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
SHOTGUN SEQUENCING FOR MICROSATELLITE IDENTIFICATION IN

**Ilex paraguariensis** (Aquifoliaceae)

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

*Applications in Plant Sciences* 2013 1(3): 1200245

1 Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Rua Mumuri, Lote 1-A, Aparecida de Goiânia, 74968-755 Goiás, Goiás, Brazil; 2 Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, W5 Norte, CP 02372, 70770-917 Brasília, Distrito Federal, Brazil; and 3 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. The authors thank Natura Inovação e Tecnologia de Produtos Ltda. for financial support. We also thank the Embrapa Florestas group for partnership.

**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.

*Applications in Plant Sciences* 2013 1(3): 1200245; http://www.bioone.org/loi/apps © 2013 Botanical Society of America

---

**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 QAquick 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. 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 Skaltsky, 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 MgCl₂, 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 6% 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 Jaguaireta (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 to the population level.

### 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. |
|-------|--------------------------|-----------------|--------------|---------------|---------|---------------------|
| lpg_01 | F: CTCTACCTTTTCGGCGGCTTAGA | HEX | (AC)$_1$(CT)$_1$ | 280–340 | 60 | GQ227560 |
|        | R: GCAAGTTGAGAAAATCATACAGGTGTC | | | | | |
| lpg_02* | F: TTTACCCGAGGGAGTCTCTTACA | — | (AC)$_8$ | 224 | 60 | GQ227561 |
|        | R: GGCTTAGCGGAGGACATATGAG | | | | | |
| lpg_03 | F: TGCTATGGCTTCCTTCAAATGCTTC | 6-FAM | (ACC)$_{10}$ | 350–380 | 58 | GQ227562 |
|        | R: CATGCGTTGCTCTCACAATAAAC | | | | | |
| lpg_06 | F: GAGAAGCCGCAACAGTGTC | HEX | (AG)$_1$ | 240–260 | 60 | GQ227564 |
|        | R: CACACCTCTCTACACACTCTCCA | | | | | |
| lpg_07 | F: CTAGTGCGCTCGCCAGCTTCTCC | 6-FAM | (AG)$_1$ | 160–190 | 58 | GQ227565 |
|        | R: TGACGACGCTGTTATTTTGAGT | | | | | |
| lpg_08 | F: GATTGGCTTTTATGGGCTGAGA | HEX | (AG)$_8$ | 260–290 | 58 | GQ227566 |
|        | R: GGTATCAATAATGGGCTTGGC | | | | | |
| lpg_10 | F: TCTTCCTGCAAAAGGGACTCTT | 6-FAM | (AG)$_1$ | 320–360 | 60 | GQ227567 |
|        | R: GAGGAATACGAGGCCATCAAC | | | | | |
| lpg_17 | F: GGCTATCTTCAGGCTCAA | 6-FAM | (AT)$_3$ | 320–360 | 56 | GQ227573 |
|        | R: TGTCATATATAGTGCCATCTATTT | | | | | |
| lpg_19 | F: TGAACATGGGATCTCGATAGCC | 6-FAM | (GT)$_3$ | 190–195 | 60 | GQ227575 |
|        | R: CGGTATACCTTAAAGGCTCAA | | | | | |
| lpg_21 | F: GTGGAACGGGCTGCTCACTATTG | 6-FAM | (AT)$_3$ | 275–290 | 56 | GQ227577 |
|        | R: ACGTACCATCATCAGGTGAGT | | | | | |
| lpg_22 | F: AAATCCCGGAAGAGGTGAGG | HEX | (AT)$_3$ | 145–155 | 56 | GQ227578 |
|        | R: TAGACCCTCTCCACACAGTCA | | | | | |
| lpg_23 | F: ATTAAGAAGACGACAGCATGATG | HEX | (AT)$_3$ | 250–280 | 62 | GQ227579 |
|        | R: TCAATGGAATTTAAGGATG | | | | | |
| lpg_27* | F: GTTCAGGTAGTTGGGATCTTTC | — | (CA)$_3$ | 340 | 62 | GQ227582 |
|        | R: GTCACCTCTCACTCCGGGTT | | | | | |
| lpg_28 | F: AAATCCCTATAGCATCTTTGGG | HEX | (CA)$_1$ | 290–320 | 56 | GQ227583 |
|        | R: TGCTGGTTCTCATAGCCCTTCTT | | | | | |
| lpg_30* | F: TGGTTGCCTCTTTCCTGGCTC | — | (CT)$_3$ | 296 | 56 | GQ227585 |
|        | R: TCAATGGAATTTAAGGATG | | | | | |
| lpg_31* | F: TCAATCTCCGAGATATCGCTTCA | HEX | (GA)$_1$ | 180 | 56 | GQ227586 |
|        | R: GCCATGCTGATTTATGTTG | | | | | |
| lpg_33* | F: AAGAGATCTGATGATGACAC | — | (GA)$_1$ | 175 | 60 | GQ225878 |
|        | R: CTACACCTTCCTCCCTCTC | | | | | |
| lpg_37 | F: TCTATGCTTGGTTGGTGAGAGA | 6-FAM | (GT)$_3$(AG)$_3$ | 150–180 | 56 | GQ227590 |
|        | R: GACGTGCTTTCTTCTGATCTCA | | | | | |
| lpg_41 | F: AAGCGGGCTGATCTAAATCCTCAT | 6-FAM | (TC)$_3$ | 130–160 | 62 | GQ227593 |
|        | R: CAACTGCGAGTTGTTTGTG | | | | | |
| lpg_44 | F: TAGAGGGGCTGTCATTCTTCA | HEX | (TC)$_3$ | 160–180 | 56 | GQ227596 |
|        | R: TTTTCACTGCTCTGCTGCTG | | | | | |
| lpg_46 | F: TTAGTTCAGCTATCCCATCAGACAA | 6-FAM | (TC)$_3$ | 160–210 | 62 | GQ227597 |
|        | R: GTAGTCCGAGTAACTCATAAAA | | | | | |
| lpg_49 | F: ATTGCACATAGCTGGAAGAGGGA | HEX | (TC)$_3$ | 120–150 | 58 | GQ227598 |
|        | R: TTTTCCCTCATTTCTCATTCA | | | | | |
| lpg_50 | F: ATATTCCCATACAATTAGAGGCC | HEX | (TC)$_3$ | 150–170 | 56 | GQ227599 |
|        | R: CATAGGCGAGTGTGATCAGTG | | | | | |
| lpg_52 | F: GGATGCGTCTATAGGGAAGTAGA | HEX | (TG)$_3$(CA)$_2$ | 140–170 | 62 | GQ227600 |
|        | R: CGAACCACGATACTACAGACG | | | | | |

**Note:** $T_a$ = optimal annealing temperature.

* Monomorphic.

http://www.bioone.org/loi/apps
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 \textit{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

\textbf{Alikareidis, F.} 1987. Natural constituents of \textit{Ilex} species. \textit{Journal of Ethnopharmacology} 20: 121–144.

\textbf{Bröndom, H. C., and J. Doly}. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. \textit{Nucleic Acids Research} 7: 1513–1523.

\textbf{Brondani, R. P. V., and D. Grattaglia}. 2001. Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. \textit{BioTechniques} 7: 493–800.

\textbf{Castello, A. T., W. Martins, and G. R. Gao}. 2002. TROLL: Tandem repeat occurrence locator. \textit{Bioinformatics} (Oxford, England) 18: 634–636.

\textbf{Goldstein, B. D., and C. Schlotterer}. 2001. Microsatellites: Evolution and applications. Oxford University Press, Oxford, United Kingdom.

\textbf{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 \url{http://hydridocton.eeb.uconn.edu/people/plewis/software.php} [accessed 23 January 2013].

\textbf{MacCari Junior, A., and J. Z. Mazuchowski}. 2000. Produtos alternativos e desenvolvimento da tecnologia industrial na cadeia produtiva da erva-mate. Série PDAC, 1. Câmara Setorial de Cadeia Produtiva da Erva-mate, Curitiba, Paraná, Brazil.

\textbf{Mazuchowski, J. Z.} 1989. A cultura da erva-mate. Empresa de Assistência Técnica e Extensão Rural (EMATER), Curitiba, Paraná, Brazil.

\textbf{Rozen, S., and H. J. Skaltsky}. 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.

\textbf{Staden, R.} 1996. The Staden sequence analysis package. \textit{Molecular Biotechnology} 5: 233–241.