Plant Genetic Resources: Characterization and Utilization

cambridge.org/pgr

Development of nuclear microsatellite markers in Yerba mate (*Ilex paraguariensis* A. St. Hil.) from whole-genome sequence data

Carolina Tassano[1], Rodrigo A. Olano[2], Paola Gaiero[1], Magdalena Vaio[1] and Pablo R. Speranza[1]

[1]Department of Plant Biology, Faculty of Agronomy, UdelaR, Montevideo, Uruguay and [2]Department of Environmental Systems, Faculty of Agronomy, UdelaR, Montevideo, Uruguay

Short Communication

Cite this article: Tassano C, Olano RA, Gaiero P, Vaio M, Speranza PR (2023). Development of nuclear microsatellite markers in Yerba mate (*Ilex paraguariensis* A. St. Hil.) from whole-genome sequence data. *Plant Genetic Resources: Characterization and Utilization* **21**, 384–387. https://doi.org/10.1017/S1479262123000758

Received: 23 December 2022
Revised: 20 September 2023
Accepted: 21 September 2023
First published online: 27 October 2023

Keywords:
Genetic diversity; highly informative nuclear markers; multiplexing; population genetics; SSR markers

Corresponding author:
Paola Gaiero;
Email: pgaiero@fagro.edu.uy

Abstract

Ilex paraguariensis A. St.-Hil. (*yerba mate*) (Aquifoliaceae Bercht. & J. Presl) is a plant species with great economic and cultural importance because its leaves are processed and ground to make infusions like *mate* or *tereré*. The species is distributed in a continuous area that includes Southern Brazil and part of Paraguay and Argentina. Uruguay represents the Southern distribution limit of the species, where small populations can be found as part of ravine forests. Although there are previous reports of molecular markers for this and other species in the genus, the available markers were not informative enough to represent the intra- and interpopulation genetic diversity in marginal Uruguayan populations. In this study, we developed highly informative polymorphic microsatellite markers to be used in genetic studies in *I. paraguariensis*. Markers were identified in contigs from the genome sequence of two individuals and then tested for amplification and polymorphism content in a diverse panel. Markers which passed these filters were tested on populations from Uruguay. They detected higher diversity within populations (in terms of number of alleles and heterozygosity) than previously reported, and levels of heterozygosity similar to those reported for two Brazilian populations. This subset of seven markers were successfully multiplexed, substantially reducing the costs of the analysis. Combined with previously reported nuclear and plastid markers, they can be used to evaluate the genetic diversity of rear-edge populations, identify genotypes for paternity studies and provide relevant information for the conservation and management of germplasm.

Introduction

Ilex paraguariensis (*yerba mate*) is a plant species with great economic and socio-cultural importance, because of the infusions made with its leaves (*mate* and *tereré*). *I. paraguariensis* is a perennial subtropical tree, distributed in southern Brazil, part of Paraguay and Argentina. Uruguay represents its Southern distribution limit, where small populations are found in ravine forests (Grela, 2004; Hernández, 2019).

Microsatellite markers developed for *I. paraguariensis* (Pereira *et al.*, 2013) detected regional differentiation but a subset of those markers showed reduced genetic variation in populations from Uruguay (Cascales *et al.*, 2014), which, being marginal populations, may show high differentiation and alleles not found in the central area of distribution (Hampe and Petit, 2005). Species-specific plastidic microsatellite markers were mostly monomorphic or showed very low polymorphism (Hernández, 2019). Therefore, microsatellite markers specifically designed to maximise the representation of diversity in marginal populations, combined with those previously reported, will be useful to evaluate genetic diversity of rear-edge populations, for paternity studies and germplasm conservation and management. In this study, we developed polymorphic species-specific microsatellite markers to implement genetic studies of marginal populations in *I. paraguariensis*.

Experimental

Plant material

To characterise the markers, we used a diverse panel from 12 populations of *I. paraguariensis* from Uruguay and Paraguay (one individual per population, Figure S1). To characterise a subset of these markers at the population level, a sample of 15 individuals from three populations was used (Figure S1). Leaves were collected and dried in silica gel. DNA extraction was performed using a standard 2X CTAB protocol (Doyle and Doyle, 1987).

Sequences

Intact genomic DNA (>1.0 µg) from one individual plant was used for library prep and low pass whole-genome sequencing (DNBseq Illumina platform, 150-bp reads, pair-end sequencing) at BGI Genomics (Hong Kong). A total of 1.41 Gbp of sequences were assembled into contigs using SOAPdenovo2 (Luo *et al.*, 2012). From this dataset, only marker *Ip100.4* (Table 1) met the selection criteria. The rest of the markers were developed from the whole-genome sequence assembly for *I. paraguariensis* deposited in GenBank (Sosa and Modenutti, 2021). In both cases, microsatellite-like nuclear sequences were identified with Phobos 3.3.11 and primers complementary to their flanking regions were designed using Primer3 (Rozen and Skaletsky, 2000), both in Geneious 9.0 (Kearse *et al.*, 2012). Sequences containing perfect repeats of at least 15 units were selected for primer design.

Primer design

Primers were designed to obtain two sets of product sizes, 100–200 bp and 250–300 bp. The target annealing temperature was set to 60 °C for all primers. We selected primers without repetitive sequences or neighbouring microsatellites within the flanking regions. A total of 40 primer pairs (Table S1) were synthesised by the Custom DNA oligosynthesis service at Macrogen, South Korea (https://dna.macrogen.com/). Following Ge *et al.* (2014), forward primers were extended with one of the following sequences complementary to oligonucleotides labelled with FAM, VIC, NED and PET, respectively: 5'-AATACAACGCGAT CGACTCC-3'; 5'-AATCCCCACACAAACACACC-3'; 5'-TCCCC TTTCAAACCTAATGG-3'; 5'-TGATCTTGAGAAGGCATCCA-3'.

Amplification

Amplifications were performed in a Vertity 96-well thermal cycler (Applied Biosystems™) and products were run in a ABI3500 XL sequencer (Applied Biosystems™). PCR cycling conditions consisted of an initial denaturation at 95 °C for 15 min; 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C −53 °C (touchdown) for 90 s, and extension at 72 °C for 30 s; and a final extension cycle at 60 °C for 30 min. For the population analysis, the seven most informative markers were combined in a multiplex reaction (Table 1). PCR multiplex amplifications contained 10 ng of genomic DNA, 0.75 µM of each forward primer and 3 µM of each reverse primer (for product sizes 100–200 bp) or 1 µM of each forward primer and 4 µM of each reverse primer (for product sizes 250–300 bp), 1.25 µl 2X Platinum Multiplex PCR Master Mix and 0.3 mL GC Enhancer (Applied Biosystems™), 10X primer mix and ultrapure water to a final volume of 8.3 µl. Analyses were performed at Genexa (https://www.genexa.com.uy/).

Table 1. Characteristics of microsatellite markers developed for *Ilex paraguariensis* from genomic data, including repeat motif and observed allele size range

Locus	Repeat motif	Allele size range (bp)	*A*	H_o	H_e	Locus	Repeat motif	Allele size range (bp)	*A*	H_o	H_e
Ip100.1	$(CT)_{19}$	92–174	5	0.330	0.640	***Ip200.1***	$(AG)_{24}$	241–269	7	0.200	0.670
Ip100.2*	$(CT)_{20}$	130–162	5	0.200	0.780	***Ip200.3****	$(AG)_{25}$	241–277	6	0.670	0.730
Ip100.3	$(AG)_{15}$	134–164	5	0.270	0.640	***Ip200.4***	$(AT)_{23}$	250–282	5	0.270	0.640
Ip100.4	$(AG)_{21}$	122–150	7	0.470	0740	***Ip200.5***	$(AG)_{19}$	246–262	4	0.530	0.610
Ip100.5	$(AT)_{17}$	119–147	5	0.330	0.580	***Ip200.6***	$(AG)_{25}$	244–280	8	0.670	0.640
Ip100.6*	$(AT)_{19}$	117–141	6	0.470	0.820	***Ip200.7***	$(AT)_{18}$	276–302	6	0.470	0.710
Ip100.7	$(AG)_{17}$	114–166	6	0.200	0.750	***Ip200.8****	$(CT)_{23}$	258–301	13	0.270	0.780
Ip100.8	$(AG)_{25}$	137–165	5	0.400	0.640	***Ip200.9***	$(AT)_{23}$	255–281	4	0.620	0.570
Ip100.9	$(AT)_{15}$	156–214	5	0.070	0.700	***Ip200.11***	$(AG)_{23}$	258–301	4	0.140	0.580
Ip100.10	$(AT)_{16}$	132–174	5	0.270	0.720	***Ip200.12***	$(AG)_{19}$	282–304	5	0.200	0.600
Ip100.11*	$(AT)_{16}$	167–215	6	0.270	0.760	***Ip200.13***	$(AT)_{22}$	276–310	9	0.530	0.830
Ip100.12	$(AAT)_{18}$	164–202	6	0.470	0.660	***Ip200.16***	$(AG)_{20}$	288–300	5	0.330	0.720
Ip100.13	$(AG)_{16}$	161–187	5	0.870	0.740	***Ip200.17****	$(AG)_{22}$	248–302	5	0.870	0.800
Ip100.14	$(AG)_{23}$	190–208	4	0.200	0.660	***Ip200.20***	$(AG)_{21}$	308–328	6	0.290	0.760
Ip100.15	$(CT)_{24}$	163–183	3	0.070	0.610						
Ip100.16	$(AG)_{26}$	179–199	4	0.270	0.590						
Ip100.17	$(CT)_{24}$	168–188	4	0.200	0.640						
Ip100.18	$(AG)_{20}$	196–226	6	0.800	0.730						
Ip100.19	$(AG)_{21}$	194–208	4	0.360	0.670						
Ip100.20*	$(AG)_{26}$	144–182	5	0.140	0.820						

A, number of alleles; H_o, observed heterozygosity, H_e, expected heterozygosity.
*Locus used in multiplex amplifications. Genbank accession number OP946888 for *Ip100.4* marker contig. The rest of the markers were designed from the Genbank genome sequence GCA_905181385.1. Forward and reverse primers sequences provided in Table S1. Annealing temperature for all primer pairs was 60 °C.

https://doi.org/10.1017/S1479262123000758 Published online by Cambridge University Press

Table 2. Genetic diversity properties of seven polymorphic microsatellite markers developed for *Ilex paraguariensis*, characterised in three populations from Uruguay: Demicheli (DM), Gruta de los Helechos (GH) and Tapera de Ayala (TA), with locations depicted in Figure S1. n, sample size for each population

	DM ($n = 14$)					GH ($n = 15$)					TA ($n = 15$)				
Locus	*A*	H_o	H_e	*HWE*		*F*	*A*	H_o	H_e	*HWE*		*F*	*A*	H_o	H_e	*HWE*		*F*
Ip100. 2	3	0.429	0.406	1.238	ns	−0.057	2	0.400	0.391	0.008	ns	−0.023	4	0.667	0.593	10.484	ns	−0.124
Ip100.6	3	0.538	0.577	0.554	ns	0.067	3	0.500	0.622	7.202	ns	0.196	3	0.333	0.531	16.445	***	0.372
Ip100.11	3	0.429	0.523	4.300	ns	0.180	1	0.000	0.000	-	-	-	2	0.600	0.464	1.278	ns	−0.292
Ip100.20	3	0.462	0.577	5.258	ns	0.200	3	0.000	0.611	26.000	***	1.000	2	0.400	0.444	0.150	ns	0.100
Ip200.3	4	0.643	0.740	16.918	**	0.131	6	0.933	0.713	28.163	*	−0.308	5	0.933	0.749	14.225	ns	−0.246
Ip200.8	2	0.214	0.191	0.202	ns	−0.120	4	0.533	0.660	2.523	ns	0.192	4	0.800	0.640	4.289	ns	−0.250
Ip200.17	3	0.500	0.559	7.011	ns	0.105	4	0.667	0.631	2.985	ns	−0.056	3	0.600	0.558	2.792	ns	−0.076
Average	3	0.459	0.510	-	-	0.072	3	0.433	0.518	-	-	0.167	3	0.619	0.569	-	-	−0.074

Parameters detailed for each marker are: *A*, number of alleles; H_o, observed heterozygosity, H_e, expected heterozygosity; *HWE*, X^2 values for the test of Hardy–Weinberg equilibrium; *F*, fixation index. Statistical deviation from HWE is indicated as: ns = not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Data analysis

Electropherograms were analysed individually with PeakScanner 1.0 © (Applied Biosystems, 2006). Data were analysed using GenAlEx 6.5 (Peakall and Smouse, 2006).

Discussion

Almost all markers designed were dinucleotides, with only one trinucleotide (AAT). Of the 40 markers, 34 were successfully amplified and were polymorphic in the 12-plant panel (Table 1). Among the dinucleotide markers, the AG repeats were the most abundant (57.6%), followed by AT (27.3%) and CT (15.2%). Allele numbers ranged between 3 and 13 (mean 5.5). Allele sizes ranged from 92 bp (*Ip100.1*) to 226 bp (*Ip100.18*) and between 241 bp (*Ip200.1* and *Ip200.3*) and 328 bp (*Ip200.20*). Non-overlapping size ranges allowed easy scoring of two loci labelled with the same fluorescent dye.

We used the seven most informative markers for the population analysis and they displayed different levels of polymorphism and frequencies in the three populations. Almost all were polymorphic, with 1 to 6 (mean 3) alleles per locus (Table 2). The level of H_o and H_e ranged from 0 to 0.933 (mean 0.504) and from 0 to 0.749 (mean 0.532), respectively (Table 2). Significant deviations from HWE based on Fisher's exact test ($P < 0.05$) were detected for one locus in population DM, two loci in population GH and one locus in population TA (Table 2). Our markers allowed comparisons among Uruguayan populations, with pairwise population Fst values of 0.325 (DM *vs* GH), 0.322 (GH *vs* TA) and 0.263 (DM *vs* TA) consistent with geographic distances (Figure S1). We detected levels of heterozygosity similar to those reported in two Brazilian populations (Pereira *et al.*, 2013) and to those reported within Uruguayan populations by Cascales *et al.* (2014), with maximum H_e 0.749 *vs* 0.742; mean 0.532 *vs* 0.459, respectively. High levels of fixation of genetic diversity were detected among Uruguayan populations (Table 2). Additionally, because primers had the same Tm and two size ranges, they were successfully multiplexed, substantially reducing costs. Our results show the reliability of the markers presented here and the utility of a subset of seven markers to evaluate genetic diversity and population structure in *Ilex paraguariensis*.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1479262123000758.

Acknowledgements. This research was funded by Comisión Sectorial de Investigación Científica from the University of the Republic (22320200200217UD and 22320200200248UD) and by Agencia Nacional de Investigación e Innovación (ANII, grant number FCE_1_2021_1_166709). In addition, CT and RAO thank ANII and the Comisión Académica de Posgrados from the University of the Republic (CAP-Udelar) for scholarships funding. We also thank Pablo Hernández, Luis Rodriguez and the land owners for collaborating in the field collections and providing plant material.

Competing interests. None.

References

Applied Biosystems (2006) PeakScanner 1.0 ©. Available at https://www.thermofisher.com/order/catalog/product/4381867.

Cascales J, Bracco M, Poggio L and Gottlieb A (2014) Genetic diversity of wild germplasm of "yerba mate" (*Ilex paraguariensis* St. Hil.) from Uruguay. *Genetica* **142**, 563–573.

Doyle J and Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* **19**, 11–15.

Ge C, Cui Y-N, Jing P-Y and Hong X-Y (2014) An alternative suite of universal primers for genotyping in multiplex PCR. *PLoS ONE* **9**, e92826.

Grela I (2004) *Geografía florística de las especies arbóreas de Uruguay: propuesta para la delimitación de dendrofloras* (Tesis de Maestría en Ciencias Biológicas, programa PEDECIBA). Montevideo, Uruguay. Universidad de la República. 103 pp.

Hampe A and Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8**, 461–467.

Hernández P (2019) *Estructuración geográfica de la variabilidad genética de Ilex paraguariensis St. Hil en el Uruguay* (Tesis de Maestría en Facultad de Agronomía). Montevideo, Uruguay. Universidad de la República. 60 pp.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P and Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**, 1647–1649.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung D, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W and Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler, *GigaScience* **1**, 18, https://doi.org/10.1186/2047-217X-1-18.

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–295.

Pereira MF, Ciampi AY, Inglis PW, Souza V and Azevedo VC (2013) Shotgun sequencing for microsatellite identification in *Ilex paraguariensis* (Aquifoliaceae). *Applications in Plant Sciences* **1**, 1200245.

Rozen S and Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In Krawetz S and Misener S (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ: Humana Press, pp. 365–386. Source code available at http://sourceforge.net/projects/primer3/.

Sosa JE and Modenutti C (2021) Direct Submission. BioProject PRJEB36685. University of Buenos Aires. Available at https://www.ncbi.nlm.nih.gov/nuccore/CAJIXV000000000.

https://doi.org/10.1017/S1479262123000758 Published online by Cambridge University Press