Genetic Structure and Gene Flow of *Eugenia involucrata* DC. Populations and Collections from Rio Grande do Sul, Brazil

Lia Rejane Silveira Reiniger[1], Diego Pascoal Golle[2], Caetano Miguel Lemos Serrote[3], Leonardo Severo da Costa[1], Caroline Borges Bevilacqua[1] & Charlene Moro Stefanel[1]

Recebido em 11/06/2021 – Aceito em 07/01/2022

[1] Universidade Federal de Santa Maria, Brasil. <liarsr@ufsm.br, leonardosev@gmail.com, carolinebevi@gmail.com, chastefanel@gmail.com>.

[2] Universidade de Cruz Alta, Brasil. <dgolle@unicruz.edu.br>.

[3] Universidade Lúrio, Moçambique. <serrotec@yahoo.com.br>.

ABSTRACT – *Eugenia involucrata* DC. is a Brazilian tree species with economic potential. However, informations on the distribution of its genetic variability are scarce. This knowledge is fundamental knowledge for planning a genetic conservation program. This study aimed to evaluate the genetic structure and gene flow in accessions of *E. involucrata* by using RAPD markers. Samples were collected from three *in situ* populations and two *ex situ* collections located in the Brazilian state of Rio Grande do Sul. Eight Random Amplified Polymorphic DNA (RAPD) primers were used to access the genetic variability. Sixty bands were generated of which 56 were polymorphic. There is low gene flow among the populations (*Nm* values ranged from 0.2752 to 1.384) which resulted in very high levels of genetic differentiation (F_{ST} values ranged from 0.153 to 0.476). The analysis of molecular variance (AMOVA) revealed that the five studied populations/collections are structured in two main groups, and the *in situ* populations have more genetic diversity (*He* values ranged from 0.19 to 0.24) in relation to *ex situ* collections (*He* values ranged from 0.08 to 0.10). These results will contribute to the advance of conservation programs for the species.

Keywords: Forest genetic resource; genetic variability; germplasm conservation; molecular markers.

Estrutura Genética e Fluxo Gênico de Populações e Coleções de *Eugenia involucrata* DC. do Rio Grande do Sul, Brasil

RESUMO – *Eugenia involucrata* DC. é uma espécie arbórea brasileira com potencial econômico. Porém, são escassas as informações a respeito da distribuição de sua variabilidade genética, cujo conhecimento é fundamental para o planejamento de programas para a conservação genética. Este estudo visou avaliar a estrutura genética e o fluxo gênico em acessos de *E. involucrata* usando marcadores RAPD. Foram amostrados indivíduos de três populações *in situ* e de duas coleções *ex situ* localizadas no estado brasileiro do Rio Grande do Sul. Foram usados oito *primers* RAPD para avaliar a variabilidade genética. Foram geradas 60 bandas, 56 das quais eram polimórficas. Foi estimado baixo fluxo gênico entre as populações (*Nm* de 0,2752 a 1,384), o que resultou em alta diferenciação genética (F_{ST} de 0,153 a 0,476). A análise de variância molecular (AMOVA) revelou que as cinco populações/coleções estudadas estão estruturadas em dois grupos principais, e as populações *in situ* possuem maior diversidade genética (*He* de 0,19 a 0,24) em relação às coleções *ex situ* (*He* de 0,08 to 0,10). Sugere-se o enriquecimento dessas coleções com germoplasma introduzido visando aumentar a variabilidade genética.

Palavras-chave: Recurso genético florestal; variabilidade genética; conservação de germoplasma; marcadores moleculares.

Estructura Genética y Flujo Génico de Poblaciones y Colecciones de *Eugenia involucrata* DC. de Rio Grande do Sul, Brasil

RESUMEN – *Eugenia involucrata* DC. es una especie arbórea brasileña con potencial económico. Sin embargo, existe poca información sobre la distribución de su variabilidad genética, conocimiento que es fundamental para planificar programas de conservación genética. Este estudio tuvo como objetivo evaluar la estructura genética y el flujo génico en accesiones de *E. involucrata* utilizando

marcadores RAPD. Se muestrearon individuos de tres poblaciones *in situ* y de dos colecciones *ex situ* ubicadas en no estado brasileiro do Rio Grande do Sul. Se utilizaron ocho cebadores RAPD para evaluar la variabilidad genética. Se generaron 60 bandas, 56 de las cuales eran polimórficas. Se estimó un bajo flujo génico entre poblaciones (*Nm* de 0,2752 a 1,384), lo que resultó en una alta diferenciación genética (F_{ST} de 0,153 a 0,476). El análisis de la varianza molecular (AMOVA) reveló que las cinco poblaciones/colecciones estudiadas se estructuran en dos grupos principales, y las poblaciones *in situ* tienen mayor diversidad genética (*He* de 0,19 a 0,24) en comparación con las colecciones *ex situ* (*He* de 0,08 a 0.10). Se sugiere el enriquecimiento de estas colecciones con germoplasma introducido para incrementar la variabilidad genética.

Palabras clave: Recurso genético forestal; variabilidad genética; conservación de germoplasma; marcadores moleculares.

Introduction

Eugenia involucrata DC. is a brazilian native fruit species belonging to the Myrtaceae family (Carvalho, 2008), which is still underexploited its great potential to integrate productive systems. The species, which is dispersed by birds, produces tasty and succulent fruits, which can be consumed *in natura* or processed, has the potential for recovery of degraded areas and for landscape interest, its leaves have phytotherapic properties, its good quality wood is used in the manufacture of agricultural tools (Lorenzi, 2002; Carvalho, 2008). In a recent study (Sardi *et al*., 2017), the extracts from the seeds and leaves of Eugenia species were found to have strong antifungal activity against *Candida albicans*, causer of candidiasis, probably due to the presence of phenolic compounds such as epicatechin and gallic acid in their composition. Like many other native forest species to Brazilian biomes, it is not yet commercially exploited, despite its potential to integrate productive systems (Eibl *et al*., 2017), due to a lack of scientific and technological information, such as those related to the genetic structure of its populations, considering the ecological complexity and the diversity they present (Kageyama *et al*., 2003). This knowledge is indispensable for planning eventual *in situ* and *ex situ* conservation projects, for pre-breeding and improvement activities of this forest germplasm, and for producing seedlings of adequate genetic quality for forest restoration and recovering degraded areas (Botrel *et al*., 2006). Genetic structure can be accessed using DNA markers which are sequences or fragments of DNA that show polymorphism within a set of individuals. They can be accessed independently of the plant's development stage and the environment in which it grows and do not compromise the viability of the specimens under study as they require small amounts of plant tissue, allowing additional analyzes to be performed (Schulman, 2007).

However, the most frequently found limitation for using DNA markers in genetic diversity analyzes of native forest tree species in Brazilian biomes (in particular) has been the cost to develop specific primers. This economic obstacle can be overcome by the use of marker techniques that make use of universal primers such as RAPD (Random Amplified Polymorphic DNA) markers, despite some disadvantages they have: they are dominant (do not distinguish between heterozygotes and dominant homozygotes), have low reproducibility across experiments and/or laboratories (the simple use of different polymerase brands is enough to generate quite different patterns of amplification) and have limitation in assessing character homologies in systematic studies (Ali *et al*., 2004).

The successful application of RAPD markers has been reported in several forest species: in populations of *Eugenia pyriformis* Cambess (Sganzerla *et al.,* 2021), in populations of *Calycophyllum spruceanum* Benth. from the Peruvian Amazon (Saldaña et al., 2021), in natural populations of *Cedrelinga cateniformis* (Cruz *et al.,* 2020), in individuals of *Sterculia quadrifida* (Uslan, 2020), in five species of the genus *Eugenia* (Brunchault *et al*., 2014), in forest remnants of *Eugenia uniflora* L. (Aguiar *et al.,* 2013), in selected individuals from a natural population of *Genipa americana* (Rabbani *et al*., 2012), in accessions of *Hancornia speciosa* Gomes of the Active Germplasm Bank of the National Research Institute of Amazônia (*INPA*) (Costa *et al*., 2011), in an *ex situ* collections of *Bactris gasipaes* Kunth (Cristo-Araújo *et al*., 2010), in individuals of *Dimorphandra mollis* from seven locations in Minas Gerais State (Oliveira *et al*. 2008), in

natural populations of *Eremanthus erythropappus* in two different successional stages (Freitas *et al*., 2008), in individuals from a remnant population of *Enterolobium contortisiliquum* from a riparian forest of the Lower São Francisco River (Santana *et al*., 2008), in a natural population of *Aniba rosaeodora* in Manaus (Santos *et al*., 2008), in populations of *Maytenus ilicifolia* Mart. ex Reiss (Mossi *et al*., 2007), among other less recent works.

In the context of the present study, it is necessary to highlight the only work carried out with *Eugenia involucrata* thus far, in which genetic diversity of accessions collected in Rio Grande do Sul were evaluated by employing RAPD markers (Guerra *et al*., 2016). However, the referred study was carried out in different municipalities and used less primers than in our study.

Another aspect to be highlighted in the present study is the use of bulk samples, which can be constituted by sets of plant samples (leaves, seeds, exchanges, dilutions of DNA solutions). Since the groups of plants to be evaluated are often extremely large, which requires a high number of analyzes and consequently implying in increased costs, it is an alternative way to optimize the analysis of the genetic structure of populations. In a study of the genetic diversity of 18 natural populations of *Maytenus ilicifolia* collected in three Brazilian states, each population was represented by three bulk samples of 10 individuals, respecting the minimum distance of 10 m between each two plants (Mossi *et al*., 2007). However, the use of bulking samples leads to loss of information at the individual level, not being suitable for the selection of plus-plants for breeding programs.

In view of the above, the objective of the present study was to analyze the distribution of genetic variability between and within natural populations and *ex situ* collections of *Eugenia involucrata* DC. using RAPD markers in order to support programs for the genetic conservation of this species.

Material and Methods

Plant material and genomic DNA extraction

In this study, were analyzed individuals grouped in 27 bulk samples from three natural populations and two *ex situ* collections located in three municipalities in the state of Rio Grande do Sul (RS). Individuals in Frederico Westphalen (FW) and Santa Maria belonged to populations preserved *in situ*, while in Pelotas they were part of *ex situ* collections stored in the Active Germplasm Bank (AGB) and the Seedling Bank (SB) of the Brazilian Agricultural Research Corporation (Embrapa Temperate Climate). The individuals sampled in Frederico Westphalen were considered to belong to the same population, since the two fragments are only 1.3Km apart. The individuals sampled in Santa Maria are part of two populations, 10Km apart. The AGG and the SB in Pelotas are less than 100m apart.

Frederico Westphalen is located in the north of the state at 27°21'25"S latitude and 53°22'27"W longitude, at an altitude of 566m above sea level (ASL), where samples were randomly performed in two locations: at Vila Faguense, a largely anthropized location in which conservation of some forest fragments has occurred; and in the forest population of the Centre of Higher Education of Northern Rio Grande do Sul (*CESNORS*) of the Federal University of Santa Maria (*UFSM*).

Likewise, random collections were carried out at two sites in Santa Maria, which is located in the center of the state at 29°41'02"E latitude and longitude 53°48'25"W, at 95m ASL. The specific sites were in the area of the Army Military Training Center (CISM), located in the Boi Morto district, and in the Forestry Research Unit of the State Foundation for Agricultural Research (Fepagro Florestas), located in the Boca do Monte district.

Next, samples were collected from all existing nine accessions at the Embrapa AGB, located in Pelotas in the south of the state, at 3°46'19"S latitude and 52°20'33"W longitude, at 17m ASL, originating from planted seeds donated by the Brazilian Association of Tobacco Growers (Afubra). There was also performed a systematized sampling in the Seedling Bank originated from the seeds of the aforementioned accessions and collected plant material from one in five individuals from the planting line.

For the analyzes, several bulk samples were formed, each containing samples of three individuals from the populations/collections located in the three municipalities, from which were removed an equal aliquot of fresh (frozen) vegetable tissue (leaves or cambium) to isolate genomic DNA. Samples of the trees were used, which were located in the sequence within the

same radius of up to 60m to compose the bulk samples. Trees that were more than 100m apart made-up distinct bulk samples. Additional bulk samples were composed when there were eventually more than three trees within a radius of 60m; however, a bulk was created with two individuals when the specimen was isolated from the others by the distance and there were not three trees; or individual sampling was alternatively used, but this situation only occurred once.

To isolate genomic DNA, the method described by Zucchi (2002) was used with some modifications, especially in the initial stage of tissue maceration and homogenization in the presence of the extraction buffer. The foliar samples (600mg) were macerated in Gral with the aid of liquid nitrogen (-196°C), and then 2ml of extraction buffer (10% CTAB, 1.4 M NaCl; mM EDTA, pH 8.00, 100mM Tris-HCl, pH 8.00, 1% PVP-40), preheated at 65°C and containing 5% β-mercaptoethanol, were added. After initial homogenization, 1ml of extraction buffer (2% CTAB, 1.4M NaCl, 20mM EDTA, pH 8.00, 100mM Tris-HCl, pH 8.00, 1% PVP-40), also containing 5% β-mercaptoethanol, were added. Then, 1.2ml of the sample was transferred into a 2ml capacity eppendorf reaction tube for DNA isolation. For the cambium samples, 2ml of the 10% extraction buffer were added into a crucible containing 1200g of vegetal tissue, which were then macerated for approximately 15 minutes. Afterwards, 1.2ml of the macerate (mainly the liquid part) was transferred into a 2ml capacity eppendorf reaction tube for DNA isolation.

RAPD reactions

Eight RAPD primers (Operon Technologies Inc., Alameda, CA), selected according to the results obtained by Zucchi *et al.* (2005) with *Eugenia dysenterica* DC, were used to amplify the DNA samples. The amplification was carried out in 25μl reaction mix containing about 5ng of genomic DNA, 10mM of Tris-HCl, pH 8.3, 50mM of KCl, 2mM of $MgCl_2$, 0.4mM of each dNTP (Life Technologies, São Paulo, Brazil), 0.25μM of primer, and 1U of Taq DNA polymerase (Life Technologies, São Paulo, Brazil).

The DNA amplification was performed in a MJ Research PTC 100 thermal cycler according to the following thermal profile: initial denaturation at 94°C for 5min, followed by 48 cycles of denaturation at 92°C for 30s, annealing at 37°C for 1.5min and extension at 72°C for 1.5min, followed by final extension at 72°C for 5min. The amplification products were separated in 1.5% agarose gel and detected by staining with ethidium bromide. The gels were photographed under UV light. Each PCR reaction was run in duplicate, and only well-defined and reproducible bands were analyzed.

Data analysis

A binary matrix was elaborated based on the gels' reading, in which the bulk samples were genotyped regarding the presence (1) or absence (0) of bands in the gel. The polymorphism percentage generated by each of the primers used was calculated using the expression:

$$P = \frac{npb}{tnb}$$

where: P = polymorphism percentage; npb = number of polymorphic bands; and tnb = total number of bands. From the data matrix, the Jaccard similarity coefficient (Jaccard, 1908) was calculated for all bulk sample pairs, using the expression:

$$S_{ij} = \frac{a}{a+b+c}$$

where: a = number of bands occurring simultaneously in both individuals; b = number of bands occurring exclusively in the individual i; c = number of bands occurring exclusively in the individual j.

A genetic diversity analysis of the populations/collections was carried out using Popgene software v.1.31 (Yeh *et al.*, 1997) by estimating the average number of alleles per locus (A), average number of effective alleles per locus (Ae), as well as the polymorphic loci percentage (P) for each population and for the set of populations.

Using estimates of Jaccard's similarity coefficients, a dendrogram was constructed following the unweighted pair group method with arithmetic mean – UPGMA (Sokal & Michener, 1958). The consistency of the dendrogram was tested by the cophenetic correlation, whose significance was tested by the Mantel test with 1000 random permutations. The analysis of molecular

variance (AMOVA) was calculated by the total decomposition of its components between and within the sampled populations/collections, using the squared Euclidean distances (Excoffier *et al*., 1992) estimated by the Arlequin v.3.11 software (Excoffier *et al*., 2006).

There was also used a clustering method for inferring population structure in STRUCTURE v.2.3.4 (Pritchard *et al*., 2000; Falush *et al*., 2003, 2007; Hubisz *et al*., 2009) based on the Markov Chain Monte Carlo (MCMC) method. The model was set for 100000 replicates followed by 50000 burn-in length in order to determine the most likely number of clusters of individuals (K). Values for K varied from one to five ensuring five repetitions for each K. Both the posterior probability of the data for the given value of K (Ln Pr(X|K)) and its rate of change (**Δ**K) were used in detecting subpopulations (Evanno *et al*., 2005). A new AMOVA was run considering the genetic clusters found in the STRUCTURE analysis.

Finally, a third AMOVA was run considering all five sampled populations. The best suited population subdivision was selected by comparing the F_{CT} values from the three performed analyzes.

The Wright's differentiation index between the populations/collections (F_{ST}) obtained from the population structure was used to estimate the gene flow in terms of the average number of migrants per generation (*Nm*) through the expression (Nagel *et al*., 2015):

$$Nm = \frac{1}{4}\left(\frac{1}{F_{ST}}\right) - 1$$

where: F_{ST} = Wright's differentiation index between the populations/collections; *Nm* = number of migrants per generation.

Results

RAPD information content

Amplification of DNA sequences occurred by using the eight tested primers, thereby generating 60 bands, of which 56 (93.33%) were polymorphic (Table 1). The number of polymorphic bands ranged from four (OPA 01 and OPA 03 primers) to 10 (OPA 13 and OPA 19 primers), resulting in an average of 7.5 bands per primer, and a ratio of seven polymorphic bands per RAPD primer employed in the analyzes.

Table 1 – Polymorphism pattern for RAPD markers according to the primers used in molecular analysis of 27 bulk samples of *Eugenia involucrata* DC. accessions collected in Rio Grande do Sul, Brazil.

Primer code	Sequence (5' – 3')	Total number of observed bands	Number of polymorphic bands	Polymorphism (%)
OPA 01	CAGGCCCTTC	06	04	66.67
OPA 02	TGCCGAGCTG	07	07	100.00
OPA 03	AGTCAGCCAC	05	04	80.00
OPA 04	GGAAGTCGCC	06	05	83.33
OPA 11	CAATCGCCGT	08	08	100.00
OPA 13	CAGCACCCAC	10	10	100.00
OPA 19	CAAACGTCGG	10	10	100.00
OPM	GAAACACCCC	08	08	100.00
Total		**60**	**56**	**-**
Average		**7.5**	**7**	**93.33**

Phenetic diversity

The dendrogram generated from the similarity matrix elaborated from the Jaccard coefficients (Figure 1) resulted in three main distinct groups. The cophenetic correlation confirmed the consistency of the dendrogram, with a coefficient value of r = 0.95291. The first group (Group I) formed by the bulk samples from Fepagro Florestas (F2, F7, F8 and F9), and C4 and C8 from CISM

Biodiversidade Brasileira, 12(2): 1-13, 2022
DOI: 10.37002/biobrasil.v12i2.2012

(both populations located in Santa Maria) presented the lowest similarity. With regard to similarity, the intermediate grouping (Group II) is constituted by the other bulk samples of Fepagro (F1, F4 and F5), some of CISM (C3, C5 and C6) and all those of Frederico Westphalen (FW1 to FW4). In analyzing the subgroups of this grouping by origin, F1 and F4 are very similar, and considerably diverging from F5. Likewise, FW1 and FW2 are similar, showing a reasonable dissimilarity in relation to the FW4 bulk sample (of CESNORS) and with FW3 presenting intermediate dissimilarity. C3 and C5 are more similar to each other than to C6. However, F1 and F4 are more similar to C5 than to F5; while FW1, FW2 and FW3 are more similar to F5 in relation to FW4; and C3 is but similar to FW4 over C5. The third grouping (Group III) is the most uniform, has the greatest genetic similarity, and is composed by the accesses and the Seedling Bank conserved *ex situ* at Embrapa in Pelotas.

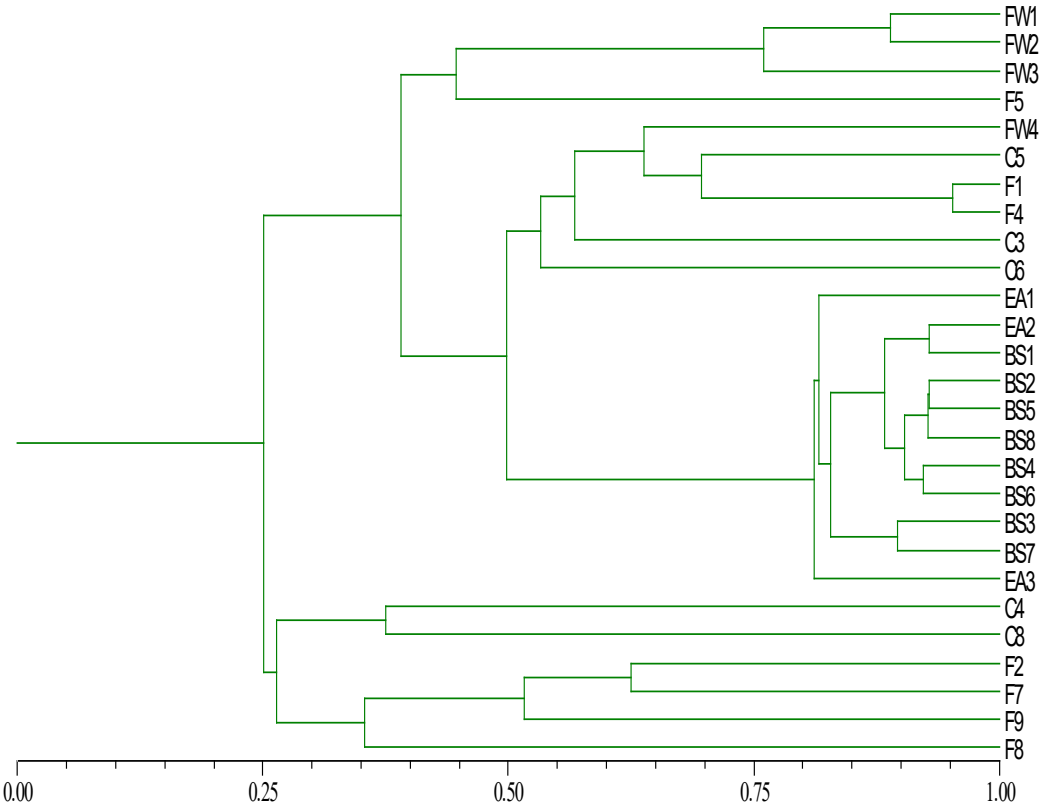

Figure 1 – Dendrogram elaborated from estimates of genetic similarity (Jaccard index), calculated by UPGMA pooling criterion, from 27 different bulk samples of *Eugenia involucrata* DC. accessions collected in the state of Rio Grande do Sul, Brazil. The acronyms represent the collected populations: Frederico Westphalen (FW), Fepagro (F), CISM (C), Embrapa accessions (EA) and Embrapa Seedling Bank (SB). Cophenetic correlation (r) = 0.95291.

Gene flow and genetic differentiation

The F_{ST} statistics, genetic differentiation among population pairs, and their geographic distances are summarized in Table 2. It should be noted that only the natural populations (FW, Fepagro and CISM) exchange genes, unlike the collections (EA and SB) which are active germplasm banks. The lowest genetic differentiation was observed between the Fepagro and CISM populations (F_{ST} = 0.153), which are also the closest ones (10Km). Considering the three natural populations, the correlation coefficient between genetic distance (F_{ST}) and geographic distances matrices was 0.999. The highest genetic differentiation levels were observed when collections were compared to natural populations, with F_{ST} values varying between 0.532 (EA and CISM) Rio Grande do Sul – Brazil and 0.712 (SB and FW). The F_{ST} value between the Embrapa AGB and the seedling bank was 0.277.

Table 2 – F_{ST} values (above the 0 diagonal) and straight-line geographic distances (below the 0 diagonal, in Km) based on RAPD markers between five *Eugenia involucrata* DC. populations/collections from the state of Rio Grande do Sul, Brazil.

	FW	Fepagro	CISM	EA	SB
FW	0	0.476	0.465	0.627	0.712
Fepagro	260	0	0.153	0.570	0.636
CISM	266	10	0	0.532	0.608
EA	501	281	271	0	0.277
SB	501	281	271	0.1	0

FW = Frederico Westephalen; EA = Embrapa accessions; SB = Seedlings Bank of Embrapa

Gene flow among natural populations in terms of number of migrants per generation (Table 3) was higher among the populations of Fepagro and CISM (Nm = 1.384), a fact that ratifies the importance of gene flow in reducing the genetic differentiation between populations. The number of migrants per generation for the other pairs was 0.2752 (FW and Fepagro) and 0.288 (FW and CISM). The dendrogram in Figure 1 ratifies this analysis, grouping the Santa Maria populations (CISM and Fepagro) and the collections (EA and SB) together, in addition to considering these two groups more similar to each other and more distant from FW.

Table 3 – F_{ST} values (above the diagonal) and number of migrants per generation (below diagonal, in Km) based on RAPD markers among three *Eugenia involucrata* populations from the state of Rio Grande do Sul, Brazil.

	Frederico Westephalen	Fepagro	CISM
Frederico Westephalen	0	0.476	0.465
Fepagro	0.2752	0	0.153
CISM	0.288	1.384	0

Genetic diversity and structure

The number of alleles (A) varied from 1.13 in the Embrapa accesses to 1.62 in Fepagro (Table 4). In turn, the number of effective alleles ranged from 1.10 in the Embrapa Seedling Bank to 1.43 in Fepagro. The Shannon diversity index was 0.08 in the Embrapa collections and 0.35 in the conserved accesses in Fepagro, constituting the minimum and maximum values, respectively. The minimum polymorphism percentage observed was 13.33% in the Embrapa collections, while the maximum was 61.67% in Fepagro, with an average of 96.67%.

Biodiversidade Brasileira, 12(2): 1-13, 2022
DOI: 10.37002/biobrasil.v12i2.2012

Table 4 – Mean genetic diversity between loci obtained through RAPD markers for bulk samples of five *Eugenia involucrata* DC. populations/collections from the state of Rio Grande do Sul, Brazil.

Parameter	EA	SB	CISM	Fepagro	FW	Total
N	3	8	5	7	4	27
A	1.13±0.34	1.22±0.41	1.53±0.50	1.62±0.49	1.47±0.5	1.97±0.18
Ae	1.11±0.27	1.10±0.22	1.36±0.38	1.42±0.40	1.33±0.33	1.51±0.31
I	0.08	0.10	0.31	0.35	0.28	0.47
P (nº)	8	13	32	37	28	58
P (%)	13.33	21.67	53.33	61.67	46.67	96.67

N = number of sampled individuals; *A* = average number of alleles per *locus*; *Ae* = effective average number of alleles per *locus*; I = Shannon Diversity Index; *P (nº)* = number of polymorphic loci; *P (%)* = percentage of polymorphic loci.

Unlike the phenetic analysis result, Bayesian clustering analysis indicated the presence of only two groups (Figure 2) when **Δ**K analysis was performed. The first group consists of the populations from Frederico Westphalen, CISM, and Fepagro, and the second from the Embrapa accesses and the Seedling Bank.

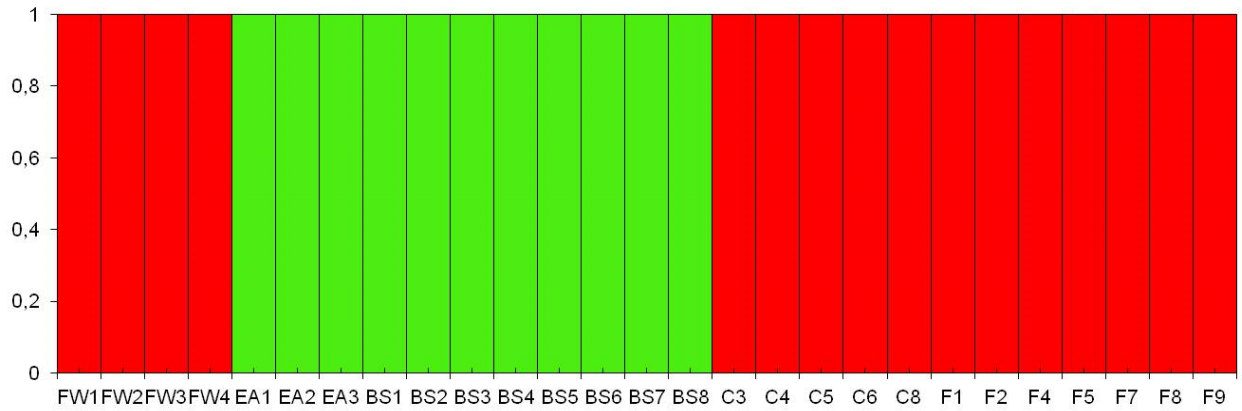

Figure 2 – Structure of the *Eugenia involucrate* DC. populations/collection from Rio Grande do Sul according to the Bayesian cluster analysis (K=2). FW = Frederico Westphalen; EA = Embrapa Accessions; SB = Embrapa Seedling Bank; C = CISM; F = Fepagro.

The results of the AMOVA analysis (Table 5) considering all five groups yielded a high and significant genetic differentiation between them (F_{ST} = 0.5438), revealing the existence of population pairs with significant heterogeneity. Considering the hierarchical model of three groups identified by the dendrogram generated from the similarity matrix elaborated from the Jaccard coefficients (Figure 1), 50.24% of the variation was distributed among those groups, however it was not significant. Therefore, the hierarchical model of two groups identified by the STRUCTURE program (Figure 2) was considered to be the most suitable to explain the genetic structure of the studied *E. involucrata* populations, with 33.993% of the variation being distributed among groups, 26.018% among populations within groups, and 39.988% within populations.

Table 5 – Analysis of molecular variance (AMOVA) for *Eugenia involucrata* DC. accessions from the state of Rio Grande do Sul, Brazil, using RAPD markers between: (1) two groups, (2) three groups and (3) all populations.

Source of variation	DF	SS	Components of variation	Percentage of variation (%)	Fixation indices
(1) Among groups	1	206.546	4.00334	33.993**	F_{CT} = 0.33993
Among populations within groups	3	138.141	3.06418	26.018**	F_{SC} = 0.39418
Within populations	70	329.660	4.70942	39.988**	F_{ST} = 0.60012
(2) Among groups	2	306.446	5.78655	50.24	F_{CT} = 0.50237
Among populations within groups	2	38.241	1.02250	8.88	F_{SC} = 0.17839
Within populations	70	329.660	4.70942	40.89	F_{ST} = 0.59114
(3) Among populations	4	344.687	5.61293	54.376**	
Within populations	70	329.660	4.70942	45.624	

(1) two groups (one constituted by the Frederico Westphalen, CISM and Fepagro collections and another formed by Embrapa Accessions and Seedling Bank collections) identified by STRUCTURE; (2) three groups (I. F2, F7, F8 and F9, from Fepagro; II. F1, F4 and F5, from Fepagro, C3, C5 and C6, from CISM; and all bulk samples from Frederico Westphalen; III. Embrapa Accessions and Seedling Bank collections), identified by the dendrogram; (3) all populations.

** $P < 0.01$ with 1023 permutations

Discussion

The RAPD primers used in the present study provided good results in the amplification of *Eugenia involucrata* DNA, with an average of 7.5 bands per primer and a polymorphism of 93.3%. Guerra (2016) obtained an average of 8.93 bands per primer and a polymorphism of 88.08% using 15 RAPD primers to analyze the genetic diversity in accessions of *Eugenia involucrata* in Rio Grande do Sul. In a study on genetic structure and gene flow of natural populations of *Eugenia dysenterica* in Goiás by Zucchi *et al*. (2005), 42 bands (35 of which polymorphic) were generated through six selected RAPD primers (mean of seven bands per primer and a polymorphism of 83.3).

The polymorphism percentage ranged from 13.33% to 61.67%, with an average of 96.67%. It should be noted that the average value of *P* is higher than the maximum because it computes all populations as one and includes alleles from high polymorphic populations and those from the less ones. In addition, this value is quite different from the average number of polymorphic bands (93.33%). This difference may be due to the visual-based screening of bands in the electrophoresis photo which criteria to consider a band may differ among researches. The accessions of the two collections of Embrapa presented the lowest values of genetic variability (number of alleles, effective number of alleles, Shannon diversity index, percentage of polymorphic loci) compared to natural populations. *Ex situ* conservation involves non-natural populations, which consist of a small sample of the genetic variability in the population of origin. According to Boef *et al*. (2007) *ex situ* conservation freezes evolution from occurring, since genotypes are removed from their original environments and are no longer subject to continuous adaptation to environmental changes. Likewise, Hawkes *et al*. (2000) emphasize *ex situ* conservation as a photograph of the moment the germplasm was deposited in the bank. Thus, less genetic variability is expected in these collections.

Gene flow among populations was determinant for their genetic differentiation, defined by the F_{ST} statistic. This statistic estimates the gene flow between populations or species, with low values indicating introgression between populations and consequently high genetic variability. On the other hand, low gene flow leads to high F_{ST} values indicating that populations became endogamic, and consequently (for allogamous species) more vulnerable to gene erosion (Neto & Morya, 2010). A scale for F_{ST} is used to qualify the genetic differentiation populations, which

considers values between zero and 0.05 as low; between 0.05 and 0.15 intermediate; between 0.15 and 0.20 as high; and above 0.20 as super high (Hartl & Clark, 2010). Based on this scale, all pairs of populations/collections presented a very high level of genetic differentiation, with the sole exception of Fepagro and CISM, whose geographic distance is 10km and the gene flow expressed in terms of number of migrants per generation was 1.384.

According to Slatkin (1985), numbers of migrants per generation (*Nm*) greater than one are sufficient to counteract the effects of genetic drift, and thus minimize genetic differentiation between populations. On the other hand, the super-high level of genetic differentiation for the FW – Fepagro (geographic distance = 260km) and FW – CISM (geographic distance = 266km) pairs, which had *Nm* values lower than 1, is justified. This finding ratifies the importance of the connectivity between geographically isolated populations via gene flow in order to reduce the genetic difference between them. Gene flow corresponds to the movement of genes (gametes, propagules and individuals) between and within populations, capable of altering gene frequencies in these populations. According to Budke *et al.* (2005) zoochoria is the main dispersion strategy in *Eugenia involucrata*. The importance of gene flow in conservation is supported by its ability to counteract the effects of genetic drift and inbreeding, the effect of which is greater in small populations and may lead to extinction (Neigel, 1997). In a similar study, Salgueiro *et al.* (2004) had low gene flow (*Nm* = 0.937) and a very high genetic differentiation (F_{ST} = 0.937) in *Eugenia uniflora* populations in the Brazilian coastal region of the Atlantic Forest.

The strong correlation between the genetic distance and geographic distance matrices (r = 0.999) is in line with the first law of geography (Tobler, 1970), which states that "everything is related to everything else, but near things are more related than distant things". According to this statement, closer populations are expected to be genetically more similar than distant ones. In fact, CISM and Fepagro have the largest gene flow between them and the germplasm used to constitute the Embrapa accession and Seedling Bank comes from Santa Maria.

The clustering analysis yielded three main distinct groups with a cophenetic correlation coefficient of 0.95291. The first group is formed by bulk samples from Santa Maria, indicating that there is gene exchange between these two populations despite the 10km distance between Fepagro Florestas and CISM. However, this flow is limited by distance, resulting in less similarity among populations, but sufficient to contribute with the lowest genetic differentiation among all pairs (F_{ST} = 0.153).

The second group, which also includes individuals from Fepagro and CISM, confirms the existence of gene flow between the two populations. However, the group is also composed of individuals from Frederico Westphalen, who are very homogenous, resulting in an increase of similarity in this group compared to the first.

The third group composed by the AGB and the Seedling Bank conserved *ex situ* at Embrapa in Pelotas, was the most uniform and had the highest genetic similarity. The high similarity between these populations may be explained by the fact that the Seedling Bank was created from seeds collected in the AGB. Thus, the enrichment of these populations with genetic material from other sites is suggested in order to increase genetic variability. The hierarchical AMOVA for Eugenia uniflora populations of the Atlantic Forest revealed significant differentiation between northern and southern groups (FCT = 0.585) due to distinct vegetational responses to climate change between both sides (Turchetto-Zolet *et al.*, 2016).

Conclusion

The RAPD primers used in this study provided good results and enabled studying the genetic structure and gene flow of the populations and collections of *Eugenia involucrata*. Very high levels of genetic differentiation were observed due to low gene flow among the populations. The five populations/collections are structured in two main groups, and the populations of *in situ* conservation have more genetic diversity in relation to *ex situ* conservation collections. The enrichment of these collections with introduced germplasm would guarantee an additional reserve of more consistent genetic variability for future use due to the continuous erosion of forest genetic resources.

Acknowledgements

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing the research grants.

References

Aguiar RV *et al*. Variabilidade genética de *Eugenia uniflora* L. em remanescentes florestais em diferentes estádios sucessionais. Revista Ceres, 60(2): 226-233, 2013.

Ali BA, Huang T-H, Qin D-N & Wang X-M. A review of random amplified polymorphic DNA (RAPD) markers in fish research. Reviews in Fish Biology and Fisheries, 14: 443-453, 2004.

Boef WS, Thijssen MH, Ogliari JB & Sthapit BR. Biodiversidade e agricultores: fortalecendo o manejo comunitário. Porto Alegre: L&PM, p.45-52, 2007.

Botrel RT, Rodrigues LA, Gomes LJ, Carvalho DA & Fontes MAL. Uso da vegetação nativa pela população local no município de Ingaí, MG, Brasil. Acta Botanica Brasilica, 20(1): 143-56, 2006.

Budke JC, Athayde EA, Giehl ELH, Záchia RA & Eisinger, SM. Composição florística e estratégias de dispersão de espécies lenhosas em uma floresta ribeirinha, arroio Passo das Tropas, Santa Maria/RS, Brasil. IHERINGIA, Série Botânica, 60(1): 17-24, 2005.

Brunchault RV, Soulange JG, Sanmukhiya VMR & Sevathian JC. Molecular and bioactive profiling of selected Eugenia species from Mauritius Island. International Journal of Plant Biology, 5(1): 4728, 2014.

Carvalho PE. 2008. Espécies arbóreas brasileiras, 1st ed., Brasília: Embrapa Informação Tecnológicas; Colombo: Embrapa Florestas, v.3, 593p.

Costa TS, Silva AVC, Lédo AS, Santos ARF & Júnior JFS. Diversidade genética de acessos do Banco de Germoplasma de mangaba em Sergipe. Pesquisa Agropecuária Brasileira, 6: 499-508, 2011.

Cristo-Araújo M, Rodrigues DP, Astolfi-Filho S & Clement CR. Genetic variability in the peach palm genebank with RAPD markers. Crop Breeding and Applied Biotechnology, 10(3): 211-217, 2010.

Cruz W, Saldaña C, Ramos H, Baselly R, Loli JC & Cuellar E. Estructura y diversidad genética de poblaciones naturales de *Cedrelinga cateniformis* "tornillo" en la región oriental del Perú. Scientia Agropecuaria, 11(4): 521-528, 2020.

Eibl BI, Montagnini F, López MA, López LN, Montechiesi R & Esterche E. 2017. Organic yerba mate, *Ilex paraguariensis*, in association with native species: a sustainable production alternative, p. 261-281. In: Montagnini F (org.). Integrating landscapes: agroforestry for biodiversity conservation and food sovereignty. Cham, Switzerland: Springer International Publishing. 501p.

Evanno G, Regnault S & Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology,** 14: 2611-2620, 2005.

Excoffier LGL & Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, 1: 47-50, 2006.

Excoffier L, Smouse P & Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial – DNA restriction data. Genetics, 13: 479-491, 1992.

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

Falush D, Stephens M & Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes, 7: 574-578, 2007.

Freitas VLO, Lemos-Filho JP & Lovato MB. Contrasting genetic diversity and differentiation of populations of two successional stages in a neotropical pioneer tree (*Eremanthus erythropappus*, Asteraceae). Genetics and Molecular Research, 7: 388-398, 2008.

Guerra D, Souza PVD, Schwarz SF, Schifino-Wittmann MT, Werlang CA & Veit PA. Genetic and cytological diversity in cherry tree accessions (*Eugenia involucrata* DC) in Rio Grande do Sul. Crop Breeding and Applied Biotechnology, 16: 219-225, 2016.

Hartl DL & Clark AG. 2010. Princípios de genética de populações. 4 ed. Porto Alegre: Artmed. 659p.

Hawkes JG, Maxted N & Ford-Lloyd BV. The ex-situ conservation of plant genetic Resources. London: Kluwer Academic Publishers, 1-250, 2000.

Hubisz MJ, Falush D, Stephens M & Pritchard JK. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources, 9: 1322-1332, 2009.

Jaccard P. Nouvelles recherches sur la distribu- tion florale. Bulletin de la Société Vaudoise des Sciences, 44: 223-270, 1908.

Kageyama PY, Sebbenn AM, Ribas LA, Gandara FB, Castellen M, Perecim MB & Vencovsky R. Diversidade genética em espécies arbóreas tropicais de diferentes estágios sucessionais por marcadores genéticos. Scientia Forestalis, 64: 93-107, 2003.

Kamada T, Picoli E AT, Alfenas AC, Cruz CD, Vieira RF & Otoni WC. Diversidade genética de populações naturais de *Pfaffia glomerata* (Spreng.) Pedersen estimada por marcadores RAPD. Acta Scientiarum. Agronomy, 31(3): 403-409, 2009.

Lorenzi H. 2002. Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. 1 ed. Nova Odessa: Plantarum. 384p.

Mossi AJ *et al*. Intra and inter populational genetic variability in *Maytenus ilicifolia* Mart. Ex Reiss, through RAPD Markers. Brazilian Journal of Biology, 67(4): 957-961, 2007.

Nagel JC *et al*. Historical gene flow within and among populations of *Luehea divaricata* in the Brazilian Pampa. Genetica, 143(3): 317-29, 2015.

Neigel JE. A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Review of Ecology, Evolution, and Systematics, 28(1): 105-128, 1997.

Neto NBM & Moryia AT. Variability in Variability in *Macadamia integrifolia* by RAPD markers by RAPD markers. Crop Breeding and Applied Biotechnology, 10: 266-270, 2010.

Oliveira DA, de Paula MFB, Pimenta MAS, Braga RF, Ferreira MFM & Rodrigues LA. Variabilidade genética de populações de fava-d'anta (*Dimorphandra mollis*) da região norte do Estado de Minas Gerais. Revista Árvore, 32: 355-363, 2008.

Oliveira VD, Rabbani ARC, Silva AVC & Lédo AS. Genetic variability in physic nuts cultivated in Northeastern Brazil. Ciência Rural, 43(6): 978-984, 2013.

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

Rabbani ARC, Silva-Mann R & Ferreira RA. Genetic variability of *Genipa americana* L belonging to the lower course of São Francisco river. Revista Árvore, 36: 401-409, 2012.

Saldaña CL *et al*. Genetic diversity and population structure of Capirona (*Calycophyllum spruceanum* Benth.) from the Peruvian Amazon revealed by RAPD markers. Forests, 12: 1125, 2021.

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

Santana GC *et al*. Diversidade genética de *Enterolobium contortisiliquum* (Vell.) Morong. no baixo Rio São Francisco, por meio de marcadores RAPD. Revista Árvore 32(3): 427-433, 2008.

Santos RP, Ângelo PCS, Sampaio PTB, Quisen RC, Leite AMC & Oliveira CL. Geographic pattern of genetic diversity in natural populations of Rosewood (*Aniba rosaeodora*), in the Central Amazonia. Acta Amazonica, 38(3): 459-466, 2008.

Sardi JC, Freires IA, Lazarini JG, Infante J, Alencar SM & Rosalen PL. Unexplored endemic fruit species from Brazil: antibiofilm properties, insights into mode of action, and systemic toxicity of four *Eugenia* spp. Microbial Pathogenesis, 105: 280-287, 2017.

Schulman L, Toivonen T & Ruokolainen K. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. Journal of Biogeography, 34(8): 1388-1399, 2007.

Sganzerla WG *et al*. Variability in the molecular, phenological, and physicochemical characteristics of uvaia (*Eugenia pyriformis* Cambess – Myrtaceae) accessions from the Brazilian Atlantic rainforest. Biocatalysis and Agricultural Biotechnology, 35: 102082, 2021.

Slatkin M. Rare alleles as indicators of gene flow. Evolution, 39(1): 53-65, 1985.

Sokal RR & Michener D. A statistical method for evaluation systematic relationships. The University of Kansas Science Bulletin, 38(22): 1409-1438, 1958.

Tobler WR. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2): 234-240, 1970.

Turchetto-Zolet *et al*. Phylogeography and ecological niche modelling in *Eugenia uniflora* (Myrtaceae) suggest distinct vegetational responses to climate change between the southern and the northern Atlantic Forest. Botanical Journal of the Linnean Society, 182(3): 670-688, 2016.

Uslan PM. Genetic diversity of *Sterculia quadrifida* in Kupang, Indonesia based on RAPD (Random Amplified Polymorphic DNA) markers. Biodiversitas, 21(7): 3407-3414, 2020.

Zucchi MI *et al*. Genetic structure and gene flow of *Eugenia dysenterica* natural populations. Pesquisa Agropecuária Brasileira, 40: 975-980, 2005.

Biodiversidade Brasileira – BioBrasil.
Fluxo Contínuo
n. 2, 2022

http://www.icmbio.gov.br/revistaeletronica/index.php/BioBR

Biodiversidade Brasileira é uma publicação eletrônica científica do Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) que tem como objetivo fomentar a discussão e a disseminação de experiências em conservação e manejo, com foco em unidades de conservação e espécies ameaçadas.

ISSN: 2236-2886