CROP BREEDING AND APPLIED BIOTECHNOLOGY

ARTICLE

Assessing genetic diversity and population structure in a *Dipteryx* alata germplasm collection utilizing microsatellite markers

Rejane Araújo Guimarães*[1,2], Kássia Marques Corrêa Miranda[1,2], Elias Emanuel Silva Mota[2], Lázaro José Chaves[2], Mariana Pires de Campos Telles[1,3] and Thannya Nascimento Soares[1,2]

Crop Breeding and Applied Biotechnology
19:3, 329-336, 2019
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332019v19n3a45

Abstract: *Dipteryx alata Vogel (Leguminosae) is a native Neotropical tree with a wide distribution in the Brazilian Cerrado that is commonly known as the baru tree. The genetic diversity of 150 D. alata progeny from a germplasm collection was characterized using nine microsatellite markers. Genetic diversity analysis detected 50 alleles ranging from 2 to 14 alleles per locus. The genetic differentiation among populations (θ_p = 0.097) suggests moderate genetic structuring and high genetic differentiation among progenies (θ_s = 0.169). The intrapopulation index (f = 0.122) indicates the presence of low endogamy. The effective population size (N_e = 96) shows that the germplasm collection has sufficient representativeness for use as a base population for breeding programmes. These results are useful for the exploitation of the genetic resources of D. alata for future conservation efforts and breeding programmes.*

Keywords: *Baru, cerrado, ex situ conservation, effective size, genetic resources.*

INTRODUCTION

Dipteryx alata Vogel is a Neotropical tree native to the Brazilian Cerrado biome commonly known as the baru tree. This species is widely distributed in the Cerrado and mainly grows in more fertile soils (Nabout et al. 2010).The potential uses of the species are very broad the baru tree can be used as a food product (Paglarini et al. 2018), for timber (Venturoli et al. 2015), and for industrial and medicinal purposes (Ribeiro et al. 2014). The main product of interest is the baru nut, which has great added value in the regional market. D. alata was included in a list of eight priority native fruit species in the Central-West region of Brazil compiled by the Brazilian government to advise on their short-term sustainable use (Vieira et al. 2016). Thus, cultivation is an efficient way to exploit this species, which is an important plant genetic resource of the Brazilian Cerrado.

For the cultivation of a native species, a broad knowledge of its biological and silvicultural characteristics is necessary. To this end, several studies have been conducted on D. alata, most of which aimed to obtain information on food and nutrition (Sousa et al. 2018), agronomic and silvicultural aspects (Santos et al. 2014, Sano et al. 2016), ex situ and in situ conservation (Melo et al. 2011, Tambarussi et al. 2017, Guimarães et al. 2019), and genetic variability (Soares et al. 2008a, Soares et al. 2008b, Soares et al. 2015).

***Corresponding author:**
E-mail: rejanearaujog@gmail.com
ORCID: 0000-0003-0186-6589

Received: 08 April 2019
Accepted: 29 May 2019

[1] Universidade Federal de Goiás, Laboratório de Genética & Biodiversidade, 74.690-900, Goiânia, GO, Brazil
[2] Universidade Federal de Goiás, Escola de Agronomia, 74.690-900, Goiânia, GO, Brazil
[3] Pontifícia Universidade Católica de Goiás, Escola de Ciências Agrárias e Biológicas, 74.690-900, Goiânia, GO, Brazil

To quantify or predict the level of total variability in an ex situ conservation programme, different types of genetic characterization can be used, such as morphological, agronomic, cytological, biochemical, physiological or molecular characterizations (Santonieri and Bustamante 2016). Molecular genetic characterization of the available D. alata germplasm could provide valuable information to ensure the proper use of this genetic resource and assist in controlling the genetic stability in the germplasm collection (Santonieri and Bustamante 2016). In the molecular genetic characterization, microsatellite markers are efficient tools for the identification and differentiation of individuals (Buso et al. 2003). Analysis of these markers can correctly classify accessions, identify duplicates, and quantify the level of variability present in a gene pool as well as its flow through time (Buso et al. 2003). Thus, the aim of the present work was to quantify and evaluate the molecular genetic diversity preserved in vivo and ex situ in the baru tree germplasm collection of the Federal University of Goiás (UFG) to contribute to conservation and breeding strategies.

MATERIAL AND METHODS

Plant material

The baru germplasm collection of the UFG is maintained in vivo and ex situ in the experimental area of the School of Agronomy of the Federal University of Goiás in the city of Goiânia-Goiás (lat 16° 35' 58.96" S, long 49° 16' 49.55" W, alt 736 m asl). The collection was planted in 2011 and consists of 600 individual accessions from 25 wild local populations sampled in the Brazilian Cerrado. In each population, six progeny arrays were sampled (c. 25 seeds per mother tree) and grown in a nursery in a completely random block design. Four seedlings per progeny from 150 mother trees were planted in the field with one plant per plot in a random block design (four blocks and 150 treatments) with 3 m x 2 m spacing in an area of approximately 3,276 m^2. Leaf samples from 575 remaining individuals in the germplasm collection were collected for DNA extraction.

Microsatellite analyses

Total genomic DNA was extracted from leaf tissue using the Cetyl Trimethyl Ammonium Bromide (CTAB) 2% protocol (Doyle and Doyle 1987). The quality of the DNA was examined on a 1% agarose gel, and its size was determined using visual comparison with an Invitrogen™ Low DNA Mass molecular marker. A portion of the DNA was diluted to a working concentration of approximately 2.5 ng mL^{-1} and stored at -20 °C. PCR was performed using a set of nine microsatellite loci, of which eight (DaE06, DaE12, DaE20, DaE34, DaE41, DaE46, DaE63 and DaE67) were developed and standardized for D. alata (Soares et al. 2012) and one (BM164) was developed for Phaseolus vulgaris and transferred to D. alata (Garcia et al. 2011).

PCR was performed in a final volume of 10 µL using 7.5 ng of template DNA and 0.22 µM primers (forward + reverse), 0.23 µM dNTP, 3.25 mg of bovine serum albumin (BSA), 1× reaction buffer (10 mM Tris-HCl [pH 8.3], 50 mM KCl, 1.5 mM $MgCl_2$) and 0.75 units of Taq DNA polymerase (5U; Phoneutria). The PCR conditions were as follows: 94 °C for 5 min; 30 cycles of 94 °C for 1 min, 58-62 °C (Table 1) for 1 min, and 72 °C for 1 min; and one cycle of 72 °C for 45 min.

Table 1. Characteristics of the loci and estimates of genetic diversity parameters *Dipteryx alata* based on nine microsatellites loci

Locus	Repeat motif	T_a (°C)	*A*	A_e	H_e	H_o
DaE06	$(AAAT)_4$	58	3	1.19	0.162	0.102
DaE12	$(ATTTTT)_3$	58	3	1.24	0.194	0.133
DaE34	$(GA)_{15}$	58	9	4.23	0.764	0.469
DaE63	$(TC)_6$	52	4	2.54	0.606	0.315
Bm164	$(GT)_9$ $(GA)_{21}$	52	8	1.47	0.322	0.241
DaE41	$(CA)_5$	60	14	5.22	0.800	0.711
DaE46	$(TTA)_5$	54	4	3.18	0.686	0.612
DaE20	$(AG)_8$	54	3	1.74	0.424	0.291
DaE67	$(GATACA)_4$	62	2	1.31	0.238	0.181
Average			5.5	2.457	0.467	0.339

T_a: annealing temperature; *A*: number of alleles per locus; A_e: number of effective alleles; H_e: expected heterozygosity based on the Hardy-Weinberg equilibrium; H_o: observed heterozygosity.

The amplified PCR fragments were sequenced using an ABI PRISM® 3100 automated DNA sequencer and a GeneScan-500 ROX size standard (Applied Biosystems). Allele calling was performed using GeneMapper 5.0 software (Applied Biosystems). Genotypes were visually reviewed, and Micro-Checker software (Oosterhout et al. 2004) was used to detect errors due to stutter bands, allele dropout, and null alleles.

Data analysis

Genetic diversity

The descriptive genetic parameters were as follows: number of alleles per locus (A), effective number of alleles (A_e), intrapopulation fixation index (f), observed heterozygosity (H_o) and expected heterozygosity under Hardy–Weinberg equilibrium (H_e), using Genetic Data Analysis ver. 1.0 (GDA) software (Lewis and Zaykin 2001). The power of individual discrimination for each locus and for the total set of loci was estimated by the probability of genetic identity and the probability of exclusion paternity, using Identity 1.0 software (Wagner and Sefc 1999).

We investigated the mating system based on mixed-mating models implemented in the software MLTR (Ritland 2002). The estimated parameters were outcrossing rate (t_m) and (t_s) obtained for global families, as well as for each population (using only sampled families of each population), biparental inbreeding rates (t_m–t_s) and multilocus paternity correlation (r_p). The confidence intervals (95%) of the parameter estimates were obtained with 10,000 randomizations.

Genetic structure

Genetic structure analyses were performed using the Weir and Cockerham (1984) method to estimate hierarchical F-statistics in the program GDA ver. 1.1 (Lewis and Zaykin 2001), which subdivides the components of variance into different levels (populations, progenies within populations and individuals within progenies). The following coefficients were estimated: genetic differentiation among populations (θ_p), genetic differentiation among progenies (θ_s), total fixation index (F) and intrapopulation fixation index (f). From θ_p and $\theta_{s,}$ we calculated the fixation index among progenies within populations, θ_m, making $\theta_m = \frac{\theta_s - \theta_p}{1 - \theta_p}$. The 95% confidence limits for all F-statistics (θ_p, θ_s, θ_m, F and f) were obtained using 10000 bootstrap resampling over loci.

Bayesian clustering approaches implemented in STRUCTURE 2.3.1 (Pritchard et al. 2000) were used to infer the genetic structure of the populations through an assignment test. The number of populations (K) was estimated with ten replicates each for K=1 to K=10 using 100,000 iterations of Markov chain after 100,000 burn-in period iterations using the admixture model. The K value was used to detect the most likely number of clusters (Evanno et al. 2005) using the STRUCTURE HARVESTER program (Earl and VonHold 2012). Due to progeny structure in the germplasm collection, only one individual of each progeny was randomly taken for this analysis, i.e., only one sib per progeny for each of the 150 mother trees was considered.

The genetic representativeness of the D. alata germplasm collection was verified based on the estimation of the effective population size (N_e). The calculation of N_e can be performed in different ways to ensure that the sample is representative of the population (Crossa and Vencovsky 2011). In this case, N_e of the germplasm collection was estimated using the equation adapted from Vencovsky et al. (2007), which is adequate for samples structured in populations, families within populations and individuals within families and assumes an infinite number of local populations in nature as follows:

$N_e = \frac{1}{2D}$, in which, $D = \theta_p \left(\frac{1 + C_s^2}{s} + \frac{1 + C_m^2}{M} \right) + \theta_s \left(\frac{1 + C_m^2}{M} + \frac{1}{N} \right) + \frac{1 + F}{2N}$, where:

θ_p, θ_s and F: population genetic structure parameters defined above; N, M, and S: the total number of plants, families, and populations, respectively; C_s and C_m: coefficients of variation for the number of plants per population and per family, respectively.

RESULTS AND DISCUSSION

Analysis of the genotyping quality did not detect errors due to stutter bands and allele dropout, but the DaE46 (pop 2), DaE34 (pop 6, 10, 15, 19), DaE63 (pop 8, 9, 14), and DaE20 (pop 20) loci presented null alleles. This may be because of the excess of apparently homozygous individuals found in these populations. Most loci showed high polymorphism,

as there were 50 alleles for the nine microsatellite loci (Table 1). The number of alleles per locus ranged from 2 to 14, with an average of 5.5 (Table 1). The highest number of alleles was detected at locus DaE41 with 14 alleles; however, the number of effective alleles was 5.22 (Table 1). The mean number of alleles per locus found in the present study was similar to that reported by Melo et al. (2011), (4.4), who evaluated a germplasm collection of D. alata from Goiás state, Brazil.

Moderate levels of expected heterozygosity (H_e) were observed in the D. alata populations equal the 0.467, while the observed heterozygosity (H_o) was lower than H_e at an average of 0.339. The difference between H_e and H_o indicates a tendency toward inbreeding and includes the effect of population subdivision, as deviations from panmixia within populations were observed. Other works in D. alata using microsatellite markers showed similar results, in which the expected average heterozygosity was higher than the observed heterozygosity (Tarazi et al. 2010, Melo et al. 2011, Tambarussi et al. 2017) (Table 1). However, the values of probability for genetic identity (I) and of the probability of exclusion paternity (Q) were 8.847×10^{-6} and 0.971, respectively, showing that the nine loci were suitable for genetic analysis.

The mean number of alleles per population ranged from 1.74 (NTO) to 2.90 (PGO), with an average of 2.12, while the number of effective alleles ranged from 1.64 to 2.81. The expected heterozygosity ranged from 0.278 (NTO) to 0.484 (VGMT), and the observed heterozygosity ranged from 0.225 (ARTO) to 0.420 (VGMT). The observed heterozygosity (H_o) and expected heterozygosity (H_e) values were 0.340 and 0.383, respectively. Except for the STGO population, all observed heterozygosity (H_o) values were lower than the expected heterozygosity (H_e) values, indicating an excess of homozygotes in these populations (Table 2).

Table 2. Estimates of genetic diversity parameters estimated based on nine microsatellite loci, evaluated in 25 populations of *Dipteryx alata*

Population	Origin	Population code	A	A_e	H_e	H_o	f
1	Cocalinho-MT	CMT	2.185	2.17	0.357	0.333	0.080
2	Água Boa-MT	ABMT	2.166	2.15	0.380	0.355	0.076
3	Pirenópolis-GO	PGO	1.907	2.01	0.311	0.262	0.169
4	Sonora-MS	SMS	2.277	2.54	0.442	0.397	0.119
5	Alcinópolis-MS	ALC	2.000	1.83	0.373	0.333	0.121
6	Alvorada-TO	ATO	2.018	2.28	0.410	0.355	0.154
7	São Miguel do Araguaia-GO	SMGO	2.148	2.00	0.391	0.360	0.095
8	Luziânia-GO	LGO	2.000	1.86	0.355	0.293	0.190
9	Icém-SP	ISP	2.074	2.08	0.358	0.318	0.127
10	Monte Alegre de Minas-MG	MAMG	2.000	2.00	0.384	0.315	0.210*
11	Estrela do Norte-GO	ENGO	2.000	2.17	0.347	0.327	0.006
12	Santa Terezinha-GO	STGO	2.092	2.09	0.370	0.381	-0.035
13	Arinos-MG	AMG	2.462	2.75	0.442	0.395	0.120
14	Pintópolis-MG	PMG	1.777	1.64	0.329	0.310	0.069
15	Paraíso-MS	PMS	2.333	2.41	0.409	0.360	0.138
16	Terenos-MS	TRMS	2.203	2.42	0.402	0.375	0.007
17	Camapuã-MS	CMS	2.500	2.42	0.473	0.383	0.216*
18	Indiara-GO	IGO	2.074	1.96	0.385	0.329	0.176*
19	Barra do Garças-MT	BGMT	2.148	2.45	0.369	0.343	0.088
20	Várzea Grande – MT	VGMT	2.462	2.62	0.484	0.420	0.159
21	Jandaia-GO	JGO	2.148	2.09	0.367	0.333	0.106
22	Natividade-TO	NTO	1.740	1.74	0.278	0.276	0.012
23	Arraias-TO	ARTO	1.814	1.88	0.281	0.225	0.224*
24	Aquidauana- MS	AQMS	2.092	2.41	0.386	0.307	0.230*
25	Cárceres- MT	CAMT	2.425	2.81	0.492	0.412	0.190*
Average			2.122	2.46	0.383	0.340	0.122*

A: number of alleles; A_e: number of effective alleles; H_e: expected heterozygosity based on the Hardy-Weinberg equilibrium; H_o: observed heterozygosity; *f*: intrapopulation fixation index. * Significant at the 95% probability level, based on the bootstrap confidence interval, using 10000 resamples.

The intrapopulation fixation index (f) was non-significant different from zero in 80% of the populations (Table 2), while the other populations (MAMG, CMS, IGO, AQMS, and CAMT) exhibited significant values for f. Most populations did not exhibit inbreeding due to the deviation from panmixia within them. The populations with significant values for f contributed to the significance of the mean value (f = 0.122*). The works of Tarazi et al. (2010) and Collevatti et al. (2013) evaluating natural populations of D. alata found values for f similar to those found in this study, which also indicates the presence of intrapopulation inbreeding.

The outcrossing rate (t_m) obtained for global families was 0.881, suggesting that the species has a mixed mating system with the predominance of allogamy. Tarazi et al. (2010), Tambarussi et al. (2017) and Guimarães et al. (2019), evaluating isolated populations, reinforces the hypothesis that this species utilizes a mixed mating system and is predominantly allogamous. The rate of crossing between populations varied from (t_m = 0.718) to (t_m = 0.979). The t_m observed for the populations agrees with other works performed with D. alata (Tarazi et al. 2010, Tambarussi et al. 2017, Guimarães et al. 2019). The t_m-t_s at the population level was significantly greater than zero in fifteen populations, ranging from 0.015 to 0.097 (table 3). The multilocus r_p obtained for global families was of r_p=0.052, suggesting that for each mother tree, the chance that the same pollen donor sired two random sibs was 5.2%. The r_p the population level was significantly greater than zero in all the populations, ranging from 0.030 to 0.152 (table 3). These populations have a high number of seeds that shared pollen donors, indicating a low number of pollen donors per tree.

The average co-ancestry between individuals of the same progeny (θ_s = 0.169) showed high diversity among progenies in the germplasm collection (Table 4). This parameter accumulates the co-ancestry due to recent relationship between individuals and the effect of subdivision of the meta-population in local populations. The estimated value is higher than what was expected for half-sib progenies (0.125), although the confidence interval contains this expected value. The total fixation index (F = 0.276, Table 4) showed global inbreeding both by the effects of the reproductive system

Table 3. Mating system parameters using MLTR in 25 population of *Dipteryx alata*

Population	***N***	t_m (SE)	t_s (SE)	t_m-t_s (SE)	r_p
1	24	0.979 (0.006)	0.881 (0.017)	0.097 (0.019)	0.057 (0.019)
2	23	0.987 (0.001)	0.879 (0.016)	0.108 (0.016)	0.087 (0.021)
3	21	0.755 (0.116)	0.843 (0.033)	-0.087 (0.096)	0.063 (0.012)
4	23	0.916 (0.041)	0.844 (0.028)	0.072 (0.050)	0.083 (0.036)
5	23	0.923 (0.044)	0.8584 (0.027)	0.066 (0.034)	0.061 (0.019)
6	23	0.800 (0.078)	0.811 (0.045)	-0.011 (0.073)	0.159 (0.057)
7	24	0.954 (0.032)	0.869 (0.017)	0.085 (0.027)	0.075 (0.029)
8	23	0.945 (0.033)	0.864 (0.033)	0.081 (0.042)	0.053 (0.013)
9	24	0.855 (0.035)	0.840 (0.033)	0.015 (0.040)	0.104 (0.059)
10	24	0.718 (0.048)	0.795 (0.044)	-0.077 (0.047)	0.090 (0.036)
11	23	0.954 (0.033)	0.843 (0.032)	-0.111 (0.039)	0.078 (0.055)
12	23	0.856 (0.043)	0.872 (0.018)	-0.016 (0.037)	0.048 (0.009)
13	24	0.863 (0.041)	0.824 (0.023)	0.039 (0.033)	0.038 (0.008)
14	21	0.855 (0.041)	0.859 (0.021)	-0.004 (0.042)	0.105 (0.029)
15	23	0.891 (0.064)	0.875 (0.027)	0.015 (0.057)	0.043 (0.009)
16	22	0.954 (0.034)	0.832 (0.024)	0.122 (0.037)	0.111 (0.028)
17	24	0.952 (0.033)	0.846 (0.031)	0.105 (0.038)	0.030 (0.006)
18	23	0.898 (0.062)	0.840 (0.028)	0.058 (0.053)	0.144 (0.061)
19	23	0.856 (0.043)	0.840 (0.022)	0.016 (0.044)	0.044 (0.008)
20	22	0.908 (0.064)	0.836 (0.024)	0.073 (0.056)	0.069 (0.022)
21	23	0.881 (0.047)	0.884 (0.014)	-0.003 (0.037)	0.073 (0.036)
22	23	0.750 (0.093)	0.836 (0.038)	-0.086 (0.076)	0.073 (0.015)
23	24	0.897 (0.041)	0.816 (0.041)	0.080 (0.042)	0.121 (0.045)
24	22	0.800 (0.042)	0.030 (0.042)	-0.030 (0.062)	0.044 (0.008)
25	20	0.954 (0.031)	0.785 (0.029)	0.169 (0.039)	0.152 (0.033)

N: sample size; t_m: multilocus outcrossing rate; ts: single locus outcrossing rate; $t_m - t_s$: mating among relatives, r_p: paternity correlation; SE: standard error.

Table 4. Estimates of hierarchical *F*-statistics using the method of Weir and Cockerham (1984)

Locus	f	F	θ_s	θ_p	θ_m
DaE06	0.325	0.370	0.067	0.035	0.033
DaE12	0.228	0.312	0.109	0.039	0.073
DaE34	0.280	0.388	0.150	0.069	0.087
DaE63	0.229	0.484	0.331	0.189	0.175
Bm164	0.075	0.256	0.196	0.141	0.064
DaE41	-0.108	0.125	0.211	0.118	0.105
DaE46	0.043	0.110	0.070	0.042	0.029
DaE20	0.220	0.316	0.123	0.095	0.031
DaE67	0.126	0.240	0.131	0.081	0.054
Average	0.129	0.276	0.169	0.097	0.080
Upper CI (95%)	0.239	0.385	0.231	0.134	
Lower CI (95%)	0.021	0.177	0.110	0.063	

f: Intrapopulation fixation index; *F*: Total fixation index; ϑs: genetic differentiation among progenies; ϑp: genetic differentiation among populations. CI: Confidence Interval obtained from 100,000 bootstrap replicates.

as a function of the significant f value and by the genetic drift effect within the populations associated with a certain restriction of the gene flow between them.

Although the Bayesian clustering analysis indicated the formation of two clusters, the allocation of individuals into each cluster was not clear. Clusters of trees separated by up to 1,000 km (ALC - NTO) contained genetically similar individuals (Figure 1a). These results indicate the intense gene flow between populations from the north and south regions, leading to moderate genetic differentiation. However, clusters 1 and 2 were distributed in all of the sampled areas, although in different proportions. The greater proportion of individuals in cluster 2 was mainly located in the northern part of the species distribution, whereas cluster 1 was found in the remaining part of the species distribution (Figure 1a). The admixture of both clusters was observed in all individuals (Figure 1b).

The effective size of the germplasm collection (N_e) was 96.0 (Table 5) considering the 575 plants remaining in the collection (4.2% of missing plants). This value is less than 1% lower than the value obtained when considering the planned collection with 600 plants (N_e = 96.9). Thus the effective population size found shows that the germplasm collection has

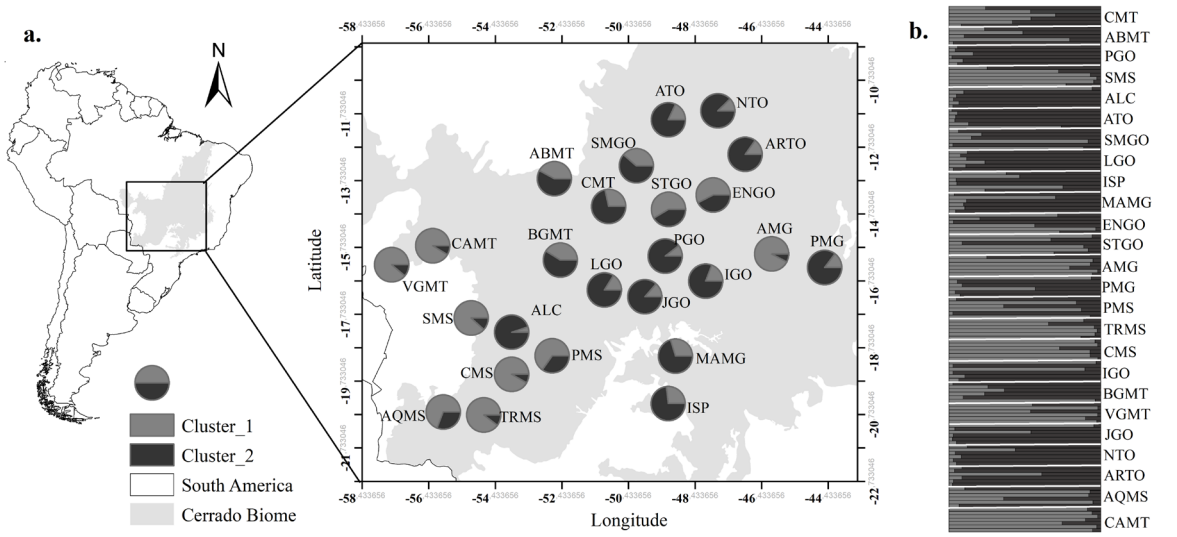

Figure 1. a) Geographical representation of the two clusters observed in *Dipteryx alata* germplasm collection using STRUCTURE software. b) Each individual is represented by a thin horizontal segment, which represent the estimated individual.

sufficient representativeness for use as a base population for breeding programmes (Jamieson and Allendorf 2012). It should be noted that the N_e was calculated in reference to the whole D. alata species and is adequate for the purpose of ex situ conservation. However, for long-term conservation purposes, to reach the ideal N_e of 500 proposed by Vencovsky and Crossa (2003), it would be necessary to complement this collection with accessions from several other regions of the Brazilian Cerrado. Therefore, a comprehensive conservation programme must integrate in situ and ex situ strategies to reach ideal representativeness.

Table 5. Effective size of the *Dipteryx alata* germplasm collection, considering the structuring in populations, families within populations and individuals within families

Sample data	Planned collection	Actual collection
N	600	575
S	25	25
M	150	150
θ_p	0.097	
θ_s	0.169	
F	0.276	
N_e	96.9	96.0

N: Total number of plants; *S*: number of sampled populations; *M*: total number of families. θ_p, θ_s and *F* are estimates of populations statistics.

In conclusion, the D. alata germplasm collection at the School of Agronomy of the Federal University of Goiás exhibits moderate molecular genetic variability, as observed in other studies with the same species. The populations that comprise the germplasm collection exhibit a low magnitude of inbreeding, and their genetic structure can be considered moderate among populations and among progenies within populations. The germplasm collection contains an appropriate effective size for initiating a long-term breeding programme. To evaluate the potential of breeding for selection, molecular and agronomical characterization is recommended. The molecular data would also be useful to identify more variable and divergent populations in a future collection for breeding purposes.

REFERENCES

Buso GSC, Ciampi AY, Moretzsohn MDC, Amaral PZDS and Brondani RV (2003) Marcadores microssatélites em espécies vegetais. **Revista Biotecnologia Ciência e Desenvolvimento 30**: 43-50.

Collevatti RG, Telles MPC, Nabout JC, Chaves LJ and Soares TN (2013) Demographic history and the low genetic diversity in Dipteryx alata (Fabaceae) from Brazilian Neotropical savannas. **Heredity 111**: 97-105.

Crossa J and Vencovsky R (2011) Basic Sampling Strategies: Theory and Practice. Available at: <http://www.mendeley.com/research/effective-population-size-under-random-mating-with-a-nite-number-of-matings/>. Accessed on Jan 02, 2019.

Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin 19**: 11-15.

Earl DA and VonHold BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conservation Genetics Resourses 4**: 359-361.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. **Molecular Ecology 14**: 2611-2620.

Garcia RV, Rangel PN, Brondani C, Martins WS, Melo LC, Carneiro MS, Borba TCO and Brondani RPV (2011) The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of Phaseolus vulgaris. **BMC Genetics 12**: 41.

Jamieson IG and Allendorf FW (2012) How does the 50/500 rule apply to MVPs? **Trends in Ecology and Evolution 27**: 580-584.

Lewis PO and Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data. Version 1.0. Available at: <http://lewis.eeb.uconn.edu/lewishome/ software.html>. Accessed on March 23, 2019.

Melo DB, Diniz-Filho JAF, Oliveira G, Santana LL, Soares TN, Chaves LJ, Naves R V, Collevatti RG and Telles MPC (2011) Optimizing sampling efforts for ex situ conservation of genetic variability of Dipteryx alata Vogel. **BMC Proceedings 5**: 1-18.

Nabout JC, Soares TN, Diniz-Filho JAF, Marco Júnior P de, Telles MPC, Naves R V and Chaves LJ (2010) Combining multiple models to predict the geographical distribution of the Baru tree (Dipteryx alata Vogel) in the Brazilian Cerrado. **Brazilian Journal of Biology 70**: 911-919.

Oosterhout C Van, Hutchinson WF, Wills DPM and Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. **Molecular Ecology Notes 4**: 535-538.

Paglarini CS, Queirós MS, Tuyama SS, Moreira ACV, Chang YK and Steel CJ (2018) Characterization of baru nut (Dipteryx alata Vog) flour and its application in reduced-fat cupcakes. J**ournal of Food Science and Technology 55**: 164-172.

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

Guimarães RA, Corrêa Miranda KM, Chaves LJ, Naves RV, de Campos Telles MP and Soares TN (2019) Mating system and pollen dispersal in Dipteryx alata Vogel (Leguminosae): comparing in situ and ex situ conditions. **Tree Genetics & Genomes 15**: 28.

Ribeiro TG, Chávez-Fumagalli MA, Valadares DG, Franca JR, Lage PS, Duarte MC, Andrade PHR, Martins VT, Costa LE, Arruda ALA, Faraco AAG, Coelho EAF and Castilho RO (2014) Antileishmanial activity and cytotoxicity of Brazilian plants. **Experimental Parasitology 143**: 60-68.

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. **Heredity 88**: 221-228.

Sano SM, Brito MA and Ribeiro JF (2016) Baru. In Vieira RF, Camillo J and Coradin L (eds) **Espécies nativas da flora brasileira de valor econômico atual ou potencial plantas para o futuro - região centro-oeste**. Ministério do Meio Ambiente, Brasília, p. 1-1160.

Santonieri L and Bustamante PG (2016) Conservação ex situ e on farm de recursos genéticos: desafios para promover sinergias e complementaridades. **Boletim do Museu Paraense Emílio Goeldi 11**: 677-690.

Santos AM, Rosado SCS and Oliveira AN (2014) Estimation of genetic parameters and verification of early selection efficiency in baru (Dipteryx alata). **Crop Breeding and Applied Biotechnology 14**: 238-243.

Soares TN, Chaves LJ, Telles MCP, Diniz-Filho JAF and Resende LV (2008a) Distribuição espacial da variabilidade genética intrapopulacional de Dipteryx alata. **Pesquisa Agropecuária Brasileira 43**: 1151-1158.

Soares TN, Chaves LJ, Telles MPC, Diniz-Filho JAF and Resende LV (2008b) Landscape conservation genetics of Dipteryx alata ("baru" tree: Fabaceae) from Cerrado region of central Brazil. **Genética 132**: 9-19.

Soares TN, Diniz-Filho JAF, Nabout JC, Telles MPC, Terribile LC and Chaves LJ (2015) Patterns of genetic variability in central and peripheral populations of Dipteryx alata (Fabaceae) in the Brazilian Cerrado. **Plant Systematics and Evolution 301**: 1315-1324.

Soares TN, Melo DB, Resende LV, Vianello RP, Chaves LJ, Collevatti RG and Telles MPC (2012) Development of microsatellite markers for the neotropical tree species Dipteryx alata (Fabaceae). **American Journal of Botany 99**: e72-3.

Sousa FF, Braga RM, Venturin N, Carlos L, Macedo RLG and Venturin RP (2018) Nutritional requirements of Dipteryx alata seedlings grown under nutrient-limiting conditions. **Ciência Florestal 28**: 102-114.

Tambarussi EV, Sebbenn AM, Alves-Pereira A, Vancovsky R, Cambuim J, Silva AM, Moraes MA and Moraes MLT (2017) Dipteryx alata Vogel (Fabaceae) a neotropical tree with high levels of selfing: implication for conservation and breeding programs. **Annals of Forest Research 60**: 1-19.

Tarazi R, Moreno MA, Gandara FB, Ferraz EM, Moraes MLT, Vinson CC, Ciampi AY, Vencovsky R and Kageyama PY (2010) High levels of genetic differentiation and selfing in the Brazilian cerrado fruit tree Dipteryx alata Vog. (Fabaceae). **Genetics and Molecular Biology 33**: 78-85.

Vencovsky R and Crossa J (2003) Measurements of representativeness used in genetic resources conservation and plant breeding. **Crop Science 43**: 1912-1921.

Vencovsky R, Nass LL, Cordeiro CMT and Ferreira MAJ (2007) Amostragem em recursos genéticos vegetais. In Nass LL (ed) **Recursos genéticos vegetais**. Embrapa Recursos Genéticos e Biotecnologia, Brasília, p. 231-280.

Venturoli F, Franco AC and Fagg CW (2015) Tree diameter growth following silvicultural treatments in a semi-deciduous secondary forest in Central Brazil. **Cerne 21**: 117-123.

Vieira RF, Agostini-Costa TS, Silva DB, Sano SM and Ferreira FR (2016) Espécies alimentícias nativas da região centro-oeste. In Vieira RF, Camillo J and Coradin L (eds) **Espécies nativas da flora brasileira de valor econômico atual ou potencial plantas para o futuro - região centro-oeste**. Ministério do Meio Ambiente, Brasília, p. 109-118.

Wagner HW and Sefc KM (1999) IDENTITY 1.0 Centre for Applied Genetics. University of Agricultural Sciences, Vienna. Available at: <http:/boku.ac. at/zag/forsgh/identity.htm>. Accessed on Feb 3, 2019.

Weir BS and Cockerham CC (1984) Estimating F-statistics the analysis of population structure. **Evolution 38**: 1358-1370.