Microssatellite markers for plant variety protection of clonally propagated forest trees: a case study in teak (*Tectona grandis* L.f.)

Mariana de M. Queiroz[1], Daniele A.A. Arriel[1,2], Sidney F. Caldeira[1], Antônio R. Higa[3], Stefânia P. Araújo[4], Alexandre M. Sebbenn[5], Dario Grattapaglia[6]*

[1] Programa de Pós-Graduação em Ciências Florestais e Ambientais, UFMT, Cuiabá, MT, 78060-900, Brazil. queirozmariana1@hotmail.com, ORCID iD: https://orcid.org/0000-0002-8672-9331; sidneycal@gmail.com, ORCID iD: https://orcid.org/0000-0001-6042-4313

[2] Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Campus Monte Carmelo, Monte Carmelo, MG, 38500-000, Brazil. daniarriel@ufu.br, ORCID iD: https://orcid.org/0000-0001-7887-1664

[3] Universidade Federal do Paraná, Curitiba, PR, 80060-000, Brazil. antonio.higa@gmail.com, ORCID iD: https://orcid.org/0000-0002-7275-8043

[4] Serviço Nacional de Proteção de Cultivares, Ministério da Agricultura, Pecuária e Abastecimento, Brasília, DF, 70.043-900, Brazil. stefania.araujo@agro.gov.br, ORCID iD: https://orcid.org/0000-0002-2597-2143

[5] Instituto Florestal de São Paulo, CP 1322, São Paulo, SP, 01059-970, Brazil. alexandresebbenn@yahoo.com.br, ORCID iD: https://orcid.org/0000-0003-2352-0941

[6] Plant Genetics Laboratory, EMBRAPA Genetic Resources and Biotechnology, 70770-970, DF, Brasilia, Brazil, and Graduate Program in Genomic Sciences Biotechnology and Universidade Católica de Brasília, SGAN Qd 916, 70790-160, DF, Brasília, Brazil. dario.grattapaglia@embrapa.br, ORCID iD: https://orcid.org/0000-0002-0050-970X

*Corresponding author: Dario Grattapaglia, E-mail: dario.grattapaglia@embrapa.br

Abstract

Forest trees cultivars developed by breeders have been increasingly deployed as clones, following long generations of breeding and testing. An established protocol for distinctiveness, uniformity and stability (DUS) testing becomes an essential element for protecting the intellectual property rights associated with these clones. DUS testing with morphological descriptors has, however, shown limitations in categorically distinguishing cultivars, especially with narrow genetic base. DNA fingerprinting based on microsatellite markers has been a powerful tool to discriminate clones. Teak (*Tectona grandis*) is an economically valuable exotic timber species planted in Brazil. We assessed the individual and combined performance of a selected set of 21 teak microsatellites for identity analysis and parentage testing in a sample of 50 clones planted in Brazil. The clones displayed high genetic diversity attributable to their wide provenance origin. The 21 microsatellites combined provide high power of individual identification with a combined probability of identity of 1.84E-23 for unrelated individuals, and 4.82E-09 for full-sibs, and a power of paternity exclusion higher than 99.99999 % in all testing scenarios. Variable subsets of these 21 markers still provide abundant power of discrimination, although a recommended set should include a minimum of 12 markers with higher information content and reliable genotyping performance. Relatedness and genetic distance analyses revealed unexpected identities or significant recent shared ancestry among otherwise considered unique clones. These results advocate for the importance of including DNA markers, at least as discretionary complementary descriptors to the 22 morphological traits currently adopted for plant variety protection of teak cultivars in Brazil.

Keywords: *Tectona grandis, plant variety protection, microsatellites, relatedness, probability of identity, paternity test*

Introduction

Teak (*Tectona grandis* L. f., *Lamiaceae* family) is a tropical tree that naturally occurs in India, Laos, Myanmar, and Thailand (Khaing et al. 2017; Prasetyo et al. 2020). Plantations of teak are on the rise worldwide due to the desirable characteristics of the wood, such as durability and strength, resulting in a high commercial value of the species (Kollert and Kleine 2017; Prasetyo et al. 2020). Globally, the total estimated area planted

DOI:10.2478/sg-2023-0019
edited by the Thünen Institute of Forest Genetics

with the species ranges from 4.25 to 6.89 million hectares, mostly concentrated in Asia, Africa, and South America (Kollert and Kleine 2017). In Brazil, the species was first introduced commercially in the late 60's by the company Cáceres Florestal Ltda. in Mato Grosso State (Murillo et al. 2013; Queiroz et al. 2021). Teak plantations in Brazil have obtained higher silvicultural yields than those in natural areas, reducing the rotation cycle from 80 to 25 years (Tsukamoto et al. 2003). Studies have reported higher growth in diameter at breast height (DBH) and dominant height in forest stands in Brazil when compared to stands in India, Indonesia, and Ghana (da Conceicao et al. 2012; Pelissari et al. 2013; Vendruscolo et al. 2019). These results are largely related to the advanced silvicultural methods used in forest plantations in Brazil and the adaptation of the species to the local edaphoclimatic conditions.

Commercial reforestation of teak in Brazil is expanding, increasing about 30 % between 2010 and 2018, from 65,440 to 93,957 ha of planted teak forest in the country (IBÁ 2022). Until the mid-2000s, teak propagation in Brazil was predominantly seminal (Murillo et al. 2013) (Queiroz et al. 2021). However, in 2003 teak clones from Sabah Malaysia breeding program were introduced in Brazil and the first clonal plantations of the species were established (Murillo et al. 2013). Currently, the establishment of clonal stands is a widely adopted practice for teak in Brazil (Madi et al. 2020; Queiroz et al. 2021). The selection of superior clones offers numerous advantages, such as resistance to disease and wind drop, greater plantation homogeneity, improved quality of raw materials and increased productivity. At the same time it has become increasingly essential to protect the breeding companies' intellectual property (IP) rights associated with the newly developed cultivars. Teak genetic improvement involve long generations of breeding and testing requiring significant investment.

In Brazil, protecting a clone through the National Plant Variety Protection Service (SNPC) is a legal procedure that a breeder or a business owner must follow in order to ensure intellectual property rights over the developed material. The application can only refer to a single cultivar and must contain evidence of distinctiveness, uniformity and stability (DUS) for domestic and foreign cultivars, among other requirements set forth in the Brazilian Plant Protection Law (LPC nº 9.456 of 25 April 1997). To ensure protection of the IP, the new variety must be the result of genetic improvement and be identifiable and distinguishable through morphological descriptors. Following the increasing relevance of teak as a timber crop, in August 2020 the governing body of plant variety protection of the Brazilian Ministry of Agriculture released an official set of 22 morphological descriptors (https://www.in.gov.br/en/web/dou/-/ato-n-9-de-3-de-agosto-de-2020-270222225) (Reategui-Betancourt et al. 2020).

Despite this important advancement, current DUS testing based on morphological descriptors has shown limitations in categorically distinguishing cultivars as frequently seen in other forest trees and crops in general (Jamali et al. 2019). Especially in perennial trees such as teak, DUS assessment is a time-consuming, costly and frequently imprecise process. The relatively small number of traits employed can be influenced by environmental conditions that vary along the long life of a tree. In teak it has been shown that the evaluation must be performed under the same planting conditions in which these descriptors were developed (Reategui-Betancourt et al. 2020). Furthermore, frequently commercial cultivars are selected from narrow elite parental backgrounds, such a single half- or full-sib family, leading to high individual similarity in morphological descriptors and thereby difficulty in cultivar identification. This clearly demands for the adoption of a DNA-based system (Grattapaglia et al. 2003; Kirst et al. 2005) which has been endorsed by the International Union for the Protection of New Varieties of Plants (UPOV) since 2013, and recently adopted (UPOV 2020).

DNA marker technologies for individual identification and discrimination have now an almost 40-year long history (Gill et al. 1985), not only in human forensics, but also plant breeding, population genetics and evolutionary research (Garrido-Cardenas et al. 2018). Not surprisingly DNA markers have been increasingly used in plant variety protection as well (Jamali et al. 2019). Molecular markers have several advantages over conventional morphological DUS testing, including high power of discrimination, high repeatability, fast and accessible data production at any life stage of the plant, with no impact of environmental variables on the DNA profiles.

A number of laboratory methods have been developed to reveal DNA polymorphisms among individuals, although essentially two sources are usually targeted: simple sequence repeats (SSR), best known as microsatellites, or single nucleotide polymorphisms (SNPs). Despite the rapid evolution of SNP genotyping platforms based on fixed content arrays and next generation sequencing, microsatellite markers have been and still are the workhorse for fast, precise and considerably more low-cost individual identification for major forest tree species (Telfer et al. 2015) and the vast majority of plant species (Yang et al. 2021). They are widely distributed throughout the genome, display a high degree of polymorphism in terms of the number of alleles per locus, allowing precise discrimination of individuals including cases of close relationship by recent common ancestry (Powell et al. 1996). Furthermore, not only they allow ascertaining that two individuals are different, but they also allow estimating formal likelihoods of the observed data conditional of the hypothesis of genetic identity (Kirst et al. 2005) as well as kinship relationships (Ashley 2010).

Microsatellites have been used to supplement morphological descriptors in plant variety protection procedures of a large number of plant species, including all the major crops such as soybean, maize, wheat and rice and a number of horticultural and forest trees (Jamali et al. 2019). In teak, the first and most widely used microsatellites across the various reports published to date, were developed by Verhaegen et al. (2005). Other reports followed, describing additional microsatellites (Huang et al. 2016; Mahesh et al. 2016). With the publication of a draft genome sequence for teak, several thousand microsatellites were discovered and characterized in silico, and an additional set of ten markers was validated electrophoretically (Yasodha et al. 2018). From the practical standpoint of cultivar identification in plant variety protection, it is essential to assess

the individual and combined performance of microsatellites for individual identification and parentage testing in a representative set of genetic material where they are to be applied. In this study, we carried out this assessment by estimating parameters of genetic information content for 21 commonly used teak microsatellites, including circumstances that involve the discrimination of genetically related individuals. Furthermore, by carrying out a detailed analysis of the multilocus microsatellite profiles we also evaluated the genotyping reliability of the markers and estimated the combined performance of selected subsets of markers for routine laboratory practice in teak cultivar identification and parentage testing.

Material and Methods

Plant material and sampling

Two groups of 25 clones each were provided by two teak-producing forestry companies: clones P1 through P25 from PROTECA and clones B1 through B25 from BIOTECA a subsidiary of FLORESTECA (Table 1). These 50 clones, either commercial varieties or under advanced testing, were selected from primary provenances in their center of origin or from secondary local sources in Mato Grosso, Brazil. For DNA extraction, healthy leaves were collected from each clone, placed in labelled paper bags and transported to the Forest Improvement and Biotechnology laboratory at the Faculty of Forest Engineering, Federal University of Mato Grosso (UFMT), Cuiabá, where they were stored at -20 °C until DNA extraction.

Microsatellite analysis

DNA was extracted using a modified CTAB method (Inglis et al. 2018). Twenty-one microsatellite markers out of those available in the literature were selected to be included in this study, based on a preliminary assessment of the published information regarding information content and ease of analytical interpretation of genotypes. For all microsatellites tested, forward primers were labeled with 6-FAM fluorochrome at the 5' end. Single locus PCR reactions and capillary electrophoretic separation were performed in a solution containing 20 ng of DNA, 0.2 mM from each primer and 1 X Go Taq® Master Mix 2X (Promega) and ultrapure water to a final volume of 15 µL. The amplification cycle varied for each primer, as described in the literature. For fragment analysis, 7.8 µL of Hi-DI™ formamide (Applied Biosystems) and 0.2 µL of GeneScan™LIZ-600 or 1200 Size Standard (Applied Biosystems), depending on fragment size, were added to 1.2 µL of PCR product. The mixture was denatured for 3 min at 95 °C and genotyping was performed in an Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems) at the Forest Pathology/Bioagro/UFV Laboratory. Genotypes were determined using GeneMapper® v.5.0 software (Applied Biosystems) and exported to spreadsheets with alleles called in base pairs with two decimal cases. Observed allele lengths were subject to automated binning to absolute allele sizes using TANDEM (Matschiner and Salzburger 2009).

Data analysis

To evaluate the power of the 21 microsatellites for individual identification and parentage testing, the following parameters were estimated using GENALEX 6.5 (Peakall and Smouse 2012): (i) Number of alleles per locus (N_a); (ii) Observed allele size range in base pairs; (iii) Observed (H_o) and expected (H_e) heterozygosity; (iv) Probability of Identity (PI) that corresponds to the probability of two unrelated individuals displaying the same genotype; (v) Probability of identity (PI_{sib}) that corresponds to the probability of two sibs displaying the same genotype; and (vi) Probability of paternity exclusion (PE) that corresponds to the probability with which the marker excludes an erroneously pointed individual tree as being the parent of an offspring. This last parameter was estimated taking into account different situations in paternity analysis: Probability of paternity exclusion of one candidate parent given the genotype of a known parent (PE_{kp}), a common situation when paternity is investigated in open pollinated progeny individuals with maternal control; Probability of paternity exclusion of single candidate parent (PE_{sp}) and Probability of paternity exclusion of a candidate parent pair PE_{pp}, a common situation when paternity and maternity need to be checked in progeny individuals with no prior knowledge of any of the two parents. The PI combined and PE combined were also estimated for all n = 21 markers combined for each one of the variations of the parameter as follows:

$$PI_{com} = \prod_{i=1}^{n} PI_i$$

$$PE_{com} = 1 - \left[\prod_{i=1}^{n}(1 - PE_i)\right]$$

Genetic relatedness (θ_{xy}) among the 50 clones was determined using the estimator for hypervariable molecular markers described by Lynch and Ritland (1999) and genetic distances were estimated using the method described by (Smouse and Peakall 1999) using GENALEX 6.5 (Peakall and Smouse 2012). From the matrix of genetic distances an UPGMA dendrogram was built for visualization using MEGA X (Stecher et al. 2020). To assess the extent of independent assortment of alleles at the microsatellites, we tested for their pairwise genotypic linkage disequilibrium (LD) using the log-likelihood ratio G-statistic for a nominal significance level of α=0.05 after Bonferroni correction, using FSTAT 2.9.4 (Goudet 1995), considering all 50 clones, as well as only considering clones with relatedness $\theta_{xy} <$ 0.25 and clones with relatedness $\theta_{xy} >$ 0.25.

Table 1
List of the 50 *Tectona grandis* clones sampled for the study with their respective provenance information as provided by the breeding companies. Clones P1 through P25 were provided by PROTECA and clones B1 through B25 by FLORESTECA

Clone	Provenance	Clone	Provenance
P1	Solomon Islands	B1	Brazil (Mato Grosso)
P2	Solomon Islands	B2	Brazil (Mato Grosso)
P3	Solomon Islands	B3	Brazil (Mato Grosso)
P4	Solomon Islands	B4	Brazil (Mato Grosso)
P5	Solomon Islands	B5	Brazil (Mato Grosso)
P6	Solomon Islands	B6	Southeast Asia
P7	Solomon Islands	B7	Brazil (Mato Grosso)
P8	Solomon Islands	B8	Unknown origin
P9	Solomon Islands	B9	Brazil (Mato Grosso)
P10	Brazil (Cáceres)	B10	Brazil (Mato Grosso)
P11	Brazil (Alta Floresta)	B11	Brazil (Mato Grosso)
P12	Malaysia	B12	Brazil (Mato Grosso)
P13	Malaysia	B13	Brazil (Mato Grosso)
P14	India	B14	Brazil (Mato Grosso)
P15	India	B15	Matta Ayer, Malaysia
P16	India	B16	Unknown origin
P17	Indonesia	B17	Solomon Islands
P18	Indonesia	B18	Solomon Islands
P19	Ghana	B19	Solomon Islands
P20	Ghana	B20	Solomon Islands
P21	Laos	B21	Solomon Islands
P22	Laos	B22	Malaysia
P23	Laos	B23	Malaysia
P24	Laos	B24	Malaysia
P25	Tanzania	B25	Malaysia

Results

To improve the consistency of genotypic data, the raw genotypes were subject to automated binning to absolute allele sizes using a least squares minimization procedure of rounding errors based on the expected size of the simple sequence repeat motif. The full dataset used in this study is provided as Supporting information file S1. Most of the microsatellites were dinucleotide repeats and, as expected, showed higher numbers of alleles on average, than the longer repeats (tri, tetra and pentanucleotides) markers. The sampled teak clones displayed high levels of genetic diversity for the sampled microsatellites. A total of 228 alleles were identified across the 21 microsatellites (mean of 10.9), the observed heterozygosity (H_o) ranged from 0.4 to 0.88 (mean of 0.64) and expected heterozygosity (H_e) from 0.504 to 0.905 (mean of 0.738) (Table 2).

Overall, the combined set of 21 microsatellites provides an extremely high power of individual identification with a combined PI of 1.843E-23 for unrelated individuals and 4.823E-09 for full-sib individuals, which correspond respectively, to a power of discrimination of 1 in 5.42E+22 and 1 in 2.07E+08. The set of markers combined also yields a probability of paternity exclusion higher than 99.99999 % in every testing scenario with either a parent known, a single parent tested or parent pair testing (Table 2). At the single marker level markers CIRAD3TeakDa09 and CIRAD1TeakF05 were the less informative markers both for individual identification and paternity testing due to spiky allele frequency distributions where one of the alleles has a considerably higher frequency than the others (Supporting information file S2).

From the full distribution of the 1225 pairwise relatedness (-0.0102±0.065) the following "*ad hoc*" relatedness thresholds were adopted, based on a departure of one standard deviation from the theoretical expected values, to putatively propose different levels of existing relatedness among the 50 clones. From theoretical expectations the relatedness between pairs of half-sibs, full-sibs, and clones are 0.125, 0.25, and 0.5 respectively. Pairs of individuals with $\theta_{xy} \leq 0.065$ were considered unrelated, pairs with θ_{xy} ranging from 0.066 to 0.19 (θ_{xy} = 0.125+0.065) were considered putative half-sibs (i.e., 0.065< $\theta_{xy} \leq 0.19$), pairs with θ_{xy} ranging from 0.191 to 0.315 (θ_{xy} =0.25+0.065) were considered putative full-sibs (i.e., 0.19<θ_{xy} ≤ 0.315), and pairs with $\theta_{xy} > 0.315$ were putative clonal replicates. Based on these thresholds, 1128 (92.1 %) were deemed unrelated pairs of individuals, 77 (6.3 %) pairs of half-sibs, 11 (0.89 %) pairs of full-sibs and nine (0.7 %) clonal pairs Supporting information file S3). The 77 pairs of half-sibs involved 41 different clones while the 11 pairs of full-sib involved 8 distinct clones, these latter ones themselves already involved in the half-sib relationships. The suspect clonal identities involved 12 different clones which were subject to further scrutiny.

To evaluate the hypothesis of clonality among otherwise considered distinct clones, a dendrogram (Figure 1) was built from a matrix of genetic distances (Supporting information file

Table 2
Summary of the parameters of information content of the 21 microsatellites evaluated. SSR (Simple Sequence Repeat) motif; Na: number of observed alleles; Ho: observed heterozygosity; He: expected heterozygosity; PI: Probability of Identity; PI_{sib} Probability of Identity of two sibs; PE: Probability of paternity exclusion with one known parent (PE_{kp}), single parent test (PE_{sp}) and Parent pair test (PE_{pp})

Microsatellite marker *	SSR Motif	Na	Allele size range (bp)	Ho	He	PI	Pisibs	PEkp	PEsp	PEpp
CIRAD1TeakA06	dinucleotide	13	170-210	0.740	0.850	0.040	0.335	0.702	0.538	0.870
CIRAD3TeakA11	dinucleotide	13	238-268	0.620	0.696	0.117	0.431	0.503	0.315	0.716
CIRAD3TeakB02	dinucleotide	14	194-226	0.800	0.866	0.031	0.325	0.735	0.580	0.896
CIRAD1TeakB03	dinucleotide	8	224-242	0.740	0.805	0.066	0.364	0.615	0.439	0.796
CIRAD2TeakB07	dinucleotide	11	108-156	0.660	0.780	0.080	0.380	0.577	0.399	0.763
CIRAD2TeakC03	dinucleotide	10	248-270	0.620	0.746	0.093	0.400	0.550	0.367	0.749
CIRAD3TeakDa09	dinucleotide	6	175-189	0.440	0.504	0.277	0.567	0.303	0.141	0.481
CIRAD4TeakDa12	dinucleotide	6	111-133	0.560	0.619	0.213	0.494	0.354	0.207	0.520
CIRAD3TeakE06	dinucleotide	9	181-205	0.460	0.614	0.168	0.485	0.427	0.237	0.645
CIRAD3TeakF01	dinucleotide	15	178-212	0.700	0.844	0.039	0.338	0.703	0.539	0.878
CIRAD4TeakF02	dinucleotide	8	196-220	0.600	0.626	0.179	0.482	0.404	0.230	0.598
CIRAD1TeakF05	dinucleotide	9	230-258	0.400	0.602	0.207	0.501	0.368	0.206	0.552
CIRAD4TeakH09	dinucleotide	7	125-143	0.500	0.627	0.192	0.485	0.384	0.223	0.566
CIRAD1TeakH10	dinucleotide	17	203-269	0.840	0.862	0.034	0.328	0.726	0.568	0.890
IFT14	tetranucleotide	6	261-281	0.780	0.713	0.130	0.426	0.474	0.302	0.657
IFT168	trinucleotide	7	286-313	0.500	0.648	0.162	0.467	0.425	0.247	0.621
IFT777	pentanucleotide	4	329-344	0.580	0.675	0.158	0.452	0.423	0.254	0.600
TG11	dinucleotide	12	352-380	0.520	0.869	0.031	0.323	0.737	0.581	0.895
TG12	dinucleotide	15	264-302	0.880	0.887	0.023	0.312	0.773	0.629	0.920
TG16	dinucleotide	13	151-181	0.720	0.764	0.076	0.387	0.588	0.404	0.792
TG9	dinucleotide	25	358-422	0.780	0.905	0.015	0.301	0.815	0.689	0.948
Mean	-	10.9	-	0.640	0.738	0.111	0.409	0.552	0.385	0.731
Combined	-	-	-	-	-	1.843E-23	4.823E-09	>99.9999	99.99841	>99.99999

*CIRADteak markers from Verhaegen et al. (2005); IFT markers from Yasodha et al. (2018) and TG markers from Mahesh et al. (2016)

S4), and an identity analysis was carried out allowing for genotype mismatches among the multilocus profiles. As expected, pairs of individuals that clustered in the dendrogram showed shared multilocus profiles for 16 or more of the 21 markers typed, with only up to 5 mismatches usually involving single allele differences. The additional mismatches involved four cases of differences in relative allele mobility estimates, explicable by a conserved size shift in allele size calls, three instances of a likely artefact allele 258 bp in the marker CIRAD1TeakF05, and three cases of inconsistent allele calls at marker TG9 possibly due to genotyping errors. Considering the extremely low probability (usually in the range of 1E-15) of a random match for the perfectly matching genotypes at 16 or more of the 21 markers, and assuming that the isolated mismatches are in fact genotyping errors, the following genetic identities were proposed: clones B3, P13 and B22; clones B19, P2 and B21; clones B11 and B23; clones B15 and P12.

Significant genotypic linkage disequilibrium (LD) was detected for 44 of the 210 pairwise microsatellite marker tests (Bonferroni corrected P-value≤ 0.00008) (Supporting information file S5). We compared the number of significant pairwise marker LD for individuals with relatedness coefficients $\theta_{xy} <$ 0.25 (Case A), with those for individuals that had higher relatedness $\theta_{xy} >$ 0.25 (Case B). For case A only 2 marker pairs had a significant LD while 14 marker pairs for case B, suggesting that a considerable proportion of the pairwise marker LD is likely due to the higher relatedness among particular clones.

Discussion

The sampled teak clones displayed high levels of genetic diversity. When compared with other genetic surveys in teak populations. The average H_o (0.64) for the 50 clones sampled was in the same range as the estimates reported earlier with 265 samples from India, Thailand, Laos, Indonesia, and Africa using 15 microsatellites (H_o = 0.583) (Verhaegen et al. 2005), with 721 individuals from 29 provenances from India, Laos, Myanmar, and Thailand using 6 microsatellites (H_o = 0.61) (Hansen et al. 2014), as well as for a small sample of 26 teak clones (H_o = 0.66) cultivated in China (Huang et al. 2016). Our estimate was

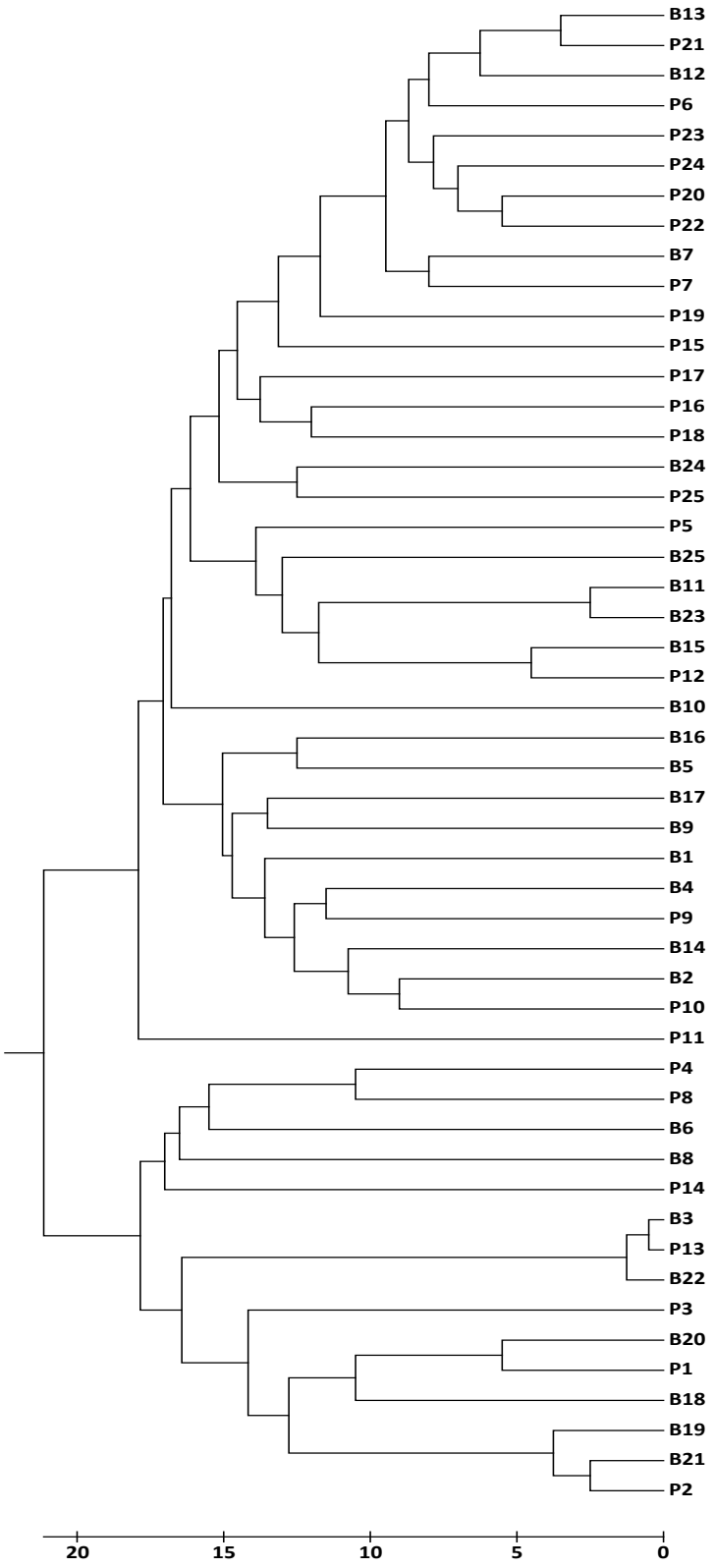


Figure 1
UPGMA dendrogram built from a matrix of genetic distances among the 50 *Tectona grandis* clones

significantly higher than the estimate reported for a sample of 166 individuals from India, Thailand, and Laos using 15 microsatellites (H_o = 0.35) (Fofana et al. 2009); also higher than the estimate for a sample of 60 clones from Mato Grosso using 9 microsatellites (H_o = 0.276) (Alcantara and Veasey 2013) and for a sample of 118 individuals from India evaluated for 12 microsatellites (H_o = 0.23) (Mahesh et al. 2016).

Although such comparisons are merely descriptive because not all microsatellites were common to these studies, and different quality control criteria were used to generate the data, the level of genetic diversity found in this relatively limited sample of 50 selected or commercial clones is nonetheless on the high end, attributable to their wide provenance origin. More importantly, the fact that the level of genetic diversity found in this set of 50 clones is essentially equivalent to what was found in much larger samples of teak trees, indicates that the parameters of genetic information content derived from this study should be widely relevant to the species at large toward applications in plant variety protection.

The 21 microsatellites displayed the expected high levels of hypervariability as described in their initial publications (Table 2). As expected, markers with a larger number of alleles are more informative for individual identification and parentage testing, having lower PIs and higher PEs. A high power of individual discrimination is reached with the use of the 21 microsatellite markers, such that the probability of declaring two individuals as identical based on chance alone are slim even in the case of highly related full-sib individuals (Table 2).

We also estimated the ability of the marker set to help identify the correct paternity of a particular teak tree reaching a power of excluding an erroneously pointed parent with over 99.999 % probability. Although this is not usually the case in plant variety protection, the results show that the marker set can be very useful to track pedigrees and help define essentially derived clonal varieties if that turns out to be the case (Jamali et al. 2019).

From theoretical expectations, the relatedness between pairs of half-sibs, full-sibs, and clones are 0.125, 0.25, and 0.5, respectively (Falconer 1989) although actual estimates obtained from molecular markers do carry variable levels of bias (Ackerman et al. 2017). Most (92.1 %) pairwise comparisons among the 50 clones showed no close relationship. However, 43 of the 50 clones were involved in the 77 putative pairwise estimates of half-sibs, 11 estimates of full-sibs and the nine suspect clonal identities (Supporting information file S3). These are relatively high numbers of related pairs of individuals, showing that despite the high genetic diversity detected, this group of clones includes a relevant proportion of trees with recent shared ancestry, suggesting that the final selection of elite clones in the breeding programs studied has been frequently practiced within the same families. This can also be visualized in the UPGMA dendrogram (Figure 1), where three major clusters of clones and several subclusters were formed with small genetic distances, despite the considerable number of microsatellites genotyped.

Significant genotypic linkage disequilibrium was detected for 20.6 % of the pairwise microsatellite marker tests. Disequilibrium is typically due to close physical linkage of markers on the same chromosome, although it can also be observed between unlinked markers (Kalinowski et al. 2007). Estimates of disequilibrium are based on allele frequencies, which are affected by several other forces including inbreeding, drift, artificial and natural selection, and population structure (Kalinowski et al. 2007; Sandler et al. 2021). Such processes can result in significant associations of alleles at different loci, even if in the absence of physical linkage.

We compared the number of significant pairwise marker disequilibria for individuals with relatedness coefficients θ_{xy} < 0.25 (Case A), with those for individuals that had higher relatedness θ_{xy} > 0.25 (Case B). For case A only 2 marker pairs had a significant disequilibrium while 14 marker pairs for case B, suggesting that a considerable proportion of the pairwise marker disequilibrium is likely due to the higher relatedness among particular clones. More importantly, however, these results indicate that lack of independent segregation will not be an issue when subsets of these 21 microsatellites are used to provide combined estimates of probability of identity in plant variety identification. Moreover, to evaluate any potential closer physical linkage among these markers, their chromosome position could be tentatively predicted based on the draft genome for *T. grandis* (Yasodha et al. 2018) although no chromosome scale assembly is yet available.

DNA profiling of microsatellite markers for individual identification in plant varietal protection requires careful data verification to minimize erroneous inferences about identity. Marker data quality may suffer allele stuttering caused by slippage of the DNA polymerase, null alleles due to allelic dropouts or mutations in the priming sites, and discrepancies in allele length due to size shifts. While false matched identities due to chance alone are extremely unlikely when many hypervariable markers are genotyped, false identity exclusions, i.e., when two cultivars are deemed different when, in fact, they are the same cultivar due to genotyping inconsistencies, could be a problem (Baldoni et al. 2009).

The naive sometimes-adopted practice of rounding allele data to the nearest whole number, frequently results in miscalls, underestimations of allelic richness and variable impact on the downstream analyses (Bonin et al. 2004). The built-in binning method of the standard GeneMapper software available in the automatic sequencer requires reference data that is usually not available for non-model organisms such as the current case of teak. In any case, fully eliminating genotyping errors cannot be achieved, mainly because molecular assays and sample handling are not 100 % reliable. In our study, to minimize such occurrences and provide the best possible assessment of the microsatellite data profiles, the GeneMapper exported genotypic data were adjusted to integer allele sizes using a statistical based method, and sample identity comparisons were based on the entire multilocus profile. Given the high information content of the markers, isolated mismatches in a few markers, considering an otherwise fully matching multilocus profile with very low probability of identity, strongly advocates for real identities as noted above.

Besides considerations regarding genetic information content, inclusion of a particular marker in a routine genotyping system for plant variety protection of teak, should also take into account the reliability of allele declaration as equally important. Dinucleotide repeat markers are more susceptible to incorrect allele interpretation due to stuttering, a PCR artefact of one repeat unit shorter in length than the parent allele, whose occurrence varies with the sequence structure and may complicate correct genotyping (Litt et al. 1993). Despite having only average information content, markers IFT14, IGT168 and IFT777, based on longer tri, tetra and pentanucleotide repeats respectively, are particularly useful in this respect, because their larger size difference between alleles allows for a more precise genotype declaration and less susceptibility to the impact of stuttering artefacts (Walsh et al. 1996). These three markers should be prioritized in routine genotyping systems for teak clone identification.

Inspecting the genotype data of the individuals that clustered in the genetic distance analysis (Supporting information file S4), 12 of the 22 total mismatches involved putative allele drop-outs likely due to single nucleotide variants in the primer annealing sites, or low DNA concentration, such that one allele may have been preferentially amplified over the other (Shestak et al. 2021). The additional mismatches involved mostly cases of relative mobility differences in the allele size. Mobility differences in microsatellite alleles between different electrophoretic runs should not be surprising and it is typically restricted to specific microsatellites that show higher susceptibility to such occurrences (see below). Across laboratories it becomes more frequent when differences in reagents, electrophoretic separation equipment and internal ladders, may affect the relative mobility estimates (This et al. 2004). Only with calibration and standardization of allelic designations together with a set of control samples analyzed across laboratories, such genotype inconsistencies can potentially be resolved.

From the standpoint of marker performance, it is also noteworthy to point out that the observed mismatches were more frequent in specific marker loci, providing further indirect evidence in support of their artefactual nature. Markers TG9, CIRAD1TeakF05 and CIRAD1TeakH10 were involved respectively in seven, three and three, i.e., 13 of the 22 total mismatches observed. Marker TG9, despite its very high information content, should probably be excluded from routine use, while the other two markers should be used with greater caution in varietal identity investigation, aware of a potentially higher susceptibility to allele dropouts. Alternatively, these markers could still be used provided that methods are implemented to estimate error rates due to allelic drop out and false alleles (Johnson and Haydon 2007).

Based on the premise that the genotype mismatches observed at the few markers are due to genotyping errors and considering the very low random match probability for the perfectly matching multilocus profiles at 16 or more markers, the genetic identities detected suggest that 10 allegedly distinct clones actually correspond to only four unique ones. Additionally, clones B13 and P21 despite having a θ_{xy} = 0.309 also showed only six mismatches, four of them as dropouts, probably also corresponding to a single clone. We are aware, however, that a limitation of our study was the lack of opportunity to further investigate these suspect identities with additional genotyping following the data analyses presented here.

In actual cases of DNA profiling for plant variety protection, it should be imperative to generate replicate profiles to assess the potential susceptibility of particular markers to genotyping inconsistencies and report exclusively fully replicated profiles for practical use as complementary molecular descriptors. Nevertheless, these results raise concerns regarding the fact that the putatively genetically indistinguishable clones at the DNA marker level found in this study, might be erroneously considered distinct based on DUS testing with morphological descriptors and therefore mistakenly subject to varietal protection. This result further underpins the importance of including DNA markers, at least as discretionary complementary descriptors, for plant protection of teak cultivars in the same manner that has been recommended for a number of crops, fruit and forest trees (Jamali et al. 2019).

Based on all the technical assessments of the 21 microsatellites in this particular set of 50 teak clones, one may wish to select and extract smaller subsets of markers to be used in routine genotyping, allowing for possibly somewhat reduced genotyping cost and optimized multiplexing. Such an exercise has been productive for other crops to provide a minimum consensus set of markers for interlaboratory comparisons and data sharing (Baldoni et al. 2009; Patocchi et al. 2009). As pointed out above, markers TG9, CIRAD1TeakF05 and CIRAD1TeakH10 could in principle be left aside due to their apparently higher susceptibility to allele drop out. Additionally, marker CIRAD3TeakDa09 could be disregarded due to its lowest power of discrimination.

The remaining 17 markers were ranked from the highest to the lowest information content, and all the cumulative combined estimates of probabilities of identity and paternity exclusion described previously were obtained as each additional marker was added to the previous set (Table 3). With a small set of the top eight informative markers, cumulative combined PI= 1.42E-11, PI_{sib}= 1.99E-04 and PE_{kp}= 99.99 % are reached. This would probably be the minimum adequate set for routine analysis, although it would require testing additional markers when a single isolated locus mismatch or suspect drop out allele is detected. Ideally a minimum set of the top 12 markers could be adopted. Such a set allows reaching two orders of magnitude higher PI_{sib}= 6.59E-06, corresponding to a discrimination power of 1 in 151,641, and a probability of paternity exclusion of 99.9994 %, while providing some initial protection against false identity exclusions due to occasional genotype mismatches.

Additionally, this set would include microsatellites IFT14 and IFT77 with longer repeats for improved precision of genotype declaration. Evidently that the additional five microsatellites could equally be included for a 17-loci routine genotyping systems if desired, further increasing the power of discrimination. Additionally, in case a laboratory already adopts exclusively the most traditionally and widely used CIRADTeak markers (Verhaegen et al. 2005), 12 of them tested in this work provide

Table 3
Summary of the cumulative combined parameters of individual discrimination and parentage testing for 17 selected microsatellite markers ranked from the most to the least informative. Combined probability of identity of unrelated individuals (PI), of sibs (PI_{sib}), probability of paternity exclusion with a known parent (PE_{kp}), single parent testing (PE_{sp}) and parent pair testing (PE_{pp})

Microsatellite marker	Cumulative combined PI	Cumulative combined PI expressed as 1 in	Cumulative combined PIsib	Cumulative combined PIsib expressed as 1 in	Cumulative combined PEkp (%)	Cumulative combined PEsp (%)	Cumulative combined PEpp (%)
TG12	2.32E-02	4.31E+01	3.12E-01	3.20E+00	77.2936	62.8669	91.9794
TG11	7.18E-04	1.39E+03	1.01E-01	9.92E+00	94.0211	84.4286	99.1579
CIRAD3TeakB02	2.26E-05	4.43E+04	3.27E-02	3.06E+01	98.4173	93.4622	99.9125
CIRAD3TeakF01	8.77E-07	1.14E+06	1.11E-02	9.04E+01	99.5298	96.9849	99.9893
CIRAD1TeakA06	3.49E-08	2.86E+07	3.71E-03	2.70E+02	99.8597	98.6071	99.9986
CIRAD1TeakB03	2.31E-09	4.33E+08	1.35E-03	7.41E+02	99.9460	99.2181	99.9997
TG16	1.76E-10	5.67E+09	5.22E-04	1.91E+03	99.9778	99.5341	99.9999
CIRAD2TeakB07	1.42E-11	7.05E+10	1.99E-04	5.03E+03	99.9906	99.7200	> 99.9999
CIRAD2TeakC03	1.32E-12	7.60E+11	7.94E-05	1.26E+04	99.9958	99.8226	> 99.9999
CIRAD3TeakA11	1.53E-13	6.52E+12	3.42E-05	2.92E+04	99.9979	99.8785	> 99.9999
IFT14	2.00E-14	5.01E+13	1.46E-05	6.86E+04	99.9989	99.9152	> 99.9999
IFT777	3.15E-15	3.17E+14	6.59E-06	1.52E+05	99.9994	99.9367	> 99.9999
IFT168	5.09E-16	1.96E+15	3.08E-06	3.25E+05	99.9996	99.9524	> 99.9999
CIRAD3TeakE06	8.55E-17	1.17E+16	1.49E-06	6.70E+05	99.9998	99.9637	> 99.9999
CIRAD4TeakF02	1.53E-17	6.52E+16	7.19E-07	1.39E+06	99.9999	99.9720	> 99.9999
CIRAD4TeakH09	2.95E-18	3.39E+17	3.48E-07	2.87E+06	99.9999	99.9782	> 99.9999
CIRAD4TeakDa12	6.28E-19	1.59E+18	1.72E-07	5.81E+06	> 99.9999	99.9827	> 99.9999

a fully adequate discrimination power with combined PI= 1.52E-12, PI_{sib}= 2.47E-5 and PE= 99.9916 %. From the practical standpoint, however, it is evident that several different combinations of the markers evaluated in this study are possible. The decision on a final set will depend on additional parameters that were not evaluated in this study, such as the compatibility of primers sequences and allele size ranges for efficient PCR co-amplification and multiplexed electrophoretic separation. Furthermore, to arrive to a final recommended consensus set of markers, it would be highly desirable to organize inter-laboratory comparison ring trials.

Conclusions

As a case study for the use of DNA markers for plant variety protection of forest trees, we have provided a detailed evaluation of a panel of 21 microsatellite markers for individual identification of teak using a set of 50 elite clones bred and planted by two separate forest tree companies. Given the very high discrimination power of the DNA markers, both for identity and parentage testing, we argue that they would add extremely useful information to the established DUS testing for a more robust protection of intellectual property of teak clonal varieties. Variable subsets of the 21 markers should be further evaluated in interlaboratory ring trials to be used for routine analysis, according to objective performance criteria and optimized protocols. Nevertheless, a recommended set should probably try to include a minimum of 12 markers with higher information content and less susceptibility to allele drop-outs and allele mobility shifts, to minimize the possibility of confounding artefactual single allele mismatches in otherwise identical multilocus profiles.

The genetic analysis of the 50 allegedly unique clones revealed some unexpected identities whose probability of occurring randomly was estimated as extremely unlikely, therefore strongly supporting the hypothesis that the same clones have been planted or might eventually be protected under different names, both within as well as between the two companies surveyed. Furthermore, the microsatellite data have shown that despite the existence of high levels of genetic variation in the sampled gene pool, this group of clones includes a relevant proportion of trees with significant shared ancestry, suggesting that the final selection of elite clones in the breeding programs has occasionally been practiced within the same families. Notwithstanding the relevance of this result from the perspective of long-term risk of low diversity plantations, these results advocate that the exclusive use of the 22 recommended morphological descriptors for adequate DUS testing of teak clones might face significant challenges in the near future. This result strengthens the importance of including DNA markers, at least as discretionary complementary descriptors, for plant protection of clonal teak cultivars in Brazil

and possibly worldwide. Finally, the data gathered in this study could provide a first step toward the recommendation of a minimum common set of markers to be used across laboratories for intercomparison purposes.

Acknowledgments

The authors thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for a scholarship to Mariana M. Queiroz and BIOTECA, PROTECA Biotecnologia Florestal Ltda and Ministério da Agricultura, Pecuária e Abastecimento (MAPA) for their collaboration in the development of the research. A.M.S. and D.G acknowledge the research productivity fellowships granted by CNPq (Brazilian National Council for Scientific and Technological Development). This work was partially supoted by FAP-DF (Foundation for Scientific Research of the Federal District) competitive grant RECGENOMICS 00193-00000924/2021-92 to D.G.

References

Ackerman MS, Johri P, Spitze K, Xu S, Doak TG, Young K, Lynch M (2017) Estimating Seven Coefficients of Pairwise Relatedness Using Population-Genomic Data. Genetics 206:105-118. https://doi.org/10.1534/genetics.116.190660

Alcantara BK, Veasey EA (2013) Genetic diversity of teak (*Tectona grandis* LF) from different provenances using microsatellite markers. Revista Arvore 37:747-758. https://doi.org/10.1590/s0100-67622013000400018

Ashley MV (2010) Plant Parentage, Pollination, and Dispersal: How DNA Microsatellites Have Altered the Landscape. Critical Reviews in Plant Sciences 29:148-161. https://doi.org/10.1080/07352689.2010.481167

Baldoni L, Cultrera NG, Mariotti R, Ricciolini C, Arcioni S, Vendramin GG, Buonamici A, Porceddu A, Sarri V, Ojeda MA, Trujillo I, Rallo L, Belaj A, Perri E, Salimonti A, Muzzalupo I, Casagrande A, Lain O, Messina R, Testolin R (2009) A consensus list of microsatellite markers for olive genotyping. Molecular Breeding 24:213-231. https://doi.org/10.1007/s11032-009-9285-8

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13:3261-3273. https://doi.org/10.1111/j.1365-294x.2004.02346.x

da Conceicao FX, Drescher R, Pelissari AL, Lanssanova LR, Favalessa CMC, Roquette JG (2012) Local productive capacity of Tectona grandis in Monte Dourado, Para State, Brazil. Ciencia Rural 42:822-827

Falconer DS (1989) Introduction to quantitative genetics, Third edn. Longman Scientific and Technical. https://doi.org/10.1017/s0016672300028573

Fofana IJ, Ofori D, Poitel M, Verhaegen D (2009) Diversity and genetic structure of teak (Tectona grandis L.f) in its natural range using DNA microsatellite markers. New Forests 37:175-195. https://doi.org/10.1007/s11056-008-9116-5

Garrido-Cardenas JA, Mesa-Valle C, Manzano-Agugliaro F (2018) Trends in plant research using molecular markers. Planta 247:543-557. https://doi.org/10.1007/s00425-017-2829-y

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318:577-579. https://doi.org/10.1038/318577a0

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. Journal of Heredity 86:485-486. https://doi.org/10.1093/oxfordjournals.jhered.a111627

Grattapaglia D, Pimenta D, Campinhos EN, Rezende GDS, de Assis TF (2003) Marcadores moleculares na proteção varietal de Eucalyptus. Proceedings of the 8º Brazilian Forestry Congress. Published in CD., pp 1-13

Hansen OK, Changtragoon S, Ponoy B, Kjær ED, Minn Y, Finkeldey R, Nielsen KB, Graudal L (2014) Genetic resources of teak (Tectona grandis Linn. f.)—strong genetic structure among natural populations. Tree Genetics & Genomes 11:802. https://doi.org/10.1007/s11295-014-0802-5

Huang GH, Liang KN, Zhou ZZ, Ma HM (2016) SSR genotyping-genetic diversity and fingerprinting of teak (*Tectona grandis*) clones. Journal of Tropical Forest Science 28:48-58

IBÁ (2022) Relatório Anual da IBÁ - Industria Brasileira de Árvores. IBÁ - Industria Brasileira de Árvores, São Paulo, p 96 pp

Inglis PW, Pappas MdCR, Resende LV, Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLOS ONE 13:e0206085. https://doi.org/10.1371/journal.pone.0206085

Jamali SH, Cockram J, Hickey LT (2019) Insights into deployment of DNA markers in plant variety protection and registration. Theoretical and Applied Genetics 132:1911-1929. https://doi.org/10.1007/s00122-019-03348-7

Johnson PCD, Haydon DT (2007) Maximum-Likelihood Estimation of Allelic Dropout and False Allele Error Rates From Microsatellite Genotypes in the Absence of Reference Data. Genetics 175:827-842. https://doi.org/10.1534/genetics.106.064618

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

Khaing N, Oo MM, Tun TN, Lwin O, Minn Y, Kleine M (2017) Natural Teak Forests – Silviculture and Stand Management. In: Kollert W, Kleine M (eds) The global teak study Analysis, evaluation and future potential of teak resources. International Union of Forest Research Organizations (IUFRO). World Series Vienna, p 108

Kirst M, Cordeiro CM, Rezende GDSP, Grattapaglia D (2005) Power of Microsatellite Markers for Fingerprinting and Parentage Analysis in Eucalyptus grandis Breeding Populations. Journal of Heredity 96:161-166. https://doi.org/10.1093/jhered/esi023

Kollert W, Kleine M (2017) The global teak study. Analysis, evaluation and future potential of teak resources, Vienna

Litt M, Hauge X, Sharma V (1993) Shadow bands seen when typing polymorphic dinucleotide repeats - some causes and cures. Biotechniques 15:280-&

Madi JPS, Calegário N, Môra R, Carvalho MPdLCe, Carvalho SdPCe (2020) Density management in clonal Tectona grandis Linn. f. plantations. Scientia Forestalis 48:e3296. https://doi.org/10.18671/scifor.v48n125.25

Mahesh S, Vaishnav V, Kumar P, Mohammad N, Ansari SA (2016) Characterization and validation of teak plus trees ramets of national teak germplasm bank, Chandrapur, Maharashtra through microsatellites. Tropical Plant Research 3:213-220

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25:1982-1983. https://doi.org/10.1093/bioinformatics/btp303

Murillo O, Wright O, Monteuuis O, Montenegro F (2013) Mejoramiento genético de la teca en América Latina. In: de Camino R, Morales JP (eds) Las plantaciones de teca en América Latina: Mitos y realidades. CATIE, Turrialba, Costa Rica, pp 86-111

Patocchi A, Fernández-Fernández F, Evans K, Gobbin D, Rezzonico F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Mathis-Jeanneteau F, Durel CE, Gianfranceschi L, Costa F, Toller C, Cova V, Mott D, Komjanc M, Barbaro E, Kodde L, Rikkerink E, Gessler C, van de Weg WE (2009) Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. Tree Genetics & Genomes 5:211-223. https://doi.org/10.1007/s11295-008-0176-7

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Pelissari AL, Caldeira SF, R. D (2013) Quantitative and qualitative development of Tectona Grandis L.f. in Mato Grosso. Floresta e Ambiente 20:371-383

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1:215-222. https://doi.org/10.1016/s1360-1385(96)86898-0

Prasetyo E, Widiyatno, Indrioko S, Na'iem M, Matsui T, Matsuo A, Suyama Y, Tsumura Y (2020) Genetic diversity and the origin of commercial plantation of Indonesian teak on Java Island. Tree Genetics & Genomes 16:34. https://doi.org/10.1007/s11295-020-1427-5

Queiroz MD, Caldeira SF, Sebbenn AM, Arriel DAA (2021) Genetic diversity of improved genotypes of Tectona grandis in the state of Mato Grosso, Brazil. Southern Forests-a Journal of Forest Science 83:120-127. https://doi.org/10.2989/20702620.2021.1877094

Reategui-Betancourt JL, Arriel DAA, Caldeira SF, Higa AR, Flores PC, Araujo SP, Marques RM, Correa BMB, Martinez DT (2020) Morphological descriptors for the characterization of teak clones (Tectona grandis L.f.) in plantations. Forest Systems 29:eRC02. https://doi.org/10.5424/fs/2020292-15634

Sandler G, Wright SI, Agrawal AF (2021) Patterns and Causes of Signed Linkage Disequilibria in Flies and Plants. Molecular Biology and Evolution 38:4310-4321. https://doi.org/10.1093/molbev/msab169

Shestak AG, Bukaeva AA, Saber S, Zaklyazminskaya EV (2021) Allelic dropout is a common phenomenon that reduces the diagnostic yield of pcr-based sequencing of targeted gene panels. Frontiers in Genetics 12. https://doi.org/10.3389/fgene.2021.620337

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82:561-573. https://doi.org/10.1038/sj.hdy.6885180

Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Molecular Biology and Evolution 37:1237-1239. https://doi.org/10.1093/molbev/msz312

Telfer EJ, Stovold GT, Li Y, Silva-Junior OB, Grattapaglia DG, Dungey HS (2015) Parentage Reconstruction in Eucalyptus nitens Using SNPs and Microsatellite Markers: A Comparative Analysis of Marker Data Power and Robustness. PLOS ONE 10:e0130601. https://doi.org/10.1371/journal.pone.0130601

This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibáñez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani N, Peterlunger E, Regner F, Zulini L, Maul E (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theoretical and Applied Genetics 109:1448-1458. https://doi.org/10.1007/s00122-004-1760-3

Tsukamoto AAF, Silva ML, Couto L, Müller MD (2003) Economic analysis of a teak plantation submitted to thinning. Revista Árvore 27:487-494

UPOV (2020) Guidance on the Use of Biochemical and Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS) (Revision). International Union for the Protection of New Varieties of Plants, Genève, Switzerland.

Vendruscolo DGS, Drescher R, Carvalho S, Medeiros RA, Mora R, Soares AAV (2019) Dominant height growth in *Tectona grandis* plantations in Mato Grosso, Brazil. Floresta E Ambiente 26:e20170703. https://doi.org/10.1590/2179-8087.070317

Verhaegen D, Ofori D, Fofana I, Poitel M, Vaillant A (2005) Development and characterization of microsatellite markers in Tectona grandis (Linn. f). Molecular Ecology Notes 5:945-947. https://doi.org/10.1111/j.1471-8286.2005.01124.x

Walsh PS, Fildes NJ, Reynolds R (1996) Sequence Analysis and Characterization of Stutter Products at the Tetranucleotide Repeat Locus VWA. Nucleic Acids Research 24:2807-2812. https://doi.org/10.1093/nar/24.14.2807

Yang Y, Tian H, Wang R, Wang L, Yi H, Liu Y, Xu L, Fan Y, Zhao J, Wang F (2021) Variety Discrimination Power: An Appraisal Index for Loci Combination Screening Applied to Plant Variety Discrimination. Frontiers in Plant Science 12:566796. https://doi.org/10.3389/fpls.2021.566796

Yasodha R, Vasudeva R, Balakrishnan S, Sakthi AR, Abel N, Binai N, Rajashekar B, Bachpai VKW, Pillai C, Dev SA (2018) Draft genome of a high value tropical timber tree, Teak (Tectona grandis L. f): insights into SSR diversity, phylogeny and conservation. DNA Research 25:409-419. https://doi.org/10.1093/dnares/dsy013