Anais da Academia Brasileira de Ciências (2019) 91(1): e20180420
(Annals of the Brazilian Academy of Sciences)
Printed version ISSN 0001-3765 / Online version ISSN 1678-2690
http://dx.doi.org/10.1590/0001-3765201920180420
www.scielo.br/aabc | www.fb.com/aabcjournal

Discovery and characterization of SSR markers in *Eugenia uniflora* L. (Myrtaceae) using low coverage genome sequencing

DEISE S. SARZI[1,2], BEATRIZ JUSTOLIN[1], CLARIANE DA SILVA[1], RAFAEL P.M. LEMOS[1,3] and VALDIR MARCOS STEFENON[1,3]

[1]Universidade Federal do Pampa, Núcleo de Ecologia Molecular e Micropropagação de Plantas, Rodovia BR 290, Campus Universitário, 97300-000 São Gabriel, RS, Brazil
[2]Universidade Federal Rural do Rio de Janeiro, Rodovia BR 465, Km 07, s/n, Zona Rural, 23890-000 Seropédica, RJ, Brazil
[3]Universidade Federal do Pampa Programa de Pós-graduação em Ciências Biológicas, Rodovia BR 290, Campus Universitário, 97300-000 São Gabriel, RS, Brazil

Manuscript received on April 27, 2018; accepted for publication on June 3, 2018

How to cite: SARZI DS, JUSTOLIN B, SILVA C, LEMOS RPM AND STEFENON VM. 2019. Discovery and characterization of SSR markers in Eugenia uniflora L. (Myrtaceae) using low coverage genome sequencing . An Acad Bras Cienc 91: e20180420. DOI DOI.

Abstract: Eugenia uniflora L. (Myrtaceae) is a tree species widely distributed in South America suffering the effects of the exploitation of natural populations. In this study, we employed low coverage sequencing of the E. uniflora genome for mining of SSR markers. The de novo assembly generated 2,601 contigs with an average length of 1139 bp and spans 3.15 Mb. A total of 76 dimer, 33 trimer and two compound SSR loci were identified. Twelve selected SSR loci were employed to genotype 30 individuals from two natural populations. A total of 73 alleles were detected (mean A= 6.1) were observed, the mean effective number of alleles was Ae = 3.91, mean H_O was 0.23 and mean H_E was 0.70). The mean Wright's within population fixation index was F_{IS} = 0.66 and significant deviation of HWE was observed in all loci, except one. The F_{ST} between populations equaled 0.27. The levels of genetic diversity and structure estimated with these 12 SSR markers are in accordance with data from genetics studies performed on other tree species of the Pampa biome, presenting moderate to high polymorphism and may be employed in studies of species conservation measures and breeding programs.

Key words: next generation sequencing, pitanga, population genetics, SSR, Surinam-cherry.

INTRODUCTION

Eugenia uniflora L. (2*n* = 22) is a tree species of the Myrtaceae family, native to the Cerrado, Atlantic Forest and Pampa biomes in Brazil, with economic and ecological importance. It has been employed in the pharmaceutical and cosmetic industries with attested anti-inflammatory functions (Auricchio and Bacchi 2003, Costa et al. 2010). Traditionally, the infusion of its leaves is used against gastrointestinal illnesses, while its fruits are consumed fresh and as juice and ice cream (Lederman et al. 1992, Ferreira et al. 1987). This species is also used in the environmental recovery of degraded areas and is an important feed source to bees (Silva and Pinheiro 2007), while its wood is widely used by populations of rural areas for heating residences

Correspondence to: Valdir Marcos Stefenon
E-mail valdirstefenon@unipampa.edu.br
ORCid: http://orcid.org/0000-0003-1091-700X

and manufacturing poles for fencing (Costella et al. 2013). Currently, there are few established orchards to economic use of this species (Almeida et al. 2012).

Based on flow cytometry analysis, *E. uniflora* has a predicted haploid genome of only 0.251 pg DNA and about 244.99 Mb (Costa et al. 2008). With the advent of the next generation sequencing (NGS) platforms, drafting such small genomes becomes an attractive initiative towards generating genomic information of huge importance for biotechnological exploitation, conservation and breeding of non-model tree species. NGS platforms are quite useful for generating low coverage genome sequencing data. With a relatively reduced cost, this strategy enabled the discovery of novel repetitive elements in barley genome (Wicker et al. 2008), the identification of homolog genes among Dipteran species (Rasmussen and Noor 2009), characterization of the whole plastidial genome of a milkeweed species (Straub et al. 2011) and the discovery of genomic SSR molecular markers (Staton et al. 2015).

In this study we report the discovery of a large set of SSR loci based on low coverage genome sequencing data, and the characterization of 12 genomic SSR markers for *E. uniflora*. The characterized markers presented moderate to high polymorphism when employed for genotyping adult individuals from two Pampean populations of *E. uniflora* and will allow accessing genetic diversity of natural populations to better understand population dynamics, to plan reliable conservation measures and to advance breeding programs for this species.

MATERIAL AND METHODS

SAMPLING AND DNA EXTRACTION

Total genomic DNA was isolated from healthy leaves of one single adult plant of *Eugenia uniflora* L. (Myrtaceae) collected in a natural population within the Pampa biome in southern Brazil (30º20'05.00"S, 54º21'44.00"W). A voucher of the collected individual was deposited in the Herbarium Bruno Edgar Irgang (HBEI) of the Universidade Federal do Pampa, Campus São Gabriel (voucher HBEI1150). Total DNA was isolated with the DNeasy® Plant mini kit (Qiagen), following the manufacturers' instructions. The quality and the amount of the isolated DNA were evaluated on a NanoVue™ Plus Spectrophotometer (GE Healthcare) and through electrophoresis on 1.0% agarose gel.

NGS SEQUENCING AND *DE NOVO* ASSEMBLY

Total genomic DNA was sheared in fragments of about 300 bp using Biorruptor® (Thermo Fisher Scientific) and the genomic libraries were built using the IonChef® (Thermo Fisher Scientific) system following the manufacturers' specifications. DNA fragments were sequenced on Ion 314™ microchip using the Ion Torrent Personal Genome Machine (Thermo Fisher Scientific) and the Ion PGM™ 200 Sequencing Kit following the manufacturers' specifications. After sequencing, the sequence reads were filtered within the PGM software, removing low quality and polyclonal sequences. All PGM filtered data were exported as a Fastq file that was used for the subsequent bioinformatics analysis.

The Fastq filtered sequences obtained from the PGM software were used for a *de novo* assembly of *E. uniflora* sequences using SPAdes 3.09 (Bankevich et al. 2012), generating contigs with a minimal size of 1,000 bp.

DISCOVERY AND CHARACTERIZATION OF SSR MARKERS

The software SSRLocator (Maia et al. 2008) was used to find di- and tri- nucleotide repeat motifs in the obtained contigs. The default parameters of SSRLocator were employed to identify SSR loci with a minimum of 6 repetitions. Primers for the identified SSR loci were designed using the Primer3 software (Untergasser et al. 2012), searching for

alleles with size ranging from 90 to 280 bp. All contigs containing SSR loci were deposited in the GenBank (ID numbers are listed in Table I and Supplementary Material – Table SI).

Potentially amplifiable SSR loci identified were tested for amplification *in silico* using the software SPCR (Cao et al. 2005). *In silico* amplification was performed using the contigs obtained from the present *E. uniflora* sequencing as template DNA (Figure 1). Using this strategy we are able to identify primer pairs that will amplify a single loci within the *E. uniflora* genome within the expected size range and discard primer pairs generating multi-loci amplifications and unfeasible band patterns.

Twelve loci with dimer and trimer motifs that revealed virtual amplification of a single locus with alleles within the expected size range (Table I) were tested in 30 individuals collected from two natural populations of *E. uniflora* located into the Pampa biome, Rio Grande do Sul State, Brazil. Populations SG (n = 12) and AL (n = 18) are about 200 km distant from each other and represent two characteristic forest formations that naturally occur in the Brazilian Pampa (Roesch et al. 2009). Population SG represents a "capão" formation (island of trees within the grassland, Roesch et al. 2009), while population AL characterizes a gallery forest. DNA was isolated from healthy leaves from each sampled plant using the DNeasy® Plant mini kit (Qiagen), following the manufacturers' instructions.

SSR markers were amplified through PCR in a final volume of 25 µL reaction mix, containing about 30 ng of DNA, 0.25 µM of buffer, 0.5 µM of $MgCl_2$, 1U of Taq DNA-Polymerase (Invitrogen®), 0.05 µM of each dNTP, 0.125 µM of each primer and 0.2 µg/µL of BSA. Amplifications were carried out with 95°C for 5 min, annealing temperature ranging from 48ºC to 51.4ºC (see Table I) for 1 min and extension at 72°C for 1 min, for a total of 30 cycles, with a final extension step of 72°C for 20 min. Alleles of each individual were resolved through electrophoresis on 6% polyacrylamide gels. Gels were stained with GelRed® and allele sizing was performed by comparison to a 100 bp ladder.

Total number of alleles (A), effective number of alleles (A_e), observed heterozygosity (H_O), expected heterozygosity (H_E), Wright's within population fixation index [$F_{IS} = (H_E - H_O)/H_E$], and deviation from Hardy-Weinberg equilibrium (HWE) were estimated for each locus in each population and overall. Differentiation between populations was estimated using the AMOVA approach (F_{ST}). All estimations were performed using the software GenAlEx 6.4 (Peakall and Smouse 2006, 2012).

RESULTS

SEQUENCING OUTPUT

The obtained reads from the low coverage genome sequencing yielded around 7.0 Gb of sequences that were used in the *de novo* assembly. After assembling and exclusion of redundant regions, a total of 2,601 contigs were generated with an average length of 1139 bp (N50 length of 1168 bp) and a cumulative length of 3.15 Mb.

DISCOVERY OF SSR LOCI

Using the selected parameters, a total of 76 di-nucleotide repeats, 33 tri-nucleotide repeats and two compound SSR loci (i.e. di- and tri-nucleotide repeats as SSR motif) were identified within the contigs of the present genome draft. After the *in silico* test for amplification, 74 out of the 111 loci were considered viable, presenting amplification of a single locus within the expected range and considered as putative informative SSR markers. Repeat motifs, forward and reverse primers, annealing temperature and GenBank ID number of the loci are listed in Table I (12 characterized SSR markers) and Table SI (62 not characterized SSR loci).

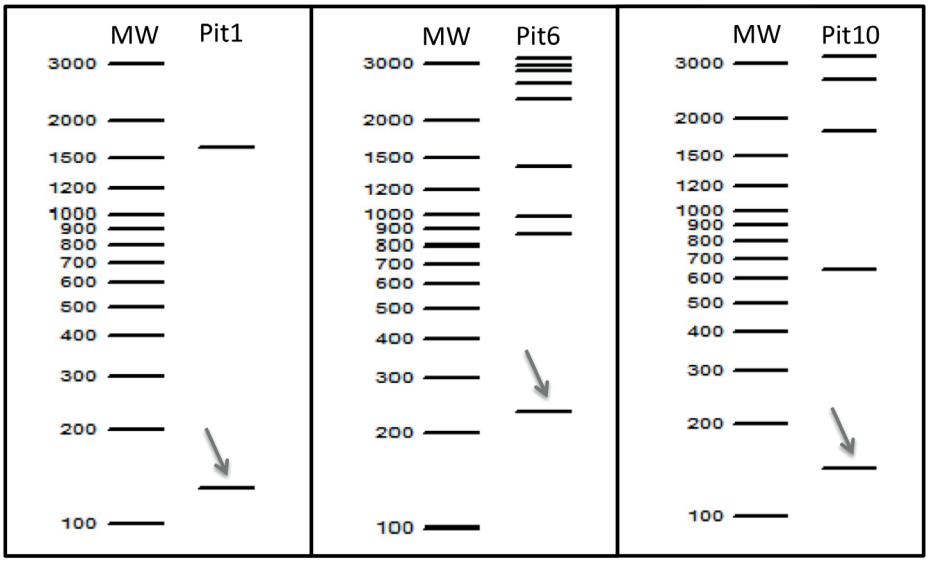

Figure 1 - Virtual electrophoresis gel from *in silico* amplification of three SSR loci discovered using low coverage sequencing of the *Eugenia uniflora* genome and selected for genotyping of 30 individuals of natural populations of the species. MW represents the molecular weight ladder. The arrow indicates the amplified SSR allele. For these three loci, a feasible amplification within the expected size is observed, leading to their selection.

CHARACTERIZATION OF SSR MARKERS

All twelve tested SSR markers were polymorphic in population AL and overall. However, amplification of loci P2, P8 and P13 failed in population SG (Table II). Overall, a total of 73 alleles, ranging from 3 to 12 (mean $A = 6.1$) alleles per locus were observed, while the mean effective number of alleles was $A_e = 3.91$, ranging from 2.27 to 8.49 (Table II). Estimations of H_O ranged from 0.00 to 0.57 (mean $H_O = 0.23$), while H_e measures ranged from 0.57 to 0.91 (mean $H_E = 0.70$). The Wright's within population fixation index (F_{IS}) ranged from 0.34 to 1.00 (mean $F_{IS} = 0.66$). A significant deviation of HWE ($p < 0.05$) was observed in all loci, except for Pit13 (Table II).

At population level, the number of alleles ranged from three to 11 in population AL and from two to seven in population SG. The effective number of alleles ranged from 2.00 to 8.76 (mean $A_e = 3.43$) in population AL and from 1.22 to 5.54 (mean $A_e = 2.67$) in population SG. Estimations of observed heterozygosity ranged from $H_O = 0.00$ to $H_O = 0.50$ (mean $H_O = 0.22$) in population AL and from $H_O = 0.00$ and $H_E = 0.67$ (mean $H_O = 0.27$) in population SG. The expected heterozigosity ranged from $H_E = 0.51$ to $H_E = 0.91$ (mean $H_E = 0.67$) in population AL and from $H_E = 0.19$ to $H_E = 0.86$ (mean $H_E = 0.60$) in population SG. The estimations of Wright's within population fixation index in population AL ranged from $F_{IS} = 0.33$ to $F_{IS} = 1.00$ (mean $F_{IS} = 0.66$), while in population SG, it ranged from $F_{IS} = 0.19$ to $F_{IS} = 1.00$ (mean $F_{IS} = 0.60$). Eleven out of the 12 loci presented significant deviation of HWE in population AL. In population SG, four out of the eight tested loci presented significant deviation of HWE. All estimations overall and for each population are summarized in Table II. The AMOVA approach revealed statistically significant ($p < 0.001$) differentiation between populations, $F_{ST} = 0.27$ (Table III).

DISCUSSION

The use of low coverage whole genome sequencing has proved to be useful to generate SSR markers

TABLE I
Characterization of 12 SSR markers for *Eugenia uniflora* including primers sequence (forward and reverse), repeat motif, annealing temperature (T_a), length of the sequenced fragment, and GenBank accession number (GenBank ID).

Locus	Primer sequence (5'→ 3')	Repeat motif	T_a (ºC)	Sequenced product size	GenBank ID
Pit1	AAACCAAATCACATGCTAA ATGTCTGTGAGGAGTAGACA	$(AAC)_7$	51.4ºC	129bp	KT768349
Pit2	TTTTAAAATAACGTCAAACC TCAAGTTTTAATTGGATTGT	$(GAA)_6$	48ºC	244bp	KT873889
Pit3	CATATTTTCCAATGTCAACT TCCACTCTGTCTATGCTAAT	$(GAG)_6$	48ºC	280bp	KT873890
Pit6	GCAAAAGCAAACATTTA CATTAGTTTCTGCTATCCAG	$(AG)_{12}$	48ºC	235bp	KT873892
Pit7	AATTGTATTGACAGAATTGG TGAAAAAGACCGTAAAATAG	$(AAT)_8$	48ºC	142bp	KT873893
Pit8	TTGACAATTCTTAGCTTCAT TTTTCTCGTATTTGAATGAT	$(GA)_6$	48ºC	129bp	KT873894
Pit9	GTAACTTTCAAAAACGAAAA GACTATGGACAAACTTGAGA	$(TTG)_6$	48ºC	263bp	KT873895
Pit10	CTCAAATTTTGTTTAGCAAT TATATTTGGACTCTGACCTG	$(AT)_6$	48ºC	144bp	KT873896
Pit12	ATTCCTAACAAAATTGGAAC GAGATAGAGCATGAGACAGA	$(AT)_8$	48ºC	242bp	KT873897
Pit13	CTTGCCCTTAGGGTTT AATGACACCATGAGTAAGAT	$(AAG)_6$	48ºC	228bp	KT873898
Pit20	CTCTCAACTCAACCACG ACTGCTTCTCTCCATACAC	$(CAC)_6$	48ºC	159bp	KT873904
Pit25	ATCTATAGGTGTGAAACGTG TCTTGGTGATTTCTATTTGT	$(AG)_6$	48ºC	280bp	KT873907

for different hardwood species, although the proportion of identified polymorphic loci with feasible interpretation of the alleles is relatively low (Khodwekar et al. 2015). In this study, 74 out of 111 SSR loci were selected based on their *in silico* single locus amplification with feasible banding pattern. Using low coverage whole genome sequencing approach for the development of SSR markers, Khodwekar et al. (2015) obtained only seven polymorphic markers out of 96 SSR loci identified in *Acer saccharum* and Owusu et al. (2013) obtained 14 polymorphic SSR markers out of 144 primer pairs tested in *Gleditsia triacanthos.* All 12 SSR markers validated in this study presented polymorphism in the genotyped individuals.

In comparison to the values of expected and observed heterozygosities (H_E and H_O, respectively) summarized by Nybom (2004) for plant species according to their life traits, the SSR markers we developed for *E. uniflora* presented overall estimations of H_E (mean H_E = 0.70, ranging from 0.57 to 0.91) within the range determined for long-lived perennial species (mean H_E = 0.68), widespread (mean H_E = 0.62), with mixed breeding system (mean H_E = 0.60), species of the early successional status (mean H_E = 0.46), and ingested seed dispersal (mean H_E = 0.73). On the other hand, estimations of H_O (mean H_O = 0.23, ranging from 0.00 to 0.57) were lower than the summarized data for long-lived perennial species (mean H_O = 0.63), widespread (mean H_O = 0.57), species of the early successional status (mean H_O = 0.39), and ingested seed dispersal species (mean H_O = 0.72). Ferreira-Ramos et al. (2008) characterized seven SSR

TABLE II

Genetic parameters estimated for *Eugenia uniflora* based on 12 SSR markers characterized in this study, overall populations and at population level. Estimations include the number of samples (*N*), number of allele per locus (*A*), effective allele number (A_e), observed (H_O) and expected (H_E) heterozigosities, Wright's within population fixation index (F_{IS}), and statistical significance of the deviation from Hardy-Weinberg equilibrium (HWE).

Population	Locus	*N*	*A*	A_e	H_O	H_E	F_{IS}	HWE
Overall	Pit1	30	12.0	7.00	0.57	0.87	0.34	***
	Pit2	16	3.0	2.51	0.25	0.62	0.58	***
	Pit3	29	4.0	2.78	0.00	0.65	1.00	***
	Pit6	30	4.0	2.27	0.30	0.57	0.46	**
	Pit7	19	11.0	8.49	0.32	0.91	0.64	***
	Pit8	18	8.0	5.23	0.11	0.83	0.86	***
	Pit9	28	3.0	2.42	0.07	0.59	0.88	***
	Pit10	30	8.0	6.10	0.30	0.85	0.64	***
	Pit12	30	9.0	2.32	0.27	0.58	0.53	***
	Pit13	13	4.0	2.50	0.39	0.63	0.36	ns
	Pit20	30	4.0	2.78	0.20	0.65	0.69	***
	Pit25	24	3.0	2.55	0.04	0.62	0.93	***
	Mean	24.75	6.1	3.91	0.23	0.70	0.66	- -
AL	Pit1	18	10.0	4.56	0.50	0.80	0.36	***
	Pit2	16	3.0	2.51	0.25	0.62	0.58	***
	Pit3	18	4.0	2.00	0.00	0.51	1.00	***
	Pit6	18	4.0	2.27	0.33	0.57	0.40	*
	Pit7	17	11.0	8.76	0.29	0.91	0.67	***
	Pit8	18	8.0	5.23	0.11	0.83	0.86	***
	Pit9	18	3.0	2.91	0.11	0.67	0.83	***
	Pit10	18	4.0	3.07	0.17	0.69	0.75	***
	Pit12	18	7.0	2.00	0.33	0.51	0.33	***
	Pit13	13	4.0	2.50	0.39	0.62	0.36	ns
	Pit20	18	3.0	2.84	0.11	0.67	0.83	***
	Pit25	18	3.0	2.56	0.06	0.63	0.91	***
	Mean	18.0	3.0	3.43	0.22	0.67	0.66	- -
SG[§]	Pit1	12	7.0	5.54	0.67	0.86	0.19	ns
	Pit3	11	3.0	1.75	0.00	0.45	1.00	***
	Pit6	12	2.0	1.88	0.25	0.49	0.47	ns
	Pit7	2	3.0	2.67	0.50	0.83	0.20	ns
	Pit9	10	2.0	1.22	0.00	0.19	1.00	**
	Pit10	12	5.0	3.79	0.50	0.77	0.32	**
	Pit12	12	5.0	2.57	0.17	0.64	0.73	**
	Pit20	12	4.0	2.59	0.33	0.64	0.46	ns
	Pit25	6	2.0	2.00	0.00	0.55	1.00	*
	Mean	9.9	3.7	2.67	0.27	0.60	0.60	- -

[§] Estimations for Pit2, Pit8 and Pit13 are not presented for population SG because amplification failed in all individual for these loci.
Statistical significance: *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; ns: not significant.

TABLE III
Summary of the analysis of molecular variance (AMOVA) for all populations, based on 12 microsatellite markers.

Source	df	SS	MS	Est. Var.	%
Among Pops	1	44.558	44.558	1.335	27%
Among Individuals	28	170.875	6.103	2.476	50%
Within Individuals	30	34.500	1.150	1.150	23%
Total	59	249.933		4.962	100%

F_{ST} = 0.27 ($p < 0.001$).

markers for *E. uniflora* and reported estimations of H_E ranging from 0.71 to 0.94 (mean H_E = 0.82), H_O ranging from 0.00 to 0.80 (mean H_O = 0.42), A ranging from five to 22 (mean A = 10.8), and F_{IS} ranging from -0.008 to 1.00 (mean F_{IS} = 0.478) genotyping 84 individuals from three populations naturally occurring in the Atlantic Forest.

Just few investigations about population genetics of tree species growing in the Brazilian Pampa have been reported. These studies reported low levels of genetic diversity and high levels of inbreeding in Pampean populations of *Schinus molle* (Lemos et al. 2015) and *Luehea divaricata* (Nagel et al. 2015). In addition, Stefenon et al. (2016) showed that this fact may have led to reduction of population fitness in these species. Thus, the comparatively lower estimations of genetic parameters obtained with the 12 SSR markers validated in this study likely are characteristics of the isolated small forest formations found in the Brazilian Pampa and reflects a trend for different tree species.

The SSR markers validated in this study are important tools that can be employed for identifying genetic control of key biotechnological and horticultural traits, for characterizing the genetic diversity and structure of the natural remnants, and will enable the wide application of marker-assisted and genomic selection that may promote the establishment of commercial orchards with improved cultivars of the species. Based on the results of this study, it is reasonable to speculate that we may obtain a large number of informative molecular markers among the 62 SSR loci we discovered for *E. uniflora* through low coverage whole genome sequencing and did not characterize in this study.

ACKNOWLEDGMENTS

We would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support and scholarships, and PROPESQ/UNIPAMPA (Edital AGP/2016) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Process 442995/2014-8) for funding the research. We specially thank to Prof. Dr. Luiz Fernando Würdig Roesch, Prof. Dr. Filipe de Carvalho Victoria and Prof. Dr. Victor Pyrlo by technical support for sequencing and data analysis.

AUTHOR CONTRIBUTIONS

DSS: sampling, sequencing, genotyping, data analysis and manuscript writing, BJ: data analysis, CS: data analysis, RPML: sampling and genotyping, VMS: study design, data analysis, funding and manuscript writing.

REFERENCES

ALMEIDA DJ, FARIA MV AND SILVA PR. 2012. Effect of forest fragmentation on microsporogenesis and pollen viability in *Eugenia uniflora*, a tree native to the Atlantic Forest. Genet Mol Res 11: 4245-4255.

AURICCHIO MT AND BACCHI EM. 2003. Folhas de *Eugenia uniflora* L (pitanga): Revisão. Rev Inst Adolfo Lutz 62: 55-61.

BANKEVICH A ET AL. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19: 455-477.

CAO Y, WANG L, XU K, KOU C, ZHANG Y, WEI G, HE J, WANG Y AND ZHAO L. 2005. Information theory-based algorithm for in silico prediction of PCR products with whoke genomic sequences as templates. BMC Bioinformatics 6: 190.

COSTA IR, DORNELAS MC AND FORNI-MARTINS ER. 2008. Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. Plant Syst Evol 276: 209-217.

COSTA DP, FILHO EGA, SILVA LMA, SANTOS SC, SILVA MRR, SERAPHIN JC AND FERRI PH. 2010. Influence of fruits biotypes on the chemical composition and antifungal activity of the essential oils of *Eugenia uniflora* leaves. J Braz Chem Soc 21: 851-858.

COSTELLA E, GARCIA LSC, CORNELEO NS, SCHÜNEMANN AL AND STEFENON VM. 2013. Anthropogenic use of gallery forests in the Brazilian Pampa. Biological Sciences 35: 211-217.

FERREIRA FR, FERREIRA SAN AND CARVALHO JEU. 1987. Espécies frutíferas pouco exploradas, com potencial econômico e social para o Brasil. Rev Bras Fruticultura 9: 11-22.

FERREIRA-RAMOS R, GUERRIERI A, KLAUS A, ROSSI A, GUIDUGLI MC, MESTRINER MA, MARTINEZ CA AND ALZATE-MARIN AL. 2008. Genetic diversity assessment for *Eugenia uniflora* L., E. *pyriformis* Cambess., *E. brasiliensis* Lam. and E. *francavilleana* O. Berg neotropical tree species (Myrtaceae) with heterologous SSR markers. Genet Resour Crop Evol 61: 267-272.

KHODWEKAR S, STATON M, COGGESHALL MV, CARLSON JE AND GAILING O. 2015. Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). Ann For Res 58: 193-204.

LEDERMAN IE, BEZERRA JEF AND CALADO G. 1992. A pitangueira em Pernambuco. IPA 19: 20.

LEMOS RPM, D'OLIVEIRA CB AND STEFENON VM. 2015. Genetic structure and internal gene flow in populations of *Schinus molle* (Anacardiaceae) in the Brazilian Pampa. Tree Genet Genomes 11: 75.

MAIA LCD, PALMIERI DA, SOUZA VQD, KOPP MM, CARVALHO FIFD AND OLIVEIRA ACD. 2008. SSR Locator: Tool for Simple Sequence Repeat discovery integrated with primer design and PCR simulation. Int J Plant Genomics 1-9.

NAGEL JC, CECONI DE, POLETTO I AND STEFENON VM. 2015. Historical gene flow within and among populations of *Luehea divaricata* in the Brazilian Pampa. Genetica 143: 317-1329.

NYBOM H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13: 1143-1155.

OWUSU SA, STATON M, JENNINGS TN, SCHLARBAUM S, COGGESHALL MV, ROMERO-SEVERSON J, CARLSON JE AND GAILING O. 2013. Development of genomic microsatellites in *Gleditsia triacanthos* (Fabaceae) using Illumina Sequencing. Applic Plant Sc 1: 1300050.

PEAKALL R AND SMOUSE PE. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288-295.

PEAKALL R AND SMOUSE PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28: 2537-2539.

RASMUSSEN DA AND NOOR MAF. 2009. What can you do with 0.1x genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). BMC Genomics 10: 382.

ROESCH LFW, VIEIRA FCB, PEREIRA VA, SCHU¨NEMANN AL, TEIXEIRA IF, SENNA AJT AND STEFENON VM. 2009. The Brazilian Pampa: a fragile biome. Diversity 1: 182-198.

SILVA ALG AND PINHEIRO MCB. 2007. Biologia floral e da polinização de quatro espécies de Eugenia L. (Myrtaceae). Acta Bot Bras 21: 235-247.

STATON M ET AL. 2015. Preliminary genomic characterization of ten hardwood tree species from multiplexed low coverage whole genome sequencing. Plos One 10: e0145031.

STEFENON VM, NAGEL JC, CECONI DE AND POLETTO I. 2016. Evidences of genetic bottleneck and fitness decline in *Luehea divaricata* populations from southern Brazil. Silva Fennica 50: 1566.

STRAUB S, FISHBEIN M, LIVSHULTZ T, FOSTER Z, PARKS M, WEITEMIER K, CRONN RC AND LISTON A. 2011. Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC Genomics 12: 211.

UNTERGASSER A, CUTCUTACHE I, KORESSAAR T, YE J, FAIRCLOTH BC, REMM M AND ROZEN SG. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res 40: 1-12.

WICKER T, NARECHANIA A, SABOT F, STEIN J, VU GT, GRANER A, WARE D AND STEIN N. 2008. Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. BMC Genomics 9: 518.

SUPPLEMENTARY MATERIAL

Table SI - List of SSR loci discovered for Eugenia uniflora using low coverage sequencing strategy and selected after in silico amplification. Sequencing was performed using an Ion PGM® platform. Table includes primers sequence (forward and reverse), repeat motif, and GenBank accession number (GenBank ID).