



## OPEN ACCESS

## EDITED BY

Yunpeng Cao,  
Chinese Academy of Sciences (CAS), China

## REVIEWED BY

Elytania Menezes,  
Universidade Estadual de Montes Claros -  
UNIMONTES, Brazil  
Carlos Silva,  
Federal Rural University of Pernambuco, Brazil  
Tian Wan,  
Shaanxi University of Technology, China

## \*CORRESPONDENCE

Francisco de Assis Alves Mourão Filho  
✉ francisco.mourao@usp.br

RECEIVED 21 July 2025

REVISED 15 November 2025

ACCEPTED 01 December 2025

PUBLISHED 13 January 2026

## CITATION

Sampaio LFS, Zucchi MI, Colombo CA,  
Jacomino AP, Figueira A and Mourão  
Filho FdAA (2026) Population genomics of  
Brazilian native fruit species of *Eugenia* spp.  
(Myrtaceae) for conservation and  
improvement.  
*Front. Plant Sci.* 16:1670349.  
doi: 10.3389/fpls.2025.1670349

## COPYRIGHT

© 2026 Sampaio, Zucchi, Colombo, Jacomino,  
Figueira and Mourão Filho. This is an open-  
access article distributed under the terms of  
the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Population genomics of Brazilian native fruit species of *Eugenia* spp. (Myrtaceae) for conservation and improvement

Laecio Fernandes Souza Sampaio<sup>1</sup>, Maria Imaculada Zucchi<sup>2</sup>,  
Carlos Augusto Colombo<sup>3</sup>, Angelo Pedro Jacomino<sup>1</sup>,  
Antonio Figueira<sup>4</sup> and Francisco de Assis Alves Mourão Filho<sup>1\*</sup>

<sup>1</sup>Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, Brazil,

<sup>2</sup>Polo Regional de Desenvolvimento Tecnológico do Centro Sul, Agência Paulista de Tecnologia dos  
Agronegócios, Piracicaba, SP, Brazil, <sup>3</sup>Instituto Agronômico de Campinas, Centro de Recursos  
Genéticos Vegetais, Campinas, SP, Brazil, <sup>4</sup>Centro de Energia Nuclear na Agricultura, Universidade de  
São Paulo, Piracicaba, SP, Brazil

Brazil is a global biodiversity hotspot, especially in the Atlantic Forest biome, which contains a high diversity of native fruit species that remain underutilized and understudied. Native fruit trees, particularly those in the Myrtaceae family, have great potential to become new fruit crops contributing to food security. The genus *Eugenia* encompasses several native species that have been little investigated, including *Eugenia brasiliensis* Lam. (grumixama), *E. pyriformis* Cambess (uvaia), and *E. involucrata* DC (Rio Grande cherry). This study investigated the genomic diversity and structure of several populations of these three native fruit species using single-nucleotide polymorphism (SNPs) markers obtained through genotyping by sequencing. We analyzed 73 accessions of *E. brasiliensis*, 93 of *E. pyriformis*, and 62 of *E. involucrata*, derived from three, four, and seven populations, respectively, maintained as living collections (due to their desiccation-sensitive seeds) in research institutions, urban afforestation projects, and small rural properties in the states of São Paulo and Minas Gerais, Brazil. The comparison among *E. brasiliensis*, *E. pyriformis*, and *E. involucrata* accessions revealed 2,299, 2,872, and 1,471 SNPs, respectively. These markers effectively characterized each species' genomic diversity and population structure, revealing levels of diversity ( $H_e = 0.22, 0.19, 0.21$  for grumixama, uvaia, and Rio Grande cherry respectively) and inbreeding ( $f = -0.06; 0.05; -0.04$ , respectively) consistent with their respective mating biology. Significant genetic structure was detected between collections ( $\Phi_{iST} = 0.29; 0.10; 0.23$  for *E. brasiliensis*, *E. pyriformis*, and *E. involucrata*, respectively), confirmed by discriminant and principal component analyses, indicating an important diversity between and within the collections. The data will serve to identify the most divergent accessions to help prioritize accessions for fruit

quality assessments and for conservation, while identifying parents to guide hybridizations to initiate a breeding program. The study highlights the importance of employing population genomics approaches to develop improved management practices for these fruit species, ultimately promoting the conservation and valorization of Brazilian native genetic resources.

#### KEYWORDS

Atlantic forest, *E. brasiliensis*, *E. involucrata*, *E. pyriformis*, genetic markers, K-means, SNP

## 1 Introduction

Brazil is a recognized global biodiversity hotspot, particularly in the Atlantic Forest biome, which harbors an exceptional wealth of native fruit species that have mainly remained underutilized, scientifically neglected, and poorly integrated into sustainable agricultural or economic systems (Araújo et al., 2019). Native fruit species represent a potential genetic resource to integrate production systems as new crops, which can contribute to food security (Souza et al., 2018). The genus *Eugenia*, Myrtaceae, for instance, contains numerous native Brazilian fruit species from this biome, most of which are poorly characterized. Our interest focused on *Eugenia brasiliensis* Lam. (grumixama), *E. pyriformis* Cambess (uvaia), and *E. involucrata* DC. (Rio Grande cherry), which are highly valued for their flavor and pulp nutritional properties (Araújo et al., 2019). These *Eugenia* species have been subjected to studies regarding their fruits' nutritional and agro-industrial characteristics (Flores et al., 2012; Infante et al., 2016; Sardi et al., 2017; Lazarini et al., 2018; Silva et al., 2019; Soares et al., 2019; Xu et al., 2020; Nehring et al., 2022; Sganzerla et al., 2022). Their fruits contain high levels of vitamins, minerals, antioxidants, together with other bioactive compounds (e.g. Silva et al., 2022; Spricigo et al., 2023), and display traditional uses in various food and popular medicinal applications (Araújo et al., 2019; Schmidt et al., 2019; Farias et al., 2020).

Therefore, we have selected these *Eugenia* species as they are promising new crops in Brazil, due to their desirable fruit sensorial and nutritional qualities, and their adaptability to local climates (Araujo et al., 2024; Saliba et al., 2025). However, little is known about their genetics and mating biology. The region of natural dispersion and center of diversity of these species have not been determined and there is little information about areas of natural occurrence. However, it is well established that the Brazilian biomes where anecdotal references to their occurrence have been reported have been continually threatened by land-use change, urbanization, and deforestation (Araujo et al., 2024). In emerging crops and in the process of domestication, as is the case of these *Eugenia* species, knowledge of genetic diversity and population structure is crucial for the development of strategies to sustainably exploit these species and to preserve genetic resources, particularly for these species that display desiccation-sensitive seeds, which largely limit the long-term conservation of accessions (Delgado and Barbedo, 2012;

Inocente and Barbedo, 2019), while starting a breeding program (Allendorf et al., 2013; Salgotra and Chauhan, 2023).

The rapid advancement of next-generation sequencing (NGS) technologies, along with their cost reduction, has enabled large-scale identification of single-nucleotide polymorphisms (SNPs) (Nadeem et al., 2017; Rasheed et al., 2017; Rasheed and Xia, 2019; Kumar et al., 2021). SNPs are the most abundant, codominant, and stable polymorphisms in plant genomes. The possibility of automating the detection and genotyping of these markers offers a notable advantage over previously used marker types (Morin et al., 2004; Helyar et al., 2011; Mammadov et al., 2012). Several strategies can be used to discover SNPs (Andrews et al., 2016). The ideal SNP genotyping would be whole-genome sequencing (Aguirre et al., 2024); however, its high cost favored the development of genotyping arrays and Reduced-Representation Sequencing (RRS) approaches for SNP identification (Andrews et al., 2016). The RRS methods grouped under the generic term 'genotyping by sequencing' (GBS) are the most cost-effective strategy currently used for non-model organisms, which combine genome reduction and sampling of both coding and non-coding regions with a genome-wide coverage, and it can be used in species lacking a reference genome (Elshire et al., 2011; Poland et al., 2012; Andrews et al., 2016; Brhane et al., 2022).

Here, we employed GBS to identify SNP markers and assess the genetic diversity and population structure of *E. brasiliensis* (grumixama), *E. pyriformis* (uvaia), and *E. involucrata* (Rio Grande cherry) accessions, which are maintained in small collections by growers and research institutions in São Paulo and Minas Gerais, Brazil. We hypothesized that there was strong genetic differentiation among collections, with clustering patterns corresponding to their geographic origins, and that individual collections would exhibit unique diversity and inbreeding profiles reflecting their specific establishment histories.

## 2 Materials and methods

### 2.1 Plant material and collection sites

Accessions of *E. brasiliensis* (grumixama -  $n = 76$ ), *E. pyriformis* (uvaia -  $n = 111$ ) and *E. involucrata* (Rio Grande cherry -  $n = 69$ )

were originally collected in several municipalities in São Paulo and Minas Gerais, Brazil (Figure 1), and were kept as living specimens (accessions) on private properties of small growers, collectors of fruit species, University collections, and experimental stations (Supplementary Table 1). Accessions of *E. brasiliensis* were collected in three locations (municipalities), while accessions of *E. pyriformis* were collected in four locations. Samples of *E. involucrata* were gathered from seven locations in five municipalities (Figure 1, Supplementary Table 1). Leaf samples from each of the three species' accessions were collected, wrapped in aluminum foil, and stored in a thermal box containing ice until arrival to the laboratory, where they were kept at  $-80^{\circ}\text{C}$  until DNA isolation. The initial sampling included 76 accessions of *E. brasiliensis*, 111 accession of *E. pyriformis*, and 69 of *E. involucrata*. After quality filtering the sequenced reads (see below), individuals with more than 40% missing data or duplicated genotypes were removed, resulting in a final dataset of 73, 93, and 62 accessions used in the analyses for *E. brasiliensis*, *E. pyriformis*, and *E. involucrata*, respectively.

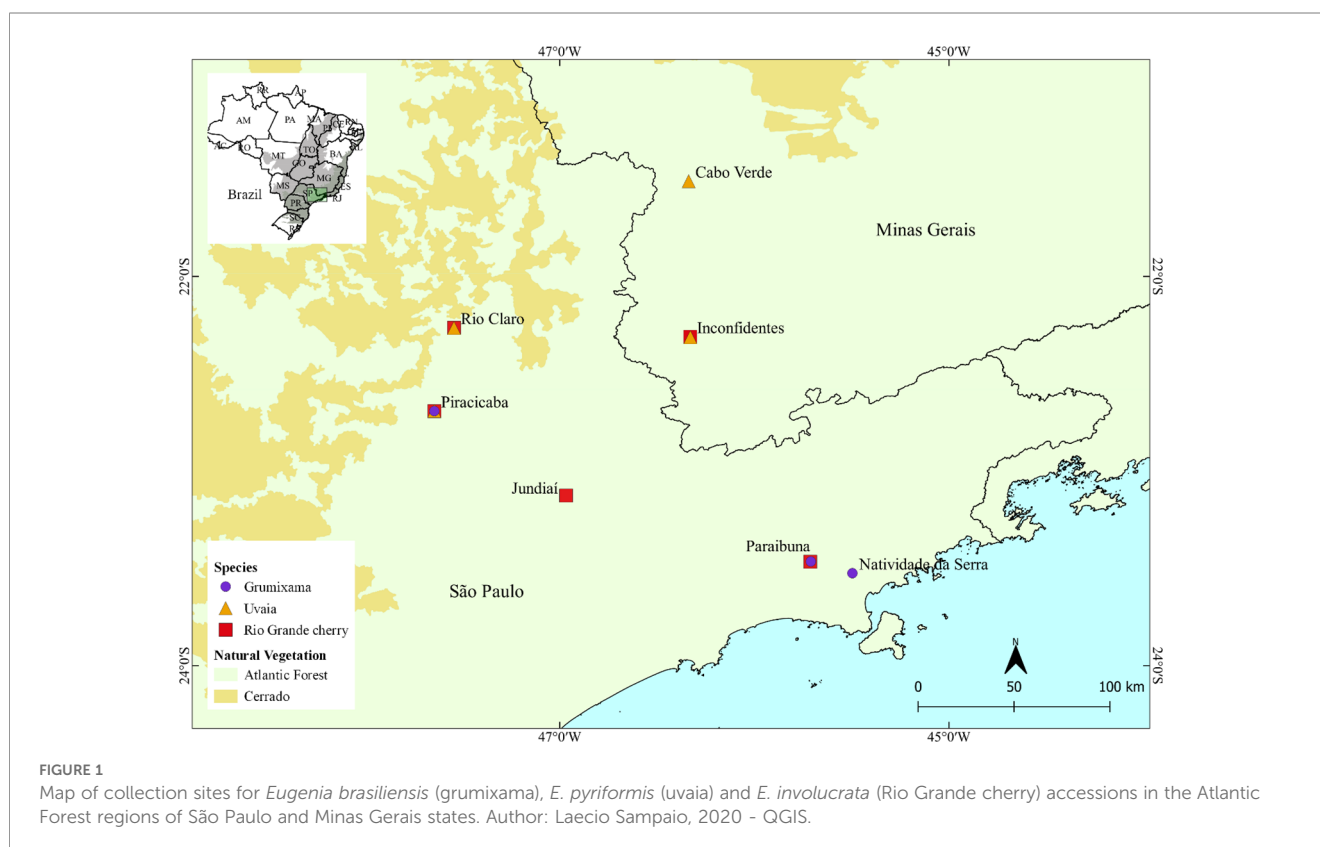
## 2.2 Extraction of genomic DNA

Genomic DNA extraction was performed using 50 mg of leaf tissue (Serenó et al., 2006). DNA quality was assessed by electrophoresis ( $1\text{ V cm}^{-1}$ ) in a 1% agarose gel stained with SYBR gold (Thermo Fisher Scientific Inc.; Waltham, MA, USA) and quantified by fluorescence using the Qubit dsDNA BR assay

(Thermo Fisher Scientific). The DNA concentration of the samples was standardized to  $25\text{ ng }\mu\text{L}^{-1}$ .

## 2.3 Preparation and sequencing of GBS libraries

The genomic libraries were prepared according to the protocol of Poland et al. (2012). For each sample, approximately 175 ng of genomic DNA ( $7\text{ }\mu\text{L}$ ) was digested at  $37^{\circ}\text{C}$  for 12 h with *NsiI* and *MseI* to reduce sequence complexity (Poland et al., 2012). After digestion, the fragments were ligated to  $0.02\text{ }\mu\text{M}$  specific Illumina adapters (Illumina Inc.; San Diego, CA, USA), which are complementary to the restriction sites. At this stage, each sample received an adapter containing barcode sequences to identify each individual after sequencing. The ligation was carried out for two h at  $22^{\circ}\text{C}$ , followed by 20 min at  $65^{\circ}\text{C}$ . The ligation reaction products from each sample were then mixed into a single pool and purified with the QIAquick PCR purification kit (Qiagen; Germantown, MD, USA). The libraries were subjected to amplification. For each library, eight replicates were made, each containing  $10\text{ }\mu\text{L}$  of purified and amplified ligation, using  $12.5\text{ }\mu\text{L}$  of Phusion High-Fidelity PCR Master Mix NEB (New England Biolabs Inc.; Ipswich, MA, USA) and two  $\mu\text{L}$  of Illumina forward and reverse primers ( $10\text{ }\mu\text{M}$ ), in a final volume of  $25\text{ }\mu\text{L}$ , using the following amplification program:  $95^{\circ}\text{C}$  for 30 s, followed by 16 cycles of  $95^{\circ}\text{C}$  for 10 s,  $62^{\circ}\text{C}$  for 20 s,  $72^{\circ}\text{C}$  for 30 s, finishing at  $72^{\circ}\text{C}$  for 5 min. Then, the libraries were purified again using the QIAquick PCR



Purification Kit. Before sequencing, the average size of the DNA fragments was verified using the Agilent DNA 12000 kit and the 2100 Bioanalyzer System (Agilent; Santa Clara, CA, USA). The libraries were quantified by quantitative PCR (qPCR) in a CFX 384 thermocycler (Bio-Rad, Hercules, CA, USA) using the KAPA Library Quantification kit (KAPA Biosystems, Wilmington, MA, USA). Finally, the libraries were normalized and diluted to  $1.8 \text{ pmol } \mu\text{L}^{-1}$  and sequenced in the *single-read* configuration and 150 base pairs, using the Illumina NextSeq 500/550 Mid Output kit v2.5 (150 cycles), on the NextSeq550 platform (Illumina). We have sequenced replicates of DNA for a sample of accessions as part of quality control.

## 2.4 Genotyping and filtering of SNPs

Read sequence quality was preliminarily checked by FastQC (Andrews, 2010). Subsequently, demultiplexing and quality filtering were performed using the `process_radtags` module based on the STACKS v.2.437 pipeline (Catchen et al., 2013). Quality filtering was performed with the default setting, and sequences were truncated to 100 nucleotides. Next, reads with adapter sequences were removed, allowing for incompatibility of up to 2 bases. The barcodes of the retained sequences were removed, and SNPs were identified. All the species investigated here lack a reference genome; therefore, the `ustacks` module was used to construct all loci *de novo*. In this step, two parameter combinations were used for filtering. For *E. pyriformis* and *E. involucrata*, the parameters were: minimum depth to form a locus ( $-m$ ) = 2, maximum intra-individual distance between stacks ( $-M$ ) = 3, and maximum distance allowed to align secondary readings to primary stacks ( $-N$ ) =  $-M + 2$  (this being the last default configuration of the `ustacks` module). To build the locus catalog in the `cstacks` module, the maximum difference allowed between the stacks ( $-n$ ) was defined as 2. Finally, the SNPs were filtered in the `populations` module, following the criteria: minimum depth of sequencing  $\geq 3X$ , minor allele frequency (MAF)  $\geq 0.05$ , filtering only SNPs that occurred in all locations/collections (four sites for *E. pyriformis*; seven sites for *E. involucrata*), and at least in 75% of samples from each location/collection. The parameters used for *E. brasiliensis* in the `ustacks` module were:  $-m = 3$ ,  $-M = 2$ , and  $-N = 2$ . Subsequently, a locus catalog (`cstacks`) was built, allowing a maximum of 2 differences between stacks from different accessions. Then, in the `populations` module, the final filtering of SNPs was carried out with a minimum sequencing depth of  $5X$ ,  $\text{MAF} \geq 0.05$  and minimum occurrence in 90% of individuals in each location/collection. The filtering parameters in the `ustacks` module were optimized for each species to maximize locus construction quality based on the sequencing output. For *E. brasiliensis*, a higher average sequencing depth ( $16.2X$ ) allowed for more stringent criteria (e.g., minimum depth of  $5X$  and 90% of individuals) compared to *E. pyriformis* and *E. involucrata* ( $6.3X$  and  $6.1X$ , respectively), for which a minimum depth of  $3X$  and presence in 75% of individuals was required. This tailored approach ensured that the final SNP dataset for each species was of the highest possible quality. The sequence data were deposited at the

BioStudies repository under the ids S-BSST2086 (*E. pyriformis*), S-BSST2087 (*E. brasiliensis*), and S-BSST2088 (*E. involucrata*).

## 2.5 Genomic and population analyses

First, the data was filtered to remove loci and individuals with high rates of missing data. SNP loci with more than 20% missing genotypes were removed from the dataset using the ‘missingno’ function in `poppr` v.2.8.5 (Kamvar et al., 2014) in R software (R Core Team, 2021). Individuals with more than 40% missing and/or duplicate data were also removed from analyses. This initial filtering resulted in the exclusion of zero locus and 3 individuals for *E. brasiliensis*, 222 loci and 18 individuals for *E. pyriformis*, and 44 loci and 7 individuals for *E. involucrata*.

Genomic diversity was characterized by the percentage of polymorphic loci (%  $P$ ), the number of alleles ( $N_a$ ), the number of alleles per locus ( $A$ ), the number of private alleles ( $P_a$ ), allelic richness ( $AR$ ), and observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities, calculated based on Nei’s (1978) gene diversity metric. The fixation indices ( $f$ ) were also estimated, and their confidence intervals were obtained with 1000 bootstraps. Genomic diversity and fixation indices were estimated using the `diveRcity` (Keenan et al., 2013) and `poppr` (Kamvar et al., 2014) packages in R software (R Core Team, 2021).

Genetic differentiation between accession groups was estimated using pairwise *PhiST* and confidence intervals with 1000 bootstraps, as implemented in the `Hierfstat` package (Goudet, 2005) within R (R Core Team, 2021). Two methods were used to explore the population structure of the species, both from the `Adegenet` v.2.1.2 package (Jombart and Bateman, 2008) in R. First, for each species, a principal component analysis (PCA) was performed using the `dudi.pca` function. Second, a Discriminant Principal Component Analysis (DAPC) was conducted (Jombart et al., 2010). For each species, the groups were *de novo* identified by the K-means clustering algorithm, from the `find.clusters` function after data transformation by Principal Component Analysis (PCA). K-means was performed with  $K$  ranging from 1 to 40, and Bayesian Information Criterion (BIC) was used to determine the most appropriate value of  $K$  (Supplementary Figures S1-S3). Individuals were assigned to groups using DAPC. To avoid retaining too many dimensions in the discriminant analysis (DPCA), the optimal number of PCs was determined using the `optim.a.score` function. Based on this analysis, 5 PCs were retained for *E. brasiliensis*, 7 for *E. pyriformis*, and 6 for *E. involucrata*. The probability of adherence of individuals from each collection to each group was determined after DAPC (subsequent assignment of DAPC analysis).

## 2.6 Core collection

The R package `CoreHunter` 3.0 (De Beukelaer et al., 2018) was used to assemble a nuclear collection that represented the maximum genetic diversity of the total number of accessions

collected for *E. brasiliensis*. We used the Entry-to-nearest-entry (E-NE) method, implemented in the CoreHunter 3.0 using the Modified Roger's distance (De Beukelaer et al., 2018). Several subsamples were generated, adjusting the size of the desired core collections to identify a subset of genotypes that could capture maximum allele diversity. Collection sizes ranged from 10 to 30% for all datasets. In each collection, genetic diversity parameters and a principal coordinate analysis (PCoA) based on the standard-covariance matrix were determined using GenAEx v. 6.5 (Peakall and Smouse, 2012).

## 3 Results

### 3.1 Overall SNP discovery in *Eugenia*

The GBS libraries were successfully sequenced for 76 *E. brasiliensis* (grumixama) accessions, 111 *E. pyriformis* (uvaia) accessions, and 69 *E. involucrata* (Rio Grande cherry) accessions. The GBS libraries of grumixama, uvaia, and Rio Grande cherry generated a total of 127,073,304, 101,022,119, and 100,281,548 reads, respectively. After quality control and filtering, the total number of accessions retained was 73 for *E. brasiliensis*, 93 for *E. pyriformis*, and 62 for *E. involucrata*, and the total number of reads retained was 122,831,851 for grumixama, 59,363,247 for uvaia, and 62,784,718 for Rio Grande cherry. A total of 2,299 SNPs was identified for *E. brasiliensis* (with an average depth per sample equal to 16.2X); 2,872 SNPs for *E. pyriformis* (with an average depth per sample equal to 6.3X); and 1,471 SNPs for *E. involucrata* (with an average depth per sample equal to 6.1X).

### 3.2 Population genomic analysis of *E. brasiliensis*

We utilized the 2,299 SNP markers identified by GBS to assess the genetic diversity and population structure of the 73 *E. brasiliensis* accessions from the three collections. The genetic diversity ( $H_e$ ) ranged from 0.21 to 0.24, with an average of 0.22 among the grumixama collections (Table 1). The Paraibuna collection displayed the largest genetic diversity ( $H_e = 0.24$ ), a higher percentage of polymorphic loci (%P), a greater number of alleles ( $N_a$ ), and a higher number of private alleles ( $P_a$ ). The Piracicaba and Natividade da Serra collections presented identical  $H_e$  values (0.21). The observed heterozygosity ( $H_o$ ) ranged from 0.22 (Piracicaba) to 0.30 (Natividade da Serra), with an average of 0.25. The Natividade da Serra collection had  $H_o$  greater than  $H_e$ , presenting a significant excess of heterozygotes ( $f = -0.31$ ). The Piracicaba and Paraibuna collections showed low fixation indices ( $f = -0.04$  and  $0.05$ , respectively), which were not significantly different from zero, suggesting that these populations were in equilibrium.

The analysis of molecular variance (AMOVA) disclosed that 29% of the total genetic variation is found between collections, while the remainder (71%) is found within collections (Table 2). The

estimated  $Phi_{ST}$  value for the *E. brasiliensis* collections was 0.29 ( $p < 0.001$ ) (Table 2), indicating high genetic divergence between the three collections. Paired  $Phi_{ST}$  estimates were high (Table 3), with values varying from 0.30 to 0.37, with the greatest divergence observed between the Piracicaba and Natividade da Serra collections ( $Phi_{ST} = 0.37$ ), and the smallest between Natividade and Paraibuna ( $Phi_{ST} = 0.30$ ). The high genetic divergence suggested by pairwise  $Phi_{ST}$  estimates was also observed in the principal component analysis (PCA). The PCA analysis, based on 2299 SNPs, explained 34.8% of the total variation in the first two components and demonstrated that the *E. brasiliensis* collections are highly distinct from each other (Figure 2A).

The population structure and genetic relationships among the 73 *E. brasiliensis* accessions were also explored by discriminant analysis of principal components (DAPC), considering the optimal number of genetic groups ( $K = 6$ ), determined by the  $K$ -means method (Supplementary Figure S1). The groups formed in the DAPC retained 81.3% of the total variation in the first two principal components (Figure 2B). The clusters formed showed great association with the collection of origin, and the grouping patterns were very similar to those observed by PCA (Figure 2A), reinforcing the high genetic structure found among the *E. brasiliensis* collections. The Paraibuna collection consisted of three genetic groups (2, 4, and 6), exhibiting substantial mixing. Conversely, the Piracicaba collection was composed of two distinct genetic groups (3 and 5). The Natividade da Serra collection contained only one genetic group (1) (Figure 2C).

#### 3.2.1 *Eugenia brasiliensis* core collections

To identify the smallest set of accessions that represent the available genetic diversity identified in this study for *E. brasiliensis*, three independent sample proportions were constructed (Collection 1, 2, and 3), with sizes varying from 10% to 30% of the entire dataset (Table 4). The largest sample, composed of 23 individuals (C3), managed to capture 99% of the 4598 alleles detected, while the smallest sample, C1, captured 4486 alleles, approximately 5% less than the total number of alleles detected. The C1 subcollection comprises one individual from Natividade da Serra, two from Piracicaba and five individuals from Paraibuna. C2 is made up of two individuals from Natividade da Serra, two from Piracicaba and 11 from Paraibuna. The C3 subcollection comprises two individuals from Natividade da Serra, two from Piracicaba, and 19 individuals from Paraibuna.

The genetic diversity indices obtained for the collections were similar or higher than those for the total number of individuals.  $H_o$  values ranged from 0.24 (C1 and C2) to 0.25 (C3), with the value for C3 being slightly above the value observed for all 73 accessions (0.24).  $H_e$  values varied from 0.28 in C3 to 0.30 in C1, which presented a  $H_e$  value equal to that of the total number of individuals ( $H_e = 0.30$ ). The C1, C2, and C3 subcollections captured between 95% and 99% of the alleles, making them suitable for use in conservation and breeding.

The representativeness of the nuclear collections was also evaluated by principal coordinate analysis (PCoA), which showed

TABLE 1 Estimates of genomic diversity and inbreeding of accessions of *Eugenia brasiliensis* (grumixama), *E. pyriformis* (uvaia), and *E. involucrata* (Rio Grande cherry).

<i>E. brasiliensis</i>									
Collections	N	%P	Na	A	Pa	Ho	He	f	f (IC 95%)
Natividade da Serra	11	61.07	1.61	1.37	90	0.30	0.21	-0.31	- 0.50/-0.31
Piracicaba	15	61.94	1.62	1.36	114	0.22	0.21	-0.04	- 0.31/0.15
Paraibuna	47	84.25	1.84	1.40	394	0.23	0.24	0.05	0.00/0.11
Average		69.09	1.69	1.38	-	0.25	0.22	-0.06	-
<i>E. pyriformis</i>									
Collections	N	%P	Na	A	Pa	Ho	He	f	f (IC 95%)
Rio Claro	13	72.8	1.72	1.29	0	0.18	0.20	0.06	- 0.04/0.15
Cabo Verde	42	92.2	1.92	1.33	193	0.20	0.22	0.09	0.04/0.13
Inconfidentes	24	67.5	1.67	1.29	49	0.19	0.18	-0.08	- 0.12/- 0.04
Esalq	14	69.5	1.69	1.29	43	0.15	0.18	0.14	0.03/0.20
Média	-	75.49	1.75	1.30	-	0.18	0.19	0.05	-
<i>E. involucrata</i>									
Collections	N	%P	Na	A	Pa	Ho	He	f	f (IC 95%)
Piracicaba 1	20	47.0	1.21	1.19	5	0.14	0.11	-0.22	-0.37/-0.06
Rio Claro	4	62.9	1.47	1.38	2	0.23	0.23	-0.02	-0.30/0.13
Jundiá 1	5	76.6	1.53	1.44	0	0.31	0.26	-0.16	-0.40/-0.05
Jundiá 2	3	53.6	1.44	1.35	0	0.27	0.20	-0.34	-0.78/-0.16
Piracicaba 2	9	76.1	1.43	1.36	0	0.22	0.23	0.08	-0.20/0.19
Inconfidentes	5	64.8	1.47	1.39	2	0.21	0.23	0.06	-0.34/0.25
Paraibuna	16	79.3	1.43	1.37	3	0.20	0.23	0.09	0.00/0.21
Average	-	65.8	1.66	1.35	-	0.22	0.21	-0.04	-

The accessions of *E. brasiliensis* were sampled in three collections and analyzed using 2299 single-nucleotide polymorphism (SNP) markers. The accessions of *E. pyriformis* derived from four collections and were analyzed using 2872 SNPs. The *E. involucrata* accessions were sampled in seven collections and were analyzed using 1471 SNPs.

Number of individuals (N); percentage of polymorphic loci (%P); number of alleles (Na); number of alleles per locus (A); Private alleles (Pa); observed (Ho) and expected (He) heterozygosities; fixation index (f); f (95% CI) = lower and upper 95% confidence intervals of fixation indices.

the distribution of the 73 accessions collected in the three sampled locations (Natividade da Serra, Piracicaba and Paraibuna) and the core collection assembled from the 2299 SNP loci is along the first two coordinates (Figure 3). The variation explained by the first two coordinates was 26.6%, indicating a general overlap of diversity between the accessions selected for the core collections and those sampled in the three locations. This result suggests that any of the three core collections would adequately represent the total genetic diversity of *E. brasiliensis* identified in this study.

### 3.3 Population genomic analysis of *E. pyriformis*

We used the 2,872 identified SNPs to evaluate the genetic diversity and population structure of the 93 accessions from the four collections sampled in São Paulo and Minas Gerais. The

genetic diversity (*He* and *Ho*) was similar between the *E. pyriformis* collections (Table 1). *He* ranged from 0.18 (Inconfidentes and Esalq) to 0.22 (Cabo Verde), with an average of 0.19. The observed heterozygosity (*Ho*) average was 0.18, and ranged from 0.15 (Esalq) to 0.22 (Cabo Verde). The highest number of private alleles (*Pa*) was observed for the Cabo Verde collection (193), and the lowest for the Rio Claro collection, which did not contain any private alleles. The Cabo Verde collection also exhibited the highest percent of polymorphic loci (%P), number of alleles per locus (*A*), and number of alleles (*Na*). In the other collections, genetic diversity estimates were similar, with %P varying from 67.5% (Inconfidentes) to 72.8% (Rio Claro), *Na* ranging from 1.67 (Inconfidentes) to 1.72 (Rio Claro), and *A* being the same for all three locations. Regarding the fixation index (*f*), the Cabo Verde and Esalq collections presented positive values and significantly different from zero, indicating the presence of inbreeding ( $f = 0.09$  and  $f = 0.14$ , respectively). The Rio Claro

TABLE 2 Analysis of molecular variance (AMOVA) considering the distribution of genomic diversity between and within the three collections of *Eugenia brasiliensis* (grumixama) accessions based on 2299 single nucleotide polymorphism (SNPs) markers; between and within the four collections of *E. pyriformis* (uvaia) accessions, based on 2872 SNPs; and between and within the seven collections of *E. involucrata* (Rio Grande cherry) accessions, based on 1471 single nucleotide polymorphism markers.

Source of variation	SS	MS	%	$Phi_{ST}$
<b><i>Eugenia brasiliensis</i> (Grumixama)</b>				
Between collections	10501.85	5250.93	29	0.29*
Within collections	45648.23	319.22	71	
<b><i>E. pyriformis</i> (Uvaia)</b>				
Between collections	7800.86	2600.28	10	0.10*
Within collections	83628.96	459.50	90	
<b><i>E. involucrata</i> (Rio Grande cherry)</b>				
Between collections	7738.25	1289.71	23	0.23*
Within collections	26009.91	222.31	77	

Sum of squares (SS); mean square (MS); percentage of variation (%); Wright fixation index ( $Phi_{ST}$ ); \*significant at  $p < 0.001$ .

collection ( $f = 0.06$ ) also showed a positive fixation index; but it was not significant. The Inconfidentes collection presented  $f = -0.08$  (with a 95% confidence interval not including zero), indicating a partiality for heterozygotes in this collection.

The analysis of molecular variance revealed that the largest proportion of genetic variation was found within the *E. pyriformis* collections (Table 2), with moderate divergence between the collections, with  $Phi_{ST} = 0.10$  ( $p < 0.001$ ). Paired  $Phi_{ST}$  estimates between *E. pyriformis* collections indicate a moderate to high level of genetic divergence (Table 5). The largest divergence was observed between the Esalq and Inconfidentes collections ( $Phi_{ST} = 0.26$ ), and the smallest, between the Cabo Verde and Rio Claro collections ( $Phi_{ST} = 0.06$ ).

The genetic structure analysis, performed using PCA based on the 2872 SNPs, explained 19.8% of the total variation in the first two principal components (Figure 4A). This analysis agreed with the paired  $Phi_{ST}$  estimates, as it is possible to observe greater genetic divergence between the Esalq and Inconfidentes collections compared to the other collections, and that individuals from Cabo Verde and Rio Claro overlapped, justifying the smaller divergence between these collections. Although individuals from the Esalq collection showed some overlap with the Rio Claro collection, they are still very distinct.

TABLE 3 Paired estimates of  $Phi_{ST}$  (Weir and Cockerham, 1984) between collections of *Eugenia brasiliensis* (grumixama) accessions based on 2299 single nucleotide polymorphism markers.

Collections	Natividade da Serra	Piracicaba
Piracicaba	0.37*	
Paraibuna	0.30*	0.35*

\*significant with  $p < 0.001$ .

The optimal number of genetic groups estimated by the *K*-means method within the set of 93 *E. pyriformis* accessions was  $K = 5$  (Supplementary Figure S2). In this way, five groups were plotted using DAPC, accounting for 75.8% of the total variation in the first two principal components (Figure 4B). Scatter plot analysis indicated the formation of three arbitrary clusters, in which groups 1 and 5 are more easily distinguishable than groups 2, 3, and 4, which overlap. This analysis, as well as the PCA, also showed greater genetic divergence between the Inconfidentes and Esalq collections, since the Inconfidentes' accessions were all assigned to group 1, while those from Esalq were mostly assigned to group 5 by DAPC. Likewise, the Cabo Verde and Rio Claro accessions were assigned to groups 2, 3 and 4 (Figure 4B). The probability of adherence for each individual ranged from 95 to 100%, indicating that the identified genetic groups showed little mixing (Figure 4C).

### 3.4 Population genomic analysis of *E. involucrata*

We utilized the 1,471 SNP markers identified by GBS to assess genetic diversity and population structure in the 62 *E. involucrata* accessions collected across seven collections. The average value of  $He$  and  $Ho$  for the seven collections was 0.21 and 0.22, respectively. Jundiá 1 showed the highest genetic diversity ( $He = 0.26$  and  $Ho = 0.31$ ), while Piracicaba 1 showed the lowest ( $He = 0.11$  and  $Ho = 0.14$ ) (Table 1). The average number of alleles per locus ( $A$ ) was 1.35. The lowest  $A$  was found in Piracicaba 1 (1.19) and the highest in Jundiá 1 (1.44). The number of alleles ( $Na$ ) ranged from 1.21 (Piracicaba 1) to 1.53 (Jundiá 1), with an average of 1.66. The Jundiá 1 and 2 and Piracicaba 2 collections did not present any private allele. The highest  $Pa$  was observed in Piracicaba 1 ( $Pa = 5$ ). The average fixation index ( $f$ ) for all seven collections was -0.04, indicating a relatively low overall level of inbreeding. The fixation indices were negative for the Piracicaba 1, Jundiá 1, and Jundiá 2 collections ( $f = -0.22$ , -0.16, and -0.34, respectively), suggesting an excess of heterozygotes in these collections. The collections from Rio Claro, Piracicaba 2, Inconfidentes and Paraibuna presented non-significant  $f$  values.

The AMOVA results revealed genetic differentiation among *E. involucrata* collections, with  $Phi_{ST} = 0.23$ , representing 23% of the total variation, with most variation occurring within collections (Table 2).  $Phi_{ST}$  values between collections ranged from 0.07 to 0.57, indicating a moderate to high level of genetic divergence (Table 6). All comparisons between the Piracicaba 1 collection and the other collections showed the highest  $Phi_{ST}$  values, indicating a higher level of genetic divergence in this collection in relation to the others. This divergence is best visualized by PCA (Figure 5A), where a clear separation of the Piracicaba 1 collection from the other collections is noticeable.

The clustering based on the *K*-means method showed  $K = 4$  (Supplementary Figure S3). The four groups formed in DAPC based on the *K*-means retained 95.8% of the total variation in the first two principal components for the 62 *E. involucrata* accessions (Figure 5B). In the scatterplot of the DAPC analysis, it is possible

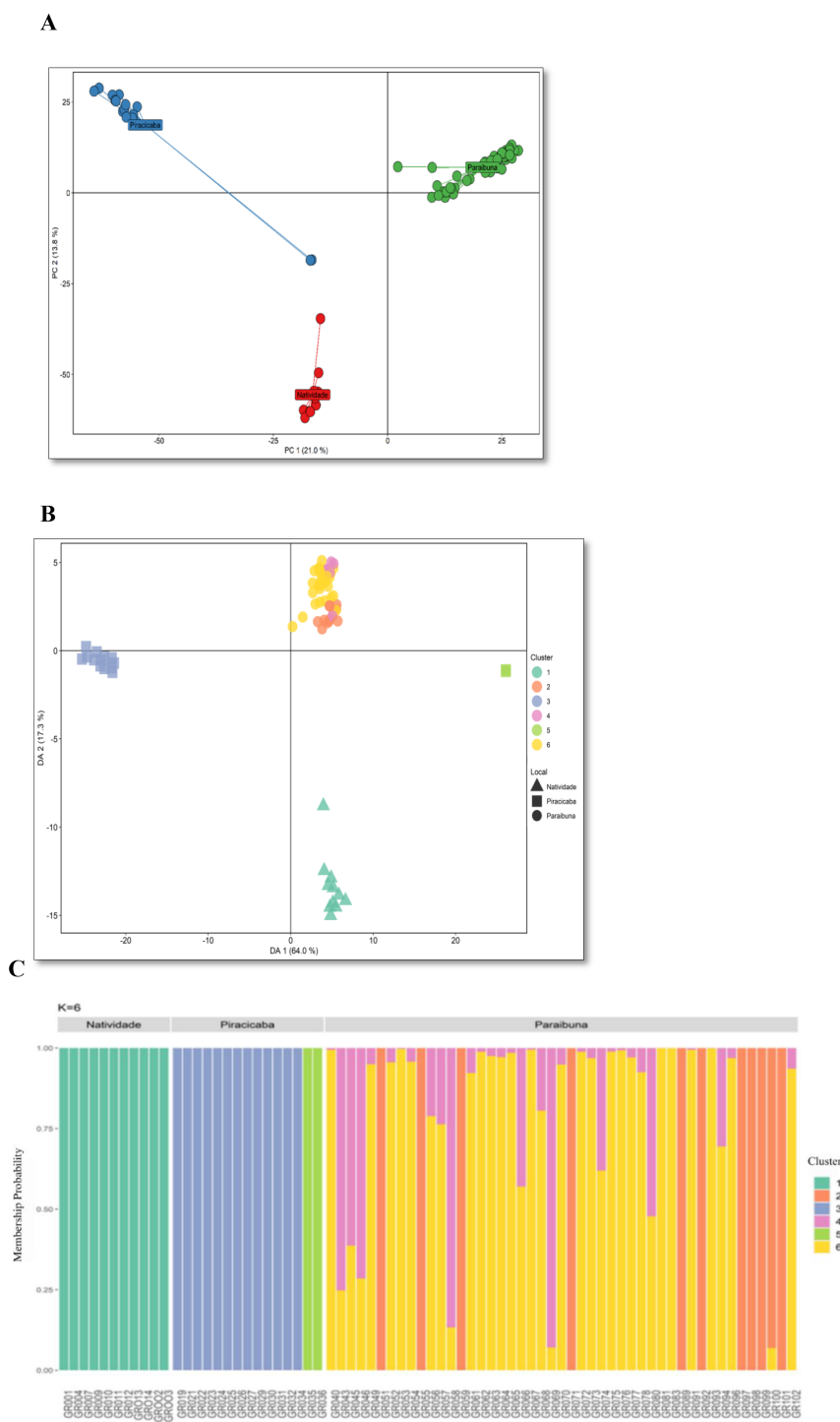


FIGURE 2

(A) Principal component analysis (PCA) representing the genetic structure of three *Eugenia brasiliensis* (grumixama) collections based on 2299 single nucleotide polymorphism markers. In red, individuals from the Natividade da Serra collection ( $n = 11$ ), in blue, individuals from the Piracicaba collection ( $n = 15$ ), and in green, individuals from the Paraibuna collection ( $n = 50$ ). (B) Discriminant analysis of principal components (DAPC) for 73 *E. brasiliensis* accessions collected in Natividade da Serra, Piracicaba and Paraibuna. The axis represents the first two main components of the Discriminant Analysis (DA). Each shape represents a site of collection and each color represents the different subpopulations identified by the DAPC analysis. (C) Graph representing the probability of membership of each individual from the three *E. brasiliensis* collections to specific genetic groups using the *K-means* method.

TABLE 4 Estimates of genetic diversity for the total number of individuals and proposed nuclear collections of *Eugenia brasiliensis* (grumixama), based on 2299 single nucleotide polymorphism markers.

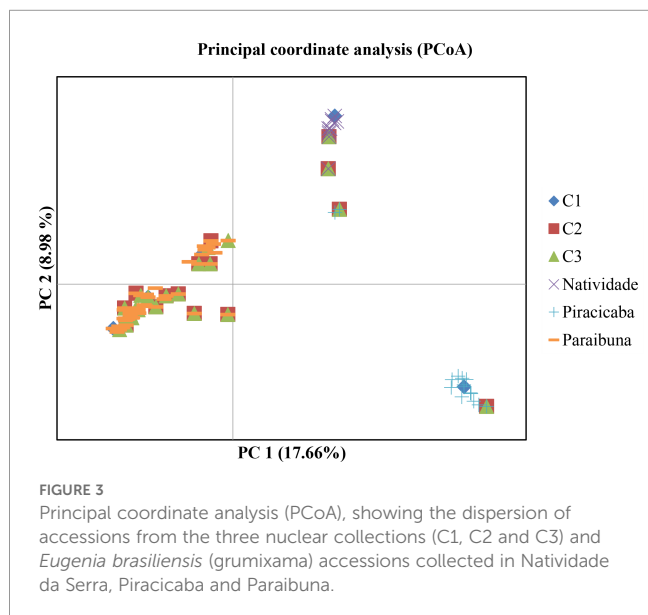
Nuclear collection	N	Na	Ho	He	f	TDSC (%)
C1	8	4486	0.25	0.30	0.12	95
C2	15	4554	0.24	0.29	0.15	99
C3	23	4567	0.24	0.28	0.16	99
Total number of individuals	73	4598	0.24	0.30	0.19	100

Number of individuals (N); number of alleles (Na); observed (Ho) and expected (He) heterozygosities; fixation index (f); total SNP diversity captured (TDSC).

to observe that some groups, such as groups 1 and 4, are more easily distinguishable than others. Group 4 was formed exclusively by individuals from the Piracicaba 1 collection, reinforcing the greater genetic divergence suggested by the  $Phi_{ST}$  and PCA values. Group 1 was formed by individuals from the Jundiá 2 and Piracicaba 2 collections, that also showed greater genetic divergence in relation to the others. There is also a small overlap of individuals between groups 2 and 3, showing less genetic divergence between these groups, which were formed by individuals from the Rio Claro and Paraibuna collections (group 2), and by individuals from Jundiá 1, Inconfidentes and Paraibuna (group 3). The association between the origin of the access (collection) and the four genetic groups formed can be better visualized by the probability graph of individuals joining each group (Figure 5C).

## 4 Discussion

This is a pioneering study on the genetic diversity of three *Eugenia* species, *E. brasiliensis* (grumixama), *E. pyriformis* (uvaia), and *E. involucrata* (Rio Grande cherry), which are promising new fruit crops, based on SNP markers. GBS has been successfully used



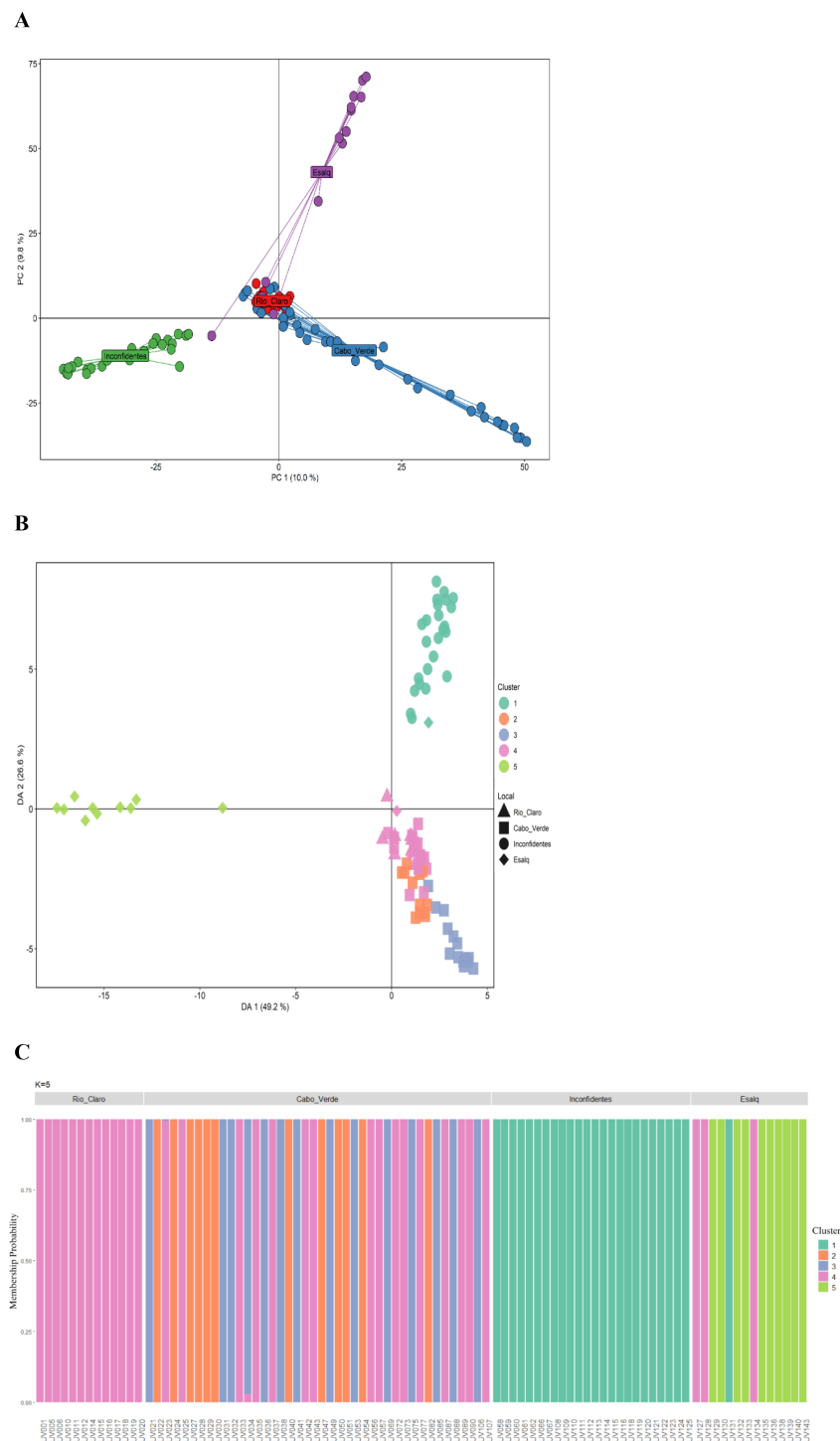
to discover SNPs for studies of genomic diversity and population structure in a wide range of species, such as *Camelina sativa* (Luo et al., 2019), *Capsicum* spp (Pereira-Dias et al., 2019), *Phaseolus vulgaris* (Delfini et al., 2021), *Acrocomia* spp (Diaz et al., 2021), *Psidium guajava* (Diaz-Garcia and Padilla-Ramirez, 2023) and *Elaeis oleifera* (Kunth) Cortés (Leão et al., 2022) to name a few. Here, GBS was used to identify SNP markers and evaluate the patterns of genomic diversity and population structure of 73 *E. brasiliensis* accessions, 93 *E. pyriformis* accessions, and 62 *E. involucrata* accessions, maintained by research institutions and small growers, used for urban afforestation and production in São Paulo and Minas Gerais. The number of SNPs revealed ranged from 1,471 for *E. pyriformis* to 2872 SNPs for *E. pyriformis*, a large number of markers to use in genetic diversity studies.

The analyses revealed that the average level of genetic diversity ( $He$ ) was similar among the three species (Table 1). The  $He$  values obtained in this study are lower than those reported by Ferreira-Ramos et al. (2014), who genotyped 26 *E. pyriformis* accessions and 16 *E. brasiliensis* accessions using microsatellite markers, and with average values of 0.75 and 0.54 for  $He$ , respectively. Stefanel et al. (2021) also found higher  $He$  values (0.62) for *E. involucrata*, using microsatellite markers. When the  $He$  values observed in this study are compared with those reported for other Myrtaceae species genotyped using SNP markers, the discrepancies between the estimates are reduced. For example,  $He = 0.22$  for *Eucalyptus camaldulensis* (Dillon et al., 2014) and  $He = 0.30$  for *E. urophylla* (Yang et al., 2020). Similar,  $He$  was also reported by Izuno et al. (2017) for *Metrosideros polymorpha*, which observed  $He$  varying from 0.19 to 0.25 using SNPs. Such discrepancies are expected and can be explained by considering the nature of the markers: SSRs are multiallelic and more polymorphic than biallelic SNP markers (Singh et al., 2013). On the other hand, SNPs occur at a much higher density throughout the genome and have lower genotyping error rates (Hamblin et al., 2007). Furthermore, SNPs more accurately reflect actual patterns of genetic diversity and genome structure than microsatellite markers (Telfer et al., 2015; Silva et al., 2020).

The observed mean heterozygosity ( $Ho$ ) for the *E. brasiliensis* ( $Ho = 0.25$ ) and *E. involucrata* ( $Ho = 0.22$ ) collections were slightly higher than the  $He$  found for these species. For the *E. pyriformis* collections, the value ( $Ho = 0.18$ ) was slightly lower (Table 1). Excess heterozygotes in *E. brasiliensis* and *E. involucrata* were also reported in previous studies using SSRs (Ferreira-Ramos et al., 2014; Stefanel et al., 2021). In the case of *E. pyriformis*, a low level of homozygotes was also reported by Ferreira-Ramos et al. (2014). These results are consistent with the  $f$  values observed in this study.

TABLE 5 Paired  $Phi_{ST}$  estimates (Weir and Cockerham, 1984) between collections of *Eugenia pyriformis* (uvaia) accessions based on 2872 single nucleotide polymorphism markers.

Collections	Rio Claro	Cabo Verde	Inconfidentes
Cabo Verde	0.06		
Inconfidentes	0.16	0.15	
Esalq	0.15	0.17	0.26



**FIGURE 4**  
**(A)** Principal component analysis (PCA) representing the genetic structure of four *Eugenia pyriformis* (uvaia) collections, based on 2872 single nucleotide polymorphism markers. In red individuals from the Rio Claro collection ( $n = 16$ ), in purple the Esalq collection ( $n = 20$ ), in blue the Cabo Verde collection ( $n = 53$ ) and in green individuals from the Inconfidentes collection ( $n = 22$ ). **(B)** Discriminant analysis of principal components (DAPC) for 93 *E. pyriformis* accessions collected in Rio Claro, Cabo Verde, Inconfidentes and Esalq. The axis represent the first two main components of the Discriminant Analysis (DA). Each shape represents a source collection and each color represents the different groups identified by the DAPC. **(C)** Graph representing the probability of membership of each individual from the four *E. pyriformis* collections to the genetic groups determined by the *K*-means method.

TABLE 6 Paired  $Phi_{ST}$  estimates (Weir and Cockerham, 1984) among 62 *Eugenia involucrata* (Rio Grande cherry) accessions based on 1471 single nucleotide polymorphism markers.

Collections	Piracicaba 1	Rio Claro	Jundiá 1	Jundiá 2	Piracicaba 2	Inconfidentes
Rio Claro	0.57					
Jundiá 1	0.36	0.09				
Jundiá 2	0.57	0.13	0.14			
Piracicaba 2	0.52	0.12	0.15	0.14		
Inconfidentes	0.45	0.15	0.09	0.16	0.14	
Paraibuna	0.47	0.07	0.09	0.13	0.17	0.15

The fixation index ( $f$ ) is one of the most important parameters in population genetics, as it measures the balance between homozygotes and heterozygotes in populations. The average  $f$  found for each of the species evaluated was -0.06 for the three *E. brasiliensis* collections, 0.05 for the four *E. pyriformis* collections, and -0.04 for the seven *E. involucrata* collections (Table 1). The relatively low values of inbreeding found may be attributed to the higher rate of cross-fertilization in these species. Although there are no detailed studies on the mating system of *E. brasiliensis*, *E. pyriformis*, or *E. involucrata*, the occurrence of higher rates of cross-fertilization has been reported for other species of the genus *Eugenia*, such as *E. uniflora*, *E. punicifolia*, *E. neonitida*, and *E. rotundifolia* (Silva and Pinheiro, 2009; Diniz and Buschini, 2016). Additionally, the fact that bees are the most important pollinators of these species (Gressler et al., 2006) may, in part, explain the relatively low level of inbreeding found.

Furthermore, the population differentiation found by AMOVA indicated greater variation within collections (Table 2), a common characteristic in cross-pollinated plants, which maintains their genetic variability distributed within populations, suggesting the predominance of cross-hybridization in these species. Sample size and/or genetic drift could also explain the negative fixation index found in some collections. However, other hypotheses for negative values of  $f$  can be highlighted, including the mixing of previously isolated populations and/or the presence of hybrids (Zalapa et al., 2010), negative selective mating, when reproduction occurs between individuals with more different phenotypes than by chance (Stoeckel et al., 2006), and unconscious selection of more heterozygous individuals to form collections.

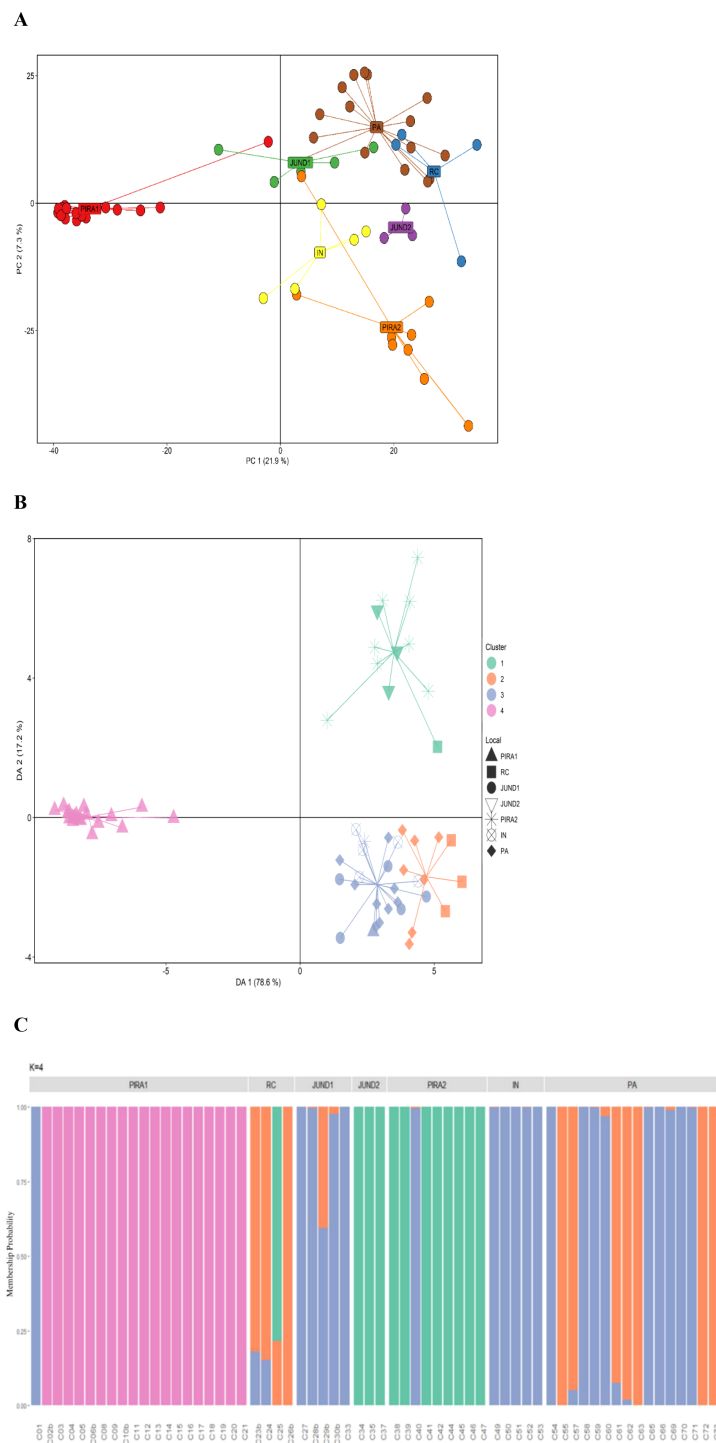
A high number of private alleles ( $Pa$ ) was observed in the *E. brasiliensis* and *E. pyriformis* collections, with emphasis on the Paraibuna (*E. brasiliensis*) and Cabo Verde (*E. pyriformis*) collections, which presented the highest number of private alleles (Table 1). The presence of many private alleles in these collections may be attributed to genetic drift and the absence of gene flow resulting from isolation. Another hypothesis to be considered in relation to the  $Pa$  rate is that, as these are managed collections, their genetic constitution may be altered by human-made introductions. The *E. involucrata* collections showed few and/or no private alleles (Table 1). The low number of private alleles in *E. involucrata* collections must be related to the restricted number of individuals that make up such collections.

AMOVA revealed that most of the genetic variation was found within the collections for all three species: 71% for *E. brasiliensis*, 77% for *E. involucrata*, and 90% for *E. pyriformis*. The remaining variation (29%, 23%, and 10%, respectively) was observed among collections within each species (Table 2). These results are consistent with the  $Phi_{ST}$  values obtained ( $Phi_{ST} = 0.29, 0.23,$  and  $0.10,$  respectively;  $P > 0.01$ ), which indicate moderate to very high genetic differentiation between collections. According to Hartl and Clark (2010),  $Phi_{ST}$  values can be classified into four categories: low differentiation (0–0.05), moderate (0.05–0.15), high (0.15–0.25), and very high (>0.25). The observed patterns are in line with previous findings for Myrtaceae species, including high within-population genetic variation in *E. involucrata* based on RAPD markers (Rejane et al., 2022), and similar results for *Campomanesia phaea* (Moreira et al., 2022), *Myrciaria floribunda* (Franceschinelli et al., 2007), and *Eugenia dysenterica* (Zucchi et al., 2003).

The moderate to high  $Phi_{ST}$  values suggest limited historical gene flow between the collections, likely due to geographic isolation and restricted genetic exchange. It is important to note that  $Phi_{ST}$  reflects the cumulative effects of past gene flow and does not capture contemporary reproductive events (Smouse and Sork, 2004). Additionally, since these are human-managed collections, the observed genetic structure may also reflect selective sampling, particularly from a small number of mother trees with desirable phenotypic traits. This may explain the modest admixture observed in some collections, such as Paraibuna (*E. brasiliensis*), Cabo Verde (*E. pyriformis*), Rio Claro, Jundiá 1, and Paraibuna (*E. involucrata*).

Overall, the results indicate strong genetic structuring among the *E. brasiliensis*, *E. pyriformis*, and *E. involucrata* collections, despite moderate levels of genetic diversity within them. These patterns, confirmed by Wright's F statistics as well as PCA and DAPC analyses, highlight the potential of these genetic resources for both conservation programs and future breeding initiatives targeting traits of agronomic interest.

The main objective of developing a core collection is to capture the maximum genetic diversity present in the germplasm in a reduced set of accessions (Egan et al., 2022). In this study, the first effort was made to create a core collection of *E. brasiliensis* accessions. To achieve the goal of representing at least 70% of the diversity present in the total sample, three possible basic *E. brasiliensis* collections were proposed (Odong et al., 2013).



**FIGURE 5**  
**(A)** Principal component analysis (PCA) representing the genetic structure of seven *Eugenia involucrata* (Rio Grande cherry) collections based on 1471 single nucleotide polymorphism markers. In red individuals from the collection of Piracicaba 1 (PIRA1,  $n = 21$ ) and in orange from Piracicaba 2 (PIRA2,  $n = 10$ ), in green Jundiá 1 (JUND1,  $n = 5$ ), in purple Jundiá 2 (JUND2,  $n = 4$ ), in blue Rio Claro (RC,  $n = 4$ ), in yellow Inconfidentes (IN,  $n = 5$ ) and in brown individuals from the Paraibuna collection (PA,  $n = 20$ ). **(B)** Discriminant analysis of principal components (DAPC) for 62 *E. involucrata* accessions collected from seven different locations. The axis represent the first two main components of the Discriminant Analysis (DA). Each shape represents a source collection and each color represents the different subpopulations identified by the DAPC analysis. **(C)** Graph representing the probability of membership of each individual from the seven *E. involucrata* collections to the genetic groups determined by the *K-means* method.

Nuclear collections account for more than 95% of the total allelic variability in sampled individuals. Furthermore, the proposed nuclear collections yielded genetic diversity estimates similar to those of the total individuals, indicating a good representation of these collections. These core collections may serve as a foundation for future studies focused on fruit quality traits and could be strategically used to initiate genetic breeding programs for the species at a lower maintenance cost.

In summary, our study is the first to evaluate the genetic diversity of *E. brasiliensis* (grumixama), *E. pyriformis* (uvaia), and *E. involucrata* (Rio Grande cherry) using SNP markers obtained by NGS technologies. As highlighted by Willing et al. (2012), the use of an extensive set of biallelic genetic markers is suitable for investigating genetic diversity, even when population samples are limited, as is the case in this study. The identified SNPs proved to be valuable even for *E. involucrata*, where sample collection was restricted in some locations and in a small geographic area. In general, the genetic diversity observed in this study tended to be lower than that reported in previous studies using second-generation markers, such as SSR, in *Eugenia* species. However, this discrepancy was expected, given that SNP markers, predominantly biallelic, have lower mutation rates than SSRs. On the other hand, SNPs are distributed at a higher density throughout the genome and are less prone to genotyping error, thus enabling more robust estimates of genetic diversity and genetic structuring.

The application of genetic evaluation based on NGS has expanded our understanding of the genetic diversity and population structure of these species, providing the basis for building a core collection for *E. brasiliensis* and opening new perspectives to assist future phenotyping studies in this species. *Eugenia brasiliensis* (grumixama), *E. pyriformis* (uvaia) and *E. involucrata* (Rio Grande cherry) stand out for their considerable economic potential. Therefore, the use of these population genomics data can be fundamental in defining best management practices for these species, to preserve and value their genetic resources.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/biostudies/studies/S-BSST2086?key=58c4f841-6648-4740-b7b1-1f28747ac341>, S-BSST2086 <https://www.ebi.ac.uk/biostudies/studies/S-BSST2087?key=22c49bcb-5085-4bfd-9dc4-866d4927bf7b>, S-BSST2087 <https://www.ebi.ac.uk/biostudies/studies/S-BSST2088?key=b5279de8-b15f-4ca7-8bd1-e45dd2f3526e>, S-BSST2088.

## Author contributions

LS: Data curation, Investigation, Writing – original draft, Writing – review & editing. MZ: Writing – review & editing, Investigation. CC: Writing – review & editing, Investigation. APJ: Project administration, Writing – review & editing, Funding

acquisition. AF: Writing – review & editing, Conceptualization. FAAMF: Project administration, Writing – review & editing, Conceptualization.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant No. 2014/126063 for financial support. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES): scholarship to the first author. Brazilian National Research and Technological Council (CNPq): scholarships to APJ, AF, and FAAMF. The authors thank the "Conselho Nacional de Desenvolvimento Científico e Tecnológico" for financial support granted to FAAMF (grant 303164/2018-2; 306677/2023-7).

## Acknowledgments

The authors also thank Sítio do Belo, Dr. Sergio Sartori, Grupo Genética e Genômica da Conservação (GGGC) and Laboratório de Melhoramento de Plantas (LAMP) for their technical support during this research.

## Conflict of interest

The authors declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declared that Antonio Figueira is an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1670349/full#supplementary-material>

### SUPPLEMENTARY TABLE 1

Location and number of accessions collected (*n*) of *Eugenia brasiliensis* (grumixama), *E. pyriformis* (uvaia) and *E. involucrata* (Rio Grande cherry). States: SP: São Paulo; MG: Minas Gerais.

### SUPPLEMENTARY FIGURE 1

Determination of the optimal number of genetic clusters (*K*) for *Eugenia brasiliensis* (grumixama). The Bayesian Information Criterion (BIC) is plotted

against the number of potential clusters (*K*). The analysis was based on 73 accessions and 2,299 SNPs. The lowest BIC value (404.0) indicates the optimal number of clusters, determined as *K* = 6.

### SUPPLEMENTARY FIGURE 2

Determination of the optimal number of genetic clusters (*K*) for *Eugenia pyriformis* (uvaia). The Bayesian Information Criterion (BIC) is plotted against the number of potential clusters (*K*). The analysis was based on 93 accessions and 2,872 SNPs. The lowest BIC value indicates (533.2) the optimal number of clusters, determined as *K* = 5.

### SUPPLEMENTARY FIGURE 3

Determination of the optimal number of genetic clusters (*K*) for *Eugenia involucrata* (Rio Grande cherry). The Bayesian Information Criterion (BIC) is plotted against the number of potential clusters (*K*). The analysis was based on 62 accessions and 1,471 SNPs. The lowest BIC value (326.1) indicates the optimal number of clusters, determined as *K* = 4.

## References

- Aguirre, N. C., Villalba, P. V., García, M. N., Filippi, C. V., Rivas, J. G., Martínez, M. C., et al. (2024). Comparison of ddRADseq and EUChip60K SNP genotyping systems for population genetics and genomic selection in *Eucalyptus dunnii* (Maiden). *Front Genet.* 15, 1361418. doi: 10.3389/fgene.2024.1361418
- Aguirre, N. C., Villalba, P. V., García, M. N., Filippi, C. V., Rivas, J. G., Martínez, M. C., et al. (2024). Comparison of ddRADseq and EUChip60K SNP genotyping systems for population genetics and genomic selection in *Eucalyptus dunnii* (Maiden). *Front Genet.* 15, 1361418. doi: 10.3389/fgene.2024.1361418
- Allendorf, F. W., Luikart, G., and Aitken, S. N. (2013). *Conservation and the genetics of populations*. Available online at: <https://www.perlego.com/book/1002490/conservation-and-the-genetics-of-populations-pdf> (Accessed April 19, 2025).
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (Accessed May 02, 2025).
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28
- Araújo, N. M. P., Berni, P., Zandoná, L. R., de Toledo, N. M. V., da Silva, P. P. M., de Toledo, A. A., et al. (2024). Potential of Brazilian berries in developing innovative, healthy, and sustainable food products. *Sustain. Food Technol.* 2, 506–530. doi: 10.1039/D3FB00130J
- Araújo, F. F., Neri-Numa, I. A., de Paulo Farias, D., da Cunha, G. R. M. C., and Pastore, G. M. (2019). Wild Brazilian species of *Eugenia* genera (Myrtaceae) as an innovation hotspot for food and pharmacological purposes. *Food Res. Int.* 121, 57–72. doi: 10.1016/j.foodres.2019.03.018
- Brhane, H., Haileselassie, T., Tesfaye, K., Ortiz, R., Hammenhag, C., Abreha, K. B., et al. (2022). Novel GBS-based SNP markers for finger millet and their use in genetic diversity analyses. *Front. Genet.* 13. doi: 10.3389/fgene.2022.848627/BIBTEX
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354
- De Beukelaer, H., Davenport, G. F., and Fack, V. (2018). Core Hunter 3: Flexible core subset selection. *BMC Bioinf.* 19, 203. doi: 10.1186/s12859-018-2209-z
- Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Ruas, P. M., Sant'Ana, G. C., Gepts, P., et al. (2021). Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. *Sci. Rep.* 11, 1–12. doi: 10.1038/s41598-021-82437-4
- Delgado, L. F., and Barbedo, C. J. (2012). Water potential and viability of seeds of *Eugenia* (Myrtaceae), a tropical tree species, based upon different levels of drying. *Braz. Arch. Biol. Technol.* 55, 583–590. doi: 10.1590/S1516-89132012000400014
- Diaz, B. G., Zucchi, M. I., Alves-Pereira, A., de Almeida, C. P., Moraes, A. C. L., Vianna, S. A., et al. (2021). Genome-wide SNP analysis to assess the genetic population structure and diversity of *Acrocomia* species. *PLoS One* 16, e0241025. doi: 10.1371/JOURNAL.PONE.0241025
- Diaz-García, L., and Padilla-Ramírez, J. S. (2023). Development of single nucleotide polymorphism markers and genetic diversity in guava (*Psidium guajava* L.). *Plants People Planet* 5, 58–69. doi: 10.1002/PPP3.10295
- Dillon, S., McEvoy, R., Baldwin, D. S., Rees, G. N., Parsons, Y., and Southerton, S. (2014). Characterisation of adaptive genetic diversity in environmentally contrasted populations of *Eucalyptus camaldulensis* dehn. (River red gum). *PLoS One* 9, e103515. doi: 10.1371/JOURNAL.PONE.0103515
- Diniz, M. E. R., and Buschini, M. L. T. (2016). Diversity of flower visiting bees of *Eugenia uniflora* L. (Myrtaceae) in fragments of Atlantic Forest in South Brazil. *Sociobiology* 63, 982–990. doi: 10.13102/SOCIOBIOLOGY.V63I3.982
- Egan, L. M., Conaty, W. C., and Stiller, W. N. (2022). Core collections: is there any value for cotton breeding? *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.895155
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, e19379. doi: 10.1371/journal.pone.0019379
- Farias, D. P., Neri-Numa, I. A., de Araújo, F. F., and Pastore, G. M. (2020). A critical review of some fruit trees from the Myrtaceae family as promising sources for food applications with functional claims. *Food Chem.* 306, 125630. doi: 10.1016/j.foodchem.2019.125630
- Ferreira-Ramos, R., Accoroni, K. A. G., Rossi, A., Guidugli, M. C., Mestriner, M. A., Martinez, C. A., et al. (2014). Genetic diversity assessment for *Eugenia uniflora* L., *E. pyriformis* Lam., *E. brasiliensis* Lam. and *E. francavilleana* O. Berg neotropical tree species (Myrtaceae) with heterologous SSR markers. *Genet. Resour. Crop Evol.* 61, 267–272. doi: 10.1007/s10722-013-0028-7
- Flores, G., Dastmalchi, K., Paulino, S., Whalen, K., Dabo, A. J., Reynertson, K. A., et al. (2012). Anthocyanins from *Eugenia brasiliensis* edible fruits as potential therapeutics for COPD treatment. *Food Chem.* 134, 1256–1262. doi: 10.1016/J.FOODCHEM.2012.01.086
- Franceschinelli, E. V., Vasconcelos, G. M. P., Landau, E. C., Ono, K. Y., and Santos, F. A. M. (2007). The genetic diversity of *Myrciaria floribunda* (Myrtaceae) in Atlantic Forest fragments of different sizes. *J. Trop. Ecol.* 23, 361–367. doi: 10.1017/S0266467407004099
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5, 184–186. doi: 10.1111/J.1471-8286.2004.00828.x
- Gressler, E., Pizo, M. A., and Morellato, L. P. C. (2006). Polinização e dispersão de sementes em Myrtaceae do Brasil. *Rev. Bras. Botânica* 29, 509–530. doi: 10.1590/S0100-84042006000400002
- Hamblin, M. T., Warburton, M. L., and Buckler, E. S. (2007). Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS One* 2, e1367. doi: 10.1371/JOURNAL.PONE.0001367
- Hartl, D. L., and Clark, A. G. (2010). *Princípios de Genética de Populações*, 4 ed. (Porto Alegre, Brazil: Artmed).
- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. L., Ogden, R., Limborg, M. T., et al. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Resour.* 11, 123–136. doi: 10.1111/j.1755-0998.2010.02943.x
- Infante, J., Rosalen, P. L., Lazarini, J. G., Franchin, M., and De Alencar, S. M. (2016). Antioxidant and anti-inflammatory activities of unexplored Brazilian native fruits. *PLoS One* 11, e0152974. doi: 10.1371/journal.pone.0152974
- Inocente, M. C., and Barbedo, C. J. (2019). Germination of *Eugenia brasiliensis*, *E. involucrata*, *E. pyriformis*, and *E. uniflora* (Myrtaceae) under water-deficit conditions. *J. Seed Sci.* 41, 76–85. doi: 10.1590/2317-1545v41n1212109
- Izuno, A., Kitayama, K., Onoda, Y., Tsujii, Y., Hatakeyama, M., Nagano, A. J., et al. (2017). The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha* (Myrtaceae). *Mol. Ecol.* 26, 1515–1532. doi: 10.1111/MEC.14016

- Jombart, T., and Bateman, A. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/BIOINFORM/BTN129
- Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94. doi: 10.1186/1471-2156-11-94
- 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* 2, e281. doi: 10.7717/peerj.281
- 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 Ecol. Evol.* 4, 782–788. doi: 10.1111/2041-210X.12067
- Kumar, P., Choudhary, M., Jat, B. S., Kumar, B., Singh, V., Kumar, V., et al. (2021). Skim sequencing: an advanced NGS technology for crop improvement. *J. Genet.* 100, 38. doi: 10.1007/S12041-021-01285-3
- Lazarini, J. G., Sardi, J., de, C. O., Franchin, M., Nani, B. D., Freires, I. A., et al. (2018). Bioprospection of *Eugenia brasiliensis*, a Brazilian native fruit, as a source of anti-inflammatory and antibiofilm compounds. *Biomedicine Pharmacotherapy*. 102, 132–139. doi: 10.1016/j.biopha.2018.03.034
- Leão, A. P., Filho, J. A. F., Pereira, V. M., Alves, A. A., and Souza Júnior, M. T. (2022). Genomic characterization of SNPs for genetic differentiation and selection in populations from the American oil palm [*Elaeis oleifera* (Kunth) cortés] germplasm bank from Brazil. *Diversity* 14, 270. doi: 10.3390/D14040270/S1
- Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., et al. (2019). Genetic diversity and population structure of a *Camelina sativa* spring panel. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00184
- Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012, 728398. doi: 10.1155/2012/728398
- Moreira, R. O., de Andrade Bressan, E., Bremer Neto, H., Jacomino, A. P., Figueira, A., and de Assis Alves Mourão Filho, F. (2022). Genetic diversity of cambuci [*Campomanesia phaea* (O. Berg) Landrum] revealed by microsatellite markers. *Genet. Resour. Crop Evol.* 69, 1557–1570. doi: 10.1007/S10722-021-01318-X/METRICS
- Morin, P., Luikart, G., and Wayne, R. (2004). SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19, 208–216. doi: 10.1016/j.tree.2004.01.009
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yildiz, M., et al. (2017). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equipment*. 32, 261–285. doi: 10.1080/13102818.2017.1400401
- Nehring, P., Katia Tischer Seraglio, S., Schulz, M., Della Betta, F., Valdemiro Gonzaga, L., Vitali, L., et al. (2022). Grumixama (*Eugenia brasiliensis* Lamarck) functional phytochemicals: Effect of environmental conditions and ripening process. *Food Res. Int.* 157, 111460. doi: 10.1016/j.foodres.2022.111460
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *genetics* 89, 583–590. doi: 10.1093/genetics/89.3.583
- Odong, T. L., Jansen, J., van Eeuwijk, F. A., and van Hintum, T. J. L. (2013). Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor. Appl. Genet.* 126, 289. doi: 10.1007/S00122-012-1971-Y
- Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Pereira-Dias, L., Vilanova, S., Fita, A., Prohens, J., and Rodríguez-Burruezo, A. (2019). Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). *Hortic. Res.* 6, 54. doi: 10.1038/s41438-019-0132-8
- 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* 7, 32253. doi: 10.1371/JOURNAL.PONE.0032253
- R Core Team (2021). *R: A language and environment for statistical computing*. (Vienna, Austria: R Foundation for Statistical Computing). Available online at: <https://www.R-project.org/>.
- R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. Available online at: <https://www.R-project.org/>.
- Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., et al. (2017). Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol. Plant* 10, 1047–1064. doi: 10.1016/j.molp.2017.06.008
- Rasheed, A., and Xia, X. (2019). From markers to genome-based breeding in wheat. *Theor. Appl. Genet.* 132, 767–784. doi: 10.1007/S00122-019-03286-4
- Rejane, L., Reiniger, S., Pascoal Golle, D., Miguel, C., Serrote, L., Severo Da Costa, L., et al. (2022). Genetic Structure and Gene Flow of *Eugenia involucrata* DC. Populations and Collections from Rio Grande do Sul, Brazil. *Biodiversidade Bras. - BioBrasil*. 12, 1–13. doi: 10.37002/BIOBRASIL.V12I2.2012
- Salgotra, R. K., and Chauhan, B. S. (2023). Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* 14, 174. doi: 10.3390/GENES14010174
- Saliba, A. S. M. C., Rosalen, P. L., Franchin, M., da Cunha, G. A., de Oliveira Sartori, A. G., and de Alencar, S. M. (2025). Fruits native to South America: a narrative review of biological properties and chemical profile. *Food Funct.* 16, 3774–3799. doi: 10.1039/D5FO00549C
- Sardi, J., de, C. O., Freires, I. A., Lazarini, J. G., Infante, J., de Alencar, S. M., et al. (2017). Unexplored endemic fruit species from Brazil: Antibiofilm properties, insights into mode of action, and systemic toxicity of four *Eugenia* spp. *Microb. Pathog.* 105, 280–287. doi: 10.1016/J.MICPATH.2017.02.044
- Schmidt, H., de, O., Rockett, F. C., Pagno, C. H., Possa, J., Assis, R. Q., et al. (2019). Vitamin and bioactive compound diversity of seven fruit species from south Brazil. *J. Sci. Food Agric.* 99, 3307–3317. doi: 10.1002/jsfa.9544
- Sereno, M. L., Albuquerque, P. S. B., Vencovsky, R., and Figueira, A. (2006). Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian amazon evaluated by microsatellite markers. *Conserv. Genet.* 7, 13–24. doi: 10.1007/s10592-005-7568-0
- Sganzerla, W. G., da Silva, A. P. G., Castro, L. E. N., da Rosa, C. G., Komatsu, R. A., Nunes, M. R., et al. (2022). Chemometric approach based on multivariate analysis for discriminating uvaia (*Eugenia pyriformis* Cambess) fruits during the ripening stages: Physicochemical characteristics, bioactive compounds, and antioxidant activity. *JSAFA Rep.* 2, 178–186. doi: 10.1002/JSF2.39
- Silva, A. L. G., and Pinheiro, M. C. B. (2009). Reproductive success of four species of *Eugenia* L. (Myrtaceae). *Acta Bot. Brasiliica*. 23, 526–534. doi: 10.1590/S0102-33062009000200024
- Silva, P. I. T., Silva-Junior, O. B., Resende, L. V., Sousa, V. A., Aguiar, A. V., and Grattapaglia, D. (2020). A 3K Axiom SNP array from a transcriptome-wide SNP resource sheds new light on the genetic diversity and structure of the iconic subtropical conifer tree *Araucaria angustifolia* (Bert.) Kuntze. *PLoS One* 15, e0230404. doi: 10.1371/JOURNAL.PONE.0230404
- Silva, A. P. G., Spricigo, P. C., Purgatto, E., de Alencar, S. M., Sartori, S. F., and Jacomino, A. P. (2019). Chemical composition, nutritional value and bioactive compounds in six uvaia accessions. *Food Chem.* 294, 547–556. doi: 10.1016/j.foodchem.2019.04.121
- Silva, A. P. G., Sganzerla, W. G., Jacomino, A. P., Da Silva, E. P., Xiao, J., and Simal-Gandara, J. (2022). Chemical composition, bioactive compounds, and perspectives for the industrial formulation of health products from uvaia (*Eugenia pyriformis* Cambess – Myrtaceae): A comprehensive review. *Journal of Food Composition and Analysis* 109, 104500. doi: 10.1016/j.jfca.2022.104500
- Silva, A. P. G., Sganzerla, W. G., Jacomino, A. P., Da Silva, E. P., Xiao, J., and Simal-Gandara, J. (2022). Chemical composition, bioactive compounds, and perspectives for the industrial formulation of health products from uvaia (*Eugenia pyriformis* Cambess – Myrtaceae): A comprehensive review. *J. Food Compos. Anal.* 109:104500. doi: 10.1016/j.jfca.2022.104500
- Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., et al. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One* 8. doi: 10.1371/journal.pone.0084136
- Smouse, P. E., and Sork, V. L. (2004). Measuring pollen flow in forest trees: an exposition of alternative approaches. *For. Ecol. Manage.* 197, 21–38. doi: 10.1016/J.FORECO.2004.05.049
- Soares, J. C., Rosalen, P. L., Lazarini, J. G., Massarioli, A. P., da Silva, C. F., Nani, B. D., et al. (2019). Comprehensive characterization of bioactive phenols from new Brazilian superfruits by LC-ESI-QTOF-MS, and their ROS and RNS scavenging effects and anti-inflammatory activity. *Food Chem.* 281, 178–188. doi: 10.1016/J.FOODCHEM.2018.12.106
- Souza, R. G., Dan, M. L., Dias-Guimarães, M. A., Guimarães, L. A. O. P., and Braga, J. M. A. (2018). Fruits of the Brazilian Atlantic Forest: allying biodiversity conservation and food security. *Acad. Bras. Cienc.* 90, 3583–3595. doi: 10.1590/0001-3765201820170399
- Spricigo, P. C., Almeida, L. S., Ribeiro, G. H., Correia, B. S., Taver, I. B., Jacomino, A. P., et al. (2023). Quality attributes and metabolic profiles of uvaia (*Eugenia pyriformis*), a native Brazilian Atlantic forest fruit. *Foods* 12, 1881. doi: 10.3390/foods12091881
- Stefanel, C. M., Reiniger, L. R. S., Serrote, C. M. L., Stefenon, V. M., and Lemos, R. P. M. (2021). Variability and genetic structure in fragments of *Eugenia involucrata* De Candolle established through microsatellite markers. *Ciec. Rural.* 51, 2021. doi: 10.1590/0103-8478CR20200008
- Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., and Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species — *Prunus avium* L. *Mol. Ecol.* 15, 2109–2118. doi: 10.1111/j.1365-294X.2006.02926.X
- Telfer, E. J., Stovold, G. T., Li, Y., Silva, O. B., Grattapaglia, D. G., and Dungey, H. S. (2015). Parentage reconstruction in eucalyptus nitens using SNPs and microsatellite markers: A comparative analysis of marker data power and robustness. *PLoS One* 10, e0130601. doi: 10.1371/JOURNAL.PONE.0130601
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641
- Willing, E. M., Dreyer, C., and van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One* 7, 42649. doi: 10.1371/JOURNAL.PONE.0042649
- Xu, K., Alves-Santos, A. M., Dias, T., and Naves, M. M. V. (2020). Grumixama (*Eugenia brasiliensis* Lam.) cultivated in the Cerrado has high content of bioactive compounds and great antioxidant potential. *Ciec. Rural.* 50. doi: 10.1590/0103-8478CR20190630

Yang, H., Liao, H., Zhang, W., and Pan, W. (2020). Genome-wide assessment of population structure and genetic diversity of *Eucalyptus urophylla* based on a multi-species single-nucleotide polymorphism chip analysis. *Tree Genet. Genomes* 16, 1–11. doi: 10.1007/S11295-020-1422-X/TABLES/5

Zalapa, J. E., Brunet, J., and Guries, R. P. (2010). ORIGINAL ARTICLE: The extent of hybridization and its impact on the genetic diversity and population structure of an

invasive tree, *Ulmus pumila* (Ulmaceae). *Evol. Appl.* 3, 157–168. doi: 10.1111/J.1752-4571.2009.00106.X

Zucchi, M. I., Vianello Brondani, R. P., Pinheiro, J. B., Chaves, L. J., Guedes Coelho, A. S., and Vencovsky, R. (2003). Genetic structure and gene flow in *Eugenia dysenterica* DC in the Brazilian Cerrado utilizing SSR markers. *Genet. Mol. Biol.* 26, 449–457. doi: 10.1590/S1415-47572003000400008