas (100 e 99, respectivamente) e MS1 emergiu como um táxon intermediário. Entre os loci outliers, temos LNG1, que promove alongamento foliar, e MYB82, um ativador de transcrição funcionalmente equivalente a GL1, no qual altos níveis de expressão resultam em diminuição da abundância de tricomas. Estudos com amostras mais representativas, juntamente com a descoberta dos SNPs encontrados aqui, constituirão uma ferramenta poderosa para entender a evolução, diversificação e conservação dessas duas espécies.

Palavras-chave: biblioteca GBS, hibridização, Myrtaceae, marcadores de SNP, co-ocorrência.

*e-mail: maramagenta@unisanta.br

Received: February 28, 2025 – Accepted: May 25, 2025

Editor: Jairo Lizandro Schmitt



This is an Open Access article distributed under the terms of the Creative Commons Attribution license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Myrcia DC. (Myrtaceae) is a genus consisting of trees, shrubs, lianas, and subshrubs occurring in all Brazilian phytogeographic domains, where is represented by 409 species (Santos et al., 2025). The genus has economic potential, either for feeding or medicinal applications (Santos et al., 2021). Among the tree species endemic to the Brazilian Atlantic rainforest domain are *Myrcia spectabilis* DC and *Myrcia guarujana* Sobral, Magenta & Caliari. *Myrcia spectabilis* is widely distributed and found in the Southeast and South regions of Brazil in ombrophilous and restinga forest (Santos et al., 2025); *M. guarujana* is known from the few specimens carried out in the state of São Paulo, in a submontane dense ombrophilous forest site in the municipality of Guarujá, SP [Magenta 883 (HUSC11287)], and also in a transition site between restinga and ombrophilous forest in Itanhaém, SP [Barretto, E. H. P. 446 (PMSP15917); 1360 (PMSP16791); 1361 (PMSP16792)]. During four years of searches in Guarujá, only four individuals of *M. guarujana* were found (all in the same area), occurring in sympatry with *M. spectabilis*. In general, the species have morphological similarities, such as habit, leaf blade shape (generally lanceolate), and inflorescence type (paniculiform).

Myrtaceae is a plant family known for its taxonomic complexity and the difficulty in determining its phylogenetic relationships due to the high similarity between its representatives, which is a challenge for morphologic interpretation (Barroso and Peixoto, 1995). High taxes of homoplasy are manifested in the family representatives, probably due to the high capacity of rapid speciation (Caliari, 2013). According to Amorim et al. (2019), the morphological singularities of each taxon should be used in the systematic definition of the groups; however, molecular characteristics should be considered together for taxonomic analysis.

Advances in genetic studies have improved genome construction quickly and affordably using next-generation sequencing (NGS) methods, such as genotyping by sequencing (GBS) libraries, which decreases the complexity of large genomes using restriction enzymes to split the DNA into certain areas (Poland et al., 2012). This procedure identifies the genetic distance between individuals or populations, regardless of the phenotype or developmental stages, by genetic variation analysis, which uses molecular markers to analyse single-nucleotide polymorphisms (SNPs) (Alipour et al., 2017).

Considering such aspects and possibilities, our study built two genomic libraries of *Myrcia spectabilis* and *M. guarujana* and, through the genomic diversity obtained, presented evidence of their taxonomic status, as well as the risk of such a relationship presented for these species, adding information about some morphological character-coding genes.

2. Methods

2.1. Plant material

The samples were collected in a submontane dense ombrophilous forest next to the SP 061 highway

(23°53'19.87"S, 46°10'12.60"O) at the Serra do Guararu Ecologic Park in the municipality of Guarujá (SP, Brazil). Leaves were obtained from only four individuals identified as *Myrcia guarujana* - the only ones found during many years of searching in the area and nearby, and from eight sympatric individuals identified as *M. spectabilis*. All samples were marked for recording. All plant material was packed in numbered plastic bags and transported in a thermal box to the Coastal Ecology Laboratory at Santa Cecília University (Santos - Brazil), where the morphological aspects of the samples were analysed using a stereomicroscope (ZEISS SteREO Discovery V8). For molecular analysis, pieces of leaves from each sample were frozen in liquid nitrogen and lyophilized for two days under 0.040 mbar pressure and -40 °C using Thermo Scientific Super Modulyo 220 system equipment (Thermo Fisher Scientific, Waltham, USA).

2.2. DNA extraction, GBS library, and sequencing

DNA extraction was performed from 30 mg of lyophilized leaves by the CTAB protocol. Total genomic DNA was quantified using a fluorescence-based method (Qubit/Thermo Fisher) and then normalized to 10ng/ µL to produce a uniform number of tag and read sequences for each sample. Genomic libraries were constructed according to the double digestion protocol described by Poland et al. (2012) using the restriction enzymes Nsil (NEB, Ipswich, USA) and Msel (NEB). The resulting libraries were pooled at 96-plex and sequenced on the Illumina NextSeq 500 sequencing platform (Illumina, Inc., USA) in mid-output mode and set to produce 150 bp single-end reads. The quality of the obtained raw reads was assessed using FastQC software (Andrews, 2010) at the Hemocentro of Ribeirão Preto facilities (Brazil). For each species, samples were demultiplexed and the raw read sequences were filtered with the script "process_radtags.pl" in the Stacks program (version 2.41) (Catchen et al., 2011). SNP calling retained only SNPs with a minimum sequencing depth of 5X, frequency of the least common allele ≥ 0.01 and occurring in at least 90% of individuals of the genera. Population genomic analyses were performed by applying additional filtering parameters to obtain the maximum-quality SNPs: (1) individual samples with $>5\%$ missing data were excluded, and (2) SNPs with missing data in 25% of the samples or a minor allele frequency (MAF) <0.05 were removed.

2.3. Statistical analysis and outlier locus discovery

The filtered data were imported as a genind object into R and were analysed mainly by using several packages for population genetics. Genetic diversity organization was performed using the diveRsity package (Keenan et al., 2013), estimating allele number (A), exclusive alleles (Ap), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficients (F_{IS}), confidence intervals at 95% for F_{IS} and the fixation index among samples (F_{ST}).

The organization of the overall genetic diversity and the clustering of samples were assessed using the neutral set of loci with PCA and with a neighbour-joining dendrogram based on Nei's (1978) genetic distances with 1,000 bootstrap

replicates. These analyses were performed in the R platform using the poppr Kamvar et al. (2014) packages, with a 1000 bootstrap confidence interval. The poppr package was also used to conduct analysis of molecular variance (AMOVA) among the two species.

Two independent methods were applied to identify outlier loci within the species, and those identified as common to both were considered. The first package used was 'fsthet' for R, which employs the distribution F_{ST} - H_T and identifies loci with excessively high or low F_{ST} estimates in comparison to the expectations under neutrality (Flanagan and Jones, 2017). Next, the PCAdapt 3.0.4 version package for R (Luu et al., 2017) was used to estimate the correlation between SNP markers and the set of SNP loci retained on each principal component analysis (PCA). The identified sequences containing outlier SNPs were researched using the tool BLASTX in the NCBI genomic dataset with blast2go (Götz et al., 2008). For sequences with significant BLASTX hits, functional annotation was applied, associated with the coding of characterized and/or described sequences, using the Gene Ontology system (GO terms).

3. Results and Discussion

3.1. Leaf morphology analysis

Leaf analysis, both by the naked eye and using a stereomicroscope, added morphology data to pre-existing descriptions of the two species (Caliari, 2013; Sobral et al., 2016) and based a questioning about possible

hybridization, given that leaves of one of the samples of *Myrcia spectabilis* (MS1) have intermediate texture and vein calibre between them.

Figure 1 allows us to see some details of the leaves and branches of these species.

3.2. Artificial key for the studied species

- a. Young branches without bracts, hirsute; long ferrugineous trichomes on terminal branches, buds, and leaves; leaves up to 30 cm long, coriaceous, blade lanceolate to oblong, marked bullate (surface with bubble-aspect air pockets), dense trichomes in the midrib of the abaxial surface; primary and secondary veins thick, 15 to 20 pairs of secondary veins; calyx hirsute, with transversely elliptical to broadly ovate lobes; trilocular ovary *Myrcia guarujana*;
- b. Young branches with bracts, setose; short yellow-brown mealy trichomes on terminal branches, buds, and leaves; leaves up to 18 cm long, chartaceous, blade lanceolate, oblong-elliptical or rarely oblanceolate, nonbullate, puberules on the abaxial surface, trichomes concentrated mainly on the midrib, thin primary and secondary veins, 20 to 25 pairs of secondary veins; calyx setous, with oval-triangular lobes; bilocular ovary *Myrcia spectabilis*.

3.3. SNP discovery, neutrality tests, and genomic diversity

The multiplexed GBS libraries of the two *Myrcia* species were sequenced in a single lane, generating a total of 21,223,129 reads (Table 1). After quality control filtering



Figure 1. Vegetative aspects of both studied species. (A) *Myrcia guarujana* (observe marked bullate blade leaves with primary and secondary veins thick and long ferrugineous trichomes on the terminal branch and leaves); (B) *M. spectabilis* (see nonbullate blade leaves with thin primary and secondary veins and short yellow-brown trichomes on the terminal branch and leave).

Table 1. Illumina sequencing from GBS libraries of 12 individuals of *Myrcia guarujana* and *M. spectabilis*: total reads, reads retained, total SNP calling (SNP loci), total missing SNP loci (SNP missing), percent of missing SNP loci (% SNP missing) and mean of depth sequencing of each sample.

Species	ID	Total Reads	Reads Retained	SNP loci	SNP missing	% SNP miss	Mean_Depth
<i>M. guarujana</i>	MG1	4,107,615	1,862,178	3,875	58	1%	16.87
<i>M. guarujana</i>	MG2	2,083,262	1,454,523	3,875	140	4%	12.19
<i>M. guarujana</i>	MG3	1,352,654	916,397	3,875	295	8%	8.73
<i>M. guarujana</i>	MG4	1,404,571	950,859	3,875	349	9%	8.44
<i>M. spectabilis</i>	MS1	2,005,306	1,473,715	3,875	15	0	8.53
<i>M. spectabilis</i>	MS2	126,351	102,240	3,875	3,775	97%	—
<i>M. spectabilis</i>	MS3	3,495,025	2,634,877	3,875	1	0	23.62
<i>M. spectabilis</i>	MS4	22,084	17,405	—	—	—	—
<i>M. spectabilis</i>	MS5	3,001,330	2,249,493	3,875	3	0	20.19
<i>M. spectabilis</i>	MS6	81,615	64,027	—	—	—	—
<i>M. spectabilis</i>	MS7	1,879,504	1,346,082	3,875	5	0	12.95
<i>M. spectabilis</i>	MS8	1,663,812	1,248,216	3,875	4	0	11.30
<i>Total</i>		21,223,129	14,320,012				1

Samples in red letters were excluded downstream analyses.

and removal of reads containing either adapter sequences or ambiguous barcodes, the total number of retained reads was 14,320,012, with 5,183,957 for *M. guarujana* and 9,136,055 for *M. spectabilis*. After SNP calling, 3,875 SNPs were identified for four individuals of *M. guarujana* and five with morphology consistent with that of *M. spectabilis* (mean depth per sample ranging from 8.4 to 23.6X). The two species showed more transitions than transversions, and the most frequent mutations were C-T and A-G transitions.

The contrast of samples from the two different species, using 3,875 SNPs, resulted in 1,040 and 1,604 outlier SNPs with the pcadapt and fsthet methods, respectively. However, only 446 SNP outlier loci were in common for either test; therefore, we adopted the conservative view that 446 SNP outliers could be consistently identified with our dataset. Population genetic analyses were conducted using 3,429 neutral SNPs that were deemed to be neutrally evolving after outlier analyses.

Different levels of genomic diversity were observed for the two species using 3,429 loci (Table 2). The total number of alleles and private alleles was higher in *M. guarujana* (A = 6,022; AP = 3,216) than in *M. spectabilis* (A = 3,641; AP = 835), despite similar sample sizes. While *M. spectabilis* exhibited a significant excess of heterozygotes ($H_o > H_e$), resulting in a negative or null (zero) inbreeding coefficient ($F_{is} = -0.152$, 95% confidence interval [CI]: -0.028 to -0.037), *M. guarujana* showed a heterozygote deficit ($H_o < H_e$) and a significantly positive F_{is} ($F_{is} = 0.802$, 95% CI: 0.787 to 0.819), suggesting a higher level of inbreeding in this species. However, broader sampling is required to confirm the contrasting inbreeding patterns observed between the two species.

The influence of sample size on SNP-based genomic analyses remains an open question, with no universally defined standard. Optimal sample size is highly dependent

on the study's objectives, the statistical power needed to test specific hypotheses, and the biological features of the taxa under investigation (Nazareno et al., 2017; Marandé et al., 2020). Recent findings indicate that reduced sample sizes may still yield reliable insights; for instance, Scaketti et al. (2025) demonstrated that a subset of only five individuals of *Casearia sylvestris* captured the genetic structure of the full dataset (23 individuals, 1,257 loci). These results emphasize the importance of tailoring sampling strategies to the specific context of each study.

An important issue to be addressed in future studies is the *M. guarujana* populations; at the time of completion of this study, the species is considered endemic to São Paulo state. (Santos et al., 2025). This could present a worrisome situation for the species as the two currently known areas of occurrence support a limited number of individuals, which increases the risk of diversity loss (Sampson et al., 2022). Furthermore, gene flow between these two populations can be considered improbable primarily due to the geographic distance (more than 100 km), which acts as a natural barrier. This limitation is associated with the restricted flight range and foraging area of bees - the main pollinators of the genus *Myrcia* (Van Nieuwstadt and Ruano Iraheta, 1996; Freitas et al., 2023; Lughadha and Proença, 1996; Gressler et al., 2006). This may contribute to increased fragility in the population structure of this forest tree species (Song et al., 2016).

Different inbreeding coefficients were observed in the two species studied here, being very high in *M. guarujana* and zero in *M. spectabilis*. For *M. spectabilis*, similar values of F_{is} were observed in a study with two populations of *Eucalyptus caesia* Benth (Myrtaceae) where the number of individuals was also similar to the present study (6 to 11) and the values ranged from -0.046 to 0.187 (Byrne and Hopper, 2008). These comparisons indicates that the

samples of *M. spectabilis*, despite being physically close, are probably not descendants of the same matrix. In addition, the high genetic similarity observed among *M. guarujana* individuals is likely due to their shared origin from a single parent tree, as three of the four individuals were regenerants located near one adult specimen. This pattern suggests a reproductive strategy involving self-pollination or apomixis, both of which have been documented in several species within the Myrtaceae (Baskin and Baskin, 2025).

3.4. Species differentiation

The overall estimate of differentiation among the two species was high and significant ($F_{ST} = 0.735$; $P < 0.05$). AMOVA showed that most of the genomic variation was

observed between (73.54%) rather than within (26.46%) groups of *M. spectabilis* and *M. guarujana* (Table 3).

Principal component analysis showed that 94.68% of the total genetic variation was explained by the first two components, which defined clusters concordant to each species, as shown by a red cluster (*M. guarujana*) and a blue cluster (*M. spectabilis*) (Figure 2). However, one individual of *M. spectabilis* was separated by the second axis, representing 19.26% of all genetic variation and suggesting hybridization between the species.

The same clustering patterns were observed for the neighbour-joining trees using 3,429 SNP loci (Figure 3). The samples of *M. spectabilis* and *M. guarujana* presented highly supported clusters each one, and confirmed the intermediate position of the MS1 individual, which may mean that it is a hybrid.

Table 2. Genomic diversity and F_{IS} of the species *Myrcia spectabilis* (MS), *M. guarujana* (MG) based on 3,429 neutral SNPs.

Species	N	A	AP	H_o	H_{o_Low}	H_{o_High}	H_e	H_{e_Low}	H_{e_High}	F_{IS}	F_{IS_Low}	F_{IS_High}
MS	5	3,641	835	0.031	0.028	0.037	0.024	0.022	0.028	-0.152	-0.223	-0.049
MG	4	6,022	3,216	0.058	0.053	0.063	0.241	0.236	0.246	0.802	0.787	0.819
Total	9	6,857		0.046			0.379			0.89	0.882	0.889

Number of samples (N), alleles number (A), exclusive alleles number (AP), observed (H_o), and expected heterozygosities (H_e) and their confidence intervals at 95%, inbreeding coefficient (F_{IS}) and its confidence interval at 95%.

Table 3. Analysis of molecular variance (AMOVA), their probabilities (P) and F_{ST} estimates based on 3,429 neutral SNPs for the species of *Myrcia guarujana* and *M. spectabilis* from Serra do Guararu (Guarujá, SP).

Source of variation	df	Sum of squares	Variance component	% variation	P	F_{ST}
Among groups	1	7,148.69	1,488.01	73.54	0.0079	0.735
Within groups	7	3,747.22	535.32	26.46		
Total	8	10,895.90	2,023.33			

df = degrees of freedom.

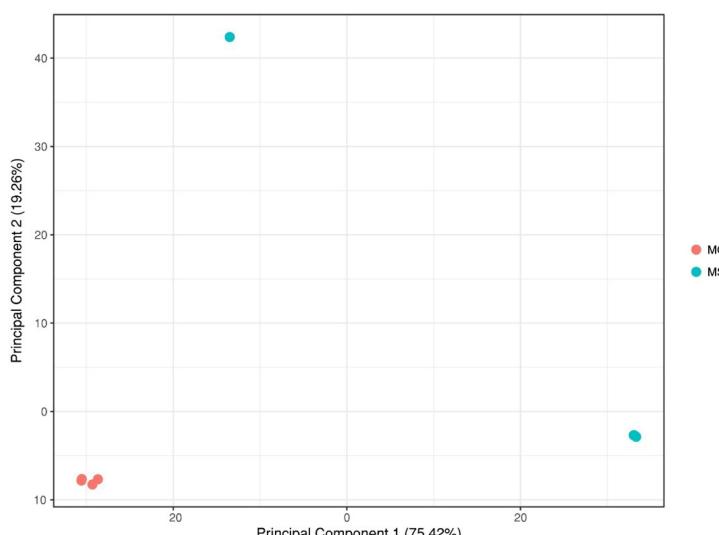


Figure 2. Dispersion plot of the first (75.42%) and second (19.26%) principal components based on the analysis of nine individuals of two *Myrcia* species, using 3,429 SNP loci. The red and blue points represent individuals of *M. guarujana* and *M. spectabilis*, respectively.

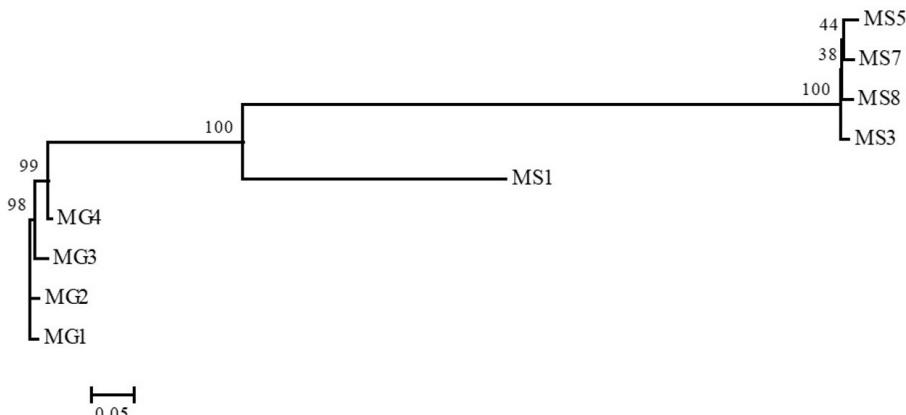


Figure 3. Neighbour-joining trees based on Nei's genetic distances using a panel of 3,429 SNP loci and nine individuals of two *Myrcia* species. Branch nodes are denoted as the percentage of bootstrap support that was generated with 1,000 replicates.

The Nei's genetic distances based on bootstrap values among 1,000 replicates within the two groups of representatives of *Myrcia* (MS and MG, Supplementary Table 2, Supplementary Material) reinforcing the condition of distinct taxa. Interestingly, the distance between the MS1 individual and the other samples previously identified as *M. spectabilis* was 1.00, ensuring total divergence; conversely, the distance between MS1 and the *M. guarujana* cluster ranged from 0.56 to 0.57. Intraspecific genetic distances were low (0.02 to 0.04 on the MG cluster; 0.02 to 0.03 on the MS cluster). This low genetic distance can be explained by the physical proximity of all analysed individuals, indicating a probability of few existing matings, particularly in the *M. guarujana*, in which all individuals were apparently generated by self-fertilization of a single individual. The possible hybrid occurrence between *M. spectabilis* and *M. guarujana* can be sustained by the fact that the most species of the genus have the same chromosome number (N) in their genomes, which facilitates chromosome pairing during meiosis (Da-Costa and Forni-Martins, 2017).

The sign of the capacity of *M. spectabilis* and *M. guarujana* to generate hybrids (MS1 individual) when sympatry occurs raises greater concern about the only known population of *M. guarujana*. Contact between two distinct species populations, one common and other rare, capable of hybridizing among themselves may cause the extinction of one of them since the formed hybrids would partially occupy the sampling effort and would form descendants of the parental species, a process called "genetic swamping" (Todesco et al., 2016). In this way, hybridization-derived lineages with the same reproductive capacity of one or both of their progenitors and that keep interacting with them may cause the extinction of the parental lineage (Hegde et al., 2006). Conversely, if these hybrids become reproductively isolated from both parental lines, a new species may arise by hybrid speciation, as recorded by Lima et al. (2015) in Myrtaceae.

The occurrence of hybrids may result in the progressively reduced presence of *M. guarujana* individuals. Genetic

analysis of hybrids can generate data that conflict with their classification based on morphological traits. Although MS1 exhibits discrete morphological characteristics that could associate it with either *M. guarujana* or *M. spectabilis*, genetic analyses clearly distinguish it from both species.

A more comprehensive investigation including samples of *Myrcia guarujana* from Itanhaém, *M. spectabilis*, and other morphologically similar taxa, such as *M. colpodes* Kiaersk and *M. longipaniculata* Caliari & V.C. Souza combined with genomic and metabomic analysis (which change among species and populations) may yield valuable insights into the evolutionary relationships and degrees of divergence among these taxa. In addition, studies that consider the influence of environmental factors on gene expression in these species may elucidate the memory mechanisms that enable them to respond to environmental stressors and transmit this adaptive capacity to subsequent generations through epigenetic modifications.

It is still unclear how the *Myrcia guarujana* has evolved on the central and southern coast of São Paulo state, and there is still no information about the evolution of this Brazilian species; new studies could provide evidence of the specimens evolution process. One of the hypotheses is that the current scenario of two locations (Guarujá and Itanhaém) with few individuals could be from a former larger population. Anthropic action may be related to the current specimens structure of *M. guarujana* in Guarujá, considering that it occurs in an area of ombrophilous forest where this kind of interference occurs, and perhaps this species needs greater spacing to reproduce.

If new research proves *M. guarujana* endemism in the Central and Southern coastal regions of the state of São Paulo, the species may be threatened by its low genetic diversity. Such genetic constraints can result in inbreeding process anomalies, further compromising its viability. In addition, the vulnerability of the species is exacerbated by the fact that few known individuals are in areas under intense anthropogenic pressure. These factors highlight the urgent need for an assessment of the species' conservation

status based on the criteria established by the International Union for Conservation of Nature (IUCN, 2012).

3.5. Annotations in BLASTx

The outlier loci indicate where natural selection is acting and wherein the species differ. SNP sequence research using the BLASTX tool resulted in 79 loci, to which ontology genetic annotations have been assigned (Supplementary Table 1). Approaches to functional analysis showed that most of the outlier loci found may be related to genes involved in cellular component molecular function, and biological processes. We observed that of the 79 outliers loci corresponding to 568 GO terms (Supplementary Table 1); 237 perform molecular functions, equivalent to 42% of these terms, being mainly catalytic activities and ATP, ion and nucleotide-binding functions; 183 (32%) act in biological processes, such as phosphorylation, biosynthetic processes and transcription regulation; and 143 (26%) are involved in cellular components, related to membrane, nucleus, cytoplasm and cytosol. Among the outlier loci, we found LNG, which promotes foliar elongation, inducing longitudinal growth on cells (Lee et al., 2006). According to this study and the observations made herein, *Myrcia guarujana* presents leaves up to 30 cm long, almost twice the length of the leaves of *M. spectabilis* (up to 18 cm long); based on these statements, this gene could be expressed in *M. guarujana*, causing morphological change. MYB82, a transcriptional activator acting as a positive regulator, is functionally equivalent to GL1 (GLABRA1), whose overexpression results in trichome number reduction (Liang et al., 2014). These regulators can act on *M. spectabilis* since its branches have short trichomes and the abaxial face of the leaf blades is puberulus, with short glabrescent trichomes, while *M. guarujana* has dense and long trichomes on the branches, especially on the terminal branches and along the main veins on the abaxial side of the leaf blades.

We also found that beta-amyrin synthase from oxidosqualene cyclization produces beta-amyrin, a pentacyclic triterpenoid oleanane (Nascimento et al., 2019). These authors reported neuropharmacological effects with anxiolytic and antidepressant actions of this compound, in addition to antimicrobial, antifungal, anti-inflammatory, anti-nociceptive, anti-arthritis and gastroprotective activities - pharmacological properties commonly observed among species of the genus *Myrcia* (Rodrigues et al., 2024). However, further studies are necessary to verify its presence in these species and quantify it for possible large-scale use.

Finally, further detailed investigation of *M. guarujana* and *Myrcia spectabilis* metabolism is needed, considering that the correlations between metabolism and genetic aspects were investigated in similar studies about the pharmacological potential of *Myrcia* (Saldanha et al., 2020).

4. Conclusions

Our results proved that the 3,875 SNP loci discovered can provide fundamental genotypic data to distinguish the two species and indicate the occurrence of hybridization,

data that can contribute to improving the management, conservation, and enhancement of genetic resources of this species.

Taking into account the, fixation index, inbreeding coefficients and the formation of two clusters based on the Nei genetic distance, we can suggest that *Myrcia spectabilis* and *M. guarujana* belong to distinct taxa and are valid as species, proving the adequacy of the morphological characters, such as leaf vein calibre, trichome size, colour on terminal branches and leaves, indumentum type, calyx lobe shape and number of ovary locules, traditionally used to differentiate species. Approaches to functional analysis showed that most of the outlier loci found may be related to genes involved in cellular and metabolic processes. Future studies using these loci, together with broader sampling of *Myrcia guarujana*, *M. spectabilis*, and other morphologically similar species across diverse geographic regions, and integrating genomic and metabolomic approaches establish a robust framework to elucidating their evolutionary relationships and to evaluating the impact of environmental factors on gene expression dynamics. Additionally, we underscore the need for focused attention on the potential conservation vulnerability of *M. guarujana*, in alignment with IUCN Red List assessment criteria."

Acknowledgements

The authors wish to thank Dr. Alessandro Alves-Pereira for contributing to the statistical analysis. This work was supported by the São Paulo Research Foundation (project FAPESP 2022/11622-1).

Data Availability Statement

Research data is only available upon request

Access to the genetic diversity

This study was registered at the National System of Genetic Resource Management and Associated Traditional Knowledge (SisGen - A4E0F73 and AC05E27, Brazil).

References

ALIPOUR, H., BIHAMTA, M.R., MOHAMMADI, V., PEYGHAMBARI, S.A., BAI, G. and ZHANG, G., 2017. Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science*, vol. 8, pp. 1293. <http://doi.org/10.3389/fpls.2017.01293>. PMid:28912785.

AMORIM, B.S., VASCONCELOS, T.N.C., SOUZA, M., ALVES, M., ANTONELLI, A. and LUCA, E., 2019. Advanced understanding of phylogenetic relationships, morphological evolution and biogeographic history of the mega-diverse plant genus *Myrcia* and its relatives (Myrtaceae: Myrtae). *Molecular Phylogenetics and Evolution*, vol. 138, pp. 65-88. <http://doi.org/10.1016/j.ympev.2019.05.014>. PMid:31132520.

ANDREWS, S., 2010 [viewed 16 July 2025]. *FastQC: a quality control tool for high throughput sequence data* [online]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

BARROSO, G.M. and PEIXOTO, A.L., 1995 [viewed 28 February 2025]. Myrtaceae da reserva florestal de Linhares, Espírito Santo, Brasil: gêneros *Calyptranthes* e *Marlierea*. *Boletim do Museu de Biologia Prof. Mello Leitão* [online], no. 3, pp. 3-38. Available from: <http://boletim.sambio.org.br/ns03.html>

BASKIN, C.C. and BASKIN, J.M., 2025. Seed dormancy and germination in Myrtaceae: a palaeohistory, tribe, life cycle and geographical distribution perspective. *Seed Science Research*, pp. 1-31. In press. <http://doi.org/10.1017/S0960258525000066>.

BYRNE, M. and HOPPER, S.D., 2008. Granite outcrops as ancient islands in old landscapes: evidence from the phylogeography and population genetics of *Eucalyptus caesia* (Myrtaceae) in Western Australia. *Biological Journal of the Linnean Society. Linnean Society of London*, vol. 93, no. 1, pp. 177-188. <http://doi.org/10.1111/j.1095-8312.2007.00946.x>.

CALIARI, C.P., 2013. *Estudos em Myrtaceae do estado de São Paulo: Myrcia seção Gomidesia*. Piracicaba: Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 128 p. Dissertação de Mestrado em Recursos Florestais.

CATCHEN, J.M., AMORES, A., HOHENLOHE, P., CRESKO, W. and POSTLETHWAIT, J.H., 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3*, vol. 1, no. 3, pp. 171-182. <http://doi.org/10.1534/g3.111.000240>. PMID:22384329.

DA-COSTA, I.R. and FORNI-MARTINS, E.R., 2017 [viewed 28 February 2025]. Chromosome studies in *Gomidesia*, *Marlierea*, *Myrcceugenia* and *Myrcia* (Myrtaceae, subtribe Myrciinae). *Kew Bulletin* [online], vol. 62, no. 1, pp. 113-118. Available from: <http://repositorio.ufc.br/handle/riufc/63295>

FLANAGAN, S.P. and JONES, A.G., 2017. Constraints on the FST-heterozygosity outlier approach. *The Journal of Heredity*, vol. 108, no. 5, pp. 561-573. <http://doi.org/10.1093/jhered/esx048>. PMID:28486592.

FREITAS, P.V.D.X., FAQUINELLO, P., ARNHOLD, E., FERRO, D.A.C., FERRO, R.A.C., LACERDA, M.L.G., LEITE, P. and SILVA NETO, C.M., 2023. Flight radius and climatic conditions affect the external activity of stingless bee *Melipona rufiventris* (Lepeletier, 1836). *Brazilian Journal of Biology = Revista Brasileira de Biologia*, vol. 83, e275645. <http://doi.org/10.1590/1519-6984.275645>. PMID:38126636.

GÖTZ, S., GARCÍA-GÓMEZ, J.M., TEROL, J., WILLIAMS, T.D., NAGARAJ, S.H., NUEDA, M.J., ROBLES, M., TALÓN, M., DOPAZO, J. and CONESA, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, vol. 36, no. 10, pp. 3420-3435. <http://doi.org/10.1093/nar/gkn176>. PMID:18445632.

GRESSLER, E., PIZO, M.A. and MORELLATO, L.P.C., 2006. Polinização e dispersão de sementes em Myrtaceae do Brasil. *Revista Brasileira de Botânica. Brazilian Journal of Botany*, vol. 29, no. 4, pp. 509-530. <http://doi.org/10.1590/S0100-84042006000400002>.

HEGDE, S.G., NASON, J.D., CLEGG, J.M. and ELLSTRAND, N.C., 2006. The evolution of California's wild radish has resulted in the extinction of its progenitors. *Evolution; International Journal of Organic Evolution*, vol. 60, no. 6, pp. 1187-1197. <http://doi.org/10.1111/j.0014-3820.2006.tb01197.x>. PMID:16892969.

INTERNATIONAL UNION FOR CONSERVATION OF NATURE – IUCN, 2012 [viewed 28 February 2025]. *IUCN Red List categories and criteria: Version 3.1* [online]. 2nd ed. Gland: IUCN, 32 p. Available from: <https://www.iucnredlist.org/resources/categories-and-criteria>

KAMVAR, Z.N., TABIMA, J.F. and GRÜNWALD, N.J., 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, vol. 2, e281. <http://doi.org/10.7717/peerj.281>. PMID:24688859.

KEENAN, K., MCGINNITY, P., CROSS, T.F., CROZIER, W.W. and PRODÖHL, P.A., 2013. DiveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, vol. 4, no. 8, pp. 782-788. <http://doi.org/10.1111/2041-210X.12067>.

LEE, Y.K., KIM, G.T., KIM, I.J., PARK, J., KWAK, S.S., CHOI, G. and CHUNG, W., 2006. LONGIFOLIA1 and LONGIFOLIA2, two homologous genes, regulate longitudinal cell elongation in *Arabidopsis*. *Development*, vol. 133, no. 21, pp. 4305-4314. <http://doi.org/10.1242/dev.02604>. PMID:17038516.

LIANG, G., HE, H., LI, Y., AI, Q. and YU, D., 2014. MYB82 functions in regulation of trichome development in *arabidopsis*. *Journal of Experimental Botany*, vol. 65, no. 12, pp. 3215-3223. <http://doi.org/10.1093/jxb/eru179>. PMID:24803498.

LIMA, D.F., MAUAD, A.V.S., SILVA-PEREIRA, V., SMIDT, E.C. and GOLDENBERG, R., 2015. Species boundaries inferred from ISSR markers in the *Myrcia larvotearna* complex (Myrtaceae). *Plant Systematics and Evolution*, vol. 301, no. 1, pp. 353-363. <http://doi.org/10.1007/s00606-014-1078-9>.

LUGHADHA, E.N. and PROENÇA, C., 1996. A survey of the reproductive biology of the Myrtoideae (Myrtaceae). *Annals of the Missouri Botanical Garden*, vol. 83, no. 4, pp. 480-503. <http://doi.org/10.2307/2399990>.

LUU, K., BAZIN, E. and BLUM, M.G.B., 2017. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, vol. 17, no. 1, pp. 67-77. <http://doi.org/10.1111/1755-0998.12592>. PMID:27601374.

MARANDEL, F., CHARRIER, G., LAMY, J.B., LE CAM, S., LORANCE, P. and TRENKEL, V.M., 2020. Estimating effective population size using RADseq: effects of SNP selection and sample size. *Ecology and Evolution*, vol. 10, no. 4, pp. 1929-1937. <http://doi.org/10.1002/ece3.6016>. PMID:32128126.

NASCIMENTO, L.O.N., AGUIAR, E.M.M.M. and AUM, Y.K.P.G., 2019. Purificação de amostras impuras de alfa e beta amina por cristalização. In: *Anais do XIII Congresso Brasileiro de Engenharia Química em Iniciação Científica*, 2019, São Paulo. São Paulo: Blucher, pp. 231-237. <http://doi.org/10.5151/cobec2019-EAT136>.

NAZARENO, A.G., BEMMELS, J.B., DICK, C.W. and LOHMANN, L.G., 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, vol. 17, no. 6, pp. 1136-1147. <http://doi.org/10.1111/1755-0998.12654>. PMID:28078808.

NEI, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, vol. 89, no. 3, pp. 583-590. <http://doi.org/10.1093/genetics/89.3.583>.

POLAND, J.A., BROWN, P.J., SORRELLS, M.E. and JANNINK, J.L., 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, vol. 7, no. 2, e32253. <http://doi.org/10.1371/journal.pone.0032253>. PMID:22389690.

RODRIGUES, G.M., MAGENTA, M.A.G., CLERIGO, H.C. and FRANCES, M., 2024. Uses and pharmacological properties of *Myrcia DC* (Myrtaceae) in Brazil: systematic review. *Revista Fitos*, vol. 18, e1505. <http://doi.org/10.32712/2446-4775.2024.1505>.

SALDANHA, L.L., ALLARD, P.-M., AFZAN, A., MELO, F.P.S.R., MARCOURT, L., QUEIROZ, E.F., VILEGAS, W., FURLAN, C.M., DOKKEDAL, A.L. and WOLFENDER, J.-L., 2020. Metabolomics of *Myrcia bela* populations in Brazilian savanna reveals strong influence of environmental

factors on its specialized metabolism. *Molecules*, vol. 25, no. 12, pp. 2954. <http://doi.org/10.3390/molecules25122954>.

SAMPSON, J., TAPPER, S., COATES, D., HANKINSON, M., MCARTHUR, S. and BYRNE, M., 2022. Genetic diversity and differentiation in south-western Australian bloodwoods (*Corymbia* section *Calophyllae*, Myrtaceae) with different ranges and abundance. *Australian Journal of Botany*, vol. 70, no. 2, pp. 146-157. <http://doi.org/10.1071/BT21081>.

SANTOS, C.V., MALLMANN, A.P., TOLEDO, A.G., BANDEIRA, D.M., COSTA, W.F., MARINS, D.M.Á., CORRÊA, J.M. and PINTO, F.G.S., 2021. Composição química, atividade antimicrobiana e antioxidante do óleo essencial de folhas *Myrcia palustris* DC. (Myrtaceae). *Research, Society and Development*, vol. 10, no. 3, e20510313303. <http://doi.org/10.33448/rsd-v10i3.13303>.

SANTOS, M.F., AMORIM, B.S., BURTON, G.P., FERNANDES, T., GAEM, P.H., LOURENÇO, A.R.L., LIMA, D.F., ROSA, P.O., SANTOS, L.L.D., STAGGEMEIER, V.G., VASCONCELOS, T.N.C. and LUCAS, E.J., 2025. *Myrcia*. In: JARDIM BOTÂNICO DO RIO DE JANEIRO, ed. *Flora e Funga do Brasil*. Rio de Janeiro: JBRJ.

SCAKETTI, M., SUJII, P.S., ALVES-PEREIRA, A., SCHWARCZ, K.D., FRANCISCONI, A.F., MORO, M.S., MORENO MARTINS, K.K., JESUS, T.A., SOUZA, G.B.F. and ZUCCHI, M.I., 2025. Sample Size Impact (SaSii): an R script for estimating optimal sample sizes in population genetics and population genomics studies. *PLoS One*, vol. 20, no. 2, e0316634. <http://doi.org/10.1371/journal.pone.0316634>. PMid:39946360.

SOBRAL, M., CALIARI, C.P., GRESSLER, E., MAZINE, F.F., MAGENTA, M. and VIANA, P.L., 2016. Seven new southeastern Brazilian species of *Myrcia* (Myrtaceae). *Phytotaxa*, vol. 247, no. 1, pp. 27-44. <http://doi.org/10.11646/phytotaxa.247.1.2>.

SONG, Z., ZHANG, M., LI, F., WENG, Q., ZHOU, C., LI, M., LI, J., HUANG, H., MO, X. and GAN, S., 2016. Genome scans for divergent selection in natural populations of the widespread hardwood species *Eucalyptus grandis* (Myrtaceae) using microsatellites. *Scientific Reports*, vol. 6, no. 1, pp. 34941. <http://doi.org/10.1038/srep34941>. PMid:27748400.

TODESCO, M., PASCUAL, M.A., OWENS, G.L., OSTEVÍK, K.L., MOYERS, B.T., HÜBNER, S., HEREDIA, S.M., HAHN, M.A., CASEYS, C., BOCK, D.G. and RIESEBERG, L.H., 2016. Hybridization and extinction. *Evolutionary Applications*, vol. 9, no. 7, pp. 892-908. <http://doi.org/10.1111/eva.12367>. PMid:27468307.

VAN NIEUWSTADT, M.G.L. and RUANO IRAHETA, C.E., 1996. Relation between size and foraging range in stingless bees (Apidae, Meliponinae). *Apidologie*, vol. 27, no. 4, pp. 219-228. <http://doi.org/10.1051/apido:19960404>.

Supplementary Material

Supplementary material accompanies this paper.

Supplementary Table 1. Gene Ontology classifications obtained in blast2go for proteins similar to translated RAD tags with SNPs putatively under selection in two species of *Myrcia* DC. (Myrtaceae).

Supplementary Table 2. Nei's genetic distance (Nei, 1978) among samples of *M. guarujana* (MG1–MG4) and *M. spectabilis* (MS1, MS3, MS5, MS7, MS8), estimated using 1,000 bootstrap replicates. Lower values indicate higher genetic similarity. The matrix highlights low genetic differentiation within *M. guarujana* and higher divergence between *M. guarujana* and *M. spectabilis* samples.

This material is available as part of the online article from <https://doi.org/10.1590/1519-6984.294682>