The traceability of *Eucalyptus* clones using molecular markers

Diego Torres-Dini[1]*, Leonardo Delgado-Cerrone[2], Lorena Luna[3], Fernando Resquin[1], Ananda Virginia Aguiar[4], Alexandre Magno Sebbenn[5]

[1] Instituto Nacional de Investigación Agropecuaria (INIA), Ruta 5 Km 386, CEP 45000, Tacuarembó, TB, Uruguay
[2] Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Av. Italia 3318, 11600 Montevideo, Uruguay.
[3] Centro Universitario de Tacuarembó (CUT), Ruta 5 Km 386, CEP 45000, Tacuarembó, TB, Uruguay
[4] Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Centro Nacional de Pesquisa de Florestas, Estrada da Ribeira, Km 111 - Bairro Guaraituba, 83411-000, Colombo, PR, Brazil.
[5] Instituto Florestal de São Paulo, CP 1322, São Paulo, SP, 01059-970, Brazil.

*Corresponding author: Diego Torres-Dini, E-mail: diego.torres.dini@gmail.com

Abstract

The improvement of Eucalyptus clones plays a crucial role in modern silviculture. This study used a set of 17 microsatellite loci to analyze the genetic diversity and structure of 107 elite clones (80 *E. grandis* and 27 *E. globulus*). All clones were cultivated in Uruguay and were sourced from three different providers. Using the fingerprinting technique, an exclusive molecular profile was assigned for each clone, and the genotyping reaction showed differences between the two species. The cumulative probability of identifying two random individuals that share the same genotype (PI) with all 17 loci, was estimated as low for *E. grandis* (1.18×10-15) and *E. globulus* (4.03×10-14). The combined $PIsibs$ was (1.05×10-5) and (2.17×10-5) for *E. grandis* and *E. globulus*, respectively. A total of 180 alleles were detected for *E. grandis* and 100 for *E. globulus*. We found a high mean number of alleles per locus (10 for *E. grandis* and 6 for *E. globulus*), and the results for mean polymorphic information content (PIC) were (0.648) and (0.548), respectively. The observed heterozygosity (H_o) ranged from 0.216 to 0.838 (mean = 0.509) for *E. grandis* and 0 to 1 (mean = 0.566) for *E. globulus*. Two core sets of seven EST-SSR loci were identified for each species. These markers revealed unambiguous fragment amplification, providing a minimum number of SSRs for effective clonal identification. The genetic structure analysis suggests that the germplasm of the *E. grandis* population is structured in four clusters, while the *E. globulus* population consists of two clusters.

Keywords: *Clone certification, clone, genotyping, identity, multiplex, nurseries, traceability*

Introduction

Commercial reforestation with Eucalyptus species exceeds 20 million hectares worldwide, in tropical and temperate climates across more than 90 countries (Rezende et al. 2014; Torres-Dini et al. 2016; Brancalion et al. 2020). The Eucalyptus genus consists of more than 800 species (Rodrigues and Faria 2021), which are characterized by rapid growth, high-quality wood with a range of end uses, wide variability, and capacity for vegetative propagation. Tree genetic improvement seeks to increase productivity by selecting genotypes with superior productive characteristics, while cloning techniques allow for the multiplication and development of elite clones. These clones are the result of decision-making in breeding programs focused on achieving differential productivity and are multiplied by the thousands in nurseries for subsequent cultivation in plantations (Grattapaglia and Kirst 2008).

However, clone labeling errors can occur during conservation, propagation, cultivation, or exchange processes (DeLucas et al. 2008; Li et al. 2011). Accurate clonal and cultivar identification is essential in vegetatively propagated plants, and it is a requirement for the registration of new cultivars (Pasqualone et al. 2015). In the past, identification was based on

DOI:10.2478/sg-2021-0019
edited by the Thünen Institute of Forest Genetics

the evaluation of morphological and agronomic characteristics. However, these parameters can be influenced by age and phenology, and similar phenotypes may represent different genetic backgrounds, leading to numerous cases of synonymy and homonymy (Bautista et al. 2002; Wunsch and Hormaza 2002; Pasqualone et al. 2015). With the use of DNA-based markers, the methods employed to identify clones and cultivars have improved (Wünsch and Hormaza 2002, Gross et al. 2018, Veloso et al. 2018), and microsatellite markers have become one of the main tools used for genotyping. These developments enable proper management of reference germplasm collections and effective traceability in the development of new genotypes (Pasqualone et al. 2015; Gross et al. 2018; Veloso et al. 2018). Worldwide, microsatellite markers are used extensively to correctly identify clones (Kirst et al. 2005; DeLucas et al. 2008; Faria et al. 2011; Li et al. 2011; Tan et al. 2015). These techniques are repeatedly requested by producers to ensure traceability during successive stages of clonal multiplication, while also protecting intellectual property rights in inter-institutional collaboration and commercial trade (Kirst et al. 2005; Li et al. 2011; Tan et al. 2015).

The hyper-variability and simple inheritance of microsatellites provide a powerful system for identifying individuals through fingerprinting, kinship analysis, and population genetics studies (Faria et al. 2011). Faria et al. (2011) presented a set of 21 transferable microsatellite markers between species based on repeats of tetra-, penta-, and hexanucleotides. This genotyping system consists of combinations of 14 and 18 markers in multiplex PCR reactions labeled with dyes, which facilitates automation and reduces the time required and costs associated with genotyping (Faria et al. 2011). The construction of genotyping profile databases are used to compare the genetic identities of elite clones. This information is analyzed electronically against reference samples, ensuring control over the traceability of cloned materials and providing a varietal protection system (Kirst et al. 2005). The power of microsatellite markers in genotypic characterization has been demonstrated and documented across a range of different studies and applications (Sumathi and Yasodha 2014). The goal of the present study was to assign a molecular profile to each of the *E. grandis* and *E. globulus* clones analyzed herein and build a database to monitor clonal identity in successive stages of vegetative propagation.

Materials and Methods

Plant material and DNA extraction

Leaf tissue samples were collected in the nursery from a total of 107 commercial Eucalyptus clones (80 *E. grandis* and 27 *E. globulus*). All clones were propagated in the nurseries and represent elite genotypes from three different providers located in Uruguay (Table 1). DNA purification was performed following a standard CTAB protocol described by Ferreira and Grattapaglia (1995). Tissue was disrupted using TissueLyser II system homogenizer (Qiagen, Valencia, CA) at a speed of 30 Hz for 5 min.

Multiplex EST-SSR genotyping

The genotyping was performed based on the cross-specific method described by Faria et al. (2011). A total of 17 primer pairs (Alpha DNA, Montreal, CA) were employed in this study, which were subdivided into two PCR reaction multiplexed systems, called A and B (Supplementary material, Table A). Each PCR reaction was carried out in a volume of 10 µl containing 1x Qiagen multiplex PCR buffer (Qiagen Inc., Valencia, CA, USA), an equal concentration of each primer (0.1 µM) for all co-amplified markers, and 20 ng of genomic DNA. The multiplex PCR cycling protocol was: 96 °C for 5 min; 10 cycles of 94 °C for 1 min, 64 °C for 1 min, 72 °C for 2 min; 20 cycles of 94 °C for 1 min, 56 °C for 1 min, 72 °C for 2 min; and 7 min 72 °C for the final extension. The forward primers were labeled with 6-FAM and HEX (Supplementary material, Table A). Genotypes were determined using Genescan Service by Macrogen (Seoul, Korea) and scored using the software Peak Scanner 1.0 (Applied Biosystems). An example of an electropherogram is shown in the Supplementary material, Figure A.

Descriptive statistics of genetic diversity

The genetic parameters for each marker were estimated separately for each species and included: number of amplified individuals (n); number of alleles per locus (k); allele size range (ASR); observed heterozygosity (H_o); polymorphism information content (PIC ; Botstein et al. 1980); probability of identity (PI ; Paetkau et al. 1995), which corresponds to the probability of two random individuals displaying the same genotype; major allele frequency (MAF); the probability of finding two full-sib individuals from a population that have the same genotype by chance ($PIsibs$; Waits et al. 2001); paternity exclusion probability (PE ; Weir 1996) which corresponds to the power with which a locus excludes an erroneously assigned individual tree from being the parent of an offspring; and the frequency of null alleles (F_{Null}). The CERVUS 3.7 software (Kalinowsky et al. 2007) was used to estimate (k), (H_o), (PIC), and (F_{Null}), while (PI), ($PIsibs$), (PE) and (*MAF*) were calculated using the IDENTITY 1.0 software (Wagner and Sefc 1999).

Genetic structure analysis

A Bayesian assignment analysis, performed in the STRUCTURE 2.2.3 software (Pritchard et al. 2000), was used to identify distinct genetic clusters in each of the two analyzed species. Bayesian analysis of population structure was performed using the admixture and allele frequency correlated allele models with a burn-in of 100,000 and 100,000 Markov chain Monte Carlo repetitions. The number of number of clusters (K) was set from 1 to 10, with 10 replicates for each K. The most probable number of clusters (K) was selected by computing the second-order rate of change of the likelihood function with respect to K (Evanno et al. 2005) using the STRUCTURE HARVESTER software (Earl 2012).

Table 1
List of clones analyzed in the study

N	Clone code	Provider	Species	N	Clone code	Provider	Species	N	Clone code	Provider	Species
1	BV3C1	Lumin	*E. grandis*	37	LAC17	Lumin	*E. grandis*	73	INIA2433	INIA	*E. grandis*
2	BV3C2	Lumin	*E. grandis*	38	LAC18	Lumin	*E. grandis*	74	INIA2436	INIA	*E. grandis*
3	BV3C3	Lumin	*E. grandis*	39	LAC19	Lumin	*E. grandis*	75	INIA2438	INIA	*E. grandis*
4	DC1	Lumin	*E. grandis*	40	LAC2	Lumin	*E. grandis*	76	INIA2707	INIA	*E. grandis*
5	DC2	Lumin	*E. grandis*	41	LAC4	Lumin	*E. grandis*	77	INIA2715	INIA	*E. grandis*
6	DC3	Lumin	*E. grandis*	42	LAC5	Lumin	*E. grandis*	78	MPHG	CEBIOF	*E. grandis*
7	DC4	Lumin	*E. grandis*	43	LAC7	Lumin	*E. grandis*	79	MPHL	CEBIOF	*E. grandis*
8	HHK1C1	Lumin	*E. grandis*	44	LAC9	Lumin	*E. grandis*	80	MP865	CEBIOF	*E. grandis*
9	HHK1C2	Lumin	*E. grandis*	45	LTC2	Lumin	*E. grandis*	81	SCDesc1	CEBIOF	*E. globulus*
10	HHK1C3	Lumin	*E. grandis*	46	LTC3	Lumin	*E. grandis*	82	SCDesc2	CEBIOF	*E. globulus*
11	HHK1C4	Lumin	*E. grandis*	47	MLPC1	Lumin	*E. grandis*	83	SC100334	CEBIOF	*E. globulus*
12	HHK2C1	Lumin	*E. grandis*	48	MLPC2	Lumin	*E. grandis*	84	SCCandon	CEBIOF	*E. globulus*
13	HHK2C10	Lumin	*E. grandis*	49	ZHC1	Lumin	*E. grandis*	85	SCMAncel1	CEBIOF	*E. globulus*
14	HHK2C12	Lumin	*E. grandis*	50	ZHC10	Lumin	*E. grandis*	86	SCMAncel2	CEBIOF	*E. globulus*
15	HHK2C2	Lumin	*E. grandis*	51	ZHC11	Lumin	*E. grandis*	87	SCMAncel3	CEBIOF	*E. globulus*
16	HHK2C3	Lumin	*E. grandis*	52	ZHC2	Lumin	*E. grandis*	88	SCMAncel4	CEBIOF	*E. globulus*
17	HHK2C5	Lumin	*E. grandis*	53	ZHC3	Lumin	*E. grandis*	89	FM11	CEBIOF	*E. globulus*
18	HHK2C6	Lumin	*E. grandis*	54	ZHC4	Lumin	*E. grandis*	90	FM12	CEBIOF	*E. globulus*
19	HHK2C8	Lumin	*E. grandis*	55	ZHC7	Lumin	*E. grandis*	91	FM21	CEBIOF	*E. globulus*
20	HHK2C9	Lumin	*E. grandis*	56	ZHC8	Lumin	*E. grandis*	92	FM22	CEBIOF	*E. globulus*
21	IC1	Lumin	*E. grandis*	57	ZH2C1	Lumin	*E. grandis*	93	FM23	CEBIOF	*E. globulus*
22	IC10	Lumin	*E. grandis*	58	INIA31	INIA	*E. grandis*	94	FM24	CEBIOF	*E. globulus*
23	IC12	Lumin	*E. grandis*	59	INIA32	INIA	*E. grandis*	95	FM25	CEBIOF	*E. globulus*
24	IC13	Lumin	*E. grandis*	60	INIA33	INIA	*E. grandis*	96	FM31	CEBIOF	*E. globulus*
25	IC14	Lumin	*E. grandis*	61	INIA34	INIA	*E. grandis*	97	FM32	CEBIOF	*E. globulus*
26	IC15	Lumin	*E. grandis*	62	INIA36	INIA	*E. grandis*	98	FM34	CEBIOF	*E. globulus*
27	IC6	Lumin	*E. grandis*	63	INIA37	INIA	*E. grandis*	99	FM35	CEBIOF	*E. globulus*
28	IC7	Lumin	*E. grandis*	64	INIA39	INIA	*E. grandis*	100	FM41	CEBIOF	*E. globulus*
29	IC8	Lumin	*E. grandis*	65	INIAJL28	INIA	*E. grandis*	101	FM42	CEBIOF	*E. globulus*
30	IC9	Lumin	*E. grandis*	66	INIA21100	INIA	*E. grandis*	102	FM43	CEBIOF	*E. globulus*
31	LAC10	Lumin	*E. grandis*	67	INIA2116	INIA	*E. grandis*	103	FM44	CEBIOF	*E. globulus*
32	LAC11	Lumin	*E. grandis*	68	INIA2133	INIA	*E. grandis*	104	FM51	CEBIOF	*E. globulus*
33	LAC12	Lumin	*E. grandis*	69	INIA2179	INIA	*E. grandis*	105	FM52	CEBIOF	*E. globulus*
34	LAC14	Lumin	*E. grandis*	70	INIA2190	INIA	*E. grandis*	106	FM53	CEBIOF	*E. globulus*
35	LAC15	Lumin	*E. grandis*	71	INIA2193	INIA	*E. grandis*	107	FM54	CEBIOF	*E. globulus*
36	LAC16	Lumin	*E. grandis*	72	INIA2198	INIA	*E. grandis*				

Phylogeny tree

Genetic relationships among clones were calculated using a matrix of dissimilarity with a simple matching coefficient in the DARWIN 6 software (Perrier and Jaquemoud-Collet 2006). The dissimilarity matrix was estimated with a setting of 1000 bootstraps, allelic data, and missing data. A dendrogram was constructed using clustering unweighted pair group method with arithmetic mean (UPGMA), again in DARWIN 6 software. The graphic representation was prepared with the software FIGTREE (Rambaut 2010).

Results

Genetic diversity parameters

The genotyping reaction showed differences between the two studied species. In the case of *E. grandis*, 13 markers exhibited a n greater than 77, while 4 markers showed low n values: EMBRA1008 ($n = 68$); EMBRA1364 ($n = 37$); EMBRA1374 ($n = 3$); and EMBRA1812 ($n = 37$). The genotyping of *E. globulus* showed a (n) value greater than 24 for 14 markers, while 3 markers showed low (n) values: EMBRA1957 ($n = 20$);

EMBRA1977 (n = 13); and EMBRA1364 (n = 13) (Table 2; Excel genotyping dataset is available in the Supplementary material). The 17 microsatellite loci were polymorphic with the number of alleles per locus (k) ranging from 4 to 16 (mean = 10) for *E. grandis* and 1 to 15 (mean = 6) for *E. globulus*, for a total of 180 and 100 alleles, respectively, for the studied species (Table 2). Based on the genotyping data, no match was detected, indicating that each of 107 clones had an exclusive fingerprint profile. The probability of identity (PI) for *E. grandis* ranged from 0.022 (EMBRA1616) to 0.859 (EMBRA1374), with a mean of 0.204. In *E. globolus* PI ranged from 0.032 (EMBRA850) to 1 (EMBRA915), with a mean of 0.228. The cumulative probability of identifying two random individuals, sharing the same genotype for the 17 loci, was estimated to be low for both species (1.18×10-15 for *E. grandis* and 4.03×10-14 for *E. globulus*). The $PIsibs$ ranged from 0.312 (EMBRA1616) to 1.357 (EMBRA1374) (mean = 0.556) for *E. grandis*, and from 0.372 (EMBRA1811) to 1.500 (EMBRA915) (mean = 0.574) for *E. globulus*. The combined $PIsibs$ was very low (1.05×10-5 and 2.17×10-5, respectively), and the PE combined for both species was 0.999.

The H_o ranged from 0.216 (EMBRA1364) to 0.838 (EMBRA1812) (mean = 0.509) for *E. grandis* and 0 (EMBRA915) to 1 (EMBRA1811, EMBRA813 and EMBRA1616) (mean = 0.566) for *E. globulus*. For *E. grandis* the PIC value ranged from 0.235 (EMBRA1040) to 0.875 (EMBRA1616) (mean = 0.648) and 0 (EMBRA915) to 0.825 (EMBRA850) (mean = 0.548) for *E. globulus*. The F_{Null} ranged respectively for each species from -0.07 (EMBRA1812) to 0.571 (EMBRA1364) (mean= 0.146) and -0.212 (EMBRA1616) to 0.671 (EMBRA915) (mean = 0.10). The MAF for each locus ranged from 0.006 (EMBRA1374) to 0.868 (EMBRA1040) (mean 0.403) for *E. grandis* and 0.129 (EMBRA1977) to 1 (EMBRA915) (mean = 0.482) for *E. globulus*.

Based on the criteria of high PIC , low PI , $PIsibs$, and *n*, we selected seven EST-SSR loci for each species. These markers gave unambiguous fragment amplification providing a minimum number of effective SSRs for clonal identification. These represent two core sets for fingerprinting the 107 Eucalyptus clones of each species. For the *E. grandis* core set, the markers EMBRA1616, EMBRA925, EMBRA954, EMBRA813, EMBRA850, EMBRA1851, and EMBRA2014 were chosen. The combined

Table 2
Results of genetic diversity per locus, as a mean and cumulative across all loci for *E. grandis* and *E. globulus* clones.

	E. grandis									*E. globulus*							
Marker	n	k	PI	$PIsibs$	PE	H_o	PIC	F_{Null}	MAF	n	k	PI	$PIsibs$	PE	H_o	PIC	F_{Null}	MAF
EMBRA1040	80	9	0.582	1.059	0.133	0.238	0.235	-0.014	0.868	27	6	0.096	0.406	0.538	0.963	0.719	-0.132	0.351
EMBRA925	77	16	0.032	0.329	0.73	0.545	0.839	0.222	0.256	27	5	0.147	0.49	0.444	0.37	0.63	0.325	0.518
EMBRA1008	68	12	0.088	0.412	0.558	0.309	0.656	0.392	0.425	27	6	0.193	0.522	0.375	0.333	0.574	0.322	0.518
EMBRA1851	79	9	0.124	0.436	0.484	0.405	0.665	0.279	0.356	27	4	0.34	0.748	0.248	0.556	0.413	-0.144	0.722
EMBRA1811	78	10	0.145	0.511	0.456	0.615	0.608	0.032	0.562	26	8	0.066	0.372	0.613	1	0.756	-0.128	0.314
EMBRA1957	78	7	0.157	0.497	0.429	0.423	0.597	0.218	0.518	20	3	0.327	0.69	0.245	0.1	0.177	0.447	0.666
EMBRA1977	77	4	0.344	0.701	0.229	0.429	0.356	0.014	0.668	13	5	0.153	0.494	0.436	0.769	0.655	-0.033	0.129
EMBRA1364	37	10	0.137	0.492	0.467	0.216	0.768	0.571	0.137	13	3	0.173	0.507	0.401	0.308	0.566	0.331	0.222
EMBRA943	80	11	0.185	0.541	0.397	0.5	0.579	0.101	0.568	27	6	0.427	0.879	0.208	0.074	0.343	0.671	0.796
EMBRA1374	3	5	0.859	1.357	0.037	0.667	0.744	ND	0.006	24	10	0.095	0.416	0.547	0.25	0.655	0.513	0.407
EMBRA850	80	11	0.077	0.391	0.584	0.738	0.749	0.025	0.375	25	15	0.032	0.334	0.032	0.52	0.825	0.256	0.296
EMBRA915	80	8	0.34	0.72	0.241	0.5	0.411	-0.04	0.693	27	1	1	1.5	0	0	0	ND	1
EMBRA954	79	13	0.055	0.364	0.647	0.468	0.787	0.272	0.331	26	6	0.169	0.49	0.413	0.846	0.574	-0.148	0.462
EMBRA2014	80	6	0.134	0.443	0.462	0.613	0.661	0.072	0.362	27	3	0.363	0.675	0.202	0.704	0.386	-0.186	0.611
EMBRA1616	79	17	0.022	0.312	0.774	0.785	0.875	0.061	0.193	26	7	0.125	0.445	0.486	1	0.643	-0.212	0.425
EMBRA1812	37	9	0.148	0.5	0.447	0.838	0.718	-0.073	0.175	24	5	0.091	0.406	0.55	0.833	0.667	-0.07	0.388
EMBRA813	77	13	0.054	0.369	0.648	0.377	0.774	0.361	0.368	27	7	0.084	0.397	0.565	1	0.738	-0.145	0.37
Mean	-	10	0.205	0.245	0.454	0.510	0.648	0.146	0.403	-	6	0.084	0.574	0.370	0.566	0.548	0.100	0.482
Cumulative	-	180	1.17× 10-15	1.05× 10-5	0.999	-	-	-	-	-	100	4.02× 10-14	2.17× 10-5	0.999	-	-	-	-

n is the sample size; k is the number of alleles per locus; PI is the probability of identity; $PIsibs$ is the probability of finding two full-sib individuals from a population that have the same genotype by chance; MAF is the paternity exclusion probability; MAF is the observed heterozygosity; MAF is the polymorphism information content; MAF is the fixation index corrected for null alleles; MAF is the major allele frequency.

PI and $PIsibs$ of the core marker set were 2.67×10^{-9} and 1.00×10^{-3}, respectively. The selected markers for *E. globulus* were EMBRA850, EMBRA1811, EMBRA813, EMBRA1040, EMBRA1812, EMBRA1374, and EMBRA1616. The combined PI and $PIsibs$ of this set were 1.84×10^{-8} and 1.50×10^{-3}, respectively. The markers EMBRA850, EMBRA813 and EMBRA1616 were common to both core sets.

Population genetic structure

The multilocus data from *E. grandis* and *E. globulus* were used separately in Bayesian assignment analysis. For *E. grandis*, we observed that (K = 4) produced a (ΔK = 6.83), while for *E. globulus* (K = 2) provided a (ΔK = 73.5). These results suggest that for *E. grandis* the germplasm is structured into four clusters, while for *E. globulus* there are two clusters (Figure 1).

Phylogeny tree

The dendrogram reflected exactly the two groups composed of individuals from *E. grandis* and *E. globulus* (Figure 1). No association was observed between the three germplasm providers used in this analysis. This dendrogram is consistent with the results shown in Table 2 and presents a clear individual identification of each clone.

Discussion

The hyper-variable nature and codominant inheritance of microsatellite markers make them an excellent option for clonal identification and relationship analysis (Faria et al. 2011; Li et al. 2011; Tan et al. 2015). However, the success of these studies is highly dependent on the quality and precision of the

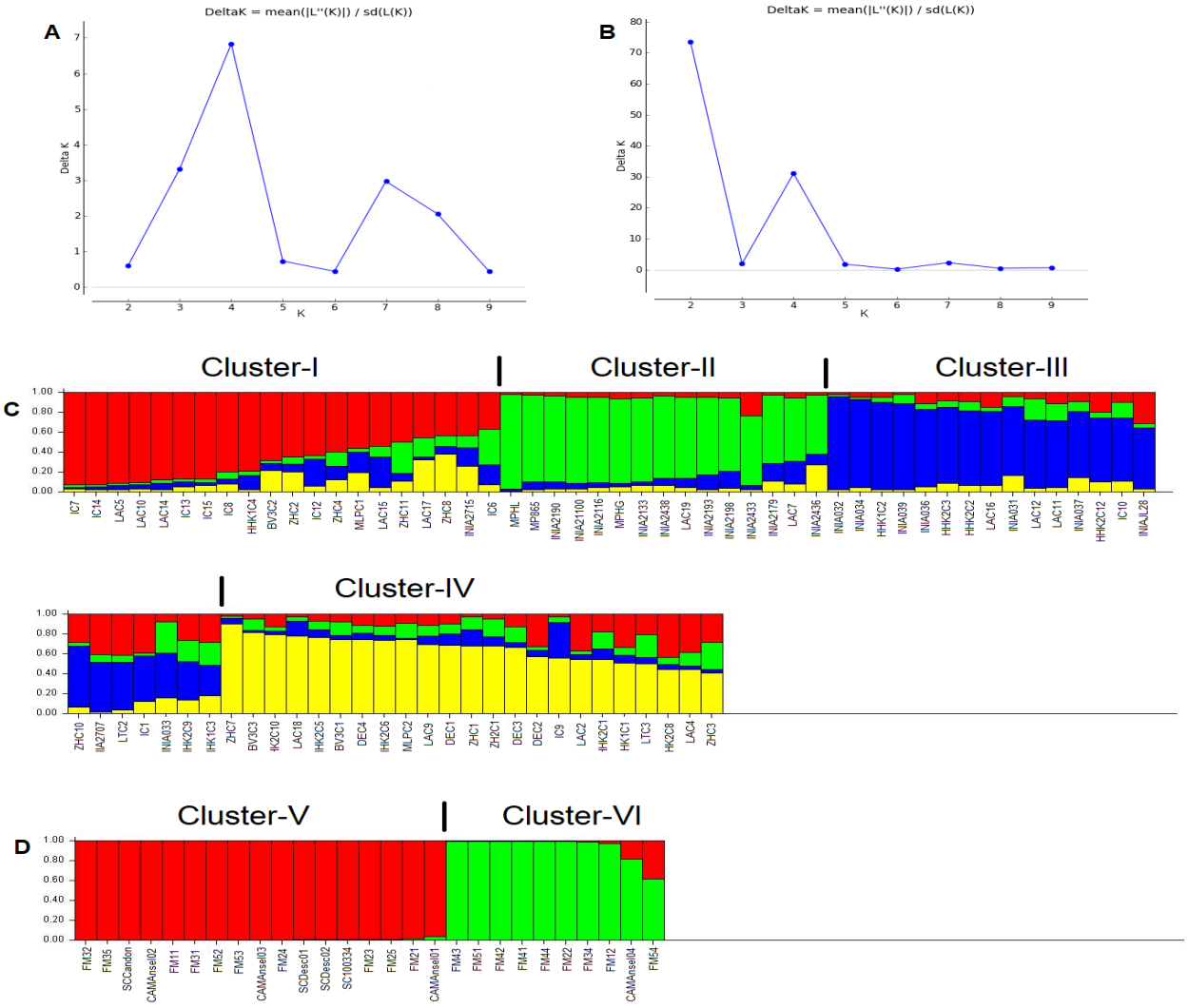

Figure 1
Identification of the optimal number of clusters in Bayesian assignment analysis for *E. grandis* (A, C) and *E. globulus* (B, D).

genotyping data (Tan et al. 2015). Herein, we demonstrate that the 107 clones sampled from Uruguay can be precisely identified by fingerprinting, assigning a unique profile for each studied clone. The alleles were detected by high resolution by capillary electrophoresis showing a high level of polymorphism. The accuracy obtained in the clonal genotyping allows the integration of the results in reference databases to control the determination of identity in future analyses, thus facilitating clonal registration and control during the stages of vegetative propagation and inter-institutional collaboration (Kirst et al. 2005; Li et al. 2011).

Power of fingerprinting

A high mean number of alleles per locus were found in the studied clones (10 for *E. grandis* and 6 for *E. globulus*). These values are lower than the (11.9) reported by Li et al. (2011) who simultaneously analyzed the germplasm of *E. urophylla*, *E. grandis*, *E. camaldulensis*, *E. tereticornis*, and *E. exserta*, which explains the greater number of alleles obtained despite only considering 24 genotypes. On the other hand, Faria et al. (2011) reported 4.3 alleles, considering only 16 individuals per species. Kirst et al. (2005) observed an average estimated value across all loci of (19.8) for 192 *E. grandis* individuals, which is greater than the results found herein. For *E. grandis*, the (PIC) for 14 (0.579-0.875) of the 17 EST-SSR markers was high (> 0.5, Botstein et al. 1980), while in *E. globulus* the (PIC) (0.566-0.825) was high for 12 markers. These results indicate that these markers are suitable for fingerprinting studies and kinship analyses. The markers that were only moderately informative (0.5 > PIC > 0.25, Botstein et al. 1980) in *E. grandis* were EMBRA915 (PIC = 0.411) and EMBRA1977 (PIC = 0.356), while in *E. globulus* they were EMBRA1851 (PIC = 0.413), EMBRA2014 (PIC = 0.386), and EMBRA943 (PIC = 0.343). The markers that showed poorly information content (PIC < 0.25, Botstein et al. 1980) were EMBRA1040 (PIC = 0.238) for *E. grandis*, and EMBRA1957 (PIC = 0.177) and EMBRA915 (PIC = 0) for *E. globulus*. High levels of variation have been reported for PIC among polymorphic molecular genetic marker loci (ISSR, SSR, SNP) in *Eucalyptus* spp., ranging from 0.262 to 0.94 (Kirst et al. 2005; Faria et al. 2010, 2011; Arumugasundaram et al. 2011; Li et al. 2011; Telfer et al. 2015; Costa et al. 2017; Lu et al. 2018; Teixeira et al. 2020). However, the comparison of PIC values across different studies, genetic markers, and loci are not robust due to the fact that PIC is strongly affected by the number of alleles (k), observed heterozygosity (H_o), and major allele frequency (MAF) in the loci. Based on our data set, the Spearman Ranking Correlation was significantly higher than zero between (PIC) and (k) (ρ = 0.621, P= 0.008) and (PIC) and (H_o) (ρ = 0.499, P= 0.041), and was significantly lower than zero between (PIC) and (MAF) (ρ = -0.887, P< 0.0001). These results indicate that samples in which loci have a high number of alleles and heterozygosity and low minor allele frequencies, will present high (PIC). Thus, (PIC) is affected by the samples used to calculate estimates, as large sample sizes of non-inbred individuals may present high (k) and (H_o) for SSR loci, resulting in a lower allele frequency across loci and an increased (PIC). Our results reflect the large number of alleles observed per marker, especially considering that some markers showed more than 10 alleles, which can also explain the differences among studies, gene markers, and loci.

The existence of null alleles can be problematic when estimating statistical parameters and conducting relationship analyses. The appearance of null alleles can be the result of mutations in one or both primer binding sites and differential amplification of allelic variants (Dakin and Avise 2004; Tan et al. 2015; Roman et al. 2020). Studies have also shown that the frequency of null alleles increases when markers that are transferable between species are used (Faria et al. 2010; Acuña et al. 2014; Roman et al. 2020). Due to high selection pressure, EST-SSR markers based on expressed sequences generally have low frequencies of null alleles compared to non-transcribed regions (Ellis and Burke 2007).

The wide variation in the size range of the alleles enables the use of different dyes, which favors clustering in Multiplex PCR reactions and reduces the time and cost of analyses. This, in turn, allows for the routine management of identity certification (Kirst et al. 2005; Faria et al. 2011). Furthermore, the tetra-, penta-, and hexanucleotide characteristics of the EST-SSR markers used in this study facilitate the detection of alleles via capillary electrophoresis, reducing the appearance of ambiguous fingerprinting profiles (Faria et al. 2011).

The (PI) and ($PIsibs$) parameters have been widely used as indicators of discrimination power by fingerprinting of molecular markers (De Lucas et al. 2008; Faria et al. 2011; Li et al. 2011; Tan et al. 2015). In this study, the pooled (PI) for the 17 markers was low for *E. grandis* (1.18×10^{-15}) and *E. globulus* (4.03×10^{-14}). However, the (PI) can be overestimated due to the substructure of the population between the tested samples and the assumption that segregation between loci are completely independent (Waits et al. 2001). Although ($PIsibs$) is considered a more conservative parameter, the results remain very low for the two studied species (1.05×10^{-5} and 2.17×10^{-5} for *E. grandis* and *E. globulus*, respectively).

The two sets of seven core markers identified herein can be used as a monitoring tool in clonal multiplication in future analyses of the Uruguayan germplasm bank, thus ensuring the exclusivity of each fingerprinting molecular profile for the 107 clones. For *E. grandis*, the results for the core set were (PI = 2.67×10^{-9} and $PIsibs$ 1.00×10^{-3}), while for *E. globulus* they were (PI = 1.84×10^{-8} and $PIsibs$ 1.50×10^{-3}). Consequently, these sets of core markers can be used to effectively monitor clonal traceability during vegetative propagation of the 107 clones. The data reported here is the first DNA fingerprint database of Eucalyptus clones in Uruguay. These primers have already been used by other laboratories and are recommended for the integration of data from different sources; however, it is always advisable to use common reference samples when working collaboratively across institutions (Faria et al. 2011; Tan et al. 2015).

Genetic structure and phylogenetic analysis

In this study we focus on the genetic information produced by fingerprinting Eucalyptus clones from three providers. However, we do not have information on clonal origin, provenance,

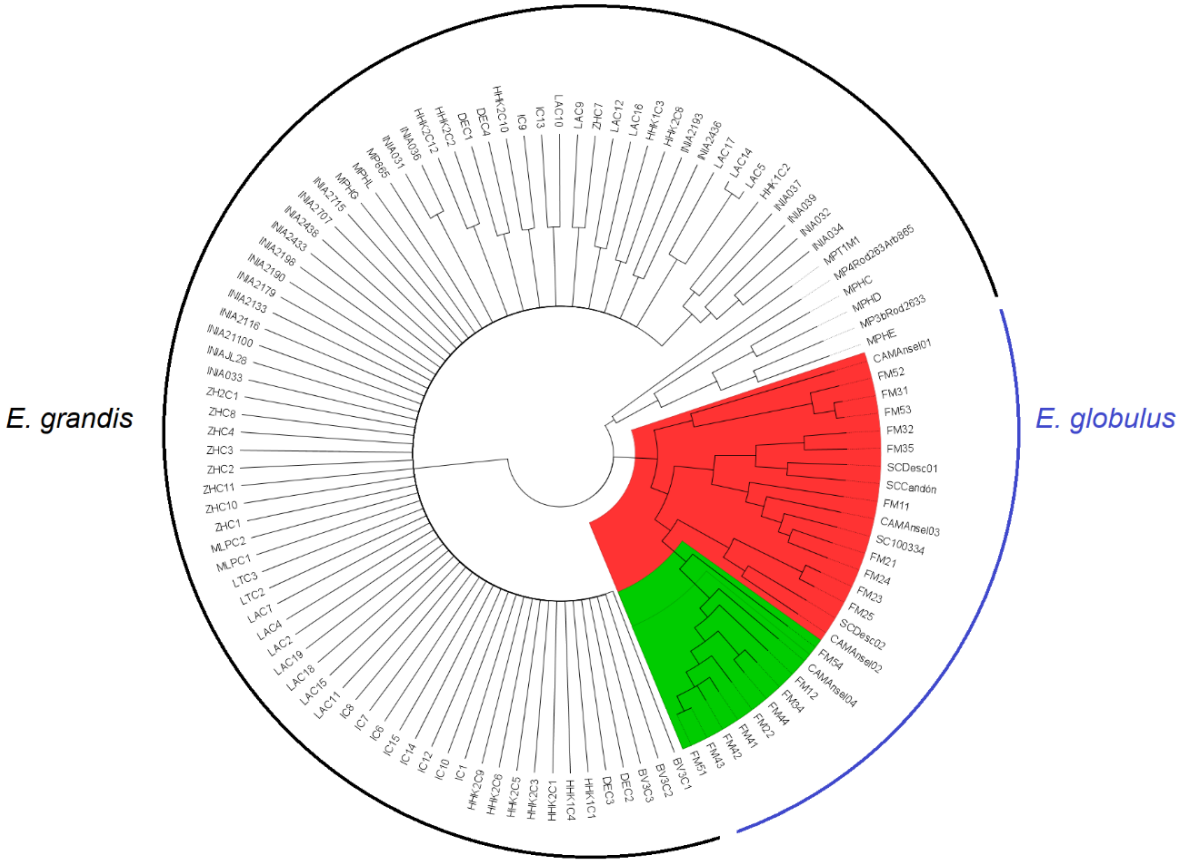

Figure 2
Dendrogram of the 107 clones obtained with 17 EST-SSR markers. The 80 *E. grandis* clones did not show clusters concordant with the results of STRUCTURE, while the 27 *E. globulus* clones are subdivided in two clusters (shown in red and green) based on the STUCTURE analysis.

productivity, or other aspects related to genetic improvement as it is the companies' intellectual property. The clonal genetic structure was inferred based on the genotyping dataset. The *E. grandis* clones were grouped in 4 clusters (K=4), numbered I to IV (Figure 1A and 1C). Cluster I, is composed of 20 clones, 19 of which belong to the provider Lumin and one (INIA2715) to INIA. Cluster II (15 clones) includes a mixture of clones from the three providers, while cluster III (22 clones) contains clones from Lumin and INIA. Cluster IV (24 clones) is composed exclusively of Lumin genotypes. The structure genetic analysis indicates that clusters I and IV are preferentially associated with germplasm from the Lumin company. No correlation is observed among clusters II and III and the three providers (Figure 1C). We also observed that there is a mixture among individuals of the clusters (Figure 1C). This is related to the fact that private genetic improvement programs in Uruguay use germplasm with diverse origins and provenances, and some common sources of germplasm with shared origins are expected considering that genotypes must be adapted to local conditions. The multilocus data for *E. globulus* show two groups (K = 2) numbered V and VI (Figure 1B and 1D); these groups are clearly defined (ΔK = 73.5) and the mixture of materials is less prevalent than that observed for *E. grandis* (ΔK = 6.83) (Figure 1A and 1B), indicating that *E. globulus* genotypes are more divergent than *E. grandis*. In the case of *E. globulus*, samples were provided by only one provider, CEBIOF.

The phylogenetic reconstruction for the 107 clones based on 17 molecular markers generated the first register of the relationship between these genotypes (Figure 2). The 80 *E. grandis* genotypes are grouped separately from the 27 *E. globulus* genotypes. No correlations were observed among the three providers and the *E. grandis* dendrogram. This supports the data obtained for genetic structure (Figure 1 and 2). For E. globulus, the dendrogram is subdivided into two clusters (Figure 2), which are consistent with the results obtained for the analysis of genetic structure (Figure 1D). It would be interesting to contrast these results with the pedigree information of the breeding programs; however, such an analysis is impossible given the politics of protecting intellectual property. This has been demonstrated by forestry companies in their efforts to protect elite genotypes, using these molecular tools to legitimize copyright in court proceedings (Campo Grande News, 2014).

Conclusions

The markers used in this study enabled us to verify that the elite clones of Uruguay can be accurately identified. The database generated applies to both *E. grandis* and *E. globulus*, and these tools can support the certification of propagated materials, thus minimizing labeling errors.

Acknowledgements

The authors thank INIA and ANII for the financial support necessary to conduct this analysis.

Conflict of Interest

The authors declare no conflict of interest.

References

Acuña CV, Villalba P, Hopp HE, Poltri SNM (2014) Transferability of microsatellite markers located in candidate genes for wood properties between *eucalyptus* species. Forest Systems 23(3):506–512. https://doi.org/10.5424/fs/2014233-05279

Arumugasundaram S, Ghosh M, Veerasamy S, Ramasamy Y (2011) Species discrimination, population structure and linkage disequilibrium in *Eucalyptus camaldulensis* and *Eucalyptus tereticornis* using SSR markers. PLoS ONE 6(12): e28252. https://doi.org/10.1371/journal.pone.0028252

Bautista R, Crespillo R, Cánovas FM, Claros MG (2003) Identification of olive-tree cultivars with SCAR markers. Euphytica 129:33–41. https://doi.org/10.1023/A:1021528122049

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32(3):314–331. https://doi.org/10.17348/era.9.0.151-162

Brancalion P, Amazonas N, Chazdon R, Melis J, Rodrigues R, Silva C, Sorrini T,Holl K (2020) Exotic eucalypts: From demonized trees to allies of tropical forest restoration? Journal of Applied Ecology 57(1):55–66. https://doi.org/10.1111/1365-2664.13513

Campo Grande News (2014) https://www.campograndenews.com.br/economia/laudo-indica-que-eldorado-usou-muda-de-eucalipto-criada-pela-concorrente. Available at [cited 01/11/2021]

Costa J, Vaillancourt RE, Steane DA, Jones RC, Marques C (2017) Microsatellite analysis of population structure in *Eucalyptus globulus*. Genome 60:770–777. https://doi.org/10.1139/gen-2016-0218

Dakin E, Avise J (2004) Microsatellite null alleles in parentage analysis. Heredity 93:504–509. https://doi.org/10.1038/sj.hdy.6800545

De-Lucas AI, Santana JC, Recio P, Hidalgo E (2008) SSR-based tool for identification and certification of commercial *Populus* clones in Spain. Annals of Forest Science 65(1):107–107. https://doi.org/10.1051/forest:2007079

Earl DA (2012) Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Ellis J, Burke J (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132. https://doi.org/10.1038/sj.hdy.6801001

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clustersof individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Faria DA, Mamani ME, Pappas MR, Pappas GJ, Grattapaglia D (2010) A Selected Set of EST-derived microsatellites, polymorphic and transferable across 6 Species of *Eucalyptus*. Journal of Heredity 101(4):512–520. https://doi.org/10.1093/jhered/esq024

Faria DA, Mamani EM, Pappas GJ, Grattapaglia D (2011) Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. Tree Genetics and Genomes 7(1):63–77. https://doi.org/10.1007/s11295-010-0315-9

Ferreira ME, Grattapaglia D (1995) Introdução ao uso de marcadores moleculares em análise genética. EMBRAPA-CENARGEN, Brasília.

Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytologist 179(4):911–929. https://doi.org/10.1111/j.1469-8137.2008.02503.x

Grattapaglia D, Mamani EM, Silva-Junior OB, Faria DA (2015) A novel genome-wide microsatellite resource for species of *Eucalyptus* with linkage-to-physical correspondence on the reference genome sequence. Molecular Ecology Resources 15(2):437–448. https://doi.org/10.1111/1755-0998.12317

Gross BL, Wedger MJ, Martinez M, Volk GM, Hale C (2018) Identification of unknown apple (*Malus* × *domestica*) cultivars demonstrates the impact of local breeding program on cultivar diversity. Genet Resour Crop Evol (65):1317–1327. https://doi.org/10.1007/s10722-018-0625-6

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16(5):1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kirst M, Cordeiro CM, Rezende GD, Grattapaglia D (2005) Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. Journal of Heredity 96(2):161–166. https://doi.org/10.1093/jhered/esi023

Li F, Gan S, Zhang Z, Weng Q, Xiang D, Li M (2011) Microsatellite-based genotyping of the commercial *Eucalyptus* clones cultivated in China. Silvae Genetica 60(5):216–223. https://doi.org/10.1515/sg-2011-0029

Lu WHL, Qi J, Lan J, Luo JZ (2008) Genetic diversity of advance generation breeding on *Eucayptus urophilla* in Chine. Journal of Tropical Forest Science 30(3):320–329. https://doi.org/10.26525/jtfs2018.30.3.320329

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Molecular Ecology 4(3):347–354. https://doi.org/10.1111/j.1365-294X.1995.tb00227.x

Pasqualone A, Montemurro C, di Rienzo V, Summo C, Paradiso VM, Caponio F (2016) Evolution and perspectives of cultivar identification and traceability from tree to oil and table olives by means of DNA markers. J Sci Food Agric 96(11):3642-3657. https://doi.org/10.1002/jsfa.7711

Perrier X, Jacquemoud-Collet P (2006) DARwin software. http://darwin.cirad.fr/darwin.

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

Rambaut, A. (2010) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Rezende GD, Resende MDV, de Assis TF (2014) *Eucalyptus* breeding for clonal forestry. In: Fenning T (eds) Challenges and opportunities for the world's forests in the 21st century. Dordrecht Springer, pp 393–424, ISBN 978-94-007-7076-8

Rodrigues A, Faria J (2021) Profiling the variability of *Eucalyptus* essential oils with activity against the *Phylum nematoda*. Proceedings of the 1st International Electronic Conference on Biological Diversity, Ecology and Evolution, MDPI: Basel, Switzerland, https://doi.org/10.3390/BDEE2021-09425

Roman MG, Gangitano D, Figueroa A, Solano J, Anabalón L, Houston R (2020) Use of *Eucalyptus* DNA profiling in a case of illegal logging. Science and Justice 60(6):487–494. https://doi.org/10.1016/j.scijus.2020.09.005

Sumathi M, Bachpai VK, Mayavel A, Dasgupta MG, Nagarajan B, Rajasugunasekar D, Sivakumar V, Yasodha R (2018) Genetic linkage map and QTL identifica-

tion for adventitious rooting traits in red gum eucalypts. 3 Biotech 8(5):242. https://doi.org/10.1007/s13205-018-1276-1

Sumathi M, Yasodha R (2014) Microsatellite resources of *eucalyptus*: Current status and future perspectives. Botanical Studies 55(1):1–16. https://doi.org/10.1186/s40529-014-0073-3

Tan LQ, Peng M, Xu LY, Wang LY, Chen SX, Zou Y, Qi GN, Cheng H (2015) Fingerprinting 128 Chinese clonal tea cultivars using SSR markers provides new insights into their pedigree relationships. Tree Genetics and Genomes 11(5):1–12. https://doi.org/10.1007/s11295-015-0914-6

Teixeira GC, Konzen ER, Faria JCT, Gonçalves DS, Carvalho D, Brondani GE (2020) Genetic diversity analysis of two *Eucalyptus* species using ISSR markers. Ciencia Florestal 30(1):270–278. https://doi.org/10.5902/1980509832804

Telfer EJ, Stovold GT, Li Y, Silva-Junior OB, Grattapaglia GD, DungeyHS (2015) Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: A comparative analysis of marker data power and robustness. PLOS ONE 10(7): e0130601. https://doi.org/10.1371/journal.pone.0130601

Torres-Dini D, Nunes ACP, Aguiar A, Nikichuk N, Centurión C, Cabrera M, Moraes MLT, Resende MDV, Sebbenn AM (2016) Clonal selection of *Eucalyptus grandis* x *Eucalyptus globulus* for productivity, adaptability, and stability, using SNP markers. Silvae Genetica 65(2):30–38. https://doi.org/10.1515/sg-2016-0014

Veloso MM, Simões-Costa MC, Carneiro LC, Guimarães JB, Mateus C, Fevereiro P, Pinto-Ricardo C (2018) Olive Tree (*Olea europaea* L.) Diversity in Traditional Small Farms of Ficalho, Portugal. Diversity 10(1):5. https://doi.org/10.3390/d10010005

Wagner HW, Sefc KM (1999) Identity 1.0. University of Agricultural Sciences, Vienna.

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. Molecular Ecology 10(1):249–256. https://doi.org/10.1046/j.1365-294X.2001.01185.x

Weir BS (1996) Genetic Data Analysis II. Massachusetts, USA: Sinauer Associates, 445 p, ISBN 0878939024

Wünsch A, Hormaza J (2002) Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. Euphytica 125:59-67. https://doi.org/10.1023/A:1015723805293