Shotgun Sequencing for Microsatellite Identification in Ilex paraguariensis (Aquifoliaceae)

Authors: Pereira, Marlei F., Ciampi, Ana Y., Inglis, Peter W., Souza, Valderês A., and Azevedo, Vânia C. R.

Source: Applications in Plant Sciences, 1(3)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200245

Applications in Plant Sciences 2013 1(3): 1200245

PRIMER NOTE

SHOTGUN SEQUENCING FOR MICROSATELLITE IDENTIFICATION IN *ILEX PARAGUARIENSIS* (AQUIFOLIACEAE)[1]

MARLEI F. PEREIRA[2], ANA Y. CIAMPI[3], PETER W. INGLIS[3], VALDERÊS A. SOUZA[4], AND VÂNIA C. R. AZEVEDO[3,5]

[2]Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Rua Mucuri, Lote 1-A, Aparecida de Goiânia, 74968-755 Goiás, Goiás, Brazil; [3]Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, W5 Norte, CP 02372, 70770-917 Brasília, Distrito Federal, Brazil; and [4]Embrapa Floresta, Estrada da Ribeira 111, CP 319, 83411-000 Colombo, Paraná, Brazil

- *Premise of the study: Ilex paraguariensis* is a native tree species from Brazil, Argentina, and Paraguay that is used in the production of beverages, medicines, and cosmetics. Primers flanking microsatellites were developed to investigate genetic parameters in the species.
- *Methods and Results:* Using microsatellites cloned from an *I. paraguariensis* shotgun genomic library, 25 pairs of primers were designed and synthesized. Levels of polymorphism were evaluated in 24 individuals from two populations. Twenty loci were polymorphic, and an average of 4.8 and 4.5 alleles per locus were detected in the two populations, respectively. The mean observed heterozygosity was lower than the expected heterozygosity (0.54 vs. 0.60), indicating a departure from Hardy–Weinberg equilibrium and suggesting endogamy in both populations.
- *Conclusions:* The reported set of markers is highly informative and constitutes a powerful tool for the development of genetic characterization studies in *I. paraguariensis.*

Key words: Aquifoliaceae; genetic conservation; genetic diversity; *Ilex paraguariensis*; microsatellites; shotgun cloning.

Ilex paraguariensis A. St. Hil., popularly known as erva mate or mate tea, is a tree species belonging to the Aquifoliaceae family and occurring in southern Brazil, Argentina, and Paraguay. This species is dioecious and can reach 30 m in height, but is kept below seven meters in height when managed for leaf extraction (Mazuchowski, 1989). *Ilex paraguariensis* has been used particularly in the production of beverages, but has also been used in medicines and cosmetics (Alikaridis, 1987; Maccari Junior and Mazuchowski, 2000). Although the cultivation of *I. paraguariensis* has been implemented with the aim of addressing a growing market demand, extraction of erva mate still largely occurs from natural populations. Knowledge of the genetic structure of a population may be applied to the conservation and management of a species, and can be acquired through the use of microsatellite, or simple sequence repeat (SSR), molecular markers. Microsatellites represent a powerful tool in determining genetic parameters, such as genetic diversity, paternity analysis, gene flow, and genetic drift (Goldstein and Schlotterer, 2001).

[1]Manuscript received 21 May 2012; revision accepted 19 July 2012.

This publication is a result of the EMBRAPA/NATURA/FUNARBE partnership. The authors thank Natura Inovação e Tecnologia de Produtos Ltda. for financial support. We also thank the Embrapa Florestas group for collecting samples and Mr. Eduardo, from Ervateira Putinguense, and his whole family for allowing us to collect samples inside his property and for help during the collection.

[5]Author for correspondence: vania.azevedo@embrapa.br

doi:10.3732/apps.1200245

METHODS AND RESULTS

A genomic library was constructed for *I. paraguariensis* by random fragmentation of DNA (shotgun) and by sample sequencing for the detection of microsatellite loci. Fragmentation of 30 µg genomic DNA was performed in 500 µL of shearing buffer in a nebulizer (TOPO Shotgun Subcloning Kit; Invitrogen, Carlsbad, California, USA) at 0.7 bar for 40 s, generating fragments between 500 and 4000 bp. The QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) was used to separately recover the largest fragments (between 2000 and 4000 bp) and smaller fragments (between 500 and 2000 bp), which were cloned into the vector PCR4Blunt-TOP (Invitrogen) and transformed into *E. coli* competent cells. Following blue/white colony selection, the plasmids were then extracted using the alkaline lysis miniprep method (Birnboim and Doly, 1979). Positive clones were sequenced in both directions in a reaction volume of 10 µL, containing 100 ng purified plasmid DNA, 3.2 µM M13 forward or M13 reverse primers, and 1 µL BigDye Terminator version 3.1 mix (Applied Biosystems, Foster City, California, USA). Cycle sequencing of clones used a program consisting of an initial denaturing step of 94°C for 2 min, followed by 35 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s, and elongation at 60°C for 4 min. Sequences of approximately 3000 random clones were resolved with an ABI PRISM 3700 sequencer (Applied Biosystems). SSRs were identified using TROLL (Castelo et al., 2002) and the Staden package (Staden, 1996), and primers were designed using Primer3 software (Rozen and Skaletsky, 2000). A total of 1434 (48%) sequences were of high quality and were subjected to reverse-strand sequencing. A total of 98 sequences contained microsatellites, which means that 3.3% of the total clones contained SSR sequences, revealing a high number of microsatellite regions in the *I. paraguariensis* genome. Of those, 46 contained dinucleotide repeats (47%), 18 tri-, 13 tetra-, 13 penta-, six hexa-, and one a heptanucleotide repeat. Forty-one SSR clones possessed adequate flanking regions for primer design. The pairs of primers were tested in PCR to verify amplification efficiency, and the annealing temperature was optimized using 12 adult individuals from two natural populations from the southern region of Brazil (Putinga, Rio Grande do Sul State, and Jaguariaíva, Paraná State; Appendix 1). The PCRs contained 3 ng DNA, 1× reaction buffer (10 mM Tris-HCl [pH 8.3], 50 mM KCl), forward and reverse primers (0.28 µM), 1.5 mM $MgCl_2$, 0.25 mg/mL bovine serum albumin (BSA), 0.25 mM each dNTP, and 1.3 units of *Taq* DNA polymerase (Invitrogen).

Thermal cycling conditions were: denaturation at 94°C for 5 min, then 30 cycles of denaturation at 94°C for 1 min, annealing temperature for 1 min (Table 1), extension at 72°C for 1 min, and a final elongation at 72°C for 15 min. Reaction products were separated on 4% denaturing polyacrylamide gels in 1× TBE buffer and visualized by silver staining. The size of the amplified alleles was estimated in comparison with marker fragments of known size (10-bp ladder; Invitrogen).

Of the 41 pairs of primers analyzed, 20 SSR loci showed polymorphism and five were monomorphic. The remaining primers did not amplify or showed nonspecific bands. Table 1 shows the forward and reverse primer sequence, repeat motif, observed amplified fragment size, annealing temperature (T_a) in °C, and GenBank accession number of the clone sequence. The forward primers of the polymorphic loci were fluorescently labeled and used to analyze 24 adult trees from two natural populations, collected from the Putinga (24 trees) and Jaguariaíva (24 trees) sites. The PCR products were analyzed by electrophoresis in an ABI PRISM 3700 sequencer (Applied Biosystems). The ROX-labeled fluorescent internal size standard used was developed by Brondani and Grattapaglia (2001).

The number of alleles per locus (A), mean observed heterozygosity (H_o), and mean expected heterozygosity (H_e) were calculated for the total number of individuals using Genetic Data Analysis (Lewis and Zaykin, 2001). All loci were individually tested for significant deviations from Hardy–Weinberg equilibrium (HWE). Significant values ($P < 0.05$) of deviation from HWE were detected in 12 and nine loci in population 1 and 2, respectively (Table 2). This result is very common for tree species. Levels of variability detected in the 20 loci were high, with the number of alleles ranging from two to 10. The average expected heterozygosity ($H_e = 0.60$) was higher than the observed ($H_o = 0.54$) for both populations, showing an increase of homozygous genotypes in relation

Table 1. Characteristics of 25 microsatellite markers in *Ilex paraguariensis*.

Locus	Primer sequences (5′–3′)	Fluorescent dye	Repeat motif	Size range (bp)	T_a (°C)	GenBank accession no.
Ipg_01	F: CTTTAACTTTTCGCGGCTTAGA R: GCAAGTGACAAAATACATACGGTC	HEX	$(AC)_{12}(CT)_{13}$	280–340	60	GQ227560
Ipg_02*	F: TTCACCCGAGGGAGTGTCTA R: GGCTTAGCGCATGAGTATGG	—	$(AC)_6$	224	60	GQ227561
Ipg_03	F: TGCCTATGTCTTCTACAATGCTTC R: CATGGGTTTGGTCTCACTAACA	6-FAM	$(ACC)_{10}$	350–380	58	GQ227562
Ipg_06	F: GAGAAACGGCAACAGTGGTC R: CACACCTCTCTACACACCCTCA	HEX	$(AG)_{12}$	240–260	60	GQ227564
Ipg_07	F: CTAGTCGACTCGGCAGTTCC R: TCGACGACGTGTATTTTGTG	6-FAM	$(AG)_{13}$	160–190	58	GQ227565
Ipg_08	F: GAATTGCCTTATATGGGTGGAA R: GTACACATAAATGGGGTTGCCT	HEX	$(AG)_6$	260–290	58	GQ227566
Ipg_10	F: TCTTCTCGATCAAAGGAACCTC R: GAGGAAATTCAGAGGCATCAAC	6-FAM	$(AG)_8$	320–360	60	GQ227567
Ipg_17	F: GGCATCTTCACAGGTCACAA R: TGGTCATTATATGAGCCTATCATTT	6-FAM	$(AT)_6$	320–360	56	GQ227573
Ipg_19	F: TGAACATTGGAATCCTAGACCC R: CCGTATATCCCTAAATGCCAAA	6-FAM	$(GA)_7$	190–195	60	GQ227575
Ipg_21	F: GGTAACGGGTGGTCATCTATGT R: ACGTACCATCATCGGTGAGTTT	6-FAM	$(AT)_8$	275–290	56	GQ227577
Ipg_22	F: AAATCCCGAAGAAGGTGGAG R: TAGACGCTTCCACAACGTCA	HEX	$(AT)_8$	145–155	56	GQ227578
Ipg_23	F: ATTAAGAAGCACGGACATGGAT R: TCATGGATGTTAATGTTGAATGTG	HEX	$(AT)_9$	250–280	62	GQ227579
Ipg_27*	F: GTTTCGAGTTGAGTTGGGATTC R: GTGTCACTCTCACTCGGGTTC	—	$(CA)_7$	340	62	GQ227582
Ipg_28	F: AAATTCCATAGCATCATTTGGG R: TGGGTGTTCATCTAGCCTCTTT	HEX	$(CA)_8$	290–320	56	GQ227583
Ipg_30*	F: TGGGTGCTCTTTTTCTGCTC R: TCATGGGATTGTTGGGAGTT	HEX	$(CTT)_7$	296	56	GQ227585
Ipg_31*	F: TCAAATCTCCAGATATCAGTCTCA R: GCCATTGTTGACTTTTAGTTGC	—	$(GA)_{10}$	180	56	GQ227586
Ipg_33*	F: AAGAGATCTATGATGAGAAACC R: CTCACCTCTCTTTCTCTCTG	—	$(GA)_6$	175	56	GQ227587
Ipg_37	F: TCTGATCCTTGTTTCGTTGAGA R: GCACTTTGCTTTCTCTCGATTC	6-FAM	$(GT)_7(AG)_6$	150–180	56	GQ227590
Ipg_41	F: AACGCGTGGATCTAATCTTCAT R: CAAGCTGCAGAGTGATTTGTGT	6-FAM	$(TC)_{21}$	130–160	62	GQ227593
Ipg_43	F: GCAACTCTCACACAAGCAATTC R: TGGGTTACAGATGACCTGAGTTT	6-FAM	$(TC)_6$	150–180	56	GQ227595
Ipg_44	F: TAGAAGGCCATCCAATTCCA R: TTTCCCATCTCTGTTCTGTGG	HEX	$(TC)_6$	165–180	56	GQ227596
Ipg_46	F: TGATCGTCGTTAACAGCATAAA R: GAGTGTCAACTAAGCTTTACCTAAGAA	6-FAM	$(TC)_7$	160–210	62	GQ227597
Ipg_49	F: ATTGCCATAGATCGAAAGGAGA R: TTTTCTCCCCATTTACTTCATCA	HEX	$(TC)_8$	120–150	58	GQ227598
Ipg_50	F: TTATTCCTCAACATCAGGAGCC R: CATAGGGAAGTGTGTGCATGTG	6-FAM	$(TC)_9$	150–170	56	GQ227599
Ipg_52	F: GCGATGGTGTAATGATTTGAA R: CGAACACCAGATACAGACACG	HEX	$(TG)_6(TA)_6$	140–170	62	GQ227600

Note: T_a = optimal annealing temperature.
* Monomorphic.

TABLE 2. Genetic characterization of polymorphic microsatellite markers in *Ilex paraguariensis*.

	Putinga population ($N = 24$)				Jaguariaíva population ($N = 24$)			
Locus	A	H_e	H_o	f	A	H_e	H_o	f
Ipg_01	7	0.804	0.524	0.355*	7	0.763	0.556	0.278
Ipg_03	5	0.456	0.417	0.089	5	0.714	0.500	0.305*
Ipg_06	5	0.642	0.381	0.413*	5	0.767	0.591	0.234
Ipg_07	5	0.717	0.864	–0.211	5	0.630	0.722	–0.151
Ipg_08	4	0.573	0.167	0.713*	2	0.418	0.571	–0.379
Ipg_10	5	0.677	0.364	0.469*	4	0.502	0.500	0.004
Ipg_17	3	0.574	0.954	–0.690*	2	0.497	0.437	0.125
Ipg_19	6	0.802	0.541	0.330*	6	0.585	0.333	0.437*
Ipg_21	2	0.467	0.700	–0.520*	3	0.576	0.636	–0.107*
Ipg_22	3	0.181	0.190	–0.053	3	0.508	0.476	0.063*
Ipg_23	7	0.694	0.625	0.102	5	0.470	0.333	0.297*
Ipg_28	4	0.384	0.417	–0.087	4	0.473	0.545	–0.159
Ipg_37	4	0.709	0.888	–0.262	3	0.663	0.850	–0.292*
Ipg_41	8	0.816	0.421	0.491*	10	0.852	0.722	0.156
Ipg_43	4	0.550	0.783	–0.437*	4	0.619	0.700	–0.134
Ipg_44	3	0.566	0.833	–0.486*	3	0.635	0.818	–0.297*
Ipg_46	5	0.764	0.400	0.483*	4	0.550	0.211	0.624*
Ipg_49	8	0.712	0.363	0.109	9	0.891	0.813	0.091
Ipg_50	4	0.496	0.524	–0.058	3	0.522	0.133	0.751*
Ipg_52	3	0.369	0.227	0.390*	4	0.385	0.313	0.194
Mean	4.8	0.598	0.543	0.094	4.5	0.601	0.538	0.107

Note: A = number of alleles; f = intrapopulation fixation index; H_e = expected heterozygosity; H_o = observed heterozygosity.

* Significant values ($P < 0.05$) for deviation from HWE.

to that expected under HWE (Table 2), and indicating deviations from random mating. Because self-fertilization is not possible in this dioecious species, occurrence of inbreeding is suggested as the cause of the reduction in heterozygosity, generated by crosses between relatives. This has probably occurred due to the reduction in native forests, combined with their ruthless exploitation, which can cause erosion and genetic drift leading to biparental inbreeding.

CONCLUSIONS

The shotgun cloning and sequencing technique proved to be efficient in the detection of microsatellite sequences in *I. paraguariensis*. The markers developed in this study represent a powerful tool for the generation of population genetic data, allowing rapid and accurate analysis of the current state of the distribution of genetic variability in the fragments of native and planted populations. This forms essential information for the conservation and sustainable management of this species.

LITERATURE CITED

ALIKARIDIS, F. 1987. Natural constituents of *Ilex* species. *Journal of Ethnopharmacology* 20: 121–144.

BIRNBOIM, H. C., AND J. DOLY. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* 7: 1513–1523.

BRONDANI, R. P. V., AND D. GRATTAPAGLIA. 2001. Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. *BioTechniques* 4: 793–800.

CASTELO, A. T., W. MARTINS, AND G. R. GAO. 2002. TROLL: Tandem repeat occurrence locator. *Bioinformatics (Oxford, England)* 18: 634–636.

GOLDSTEIN, B. D., AND C. SCHLOTTERER. 2001. Microsatellites: Evolution and applications. Oxford University Press, Oxford, United Kingdom.

LEWIS, P. O., AND D. ZAYKIN. 2001. Genetic Data Analysis (GDA): Computer program for the analysis of allelic data, Version 1.0 (d16c). Free program distributed by the authors over the internet from http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php [accessed 23 January 2013].

MACCARI JUNIOR, A., AND J. Z. MAZUCHOWSKI. 2000. Produtos alternativos e desenvolvimento da tecnologia industrial na cadeia produtiva da erva-mate. Série PDACT, 1. Câmara Setorial de Cadeia Produtiva da Erva-mate, Curitiba, Paraná, Brazil.

MAZUCHOWSKI, J. Z. 1989. A cultura da erva-mate. Empressa de Assistência Técnica e Extensão Rural (EMATER), Curitiba, Paraná, Brazil.

ROZEN, S., AND H. J. SKALETSKY. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

STADEN, R. 1996. The Staden sequence analysis package. *Molecular Biotechnology* 5: 233–241.

APPENDIX 1. Overview of the collected material of *Ilex paraguariensis* used in this study.

Location	Geographic coordinates	Extracted from	Collection year	Collector	Collection no.	Herbarium no.	N
Putinga, Rio Grande do Sul, Brazil	28°57′17″S, 52°10′35″W	Fresh leaves	2009	AYC, VCRA	P1_09	80367[a]	24
Jaguariaíva, Paraná, Brazil	24°32′17″S, 49°31′22″W	Fresh leaves	2009	VAS	J1_09	80368[a]	24

Note: AYC = Ana Yamaguishi Ciampi; N = number of samples for each population; VAS = Valderês Aparecida Souza; VCRA = Vânia Cristina Rennó Azevedo.

[a] Herbarium material deposited at Embrapa Genetic Resources and Biotechnology Herbarium–CEN; DNA conserved at Embrapa Genetic Resources and Biotechnology DNA Bank.