Molecular Biology Reports
https://doi.org/10.1007/s11033-021-06338-5

ORIGINAL ARTICLE

Detection and application of novel SSR markers from transcriptome data for *Astronium fraxinifolium* Schott, a threatened Brazilian tree species

Maiara R. Cornacini[1] · Ricardo O. Manoel[2] · Marcelo A. M. Alcantara[1] · Mário L. T. Moraes[3] · Edvaldo A. A. Silva[4] · Leonel G. Pereira Neto[5] · Alexandre M. Sebbenn[6] · Bruno C. Rossini[1,2] · Celso L. Marino[1,2]

Received: 27 December 2020 / Accepted: 7 April 2021

Abstract

Astronium fraxinifolium is an endangered tree species from Brazil. Due to its significance in environmental reforestation, as well as the continued exploitation of its wood, it is necessary to develop management programs that support the conservation of the species. Simple sequence repeats (SSR) or microsatellite markers are widely used in population genetic studies across a range of diverse organisms. In this study, we present the first SSR markers developed for *A. fraxinifolium* as well as their frequency and distribution based on transcriptome data. From transcriptome data, we identified more than 100 thousand sequences presenting microsatellites, with a predominant distribution of trinucleotide repeats. From the initial screening, we selected 20 microsatellite loci which were validated and evaluated for genetic indices in two natural populations. All loci were polymorphic, ranging from four to 11 alleles per locus. The observed and expected heterozygosities ranged from 0 to 1.0 and from 0.533 to 1.0, respectively, while the genetic differentiation (G_{ST} = 0.363) was greater within than between populations. The developed SSR loci from RNA-Seq data provides a foundation for future studies on genetic diversity and population structure, mating system, and gene flow for *A. fraxinifolium* populations and related species, aiming at conservation and management.

Keywords Anacardiaceae · Conservation genetics · Management · Microsatellite markers · Population genetics

Introduction

Forest fragmentation has a direct impact on the genetic diversity, population structure, mating system, and gene flow of tree populations. Population genetic studies based on genetic markers are key to understanding the effects of anthropogenic activities on natural populations, conservation efforts, and the improvement of trees species. The tree *Astronium fraxinifolium* Schott (Anacardiaceae) is distributed discontinuously throughout several biomes in Brazil, occurring on rocky terrain from the Cerrado to the Caatinga [1, 2]. It is a dioecious, insect pollinated tree that is often used in the restoration of degraded areas [2]. Due to the extensive fragmentation of its biomes, *A. fraxinifolium* is classified as threatened with extinction, and remaining populations are often found as isolated trees along highways or in small forest fragments [3, 4]. Thus, the development of genetic markers, such as microsatellite loci (SSR), is urgently needed to better understand the genetic diversity,

✉ Ricardo O. Manoel
rickom.isa@gmail.com

1 Departamento de Ciências Químicas e Biológicas, Instituto de Biociências, UNESP - Universidade Estadual Paulista, Botucatu, Brazil

2 Instituto de Biotecnologia (IBTEC), UNESP - Universidade Estadual Paulista, Botucatu, Brazil

3 Departamento de Fitotecnia, Tecnologia de Alimentos e Sócio Economia, Faculdade de Engenharia de Ilha Solteira, UNESP - Universidade Estadual Paulista, Ilha Solteira, Brazil

4 Departamento de Produção Vegetal, Faculdade de Ciências Agronômicas, UNESP – Universidade Estadual Paulista, Botucatu, Brazil

5 Embrapa Recursos Genéticos e Biotecnologia (Cenargen), Brasilia, Brazil

6 Instituto Florestal de São Paulo, Piracicaba, Brazil

Published online: 18 April 2021

population structure, mating system, and gene flow of the remaining populations.

Genomic studies of the *Astronium* genus are rare. Recently, the use of high throughput sequencing technologies has facilitated the development of molecular markers for a broad range of organisms [5–8]. The ability to search within these datasets has increased the chances of finding SSRs without any prior enrichment. From this, the development of SSR markers from transcriptome sequences has become an effective tool for population genetic studies [9–11], particularly for endangered species [10]. Here, we developed a set of 20 polymorphic microsatellite loci for *A. fraxinifolium* and evaluated their frequency and distribution based on Illumina generated RNA-Sequence (RNA-Seq) data. We also included an analysis of repeats and GO classification of the reads. The loci were validated for reproducibility and applicability to assess genetic diversity, mating system, and gene flow for this tree species.

Materials and methods

Sample collection and DNA extraction for validation

For validation, we collected fresh leaves from 60 *A. fraxinifolium* specimens from two natural populations. The first population included 30 individuals sampled along the SP-595 highway in the municipality of Ilha Solteira, São Paulo State (20°21′36.46″S, 51°01′15.52″W), Brazil, which is characteristic of the Semideciduous Seasonal Forest. The second population is composed of 30 adult trees located along highway BR-158 in the municipality of Selvíria, Mato Grosso do Sul State (20°12′02.30″S, 51°14′56.81″W), Brazil, which is characteristic of the Cerrado biome. Authorization to collect leaf tissue samples was provided by the Institute for Biodiversity Conservation (ICMBio), of the Ministry of the Environment (MMA), protocol numbers 41166-1 and 73998-1.

SSR loci identification and characterization

Species-specific microsatellite primers were generated from a cDNA genomic library of the species [12] (unpublished data) and submitted to new generation sequencing (HiSeq Sequencing 2500 System, Illumina) (Supplementary material S1). Briefly, using the NucleoSpin RNA Plant Kit (Macherey-Nagel, Düren, Germany), the total RNA was extracted from the fresh embryonic axis of seeds collected from the municipalities of Alto Paraíso, Cavalcante, Colinas do Sul, and Niquelândia in Goiás State, and Montes Claros, Mirabela, and Lontra in Minas Gerais State. Seed collection was authorized by ICMBio, protocol number 41166-1. The library was constructed using the TrueSeq RNA Library Prep Kit V2 (Illumina) and sequenced with a 2× 100 bp paired-end run. To develop useful microsatellite markers for population genetic studies, we considered a subset of four samples from this run. The software SSR_pipeline [13] was used to search SSRs considering the parameters of di-, tri-, tetra-, penta- and hexanucleotides for a minimum of seven, six, five, four and four repeats, respectively, with a 40 bp flanking sequence. This software uses the raw data reads as input, which is processed in a quality sorting module, followed by an aligning of reads, and a final search for SSRs. Primer design was conducted in BatchPrimer3 v1.0 [14]. Functional annotation of SSR-containing coding sequences were analyzed in Blast2GO software, using the Ensemble-Plants database from UniProt [15]. The identified sequences were deposit in Targeted Locus Study (TLS) from NCBI under BioProject PRJNA70516.

For the population analysis, genomic DNA was isolated from fresh leaves using the cetyltrimethylammonium bromide (CTAB) protocol [16] quantified with a NanoDrop ND-1000 Spectrophotometer (NanoDrop Products, DE, USA). Integrity was verified by running 1% agarose gels with a TBE (1X) at a constant voltage of 5 V/cm. We selected 20 SSR loci for population validation. Polymerase chain reactions (PCR) were performed with a Mastercyler thermocycler (Eppendorf, Hamburg, Germany) with the addition of a M13 tail for fluorescent labeling [17]. The reaction mixture contained 5.0 µL GoTaq Colorless Master Mix, 0.3 µL of forward primer (2 pmol), 0.3 µL of reverse primer (8 pmol) 0.3 µL of fluorescent primer with M13 (8 pmol; 6-FAM, VIC, PET or NED, Applied Biosystems), 0.5 µL of Bovine Serum Albumin (BSA, 0.1 mg/ml from Promega Corporation), 1.0 µL genomic DNA (approximately 50 ng), and nuclease-free water to a final reaction volume of 10 µL. The PCR program was as follows: 2 min at 96 °C; 35 cycles of 30 s at 96 °C; a primer-specific annealing temperature (Supplementary material S2) for 1 min 30 s; 72 °C for 1 min 30 s; followed by 12 cycles of 96 °C for 30 s; 53 °C for 1 min 30 s; 72 °C for 1 min 30 s; and a final extension at 72 °C for 20 min. The PCR products were genotyped in an ABI3130xl Genetic Analyzer (Applied Biosystems) with GeneScan 500 LIZ (Applied Biosystems) (Supplementary material S3). Genotypes were assigned using the software GeneMapper v.5.0 (Applied Biosystems).

Statistical population analysis

The number of alleles per locus (k), observed (H_o) and expected (H_e) heterozygosity, and polymorphism information content (PIC) were estimated using the software CERVUS 3.03 [18]. The fixation index (F) and genotypic linkage disequilibrium (LD) were estimated for each population using FSTAT [19]. To test if the F values and LD were significantly different from zero, we used Monte Carlo

permutations and a Bonferroni correction (95%, α = 0.05). The program Micro-Checker v.2.2.3 [20] was used to detect the occurrence of null alleles (*Null*) and estimate the genetic differentiation (G_{ST}) based on Hedrick's statistics [21]. The ´adegenet´ package in the R software environment was used to conduct the principal component analysis (PCA), discriminant analysis of principal components (DAPC), and the assignment probability of each individual [22]. Additionally, a Bayesian analysis was performed in the STRUCTURE software [23], assuming an admixture model with correlated allele frequencies, testing each K (1–4) with 10 independent runs using a burn-in of 100,000 and 1,000,000 generations. The identification of the optimal K was inferred using the Evanno method [24] in the Structure Harvester program [25].

Table 1 Summary of Illumina paired-end sequence data; it includes non- and perfect motifs di-, tri-, tetra-, penta- and hexa-nucleotides for *Astronium fraxinifolium*

Motif	Di-	Tri-	Tetra-	Penta-	Hexa-
Number of contigs containing microsatellites	146,182	163,062	13,274	29,540	38,160
Number of contigs with flanking sequence	37,389	60,331	5434	8852	13,202

Total number of reads = 189,057,492; all contigs = 95,768,009

Results

Identification and classification of SSR markers

In this study, the sequence run produced 189 million paired-end reads, of which more than 95 million were successfully joined. The trinucleotide motifs were the most abundant followed by di-, hexa-, penta, and tetranucleotides (approximately 41.8, 37.5, 9.80, 7.6, and 3.4%, respectively, Table 1). Of these, an average of 32% of the sequences had sufficient flanking regions to design primers, except for the tetra- and trinucleotide loci where the number of sequences with flanking regions was 40.9 and 36.9%, respectively. From an initial screening, more than 113,000 were identified with tandem repeats. Subsequently, we designed 20 primer pairs for amplification and testing through population analysis.

The SSR functional annotation was classified considering the three major categories: cellular component, molecular function, and biological process (Fig. 1). The GO classification relating to the cellular component showed that the most abundant are intracellular (GO:0005622) or intracellular

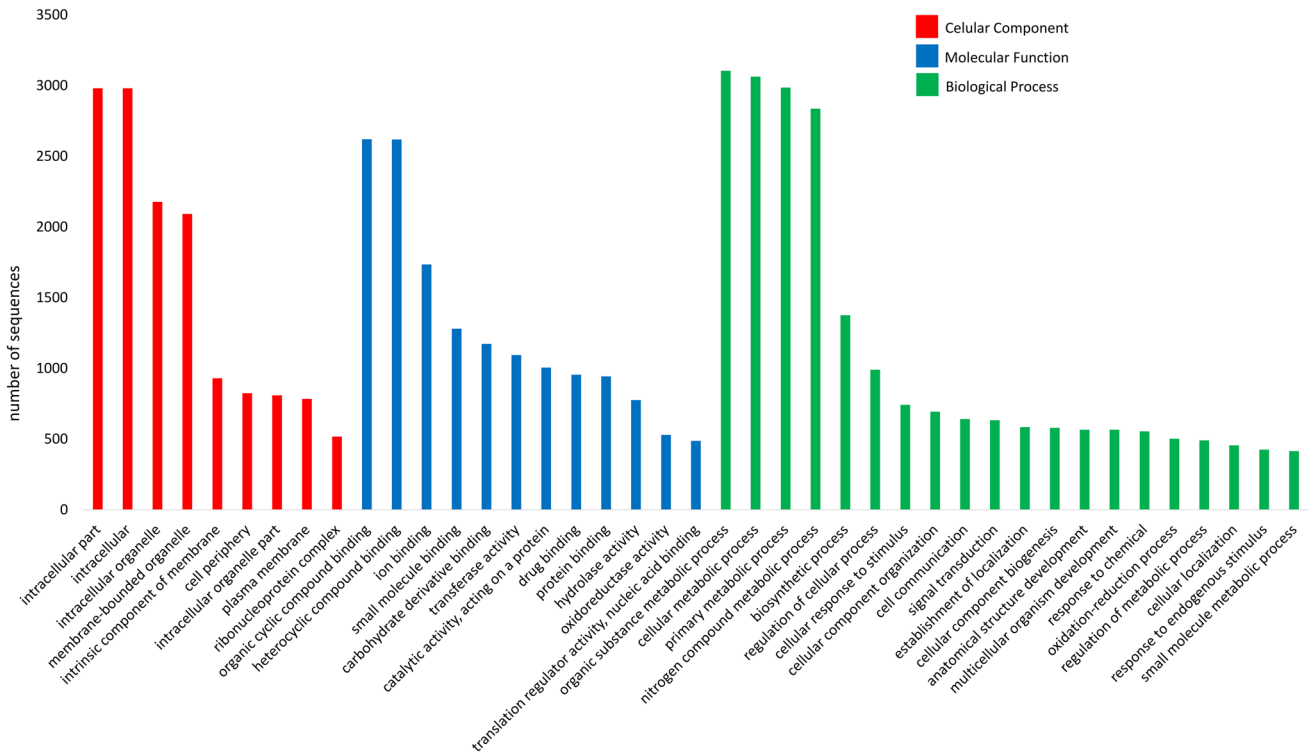

Fig. 1 Functional annotation of SSRs in coding regions of *Astronium fraxinifolium* transcriptome, including the number of genes putatively involved in different subcellular functions based on GO classification

part (GO:0044424). For molecular function, the most representative GOs were related to organic cyclic compound binding (GO:0097159) and heterocyclic compound binding (GO:1901363). The biological process category were mainly represented by genes involved in organic substance metabolic process (GO:0071704) and cellular metabolic process (GO:0044237). Considering the tissue analyzed in the transcriptome analysis, the GO classification also shows a large number of genes related to cell communication (GO:0007154), anatomical structure development (GO:0048856), and multicellular organism development (GO:0007275). These are consistent with the initial stages of development and suggest significant potential for future studies on these regions.

Genetic diversity of natural populations

The use of SSRs derived from RNA-Seq increases the success of amplification, including for related species, due to the conservation of transcribed flanking regions. All designed microsatellite primer pairs were amplified successfully and showed polymorphism for the studied populations, with four to 11 alleles detected per locus and *PIC* values ranging from 0.346 to 0.857 (Table 2), indicating that the markers are appropriate for population studies.

Despite the history of fragmentation of *A. fraxinifolium* populations, the SSR loci showed a large amount of genetic variation: for Ilha Solteira the observed heterozygosity (H_o) ranged from 0 to 0.944 (mean of 0.674) and expected heterozygosity (H_e) from 0.533 to 0.871 (mean of 0.741); in Selvíria, H_o ranged from 0.133 to 1.0 (mean of 0.668) and H_e from 0.606 to 1.0 (mean of 0.797). The fixation index (E) ranged from − 0.394 to 0.494 (mean of 0.090) for Ilha Solteira, and from − 0.069 to 0.783 (mean of 0.162) for Selvíria. Null alleles were observed in four and six loci in Ilha Solteira and Selvíria, respectively. After Bonferroni correction, genotypic linkage disequilibrium (LD) was observed in four pairs of loci in Ilha Solteira and three pairs in Selvíria (Table 3).

To test the genetic similarity between populations, we used DAPC and STRUCTURE analyses. The genetic differentiation (G_{ST}) between populations (0.363) was high suggesting that a large proportion of the genetic diversity is distributed within rather than among populations. The high genetic differentiation among the populations was expected given their geographical distance (50 km). However, this differentiation can also be explained by the fact that the species

Table 2 Results of screening in two populations of *Astronium fraxinifolium*

Locus	Ilha Solteira ($n=30$)						Selvíria ($n=30$)						G_{ST}
	k	*PIC*	H_o	H_e	*F*	*Null*	*k*	*PIC*	H_o	H_e	*F*	*Null*	
Ga01	8	0.760	0.933	0.799	− 0.171	−	9	0.754	0.654	0.792	0.177*	−	0.757
Ga02	11	0.551	0.867	0.626	− 0.394	−	7	0.784	0.808	0.824	0.105	−	0.513
Ga03	8	0.346	0.000	0.533	0.094	−	8	0.632	0.367	0.680	0.337*	+	0.091
Ga04	11	0.788	0.467	0.818	0.434*	+	10	0.804	0.533	0.836	0.366*	+	0.044
Ga05	8	0.656	0.367	0.718	0.494*	+	5	0.664	0.296	0.726	0.597*	+	0.785
Ga06	9	0.755	0.733	0.797	0.081*	−	8	0.712	0.769	0.754	− 0.020	−	0.959
Ga07	10	0.762	0.800	0.804	0.005	−	8	0.693	0.630	0.746	0.158*	−	0.484
Ga08	8	0.840	0.933	0.871	− 0.073	−	8	0.719	0.821	0.769	− 0.069	−	0.583
Ga09	10	0.805	0.667	0.833	0.202*	+	11	0.857	0.733	0.885	0.174*	+	0.337
Ga10	7	0.569	0.308	0.654	0.475*	+	4	0.725	0.600	0.785	0.106	−	0.263
Ga11	8	0.724	0.800	0.777	− 0.030	−	6	0.785	0.731	0.827	0.159*	−	0.058
Ga12	8	0.657	0.944	0.730	− 0.175	−	6	0.743	0.800	0.799	0.137*	−	0.314
Ga13	7	0.756	0.722	0.810	0.124*	−	8	0.725	0.762	0.779	0.089	−	0.194
Ga14	9	0.734	0.722	0.794	0.069	−	8	0.746	0.824	0.804	0.133*	−	0.088
Ga15	8	0.711	0.826	0.769	0.085	−	8	0.757	0.500	0.817	0.377*	+	0.013
Ga16	9	0.580	0.600	0.623	0.038	−	6	0.761	0.724	0.797	0.093	−	0.693
Ga17	5	0.669	0.828	0.721	− 0.129	−	8	0.655	0.567	0.719	0.215*	−	0.340
Ga18	8	0.690	0.655	0.743	0.125	−	11	0.815	0.733	0.849	0.138*	−	0.149
Ga19	4	0.558	0.552	0.634	0.140	−	7	0.551	0.133	0.606	0.783*	+	0.086
Ga20	8	0.710	0.759	0.760	0.010	−	9	0.781	0.733	0.821	0.109	−	0.072
Mean	8.2	0.681	0.674	0.741	− 0.029	−	7.5	0.720	0.668	0.797	0.059	−	0.363

k number of alleles per locus, *PIC* polymorphism information content, H_o observed heterozygosity, H_e expected heterozygosity, *F* fixation index, *Null* null alleles occurrence, G_{ST} genetic differentiation between populations; $P < 0.05$

Table 3 Genotypic disequilibrium between pairwise microsatellite loci in adult of *Astronium fraxinifolium*

Pairwise loci	Ilha Solteira	Selvíria
Ga01xGa05	0.00005	1.00000
Ga01xGa06	0.00005	1.00000
Ga01xGa07	0.00005	1.00000
Ga01xGa08	0.00011	1.00000
Ga01xGa19	0.89137	0.00005
Ga01xGa20	0.73405	0.00005
Ga02xGa03	0.00047	0.00005

The values represent the probability of genotypic disequilibrium after 19.000 permutations of alleles among individuals. Probability after Bonferroni's corrections: P = 0.000263 (α = 0.05)

is pollinated by bees, which have been shown to disperse pollen over limited distances [1]. The PCA showed a clear differentiation between populations, with some individuals mixed between them. Furthermore, DAPC and the results from both assignment probability tests (adegenet package and STRUCTURE) showed similar results for population structure with two distinct populations (Fig. 2).

Discussion

From *A. fraxinifolium* transcriptome data, a predominance of tri-, followed by dinucleotide motifs were identified, representing more than 79% of all identified contigs, which could not affect the protein structure [26, 27]. When analyzing the number of repeats, we identified that those with more than 10 correspond to less than 7.8% of all SSRs (Table 4). The

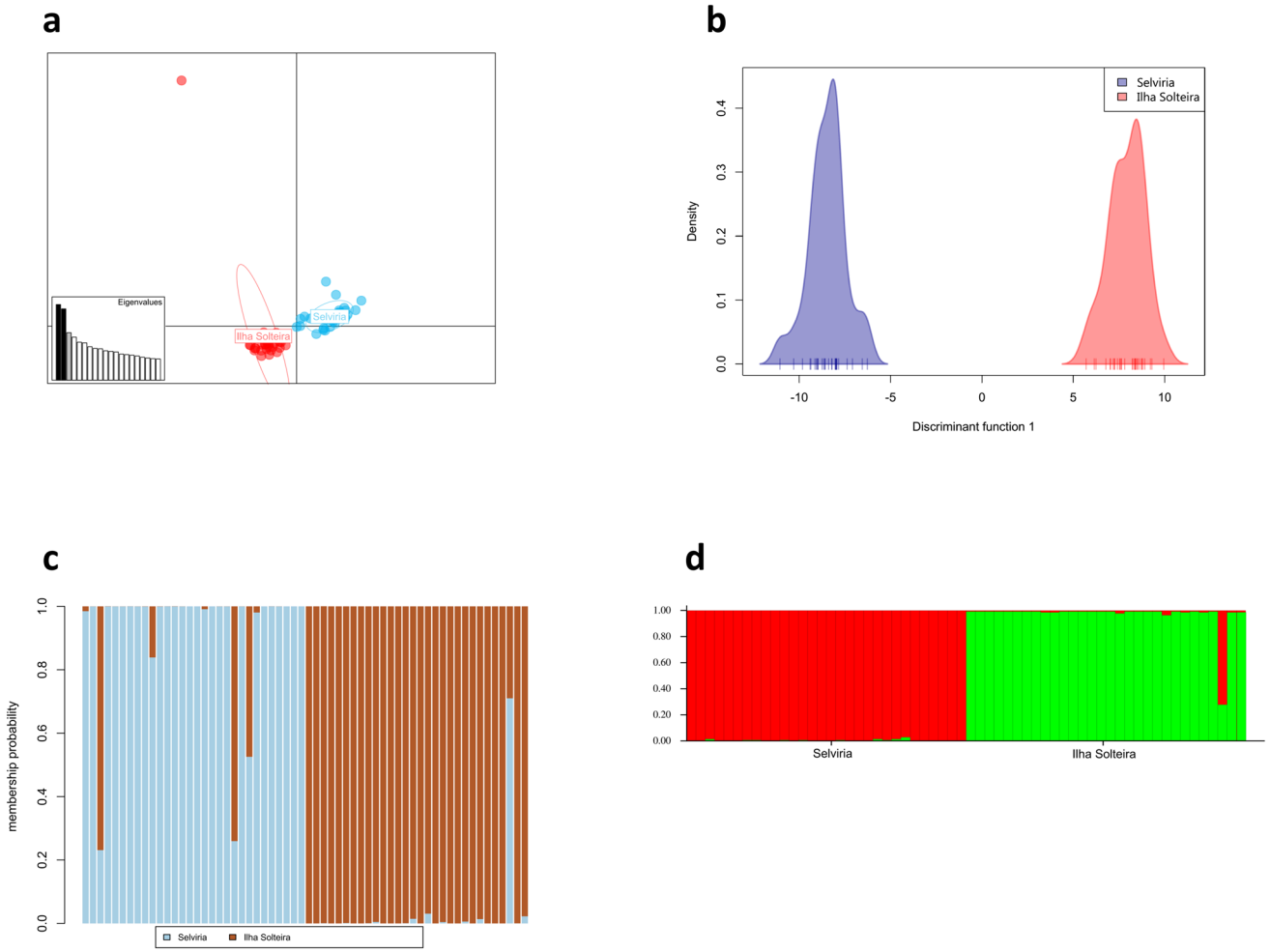

Fig. 2 Principal Component Analysis (PCA), discriminant analysis of principal components (DAPC), and assignment tests for two analyzed populations of *Astronium fraxinifolium*: **a** PCA showing the distribution of genotypes; **b** DAPC clearly showing the differences between populations. Colors in (**a, b**) indicate population: Ilha Solteira (red) and Selviria (blue). **c** Assignment test from adegenet package. **d** STRUCTURE results from analysis of optimal K = 2. **c, d** Each column and color indicates the genetic assignment of individuals: **c** Ilha Solteira (brown) and Selviria (blue); **d** Ilha Solteira (green) and Selviria (red)

Table 4 Frequency distribution of SSRs identified in the *A. fraxinifolium* transcriptome

Repeat number	Di-	Tri-	Tetra-	Penta-	Hexa-	Total
4	N/A	N/A	N/A	6685	11,008	17,693
5	N/A	N/A	3003	1740	1644	6387
6	N/A	33,155	2139	365	466	36,125
7	13,340	15,352	217	56	55	29,020
8	6167	7539	52	3	20	13,781
9	5489	2413	10	2	4	7918
10	3511	1013	4	0	2	4530
> 10	8882	859	9	1	3	9754
Total	37,389	60,331	5434	8852	13,202	125,208

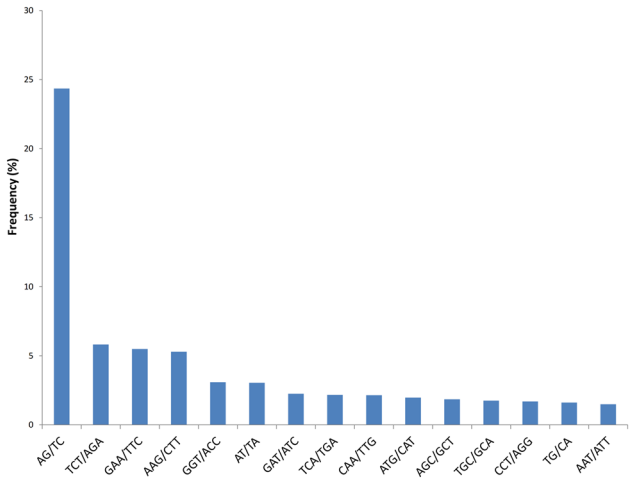

Fig. 3 Frequency distribution of the most representative SSR motif types in the *A. fraxinifolium* transcriptome

SSR frequency decreased with an increase in motif length, as was also reported for *Magnolia wufengensis* [28]. The frequency of motifs from AG/CT corresponds to more than 24.3% and is the most abundant motif in this species, followed by TCT/AGA repeats with less than 6% (Fig. 3). The frequency of AG/CT repeats are higher than those found for other species such as bamboo (17.11%) [26], but lower than that found for Magnolia (37.8%) [28]. High frequencies of AG repeats have been reported for other plant species, suggesting that those frequencies could be related to mutation mechanisms in the generation of SSRs or selective pressure on particular sequences [26–31].

The developed microsatellite markers were effective for assessing the genetic differentiation among sampled populations. The average levels of observed and expected heterozygosity were greater than that reported for populations of *Astronium graveolens* [32] and for other tropical species, such as populations of *Cedrela fissilis* (Meliaceae) [33], *Campomanesia xanthocarpa* (Myrtaceae) [34], *Myracrodruon urundeuva* [6], and *Eugenia uniflora* L. (Myrtaceae) [35], which confirms the high genetic variability found in the populations assessed herein. Therefore, these genetic markers are reliable for use in population genetic studies, such as the investigation of pollen and seed dispersal patterns. Such studies can help to understand the actual distribution of natural populations that reflects the evolutionary history of the species. Previous studies on other tree species show long-distance pollen dispersal, for example with pollen dispersal for *Hymenaea stignocarpa* reaching more than 8 km between analyzed populations [36], and for *Ceiba pentandra* reaching up to 18 km [37, 38]. However, these distances are mainly related to dispersal by bats, which have a large foraging area. *A. fraxinifolium* is pollinated by bees, and behavior analyses of insects have reported foraging distances of up to 6 km [39, 40]. Such results indicate that further analyses of pollen/seed dispersal are necessary for this species. To date, few studies have been conducted on natural populations of *A. fraxinifolium* and most have focused on silvicultural traits [4]. Recent studies have described SSR loci for *A. graveolens*, but no other analyses have been conducted on other *Astronium* species [32]. Therefore, the microsatellite markers developed herein are useful for studies on genetic diversity and structure, gene flow, and mating system, and can provide information for the conservation, breeding, and reforestation of the species. In addition, our study provides a database with more than 125,000 expressed SSR sequences in the genome that will serve as a basis for analyses of the consequences of forest fragmentation in tropical forests of Brazil. As such, it contributes to the development of appropriate strategies for the conservation of *A. fraxinifolium* and related species from the Anacardiaceae family.

This study presents the first SSR markers for *A. fraxinifolium*. The frequency and distribution of SSR motifs shows high levels of diversity, with a predominance of trinucleotides, as has been reported for other plant species. Functional annotation of SSRs can help future breeding programs in the selection of genes related to important development characteristics. Furthermore, the use of transcriptome derived SSRs can increase the rate of amplification in related species, due to the conservation of the flanking regions of these loci. At the population level, these SSR markers show

sufficient levels of polymorphism. Therefore, our results suggest that these markers can be used as tools for ecological population genetic studies, such as genetic diversity, spatial genetic structure, mating system, and gene flow, while also helping to improve the development of genetic conservation and management strategies for fragmented populations and related species.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-021-06338-5.

Acknowledgements The authors are thankful for the technical support in laboratory by Dra. Patrícia Ferreira Alves and in field by Dr. José Cambuim and Dr. Alexandre Marques Silva.

Author contributions MRC and ROM designed the research. MRC, MAA, and ROM carried out the field and laboratory phases of the research. EAAS and LGPN developed the cDNA genomic library. BCR and ROM performed analyses and wrote the manuscript. MLTM, BCR, and CLM designed the experiment. BCR and CLM supervised the research and MLTM and AMS reviewed the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP#2018/00898–0; scholarship to ROM) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ; Fellowship to EAAS and MLTM).

Declarations

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

Ethical approval All applicable international, national, and/or institutional guidelines were followed.

Informed consent All individual participants included in the study consent to this manuscript to participate and for publication.

References

1. Leite EJ (2002) State-of-knowledge on *Astronium fraxinifolium* Schott (Anacardiaceae) for genetic conservation in Brazil. Perspect Plant Ecol Evol Syst 5:63–77. https://doi.org/10.1078/1433-8319-00023
2. Carvalho PER (2010) Espécies arbóreas brasileiras. Embrapa Informação tecnológica/Embrapa Florestas, Brasília/Colombo, pp 231–240
3. Ibama (1992) Portaria Ibama nº 37-N, de 03 de abril de 1992. http://www.mma.gov.br/estruturas/179/_arquivos/179_05122008034139.pdf. Accessed 26 Nov 2019
4. Aguiar AV, Bortolozo FR, Moraes MLT et al (2003) Genetic variation in *Astronium fraxinifolium* populations in consortium. Crop Breed Appl Biotechnol 3(2):95–106. https://doi.org/10.12702/1984-7033.v03n02a02
5. González-Castellano I, Perina A, González-Tizón AM et al (2018) Isolation and characterization of 21 polymorphic microsatellite loci for the rockpool shrimp *Palaemon elegans* using Illumina MiSeq sequencing. Sci Rep 8:17197. https://doi.org/10.1038/s41598-018-35408-1
6. Souza DCL, Rossini BC, Souza FB et al (2017) Development of microsatellite markers for *Myracrodruon urundeuva* (F.F. & M.F. Allemão), a highly endangered species from tropical forest based on next-generation sequencing. Mol Biol Rep 45(1):71–75. https://doi.org/10.1007/s11033-017-4142-z
7. Fraga DA, De Carvalho AF, Santana RS et al (2020) Development of microsatellite markers for the threatened species *Coleocephalocereus purpureus* (Cactaceae) using next-generation sequencing. Mol Biol Rep 47:1485–1489. https://doi.org/10.1007/s11033-019-05213-8
8. Hou M, Gz Du, Zg Z (2020) Development of genomic microsatellite markers for *Aconitum gymnandrum* (Ranunculaceae) by next generation sequencing (NGS). Mol Biol Rep 47:727–729. https://doi.org/10.1007/s11033-019-05160-4
9. Zheng X, Pan C, Diao Y et al (2013) Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). BMC Genomics 14(1):490. https://doi.org/10.1186/1471-2164-14-490
10. Zhou T, Li ZH, Bai GQ et al (2016) Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia oliver* (Aceraceae). Molecules 21(3):166. https://doi.org/10.3390/molecules21030166
11. Lu QX, Gao J, Wu JJ et al (2020) Development of 19 novel microsatellite markers of lily-of-the-valley (Convallaria, Asparagaceae) from transcriptome sequencing. Mol Biol Rep 47:3041–3047. https://doi.org/10.1007/s11033-020-05376-9
12. Pereira-Neto LG (2016) *Astronium fraxinifolium* Schott seed longevyti: physiological, biochemical and molecular studies. Thesis in Agronomy. Faculty of Agronomic Sciences, Paulista State University, Botucatu
13. Miller MP, Knaus BJ, Mullins TD et al (2013) SSR_pipeline: A bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data. J Hered 104:881–885. https://doi.org/10.1093/jhered/est056
14. You FM, Huo N, Gu Y et al (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinform 9(1):253. https://doi.org/10.1186/1471-2105-9-253
15. Gotz S, Garcia-Gomez JM, Terol J et al (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36(10):3420–3435. https://doi.org/10.1093/nar/gkn176
16. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15
17. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234. https://doi.org/10.1038/72708
18. Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x
19. Goudet J (2002) FSTAT version 2.9.3.2, a program to estimate and test gene diversities and fixation index. Institute of Ecology, Lausanne, Switzerland
20. Van Oosterhout C, Hutchinson W, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x
21. Hedrick PW (2005) A standardized genetic differentiation measure. Evol Int J Org Evol 59(8):1633–1638. https://doi.org/10.1111/j.0014-3820.2005.tb01814.x

22. Jombart T, Almed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071. https://doi.org/10.1093/bioinformatics/btr521
23. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959
24. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2635. https://doi.org/10.1111/j.1365-294X.2005.02553.x
25. Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361. https://doi.org/10.1007/s12686-011-9548-7
26. Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007. https://doi.org/10.1093/molbev/msh073
27. Mrázek J, Guo X, Shah A (2007) Simple sequence repeats in prokaryotic genomes. Proc Natl Acad Sci USA 104:8472–8477. https://doi.org/10.1128/MMBR.62.2.275-293.1998
28. Wang L, Gong X, Jin L et al (2019) Development and validation of EST-SSR makers of *Magnolia wufengensis* using de novo transcriptome sequencing. Trees. https://doi.org/10.1007/s00468-019-01853-2
29. Cai K, Zhu L, Zhang K et al (2019) Development and characterization of EST-SSR markers from RNA-Seq data in *Phyllostachys violascens*. Front Plant Sci 10:50. https://doi.org/10.3389/fpls.2019.00050
30. Cho YG, Ishii T, Temnykh S, Chen X et al (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722
31. Wang P, Yang L, Zhang E et al (2016) Characterization and development of EST-SSR markers from a cold-stressed transcriptome of centipedegrass by illumina paired-end sequencing. Plant Mol Biol Rep 35(2):215. https://doi.org/10.1007/s11105-016-1017-8
32. Sanchez-Gomez KF, Cristóbal-Pérez EJ, Harvey N, Quesada M (2020) Isolation and characterization of microsatellite loci in *Astronium graveolens* (Anacardiaceae) and cross amplification in related species. Mol Biol Rep 47:4003–4007. https://doi.org/10.1007/s11033-020-05397-4
33. Gandara FB, Tambarussi EV, Sebbenn AM et al (2014) Development and characterization of microsatellite loci for *Cedrela fissilis* Vell (Meliaceae), an endangered tropical tree species. Silvae Genet 63:240–243. https://doi.org/10.1515/sg-2014-0031
34. Góes BD, Ruas EA, Benício LM et al (2019) Development and characterization of microsatellite loci for *Campomanesia xanthocarpa* (Myrtaceae) and cross amplification in related species. Acta Sci Biol Sci 41(1):43454. https://doi.org/10.4025/actascibiolsci.v41i1.43454
35. Sarzi DS, Justolin B, Silva CD et al (2019) Discovery and characterization of SSR markers in *Eugenia uniflora* L. (Myrtaceae) using low coverage genome sequencing. Anais Da Academia Brasileira de Ciências. https://doi.org/10.1590/0001-3765201920180420
36. Moraes MA, Kubota TYK, Rossini BC et al (2018) Long-distance pollen and seed dispersal and inbreeding depression in *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) in the Brazilian savannah. Ecol Evol 8(16):7800–7816. https://doi.org/10.1002/ece3.4253
37. Gribel R, Lemes MR (1997) Mating system and pollen flow of *Ceiba pentandra* (Bombacaceae) in Central Amazon Assessment of levels and dynamics of intra-specific genetic diversity of tropical trees. Second Annual Report to the European Commission.
38. Dick CW, Hardy OJ, Jones FA, Petit RJ (2008) Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. Tropical Plant Biology 1:20–33. https://doi.org/10.1007/s12042-007-9006-6
39. Hagler JR, Mueller S, Teuber LR, Machtley SA, Van Deynze A (2011) Foraging range of honey bees, *Apis mellifera*, in alfalfa seed production fields. J Insect Sci. https://doi.org/10.1673/031.011.14401
40. Abou-Shaara HF (2014) The foraging behaviour of honey bees, *Apis mellifera*: a review. Vet Med 59(1):1–10