Molecular Biology Reports (2018) 45:71–75
https://doi.org/10.1007/s11033-017-4142-z

SHORT COMMUNICATION

Development of microsatellite markers for *Myracrodruon urundeuva* (F.F. & M.F. Allemão), a highly endangered species from tropical forest based on next-generation sequencing

Danilla Cristina Lemos Souza[1] · Bruno César Rossini[2,3] · Francine Beatriz de Souza[4] · Alexandre Magno Sebbenn[5] · Celso Luís Marino[2,3] · Mário Luiz Teixeira de Moraes[1,4]

Received: 30 January 2017 / Accepted: 19 December 2017 / Published online: 29 December 2017

Abstract

Myracrodruon urundeuva is a tree species of high economic importance due the strength and durability of its wood. Threatened of extinction in Brazil, it is present only in a few forest remnants, mostly in conservation units. Currently, there is little information on the genetic diversity of natural populations in Brazil and even less information about the genome of this species. Here, new species-specific microsatellite loci were developed based on next-generation sequencing (Illumina). More than 100,000 loci were identified in the run, with di- to hexanucleotides motifs. Of these, 20 loci were selected for validation in 30 individuals, with 15 successfully polymorphic loci detected. The number of alleles ranged among loci from 3 to 16, with an average of 7.73, expected (H_e) and observed (H_o) heterozygosity ranged from 0.246 to 0.902 and from 0.103 to 0.867, respectively. These results point out that these new set of markers has a great potential for use in population genetic studies for genetic conservation of the species.

Keywords Aroeira · Conservation genetics · Microsatellite · Population genetics

Introduction

Within the process of human disordered occupation, tropical forests are one of the most devastated with the exploitation of native tree species for economic interests, promoting reduction of their populations and also spatial and genetic isolation between forest remnants. In a further perspective, if natural tree populations remain fragmented for many generations, possible losses in genetic diversity increase with inbreeding and the extinction of species may occur. Thus, it is necessary to understand the effects of forest fragmentation on the genetic diversity and population structure, mating system and gene flow of tropical tree species for its conservation and sustainable use [1].

Myracrodruon urundeuva (F.F. & M.F. Allemão), or "aroeira" (Anarcadiaceae), is a dioecious species [2, 3] with high wood quality and medicinal properties [4], but threatened of extinction [5] due to strong forest fragmentation of their biomes of occurrence. Native of South America, it occurs in Brazil, Bolivia, Paraguay and Argentina [6]. Currently, in Brazil, populations are mostly observed only in conservation units.

The use of molecular markers in population genetic studies of tree species has been shown to be effective. In *M. urundeuva*, previous studies were performed using isoenzyme markers [7, 8], RAPD and cpDNA [9, 10], AFLP [11, 12] and more recently by microsatellite markers [1, 3, 13–15]. In these last studies, nine microsatellite loci developed by Caetano et al. [13] were used in the analysis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-017-4142-z) contains supplementary material, which is available to authorized users.

✉ Bruno César Rossini
bu_rossini@yahoo.com.br

1 Faculty of Agronomy, UNESP - Univ Estadual Paulista, C.P. 237, CEP 18610-307 Botucatu, SP, Brazil

2 Biotechnology Institute (IBTEC), UNESP - Univ Estadual Paulista, CEP 18607-440 Botucatu, SP, Brazil

3 Department of Genetics, UNESP - Univ Estadual Paulista, CEP 18618-689 Botucatu, SP, Brazil

4 Faculty of Engineer, UNESP - Univ Estadual Paulista, CEP 15385-000 Ilha Solteira, SP, Brazil

5 São Paulo Forestry Institute, C.P. 1322, CEP 01059-970 São Paulo, SP, Brazil

of populations of *M. urundeuva* only from Paraguay and Argentina. Given the low number of available microsatellite loci and that Brazil is the center of origin of the species [2], it is of great importance that a new variable set of microsatellite loci be developed from the Brazilian populations of *M. urundeuva*, since the reduction of natural populations has led to a loss of genetic diversity that could be associated of adaptation to specific environments of occurrence of tree species [12].

With the development of new sequencing techniques, the characterization of microsatellite loci has become faster and more efficient, allowing the identification of thousands of loci only in one sequencing run. In this way, the aim of this work was to develop microsatellite loci for *M. urundeuva* from a population of Brazil, for the purpose of the implementation of genetic conservation research for this species.

Materials and methods

Sampling and DNA extraction

Thirty samples of *M. urundeuva* obtained from progenies of a natural population from Itarumã, Goiás State, Brazil (18º44′S, 51º13′W; Brazilian savanna biome) were used. The genomic DNA was isolated from fresh leaves, using CTAB protocol [16]. The genomic DNA was quantified in NanoDrop ND-1000 Spectrophotometer (NanoDrop Products, DE, USA) and its integrity was verified in 1% agarose gels in running with TBE (1X), at constant voltage of 5V/cm.

Identification, design of primers and validation of microsatellite loci

Species-specific microsatellite primers were developed based on next-generation sequencing (MiSeq Sequencing System, Illumina). For the library construction it was used Nextera DNA Library Preparation kit which uses transposome to tagment genomic DNA and ligation of adapters for sequencing. For this library, a pool of ten individuals was used to ensure variability of the genome of the species. The MiSeq Reagent v2 500 cycles, 2 × 250 pb, with an estimated 7.5 Gb of data, was used 50% of flow cell capacity in a paired-end run. After sequencing, it was used the software SSR_pipeline [17], an automated package capable of quality filtering, alignment of paired-end reads (contigs) and searching of microsatellite loci. The default parameters for searching of microsatellite loci were used, with motifs from two to six nucleotides and 40 bp flanking regions for the design of the primers. The design of primers were conducted with BatchPrimer3 v1.0 [18]. For the amplification test, each sample was prepared in a final volume of 10 µL, containing: 5 µL GoTaq® Colorless Master Mix (Promega), 0.3 µL of forward primer (2 pmol), 0.3 µL of reverse primer (8 pmol) and 0.3 µL of fluorescent primer with M13 (8 pmol, 6-FAM, VIC, PET, or NED, Applied Biosystems, as the protocol described by Schuelke [19]), 1 µL genomic DNA (approximately 50 ng). The following amplification program was used: 96 °C for 2 min, 35 cycles of 96 °C for 30 s, specific melting temperature of each primer pair (T_m) for 1 min and 30 s, 72 °C for 1 min and 30 s, followed by 12 cycles of 96 °C for 30 s, 53 °C for 1 min and 30 s, 72 °C for 1 min and 30 s, with final extension at 72 °C for 20 min (Table 1). The final PCR product was subjected to capillary electrophoresis on the ABI3130xl Genetic Analyzer (Applied Biosystems) with GeneScan 500 LIZ (Applied Biosystems). The genotypes were read using the GeneMapper v.5.0 software (Applied Biosystems).

Data analysis

The number of alleles per locus (*A*), observed heterozygosity (H_o) and expected heterozygosity (H_e) were estimated for each locus using the software CERVUS v.3.0.7 [20]. Deviations from Hardy–Weinberg (HWE) and also linkage disequilibrium were estimated using Genepop software v.4.5.1 [21]. Micro-Checker v.2.2.3 [22] was used to detect the possibility of occurrence of null alleles.

Results and discussion

The new DNA sequencing methods provide rapid screening and development of microsatellite loci. Several studies point to a high number of loci identified from next-generation sequencing [23–25], thus facilitating the development of new markers and their use in population genetic studies. In this work, more than 8.5 million pairs of reads were generated in a single run, from which it was possible to identify more than 50,000 dinucleotide loci, 24 trinucleotides, 9 tetranucleotides, 10 pentanucleotides and 5 hexanucleotides. From these, on average, 65% of the sequences had enough flanking sequence for designing of the primers, except for the dinucleotide loci where this was only 36% (Table 2).

Twenty microsatellite loci were selected for validation, but only 15 of them amplified in the expected size, all of which were polymorphic (Supplemental Figure S1). Eighteen–thirty individuals were genotyped for each locus. The number of alleles ranged from 3 to 16, with an average of 7.73. The expected (H_e) and observed (H_o) heterozygosity ranged from 0.246 to 0.902 and from 0.103 to 0.867, respectively. Several loci, such as *Aro03, Aro04, Aro05, Aro06, Aro11, Aro15* and *Aro19* loci deviated significantly from the HWE, after Bonferroni sequential correction (Table 3), possible related to the population itself, which is composed of progenies derived from a natural population,

Table 1 Characteristics of the 15 microsatellite loci developed for *Myracrodruon urundeuva*

Locus	GenBank ID	Primer sequence 5′–3′	Repeat motif	T_m (° C)	Allele size range (pb)
Aro02	KY511278	F[a]: TAAAATTGCCCTCGAGTTGA R: TCATCCGTTCGATTCCTCTA	AT(13)	54	193–215
Aro03	KY511281	F[a]: TTCCACATAAGCGTTCCTTC R: CAATCATCAAGCAATGAAACA	TG(12)	55	177–215
Aro04	KY511279	F[a]: TTGTTAGAGAGCGCGAGAAT R: ACAAGAACAGCCAACGAGAG	TC(17)	59	200–224
Aro05	KY511280	F[a]: CAAATTATTGGGCTGCAACT R: GTAGCTCCCACTTGCCATTA	AT(12)	54	214–262
Aro06	KY511277	F[a]: GTCACTGAAAGAGCCCCAAC R: GCGCCAAAGAATTTTTGTAA	TC(12)	59	201–219
Aro10	KY511282	F[a]: AAAAAGTGCAATGTTTTGAGG R: TGCAACTTCCATCCACTGTA	TTC(12)	59	184–212
Aro11	KY511283	F: CGGGGTTTGCATTTTTACTT R[a]: CTGAACGAATTGAATTTCTGG	TCT(12)	54	203–215
Aro12	KY511284	F[a]: TGTCCATGTAGGGCACATTA R: TGCCCCTACTTACAACCAAA	TAT(10)	54	196–220
Aro14	KY511285	F[a]: CATGCCAAACCTTAGCAACT R: CCCATTTTGGTCCTTATCCT	TAT(10)	55	230–239
Aro15	KY511286	F[a]: GACCATGGATTGACCTCTTG R: AGGTGATTGGAAGGTTTTTG	TTA(13)	59	169–223
Aro16	KY511287	F[a]: GGAGCCCCAGAGAGTAAAAG R: CTTCACGATCAGGATCGAAT	AGAA(7)	54	208–222
Aro17	KY511288	F[a]: CTTTCATGGACACCCCTCTT R: GGAGCCCCAGAGAGTAAAAG	TTTC(7)	59	196–208
Aro18	KY511289	F[a]: TGACACTGCATCCGTAAGTG R: CTGCCTGAAATTTGGAAAGA	TATC(7)	59	253–285
Aro19	KY511290	F: GCGTTAATCTCACTGCACAA R[a]: AATTTTTCGCCTGTTTGCTT	AACA(7)	59	222–242
Aro20	KY511291	F[a]: TTGTCTTGGTTCGAATCCTT R: TTGTTAGGGATTTCAAAGACT	TATC(8)	54	166–208

T_m Melting temperature

[a]M13-tailed primer fluorescent labeled, following Schuelke [19]

Table 2 Summary of Illumina paired-end sequence data; it includes non- and perfect motifs di-, tri-, tetra-, penta- and hexanucleotides for *Myracrodruon urundeuva*

Total number of reads	17,330.264				
Read average length bp	251				
All contigs	74,712.86				
Motif	Di-	Tri-	Tetra-	Penta-	Hexa-
Number of contigs with microsatellite	50.426	24.009	9.763	10.466	5.467
Number of contigs with flanking sequence	18.267	15.766	6.282	7.253	3.387

and also by the occurrence of null alleles, except for the locus *Aro05*. No significant linkage disequilibrium was detected between any pair of loci. Thus, the set of loci developed herein may be useful in analyses on diversity and genetic structure, kinship estimation, mating system and gene flow, information that are crucial for conservation and genetic improvement of the species.

Table 3 Results of the 15 microsatellite loci developed for *M. urundeuva*

Locus	N	*A*	H_o	H_e	PIC	HWE	Null alleles
Aro02	29	10	0.690	0.844	0.808	ns	–
Aro03	21	11	0.524	0.869	0.831	*	*
Aro04	28	9	0.429	0.748	0.678	*	*
Aro05	27	16	0.815	0.902	0.876	*	–
Aro06	30	7	0.567	0.765	0.716	*	*
Aro10	30	9	0.867	0.802	0.767	ns	–
Aro11	30	5	0.167	0.543	0.496	*	*
Aro12	28	6	0.571	0.645	0.584	ns	–
Aro14	27	4	0.481	0.511	0.452	ns	–
Aro15	30	11	0.667	0.859	0.829	*	*
Aro16	25	5	0.520	0.682	0.616	ns	–
Aro17	30	4	0.567	0.642	0.573	ns	–
Aro18	28	9	0.679	0.781	0.739	ns	–
Aro19	29	3	0.103	0.246	0.220	*	*
Aro20	18	7	0.722	0.746	0.686	ns	–

N Number of individuals genotyped, *A* number of alleles, H_o observed heterozigosity, H_e expected heterosigosity, *PIC* polymorphism information content, *HWE* Hardy–Weinberg deviation, *ns* not significant

*Significant

In a previous study conducted by Caetano [14], nine microsatellite loci developed for *M. urundeuva*, analyzing three natural populations only of Argentina and Paraguay, excluding any Brazilian population, revealed an average of 8.1 alleles per locus with genotyping performed for up to 47 individuals per locus. In this study, despite the use of only one population and reduced number of individuals genotyped per locus compared to Caetano [14], greater genetic diversity was detected. Thus, these results support the usefulness of these loci in studies of population genetics, contributing to the genetic characterization of the forest fragments of *M. urundeuva* and helping in decision making for the genetic conservation of this species of important ecological and economical role.

Acknowledgments The authors thank all the individuals who assisted in the collection of samples. This investigation was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo; Grant Number for DCLS 2013/24503-1).

Compliance with ethical standards

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

References

1. Gaino APSC, Silva AM, Moraes MA et al (2010) Understanding the effects of isolation on seed and pollen flow, spatial genetic structure and effective population size of the dioecious tropical tree *Myracrodruon urundeuva*. Conserv Genet 11:1631–1643
2. Santin DA, Leitão Filho HF (1991) Restabelecimento e revisão taxonômica do gênero *Myracrodruon* Freire Allemão (Anacardiaceae). Rev Bras Bot 14:133–145
3. Gaino APSC, Moraes MLT, Moreira JP et al (2011) Mating system in *Myracrodruon urundeuva* (Anarcadiaceae): implications for conservation genetics. Rev Bras Bot 34(4):545–551
4. Viana G et al (2014) *Myracrodruon urundeuva* Allemão, a Brazilian medicinal species, presents neuroprotective effects on a Parkinson's disease model, in rats. Eur Neuropsychopharmacol 24(2):230–231
5. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (1992) Lista oficial de espécies da flora brasileira ameaçadas de extinção. Diário Oficial, Brasília
6. Carvalho PER (1994) Espécies florestais brasileiras: recomendações silviculturais, potencialidade e uso da madeira. Empresa Brasileira de Pesquisa Agropecuária, Colombo
7. Moraes MLT, Kageyama PY, Sebbenn AM (2004) Correlated mating in dioecious tropical tree species. *Myracrodruon urundeuva* Fr All. For Genet 11(4):53–59
8. Moraes MLT, Kageyama PY, Sebbenn AM (2005) Diversidade e estrutura espacial em duas populações de *Myracrodruon urundeuva* Fr. All. sob diferentes condições antrópicas. Rev Árvore 29(2):281–289
9. Reis AMM, Grattapaglia D (2004) RAPD variation in a germplasm collection of *Myracrodruon urundeuva* (Anacardiaceae), an endangered tropical tree: recommendations for conservation. Genet Resour Crop Evol. https://doi.org/10.1023/B:GRES.0000024159.48644.e6
10. Santos CAF, Oliveira VR, Kiill LHP, Sá IIS (2007) Variabilidade genética, com base em marcadores RAPD, de três espécies arbóreas ameaçadas de extinção no semi-árido brasileiro. Sci For (74):37–44
11. Freitas MLM, Sebbenn AM, Moraes MLT, Lemos EGM (2004) Mating system of a population of *Myracrodruon urundeuva* F.F. & M.F. Allemão using the fAFLP molecular marker. Genet Mol Biol. https://doi.org/10.1590/S1415-47572004000300019
12. Freitas MLM, Aukar APA, Sebbenn AM, Moraes MLT, Lemos EGM (2005) Variabilidade genética intrapopulacional em

Myracrodruon urundeuva Fr. All. por marcador AFLP. Sci For 68:21–28
13. Caetano S, Silveira P, Spichiger R, Naciri-Graven Y (2005) Identification of microsatellite markers in a neotropical seasonally dry Forest tree, *Astronium urundeuva* (Anacardiaceae). Mol Ecol Notes 5(1):21–23
14. Caetano S, Prado D, Pennington RT et al (2008) The history of seasonally dry tropical forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). Mol Ecol 13(17):3147–3159
15. Viegas MP, Silva CLSP, Moreira JP et al (2011) Diversidade genética e tamanho efetivo de duas populações de *Myracrodruon urundeuva* Fr. All., sob conservação ex situ. Rev Árvore 35(4):769–779
16. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15
17. Miller MP, Knaus BJ, Mullins TD, Haig SM (2013) SSR_pipelin: a bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data. J Hered 104:881–885
18. You FM, Huo N, Gu YQ et al (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinform*. https://doi.org/10.1186/1471-2105-9-253. http://probes.pw.usda.gov/batchprimer3/. Accessed Nov 2016
19. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234
20. Kalinowsk ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. https://doi.org/10.1111/j.1365-294x.2007.03089.x
21. Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resour 8:103–106
22. van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538
23. Duan C-x, Li D-d, Sun S-l et al (2014) Rapid development of microsatellite markers for *Callosobruchus chinensis* using Illumina paired-end sequencing. PLoS ONE. https://doi.org/10.1371/journal.pone.0095458
24. Yang T, Fang L, Zhang X et al (2015) High-throughput development of SSR markers from pea (*Pisum sativum* L.) based on next generation sequencing of a Purified Chinese commercial variety. PLoS ONE. https://doi.org/10.1371/journal.pone.0139775
25. Zhou T, Li Z-H, Bai G-Q et al (2016) Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia* Oliver (Aceraceae). Molecules. https://doi.org/10.3390/molecules21030166