Physiol Mol Biol Plants (October 2020) 26(10):2095–2101
https://doi.org/10.1007/s12298-020-00876-1

SHORT COMMUNICATION

Development and characterization of microsatellite markers in *Stryphnodendron adstringens* (Leguminosae)

Ariany Rosa Gonçalves[1,2] · Luciana Oliveira Barateli[1,2] · Ueric José Borges de Souza[1,2] · Ana Maria Soares Pereira[3] · Bianca Waléria Bertoni[3] · Mariana Pires de Campos Telles[2,4]

Received: 4 March 2020 / Revised: 3 August 2020 / Accepted: 26 August 2020 / Published online: 16 September 2020

Abstract In this study, we report the development and characterization of 15 new microsatellite markers for *Stryphnodendron adstringens* (Leguminosae) in order to support future analyses of genetic diversity in populations of this species. In screening with 48 individuals from three different populations of *S. adstringens*, we tested the amplification of 20 microsatellite loci, of which five are not useful for population genetic studies due to the lack of polymorphisms or amplification failures. For the final set of 15 loci, the number of alleles ranged from 2 to 15, with a total of 116 alleles. The expected heterozygosity ranged from 0.1219 to 0.8965, with an average of 0.6694 per locus. The combined probability of genetic identity ($PI = 8.12 \times 10^{-15}$) and paternity exclusion ($Q = 0.99999$) estimations showed that the loci may be useful to discriminate between individuals of *S. adstringens*. Initial cross-amplification tests were satisfactory in three species of the genus *Stryphnodendron*: *S. rotundifolium*, *S. coriaceum* and *S. polyphyllum*. This new set of markers will be a useful tool for population genetic studies, contributing to the knowledge about the evolutionary history of *S. adstringens* and, additionally, correlated species.

Keywords Cerrado · Genetic diversity · Next Generation Sequencing · Medicinal plant · SSR markers

Introduction

The genus *Stryphnodendron* Mart. (Leguminosae family) has actually 25 recognized species, of which 21 occur in Brazil (Souza and Lima 2020). Among several plant species native to Brazil, *Stryphnodendron adstringens* (Mart.) Coville stands out because of its medicinal relevance, whose antioxidant, antimicrobial and antitumor properties have already been highlighted in several clinical studies (see Souza-Moreira et al. 2018; Pellenz et al. 2019). This species, popularly known as barbatimão, is widely distributed in Brazil, especially in Cerrado and Caatinga phytogeographic domains, with the highest population densities in Cerrado *lato* sensu and "campo rupestre" phytophysiognomies (Souza and Lima 2020), preferably in open areas or that have been affected by fire (Diniz and Franceschinelli 2015).

Previous population-based genetic studies using isozymes and Amplified Fragment Length Polymorphism (AFLP) markers showed moderate levels of genetic diversity (Glasenapp et al. 2014) and high differentiation (Mendonça et al. 2012) among populations of *S. adstringens*. However, despite its high potential in the herbal medicine trade, the only microsatellite markers available for this species were transferred from other species, such as

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12298-020-00876-1) contains supplementary material, which is available to authorized users.

✉ Mariana Pires de Campos Telles
tellesmpc@gmail.com

[1] Programa de Pós-graduação em Genética e Melhoramento de Plantas, Escola de Agronomia, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[2] Laboratório de Genética e Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[3] Unidade de Biotecnologia, Universidade de Ribeirão Preto, Ribeirão Preto, São Paulo, Brazil

[4] Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás, Goiânia, Goiás, Brazil

Eucalyptus spp., *Melaleuca alternifolia*, *Prunus avium*, *Schizolobium parahybum* and *Annona cherimola* (Branco et al. 2010). It is worth mentioning that the efficiency of transferring microsatellite markers between species or between genera is highly variable between taxa, especially when there are important differences in the complexity of the genome between the species for which the marker was developed and the target (Barbará et al. 2007). Thus, the development of microsatellite markers for *S. adstringens* will clarify the evolutionary mechanisms in the distribution of genetic variability in this species and will generate useful information for conservation strategies and sustainable management planning.

Advances in DNA sequencing technologies have made the cost of developing species-specific markers more affordable (Unamba et al. 2015). The identification of microsatellite regions by the large-scale sequencing approach (*Next Generation Sequencing - NGS*) has advantages over the traditional DNA enriched library. Among these advantages, the NGS approach does not require cloning before sequencing, provides higher coverage of the genome with greater data volume and enables rapid and cost-effective discovery of microsatellite loci (Zalapa et al. 2012; Taheri et al. 2018). Here, we report the development and characterization of microsatellite loci for *S. adstringens* using NGS technology, and tested the cross-amplification of 20 loci in three related species of the genus *Stryphnodendron*.

Materials and methods

Sample collection and DNA extraction

For genome sequencing, we used dehydrated leaf tissue samples from two individuals of *S. adstringens* collected in two different localities: Niquelândia (Goiás state) and Candeias (Minas Gerais) (see Supplementary Material, Table S1). For amplification and genotyping tests, we collected leaf tissue samples from 48 individuals distributed in three local populations of the Brazilian Cerrado: Candeias (Minas Gerais), Posse (Goiás), and Chapada dos Guimarães (Mato Grosso) (see Supplementary Material, Table S1). The collections were carried out between the years 2013 and 2016, under SisGen registration A64E081 and A4EE2BE. Total genomic DNA was isolated from silica-dried leaves using the CTAB protocol (Doyle and Doyle 1987), with adaptations to *S. adstringens.* We quantify the DNA using a NanoDrop spectrophotometer (Thermo Scientific) and evaluated the DNA's quality using 1% agarose gel. In addition, we quantify the DNA through fluorometry using Qubit 2.0 (Life Technologies).

Library preparation, primer design and PCR amplification

For the preparation of the genomic library, we used 50 ng of DNA from each sample, following the protocol of the Nextera DNA Sample Preparation Kit (Illumina, San Diego, USA). The resulted libraries were validated through the Bioanalyzer equipment (Agilent Technologies Inc., Santa Clara, CA, USA) and quantified by real time PCR with the Universal Library Quantification Kit (KAPA Biosystems, Wilmington, MA, USA). The library was sequenced on a single lane in paired-end mode (2 × 100 bp) using the HiSeq 2500 platform and V4 SBS kit (Illumina) at the University of São Paulo (Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo) in Piracicaba, Brazil.

We assessed the quality of reads using FastQC v0.11.5 (Andrews 2015) and process raw reads with Trimmomatic V.0.36, to remove low-quality reads and and Illumina adapter sequences (Bolger et al. 2014). The filtered reads were assembled de novo into contigs using MaSuRCA (Zimin et al. 2013). We used the Redundans pipeline to filter redundant streams due to heterozygosity (Pryszcz and Gabaldón 2016).

We used the bioinformatics pipeline QDD version 3.1 (Meglécz et al. 2014) to identify contigs possessing microsatellite motifs as well as to design primer pairs with the following parameters: amplicon size between 150 and 400 base pairs (bp), primer length between 20 and 25 bp, melting temperature between 56 and 62 °C, and CG content between 30 and 60%. We identified 20,124 microsatellite regions, from which it was possible to design 2053 primer pairs, for 140 different loci. As there are many sequences containing microsatellites, we applied some filters to choose primers from the primer table, such as: (1) the choice of pure microsatellite rather than compound (2) the avoidance of regions composed only by adenine and thymine bases, that can form hairpin, e.g. (AT)n; (3) the option of primers with a size between 20 bp and 24 bp with the annealing temperature between the primers up to 1 °C; (4) PCR product size between 150 and 360 bp; and (5) the avoidance of primers that are very close to the target microsatellite (> 20 bp). Of these set, we selected 20 primer pairs and used them for PCR amplification tests and polymorphism verification.

The amplification stage was performed in a final volume of 10 µl reaction volumes containing 2.5 ng of DNA template, 0.14 mM of both primers (forward + reverse), 0.23 µM of dNTP, 3.25 mg of bovine serum albumin (25 mg/ml), and 1 × reaction buffer (10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 0.75 U of *Taq* DNA polymerase (5U-Phoneutria®, Brazil). The PCR protocol comprised the following conditions: an initial denaturation

at 94 °C for 1 min; 30 cycles of denaturing for 1 min at 94 °C, 1 min at annealing temperature ranging from 52–58 °C (Table 1) and extension for 1 min at 72 °C; and one cycle for final extension at 72 °C for 45 min. We analyzed the amplicons on a 3% agarose gel and detected the polymorphisms by 6% denaturing polyacrylamide gel

Table 1 Description of 20 microsatellite loci and the primer pairs developed for *Stryphnodendron adstringens*

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)[a]	T_a (°C)	GenBank accession no.
SadH1	F: AGTGCCTCTACTGATTTATGACCC R: CCAGGCGATTCAACCATATC	$(GT)_5$	161–171	56	MN525570
SadH2	F: AAGCCGTAGCTTGGAAGTGA R: TGTCCATAGATCCAAACGAGG	$(AG)_6$	320[b]	56	MN525571
SadH3	F: TCGGCTACTTCTTCTGCATCT R: CAGGTTGTGGTTGGCCTATT	$(CA)_5$	248–252	56	MN525572
SadH4	F: TCAGAGAGGTGGTTTGAGCA R: CACCAAATCAATCATGTTCCA	$(GT)_6$	350[b]	56	MN525573
SadH5	F: GTGGTCTCTCTGCTGCCTTC R: ACAGGGATGAGCCAGAGCTA	$(CT)_{14}$	274–304	54	MN525574
SadH6	F: ACGATATTAAGAAGGGACTAGCAG R: AGGTTTGAAGGCTCATCAGTT	$(AG)_5$	320[b]	56	MN525575
SadH7	F: TAAAGGGTCATTATATGTGGCAA R: TTTCTGTAACCCTTCGACCA	$(CA)_{14}$	154–168	54	MN525576
SadH8	F: TACAGCTTCAGCAACAACCC R: GTGTCGCTGGAGAATCACAT	$(AG)_{14}$	154–174	58	MN525577
SadH9	F: AGTGGAAGAAGAGCCCACAG R: CCTGGAAAGGTTGGAGAGTG	$(CT)_{11}$	198–246	58	MN525578
SadH10[c]	F: GAAGAAGCAGAGGGTTGTCAG R: GAATACATGGGCAAATGATGG	$(GA)_{14}$	247–287	56	MN525579
SadH11	F: TTAAGTCACGCCTCTTCGTC R: CTGTATAGTGAAGGCATGTTTCC	$(AG)_{15}$	312–326	56	MN525580
SadH12[c]	F: AACACCTCCCTAGTCCCTCC R: TCAGAATGTGCTTCTTTGCG	$(CT)_{16}$	138–162	52	MN525589
SadH13	F: CTTCCAGGTGCCTTGCTTAC R: TGCTCATCTGTTTCTTTGGTTC	$(GT)_{14}$	203–255	56	MN525581
SadH14	F: GAGACATCGTCCGAGGCTAA R: CTGACCCAAATCAGCACAGA	$(TC)_{14}$	245–267	58	MN525582
SadH15	F: TGAGTTGGGTGCTCTACCTT R: AAGAACGAAGAAATGGCAAA	$(ACGA)_7$	214–242	56	MN525583
SadH16	F: TGGAGGAGGGAGTATAGGTGA R: ACTAGGGACACTGACGAGGC	$(AACG)_8$	329–345	58	MN525584
SadH17	F: GTCTCGGATTTGATTTCGCT R: AATTTAGACAGCATTGTGGAGC	$(AAAG)_6$	222–268	52	MN525585
SadH18	F: ATGAGCTTGGATGGTTGATG R: TGGAAGGCTACGTGGAATTA	$(TTAT)_6$	260–268	58	MN525586
SadH19	F: GGCGTGGAGAAGACAAGTTC R: AGAGGAAACCGAACGTCAAA	$(CTATT)_5$	174–180	58	MN525587
SadH20	F: TTGTGTTTGGCTATGGAGAAGA R: TGTAGAGACAAGGTGTGGCG	$(CACCT)_5$	140–168	56	MN525588

T_a, Annealing temperature

[a]Fragment size range based on 48 individuals from three populations in Brazilian Savannah

[b]Monomorphic locus not used to analyze genetic diversity

[c]Locus removed from genetic diversity analysis, because it didn't have a good amplification pattern

stained with silver nitrate (Creste et al. 2001). The allele size was confirmed by comparison to 10 bp DNA ladder standard (Invitrogen®, USA). We performed the adjustment of the annealing temperature (°C) of the loci, until the visualization of the bands presented the best resolution. Additionally, we carry out initial testing of cross-amplification on three different close related species of the genus *Stryphnodendron*: four individuals of *S. rotundifolium* species, four of *S. coriaceum* and three of *S. polyphyllum* (see Supplementary Material Table S2, for more details). The PCR amplification followed the protocol described above, and in this step, the annealing temperature used was the same for all pairs of primers (52 °C).

Microsatellite loci characterization

The loci that showed polymorphism were used to estimate genetic diversity parameters and the genetic variability in the 48 *S. adstringens* individuals, distributed in three natural populations (For details on individual and population codes, see Supporting Information Table S1). For the final set of 15 loci, we evaluated the number of alleles per locus (*A*), expected heterozygosity (*He*) under the Hardy–Weinberg equilibrium (HWE) (Nei 1978). The linkage disequilibrium (LD) was verified for all pairs of loci. Analyses and randomization-based tests were performed with the softwares FSTAT 2.9.3.2 (Goudet 2002) and Genepop v. 4.5 (Rousset 2008). To assess the discrimination power of each locus and the set of loci, we calculated the probability of genetic identity (Paetkau et al.1995) and the exclusion of paternity (Weir 1996) in the software Identity v. 4.0 (Wagner and Sefc 1999).

The population polymorphism was characterized as the number of alleles per locus (*A*). We evaluated the genetic diversity of populations based on the expected heterozygosity index (*He*) under the HWE. We also estimated the observed heterozygosity (*Ho*) and the fixation index (F_{IS}), in order to detect possible deviations from HWE. Additionally, we tested whether populations are genetically differentiated using the Wright's F_{ST} measure (Wright 1951), obtained from an analysis of variance of allele frequencies (Weir and Cockerham 1984), implemented in the software FSTAT 2.9.3.2 (Goudet 2002).

To detect the degree of genetic information of these specific markers with the transferred markers, we compared some parameters of this loci set, such as the number of alleles (*A*), the expected heterozygosity (*He*) and the fixation index (F_{IS}), with the results obtained of transferred microsatellite (Branco et al. 2010), using a *t* test ($p \leq 0.05$). This analysis was performed using the R platform (R Core Team 2020).

Results and discussion

Sequencing of the *S. adstringens* genome produced a total of 554,143,303 reads. After trimming the low-quality bases and adapter sequences, a total of 511,837,144 were assembled into 63,320 contigs with the minimum contig length of 500 bp and an N50 of 14,541 bp. Primers were successfully designed in silico for 700 sequences containing repeats. These consisted of 537 dinucleotides, 89 trinucleotides, 45 tetranucleotides, 22 pentanucleotides, and 7 hexanucleotides microsatellites primers pairs.

All the 20 primer pairs tested successfully amplified the microsatellite loci, of these 17 revealed polymorphisms and three (SadH2, SadH4, and SadH6) were monomorphic (Table 1). The SadH10 locus, despite having a high number of alleles, did not present a good pattern in the genotyping stage, presenting duplicate bands. Thus, for population studies, it would not be feasible to use it and was therefore removed from genetic diversity analysis. The SadH12 locus has also been removed from the analysis, because it had a high frequency of non-amplified alleles, i.e. null alleles.

Table 2 Characterization of 15 microsatellite loci developed for *Stryphnodendron adstringens*, based on 48 individuals

Locus	*A*	*He*	*pHWE*	*PI*	*Q*
SadH1	3	0.3840	0.6932	0.4491	0.1622
SadH3	3	0.1219	0.1998	0.7784	0.0592
SadH5	14	0.8721	0.3164	0.0333	0.7274
SadH7	7	0.7474	0.0277	0.1070	0.5184
SadH8	8	0.8142	0.1442	0.0634	0.6231
SadH9	15	0.8026	0.8092	0.0649	0.6225
SadH11	8	0.7939	0.3604	0.0724	0.5972
SadH13	14	0.8507	0.1727	0.0417	0.6946
SadH14	11	0.8965	0.6925	0.0237	0.7704
SadH15	8	0.7596	0.0332	0.0874	0.5609
SadH16	6	0.7643	0.7490	0.0973	0.5361
SadH17	6	0.7530	0.2032	0.1075	0.5144
SadH18	4	0.4193	0.0005*	0.3992	0.1972
SadH19	2	0.3208	0.0317	0.5162	0.1335
SadH20	7	0.7410	0.7700	0.1175	0.4939
Overall	116	0.6694	0.0036	8.12×10^{-15}	0.99999

A, number of alleles; *He*, expected heterozygosity under Hardy–Weinberg equilibrium; *pHWE,* probability of deviation from Hardy–Weinberg equilibrium following Bonferroni correction (*p* value = 0.00111); *PI*, Probability of Identity; *Q*, Probability of paternity exclusion

*Significant value; *p* value (HWE) = 0.0011 adjusted by Bonferroni 's correction for a nominal value of 5%

For the set of 15 loci, alleles per locus ranged from 2 (SadH19) to 15 (SadH9), with a total of 116 alleles (Table 2). The global expected heterozygosity (*He*) was 0.6694, ranging from 0.1219 (SadH3) to 0.8965 (SadH14). No significant linkage disequilibrium ($p > 0.05$) was detected among loci pairs (data not shown). When overall populations were considered, only SadH18 locus showed significant deviations ($p < 0.0011$) from the Hardy–Weinberg equilibrium (Table 2). The combined probability of genetic identity ($PI = 8.12 \times 10^{-15}$) showed that the loci set is a useful tool to discriminate between individuals of *S. adstringens*. Additionally, the combined probability paternity exclusion ($Q = 0.99999$) also indicates that these markers will permit detailed parentage and genetic structure studies in natural populations (Table 2). These values also indicate that the battery of loci is useful and accurate to detect clonality, as there are reports of clonal regeneration arising by root sprouting described for *S. adstringens* and other species of genus *Stryphnodendron* (Rizzini and Heringer 1966).

All three populations presented large numbers of alleles, ranging from 54, in the CHGMT population, to 70 in CANMG (Table 3). CANMG population also presented the largest expected heterozygosity (*He* = 0.5435). Significant deviations from HWE based on Fisher's exact test ($p < 0.05$) in *S. adstringens* were detected for four loci in the CANMG population, two loci in the POSGO population, and no locus showed significant deviations in the CHGMT population (Table 3). Deviations from HWE may be due to factors such as clonality, population subdivision, or even the presence of null alleles.

The fixation index was high in the POSGO population ($F_{IS} = 0.1602$) indicating deficiency of heterozygotes ($p < 0.001$). Positive values in inbreeding coefficients can indicate that population have non-random mating. The global value of F_{ST} was 0.2989 ($p < 0.05$), which indicates a high level of genetic differentiation between populations, since almost 30% of the genetic variation is presented in the population component.

The overall loci averages such as number of alleles ($A = 6.100$) and expected heterozygosity ($He = 0.7139$) obtained by Branco et al. (2010) and the present study ($A = 7.733$; $He = 0.6694$) were similar and there was no significant difference for either mean ($p > 0.05$), while the fixation index found for the two sets of markers were significantly different ($p < 0.001$) (see Supplementary Material Table S3). It is worth mentioning that the loci set described here, besides to complementing the SSR markers already transferred to *S. adstringens*, can more accurately detect genetic variability and how it is organized in individuals and populations of this species. A larger sample of individuals will provide more robust information on the

Table 3 Genetic characterization of 15 polymorphic microsatellite loci in three populations of *Stryphnodendron adstringens*, from the Brazilian Cerrado

	Candeias - CANMG (N = 16)				Posse - POSGO (N = 16)				Chapada dos Guimarães - CHGMT (N = 16)			
Locus	*A*	*He*	*Ho*	F_{IS}	*A*	*He*	*Ho*	F_{IS}	*A*	*He*	*Ho*	F_{IS}
SadH1	1	0.0000	0.0000	0.0000	1	0.0000	0.0000	0.0000	3	0.4335	0.4375	− 0.0096
SadH3	2	0.2391	0.1333	0.4510	2	0.0625	0.0625	0.0000	2	0.0625	0.0625	0.0000
SadH5	8	0.7621	0.7500	0.0164	9	0.7782	0.8125	− 0.0456	6	0.7540	0.6250	0.1758
SadH7	5	0.7287	0.4667	0.3677*	3	0.4859	0.2500	0.4937*	4	0.7480	0.7500	− 0.0028
SadH8	7	0.8770	0.8750	0.0024	6	0.7621	0.6250	0.1848	6	0.7218	0.6667	0.0789
SadH9	7	0.5242	0.5625	− 0.0757	11	0.9093	0.9375	− 0.0321	3	0.1794	0.1875	− 0.0465
SadH11	6	0.7278	0.7500	− 0.0315	6	0.8165	0.6875	0.1624	3	0.3306	0.3125	0.0566
SadH13	7	0.8266	0.8750	− 0.0606	7	0.7298	0.5625	0.2351	6	0.6613	0.5625	0.1536
SadH14	8	0.8347	0.9375	− 0.1278	7	0.8427	0.8125	0.0370	7	0.7460	0.6875	0.0808
SadH15	4	0.6431	0.3846	0.4118*	3	0.6452	0.5625	0.1318	2	0.3528	0.4375	− 0.2500
SadH16	4	0.5625	0.8125	− 0.4662*	5	0.6956	0.6250	0.1045	3	0.6190	0.5625	0.0940
SadH17	1	0.0000	0.0000	0.0000	3	0.5060	0.3125	0.3902	4	0.7480	0.7500	− 0.0028
SadH18	3	0.2339	0.0000	1.0000*	2	0.2258	0.2500	− 0.1111	2	0.5081	0.3750	0.2683
SadH19	2	0.4980	0.6875	− 0.3983	2	0.3145	0.0000	1.0000**	1	0.0000	0.0000	0.0000
SadH20	5	0.6943	0.7333	− 0.0584	2	0.2468	0.2727	− 0.1111	2	0.1250	0.1250	0.0000
Overall	70	0.5435	0.5312	0.0230	69	0.5347	0.4515	0.1602**	54	0.4660	0.4361	0.0662

A: number of alleles; *He*: expected heterozygosity; *Ho*: observed heterozygosity; F_{IS}: fixation index; *N*: number of analyzed individuals

*Significant *p* value ($p < 0.05$) after 20,000 randomizations

**Significant *p* value ($p < 0.001$) after 20,000 randomizations

genetic diversity of *S. adstringens*. The probability of detecting polymorphisms in other wild species of the genus *Stryphnodendron* becomes greater when the markers were developed in closely related species.

Regarding the cross-amplification tests, of the 20 loci tested, all were successfully amplified, at an annealing temperature of 52 °C, in two species: *S. rotundifolium* and *S. polyphyllum*, with an amplification rate of 95% and 97%, respectively (see Supplementary Material Table S4; Fig. S1). As they are species of the same genus, the cross-amplification rates were satisfactory for both, requiring only that the annealing temperature be optimized to check for clearer bands and, later, polymorphism tests can be performed. In *S. coriaceum*, only nine loci amplified at this temperature, with an amplification rate of 45% (see Supplementary Material Table S3 and Fig. S1). Even though the amplification rate in this case was below the expected 60% (Barbará et al. 2007), further tests with more individuals and different temperatures can be conducted for other loci that have not amplified.

Conclusions

Our results showed that the panel of microsatellite markers developed for the *S. adstringens* presented satisfactory results to support future population-based genetic studies, being a powerful genetic tool for elucidating, across a large scale, the evolutionary pattern of this species, including analyses of diversity and genetic structure of natural populations or germplasm collections, gene flow and mating system. Thereby, these markers will be relevant for the establishment of adequate conservation and genetic breeding strategies for the management of *S. adstringens* and other species at the genus *Stryphnodendron* as a valuable genetic resource.

Acknowledgements This work was supported by a PRONEX project from: PRONEX—FAPEG/CNPq (CP 07/2012) and CNPq (447754/2014-9). A.R.G, L.O.B, and U.J.B.S were supported by a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Works conducted by A.N.S.P., B.W.B, and M.P.C.T. have been continuously supported by productivity fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Our current research in Genetics and Genomics is developed in the context of National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation, supported by MCTIC/CNPq (Proc. 465610/2014-5) and Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG), which we gratefully acknowledge.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Andrews S (2015) FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MFC, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. Mol Ecol. https://doi.org/10.1111/j.1365-294X.2007.03439.x

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu170

Branco EA, Zimback L, Lima AB, Mori ES, Aoki H (2010) Estrutura Genética de populações de *Stryphnodendron adstringens* (Mart.), pp 31–37. Available online at https://smastr16.blob.core.windows.net/iflorestal/RIF/SerieRegistros/IFSR42/IFSR42.pdf

Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep. https://doi.org/10.1007/BF02772828

Diniz VSS, Franceschinelli EV (2015) Estrutura populacional e brotamento de três espécies nativas do cerrado em diferentes regimes de queimadas. J Neotr Biol. https://doi.org/10.5216/rbn.v11i2.29488

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19(1):11

Glasenapp JS, Casali VWD, Barbosa PB et al (2014) Description of genetic diversity of natural populations of barbatimão *Stryphnodendron adstringens* (Mart.) Coville in conservation areas from Minas Gerais. Rev Arvore. https://doi.org/10.1590/S0100-67622014000100010

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3.2). Available online at http://www.unil.ch/izea/softwares/fstat.html

Meglécz E, Pech N, Gilles A et al (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. Mol Ecol Res. https://doi.org/10.1111/1755-0998.12271

Mendonça PC, Bertoni BW, Amui SF et al (2012) Genetic diversity of *Stryphnodendron adstringens* (Mart.) Coville determined by AFLP molecular markers. Biochem Syst Ecol. https://doi.org/10.1016/j.bse.2011.12.007

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individual. Genetics 89:583–590

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol. https://doi.org/10.1111/j.1365-294X.1995.tb00227.x

Pellenz NL, Barbisan F, Azzolin VF, Marques LPS, Mastella MH, Teixeira CF, Ribeiro EE, da Cruz IBM (2019) Healing activity of *Stryphnodendron adstringens* (Mart.), a Brazilian tannin-rich species: a review of the literature and a case series. Wound Med. https://doi.org/10.1016/j.wndm.2019.100163

Pryszcz LP, Gabaldón T (2016) Redundans: an assembly pipeline for highly heterozygous genomes. Nucl Acids Res. https://doi.org/10.1093/nar/gkw294

Rizzini CT, Heringer EP (1966) Estudo sobre os sistemas subterrâneos difusos de plantas campestres. Anais da Acad Bras de Ciênc 38:85–112

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Souza VC, Lima AG (2020) *Stryphnodendron* in Flora do Brasil 2020 em construção. Jardim Botânico do Rio de Janeiro. Available online at http://reflora.jbrj.gov.br/reflora/floradobrasil/FB23174

Souza-Moreira TM, Queiroz-Fernandes GM, Pietro RCLR (2018) *Stryphnodendron* species known as "Barbatimão": a comprehensive report. Molecules. https://doi.org/10.3390/molecules23040910

Taheri S, Abdullah TL, Yusop MR et al (2018) Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. Molecules. https://doi.org/10.3390/molecules23020399

R Core Team (2020) R: a language and environment for statistical computing R Foundation for Statistical Computing. Vienna, Austria. Available online at http://www.r-projectorg/

Unamba CIN, Nag A, Sharma RK (2015) Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. Front Plant Sci. https://doi.org/10.3389/fpls.2015.01074

Wagner HW, Sefc KM (1999) IDENTITY 1.0. Centre for Applied Genetics, University of Agricultural Sciences, Vienna (version 4.0). Available online at http://www.uni-graz.at/~sefck

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates Inc., Sunderland

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

Wright S (1951) The genetic structure of populations. Ann Eug 15:323–354

Zalapa JE, Cuevas H, Zhu H et al (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot. https://doi.org/10.3732/ajb.1100394

Zimin AV, Marçais G, Puiu D et al (2013) The MaSuRCA genome assembler. Bioinformatics. https://doi.org/10.1093/bioinformatics/btt476