



Isolation and Characterization of Fifteen Microsatellite Loci for the Use in Breeding of *Gmelina arborea* Roxb. (Lamiaceae)

Oscar J Rocha ^{*,a}, Dawa Méndez-Álvarez ^b, Fabiana Rojas Parajeles ^b and Olman Murillo Gamboa ^b

^a Department of Biological Sciences, Kent State University, Ohio, 44242, Kent, USA

^b School of Forest Engineering, Instituto Tecnológico de Costa Rica, 159-7050, Cartago, Costa Rica

Abstract: *Gmelina arborea* (melina) is a valuable tree species throughout tropical areas, and there are extensive commercial plantations of this species in Southeast Asia, West Africa, and Latin America. As part of a research program for the genetic improvement and management of *G. arborea* at Instituto Tecnológico de Costa Rica, we developed, validated, and optimized fifteen microsatellite loci. We used 23 clones belonging to five different companies currently using clonal selection to manage their commercial plantations. Our results showed that all fifteen loci were polymorphic and together had 75 alleles (2-7 alleles/locus). We also found that eleven loci showed lower heterozygosity than expected under Hardy-Weinberg equilibrium (HWE). We calculated the genetic similarity among all clone pairs using the number of shared alleles to examine the potential of these loci for clone discrimination. Overall, pairwise similarity among clones ranged from 0.36 to 0.83, and our findings also showed that clones from the same commercial plantation tended to be more similar to each other than to clones from other plantations. These microsatellite loci will contribute toward the characterization of the genetic diversity, the identification of elite clone lines for timber production, and breeding and adequate management of commercial plantations of *G. arborea*.

Keywords: breeding, clonal forestry, gene markers, melina, population genetics

Citation: Rocha, O. J., Méndez-Álvarez, D., Rojas Parajeles, F., Murillo Gamboa, O. (2020). Isolation and Characterization of Fifteen Microsatellite Loci for the Use in Breeding of *Gmelina arborea* Roxb. (Lamiaceae). *Genetic Resources* 1 (2), 23–28. doi: [10.46265/genresj.LLTJ1737](https://doi.org/10.46265/genresj.LLTJ1737).

© Copyright 2020 the Authors.

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Gmelina arborea (melina) is a valuable timber species that grows throughout tropical areas. Native to Southeast Asia and India, it is an important commercial timber species in tropical regions worldwide, particularly in Southeast Asia, West Africa, and South America, where *G. arborea* is grown in large plantations. Its low-density wood is durable and yields reasonable quantities of relatively uniform, stable, and light color pulp (Dvorak, 2004; Wee et al, 2012). The Panel of Experts on Forest Genetic Resources of the Food and Agriculture Organization (FAO) describes *G. arborea* as an important tree

species with high potential and utility (Lauridsen and Kjaer, 2002).

G. arborea is the second most planted timber species in Costa Rica because of its rapid growth rate, easy establishment, high productivity, a wide range of tolerance to site conditions and excellent regrowth capacity (Rosero et al, 2011; Ávila Arias et al, 2015a,b; Vergara et al, 2017). This species was first introduced into Costa Rica in 1966 for pulp production by the local paper company and to serve as a seed source for the establishment of plantations in the Jari Project in Brazil in Eastern Amazonia (Rounda, 1988). This initial introduction consisted of seeds from twenty independent origins, i.e., provenances from different regions throughout its native range in Asia

*Corresponding author: Oscar J Rocha (orocha@kent.edu)

(India, Pakistan, and Bangladesh), and commercial plantations in Africa (Nigeria and Cameroon) and British Honduras (now Belize). These provenances were planted separately in >100 ha blocks, to provide a broad base for genetic improvement. More than 20 years later, seeds from this plantation were collected from healthy trees with desirable phenotypes, initiating its spread throughout the region.

Breeding efforts of *G. arborea* in Costa Rica started in the early '90s, leading to the development of highly productive genetic stock for timber production at a regional scale (Ávila Arias et al, 2014, 2015a,b). The most successful melina breeding programs in the region use clonal propagation to establish their commercial plantation, as this strategy provides a reliable stock of propagules that are easy to produce and plant and results in fast-growing trees and high productivity. Moreover, researchers have used variables such as trade volume and quality of wood and other indicators of each clone line's performance for the selection of the genetic stock to be planted in sites with different soil characteristics, flooding, and land use-history (Ávila Arias et al, 2015a,b).

Here, we describe fifteen microsatellite loci developed to support ongoing breeding programs of melina in Costa Rica using a small number of clones selected for their rapid growth and high productivity. These markers will be used for clone identification and potentially for marker-assisted breeding of *G. arborea*.

Materials and Methods

Development of Microsatellite Markers

The microsatellite markers were developed using the magnetic bead protocol described by Cullings (1992) and Li et al (1997) and modified by Glenn and Schable (2005). Genomic DNA from a sample of five *G. arborea* trees was digested using *HaeIII/PshA1* restriction enzymes (Invitrogen; Carlsbad, CA). Two linkers were added to the digested genomic DNA (M28 5'CTCTTGCTTGAATTCGGACTA 3' and M29 5'pTAGTC-GGAATTCAGCAAGAGCACACA 3') and M28 was used as a primer for subsequent polymerase chain reactions (PCR). Finally, the digested genomic DNA was amplified in multiple PCR reactions and their product concentrated to gain enough DNA for the following bead hybridization process.

Two arbitrary repeat motifs (CA20 and AG17) were selected as probes for the bead hybridization reactions based upon Cardle et al (2000). The short tandem repeat (STR) probes from Integrated DNA Technologies (Coralville, IA, USA) had a biotin label on the 5' end. The STR probes were added to a bead hybridization reaction to select for DNA fragments that contained the repeat motif of the probe. This bead hybridization process aimed to allow the fragments containing repeats to anneal to the biotin-labeled probes. After the hybridization, the selected fragments were isolated from the rest of the genomic DNA using streptavidin-coated

magnetic beads, which bind to the biotin-labeled probes. These fragments were then eluted and re-amplified using the M28 primer in additional PCR reactions. The bead hybridization and PCR pre-amplification processes were repeated one more time to enrich for genomic DNA containing the selected repeats.

After completing the bead hybridization and selection process, the repeat sequences enriched DNA was ligated into a pGEM-T vector from Promega (Madison, WI, USA) to begin the sequencing phase of this protocol. We cloned the vectors into electrocompetent *Escherichia coli* cells. We later plated transformed *E. coli* cells onto selective media containing 0.1 mg/mL ampicillin, 0.05 mg/mL X-Gal, and 1mM IPTG. All positive clones were sequenced on an ABI PRISM 377 DNA Sequencer using universal M13 forward (F) and reverse (R) primers (Schuelke, 2000). The sequencing reactions were standard 20 ml reactions using the ABI PRISM BigDye Terminator sequencing kits (Applied Biosystems, Foster City, CA, USA) and 3.2 pmol of PCR product for the template. Primers for each of the fifteen microsatellite loci were designed from sequences containing multiple copies of the repeated motif and with sufficiently long flanking regions on the 5' and the 3' end of the repeated region pairs using Primer 3.0 software (Rozen and Skaletsky, 2000).

Microsatellite Loci Characterization

All primer pairs were tested for amplification and polymorphism using DNA obtained from 23 promising genotypes (clones) of *G. arborea* belonging to five different privately operated clonal breeding programs. Two ramets from each clone were gathered from a clonal collection maintained in a greenhouse at the Instituto Tecnológico de Costa Rica to validate all alleles by genotyping them separately. As described above (Doyle and Doyle, 1987; Lodhi et al, 1994), total genomic DNA was extracted at the Forest Molecular Genetic Laboratory, in the Forest Innovation Research Center (CIF) at the Instituto Tecnológico de Costa Rica, Cartago, Costa Rica. Copies of these clones are maintained in the mini clonal garden facility and could be made accessible upon request.

PCR Amplification and Fragment Analysis

Polymerase chain reactions were performed in a final volume of 15 μ l, containing approximately 50 ng of genomic DNA, 10 mM Tris buffer, pH 8.0, 10 mM MgCl₂, 0.2 mM dNTPs, 0.4 μ M of each primer, and 1 U of Taq polymerase (Fermentas®) using an Eppendorf® Mastercycler EP thermal cycler. The PCR program used included an initial step of 2 min of denaturation at 94 °C, 30 cycles of 15 s at 94 °C, 15 s at 55 °C and 30 s at 72 °C, and a final extension cycle of 1 min at 72 °C. To genotype each individual, we conducted electrophoresis for fragment separation using a QIAxcel Advanced fragment analyzer from QIAGEN® at Centro de Investigación en Biología Celular y Molecular (CIBCM) at Universidad de Costa Rica. Once

Table 1. Development and optimization of microsatellite markers for *Gmelina arborea*. Locus name, accession number, repeat motifs, oligonucleotide primer sequences, PCR product size range, and annealing temperatures for the fifteen microsatellite loci isolated for *G. arborea*.

Locus name	Accession no.	Repeat motif	Primer sequence	Range of PCR products (bp)	Annealing Temp (°C)
Meldi1	MH249250	CT ₁₂	R: 5'-CCCCACTTTTGATGCTCTCT-3' F: 5'-TAGTTGGTGAAATGAAAATTCGC-3'	113-131	56
Meldi2	MH230082	TC ₉ .GA ₁₅ .GA ₉	R: 5'-TAGTCGCATAAGGAAAAGTTAG-3' F: 5'-CCAATTCTGTGCTGCTAAGG-3'	299-317	55
Meldi4	MH249251	TC ₁₇	R: 5'-ACTCCAGACCTTCTCGCTC-3' F: 5'-CCTTCTTCAAAATCCGATCTTTC-3'	214-237	58
Meldi5	MH249252	GA ₁₀	R: 5'-CAATTCCTTCAGTTATCCAC-3' F: 5'-CAAAATAGATTAAGCCACATC-3'	283-307	50.8
Meldi6	MH249253	TC ₁₂ TC ₉ .TC ₁₀	R: 5'-CCCATATCACCTGCTCTTAG-3' F: 5'-GGATGTCACGATGTCGGATG-3'	292-320	56
Meldi7	MH249255 MH249254	CT ₂₂ ...CT ₁₉	R: 5'-CCTAACAATCATGGAAAAGAG-3' F: 5'-CACTCAGCGTGGCAAAGAG-3'	352-376	54
Meldi10	MH249256	CT ₁₉	R: 5'-CAATTTCTCCCTAGCAAGTTC-3' F: 5'-AGTACATCTACTCAAAGGAGGT-3'	174-210	56
Meldi11	MH249257	GA ₁₂	R: 5'-GCTTGTGGTTTGGGTGACGA-3' F: 5'-AGTGATGCTCCTCATTGAGGA-3'	168-182	57
Meldi11.2	MH249257	TC ₉	R: 5'-TTCATAATGAGCAAAGGGGAC-3' F: 5'-GGCCACTTCAGTATCCACAC-3'	140-160	56
Meldi12	MH249258	TC ₁₅	R: 5'-CCTAGATTAGCCATTTATATTG-3' F: 5'-GGTGGAGCAGAAGAAGAGAGA-3'	206-214	56
Meldi16	MH249260	TG ₂₃	R: 5'-ACCAGGACACCCTCGACAC-3' F: 5'-GTTGAGGAGCAGTTTATAAGAC-3'	213-225	55.2
Meldi17	MH249261	TG ₁₀ .GA ₁₇	R: 5'-GGTTATTATAATCACTTCTGCC-3' F: 5'-CCAATGTAATAGGACAAAATC-3'	153-167	55
Meldi20	MH249263	GT ₈ .GA ₁₇	R: 5'-GGGAAAATCAGATACACTGCC-3' F: 5'-ATCAACCAAGGGGACAAGCC-3'	205-231	57
Meldi21	MH249264	GA ₁₄ ..TG _n	R: 5'-GCATGCAACAGAGAAAAGAAG-3' F: 5'-GATAATCGTACTAAAGTTGAGG-3'	231-262	55
Meldi24	MH249267	CA ₆ ..GA ₁₇	R: 5'-CCTTCCCTAAGTCCACACA-3' F: 5'-TTTTGTCCAAATGAGGCTCC-3'	211-219	53

23 clones were selected. We expected to encounter high genetic diversity among the clones used in this study because they represent a sample taken from collections of *G. arborea* selected by growers because of their performance. Moreover, the plantations where these clones were selected have different soil types, precipitation regimes, and topography.

We found that all clones from two clonal breeding programs clustered together in the dendrogram (PC and AC; Figure 1). However, this is not true for clones from all breeding programs, as clones from the same program may not group in the same cluster. For example, clone T-27 did not cluster together with the other four genotypes in the same program (T-26, T-28, T-29, and T-30). Similarly, clones N-15 and MC-1313 did not group with the other trees from their program. However, clones from the same breeding program tended to group, suggesting that the process of selecting promising clones, based on what breeders

considered desirable phenotypes, varies among breeding programs. Furthermore, this finding also implies that promising clone lines could perform well in a given environment. Therefore, it suggests that the degree of similarity of allelic composition among clones may indicate similarities in their ability to respond to environmental conditions.

Ávila Arias et al (2014) conducted a field trial using different clone lines planted in two locations in southwestern Costa Rica. They found significant differences in diameter at breast height (DBH), commercial height, commercial volume of the trunk over bark, trunk quality, and the volume and quality of the wood among clone lines two years after planting. Their analysis also showed significant genotype by environment interaction in clonal performance, as some accessions grew well in their site of origin but not in other locations. Murillo-Gamboa et al (2016) reported differences in the tolerance to melina's wilt, a critical disease in Costa Rica,

Table 2. Genetic analysis of microsatellite loci on 23 *Gmelina arborea* clones. Observed number of alleles per locus, observed and expected heterozygosity, significance value for test for the deviation from Hardy-Weinberg proportions (χ^2 test), and test for null alleles are listed for each of the fifteen polymorphic loci developed for *G. arborea*.

Locus name	Number of alleles	Heterozygosity		χ^2 test	Test for Null Alleles
		Observed	Expected	P-value	P-value
Meldi1	3	0.435	0.446	ns	†
Meldi2	5	0.348	0.712	***	***
Meldi4	6	0.652	0.671	ns	†
Meldi5	6	0.652	0.750	**	ns
Meldi6	7	0.217	0.774	***	***
Meldi7	7	0.550	0.839	***	**
Meldi10	6	0.435	0.722	***	**
Meldi11	3	0.227	0.404	***	†
Meldi11.2	6	0.500	0.792	*	**
Meldi12	2	0.609	0.423	*	†
Meldi16	3	0.565	0.542	*	†
Meldi17	4	0.348	0.661	***	***
Meldi20	6	0.522	0.705	ns	*
Meldi21	5	0.435	0.746	***	**
Meldi24	6	0.435	0.775	***	***

ns = not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, †A single allele contributed to more than 50% of the observation in this locus. Binomial analysis could not be performed.

among clone lines used in the field trial conducted by Ávila Arias *et al* (2015b,a). These findings indicate that clone selection is biased toward genotypes performing well in particular environments, thus suggesting that genetic markers could play a role in identifying promising genotypes.

In summary, the fifteen polymorphic microsatellite markers we described here have great potential use for the breeding of *G. arborea*, including genotyping the breeding collections, as well as keeping the identity and assessing the purity in clonal gardens. In this respect, there are eleven additional loci available to expand the multilocus genotype of each clonal line (Liao *et al*, 2010) to increase the possibilities for genetic analysis and marker-assisted selection of *G. arborea*.

Acknowledgements

This work was supported by the Fundación CR-USA, the Department of Biological Sciences of Kent State University and Instituto Tecnológico de Costa Rica.

Supplemental data

Supplemental Table 1: Alleles observed in each of the 23 clones of *Gmelina arborea*. Summary of the presence (1) or absence (0) of the 75 alleles found in the fifteen polymorphic microsatellite loci in each clone. This information was used to estimate genetic similarities for all pairwise comparisons among clones

Author contributions

O.J.R. and O.M-B. contributed to the conception and design of the submitted manuscript. O.J.R. isolated

and developed the fifteen microsatellite loci, D.M-A and F.R-P. validated and optimized each of the fifteen polymorphic loci. O.J.R. drafted the manuscript, and all authors contributed the drafting, revision, and final approval of the submitted manuscript.

Conflict of interest statement

The Authors declare no conflict of interest.

References

- Ávila Arias, C., Murillo-Cruz, R., Murillo-Gamboa, O., and Sandoval-Sandoval, C. (2014). Desarrollo juvenil de clones de *Gmelina arborea* Roxb. de dos procedencias, en sitios planos del Pacífico Sur de Costa Rica. *Revista Forestal Mesoamericana Kurú* 12(28), 23–23. (in Spanish). doi: <https://doi.org/10.18845/rfmk.v12i28.2097>
- Ávila Arias, C., Murillo-Cruz, R., and Murillo-Gamboa, O. (2015a). Selección de clones superiores de dos conjuntos genéticos de *Gmelina arborea* en el Pacífico Sur de Costa Rica. *Revista de Ciencias Ambientales* 49(1), 17–35. (in Spanish). doi: <https://doi.org/10.15359/rca.49-1.2>.
- Ávila Arias, C., Murillo-Cruz, R., Murillo-Gamboa, O., and Sandoval-Sandoval, C. (2015b). Interacción genotipo sitio para dos conjuntos clonales de *Gmelina arborea* Roxb., en sitios planos del Pacífico Sur de Costa Rica. *Revista Forestal Mesoamericana Kurú* 12(29), 02–14. url: <https://revistas.tec.ac.cr/index.php/kuru/article/view/2250/2033>.
- Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., and Waugh, R. (2000). Computational and experimental characterization of physically clustered

- simple sequence repeats in plants. *Genetics* 156, 847–854.
- Cullings, K. W. (1992). Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology* 1(4), 233–240. doi: <https://doi.org/10.1111/j.1365-294x.1992.tb00182.x>
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology* 26(3), 297–302. doi: <https://doi.org/10.2307/1932409>
- Doyle, J. J. and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11–15. url: <https://webpages.uncc.edu/~jweller2/pages/BINF8350f2011/BINF8350 Readings/Doyle plantDNAextractCTAB 1987.pdf>.
- Dvorak, W. S. (2004). World view of *Gmelina arborea*: opportunities and challenges. *New Forests* 28(2/3), 111–126. doi: <https://doi.org/10.1023/b:nefo.0000040940.32574.22>.
- Glenn, T. C. and Schable, N. A. (2005). Isolating microsatellite DNA loci. In *Molecular Evolution: Producing the Biochemical Data*, ed. Zimmer, E. A. and Roalson, E. H. (Academic Press), volume 395 of *Methods in Enzymology*, 202–222.
- Lauridsen, E. B. and Kjaer, E. D. (2002). Provenance research in *Gmelina arborea* Linn., Roxb. A summary of results from three decades of research and a discussion of how to use them. *The International Forestry Review* 4(1), 20–29. url: <https://www.jstor.org/stable/43740942>.
- Lega Rounda, F. F. (1988). Estudios de la forma de *Gmelina arborea* (Roxb). Análisis de las plantaciones de Manila, Siquirres. MSc Thesis, CATIE; UCR, Turrialba, Costa Rica (in Spanish). url: <http://hdl.handle.net/11554/4991>
- Li, S.-H., Huang, Y.-J., and Brown, J. L. (1997). Isolation of tetranucleotide microsatellites from the Mexican jay *Aphelocoma ultramarina*. *Molecular Ecology* 6(5), 499–501. doi: <https://doi.org/10.1046/j.1365-294x.1997.00215.x>
- Liao, M., Li, C., Qi, J., and Hong, Y. (2010). Isolation and characterization of 11 polymorphic microsatellite loci for *Gmelina arborea*. *Molecular Ecology Resources* 10, 1106–1108. doi: <https://doi.org/10.1111/j.1755-0998.2010.02916.x>
- Lodhi, M. A., Ye, G.-N., Weeden, N. F., and Reisch, B. I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter* 12(1), 6–13. doi: <https://doi.org/10.1007/bf02668658>
- Murillo-Gamboa, O., Salas-Rodríguez, A., Murillo-Cruz, R., and Ávila Arias, C. (2016). Tasa de avance de la pudrición del tronco en melina *Gmelina arborea* Roxb. y posibilidades de manejo. *Revista Forestal Mesoamericana Kurú* 13(32), 40–40. (in Spanish). doi: <https://doi.org/10.18845/rfmk.v0i0.2551>
- Peakall, R. and Smouse, P. E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6(1), 288–295. doi: <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Rohlf, F. J. (1993). NTSYS-pc : numerical taxonomy and multivariate analysis system.
- Rosero, C., Argout, X., Ruiz, M., and Teran, W. (2011). A drought stress transcriptome profiling as the first genomic resource for white teak - Gamhar - (*Gmelina arborea* Roxb) and related species. *BMC Proceedings* 5 (Suppl 7), P178–P178. doi: <https://doi.org/10.1186/1753-6561-5-s7-p178>
- Rozen, S. and Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols*, ed. Misener, S. and Kravetz, S., (Totowa, NJ: Humana Press), volume 132 of *Methods in Molecular Biology*TM, 365–386.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18(2), 233–234. doi: <https://doi.org/10.1038/72708>
- van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4(3), 535–538. doi: <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Vergara, A., Mercado, T., Jarma, A., and Gatti, K. (2017). Requerimiento Hídrico de *Gmelina arborea* en Etapa de Vivero Bajo Condiciones Controladas (Water Requirement of *Gmelina arborea* on Nursery Stage Under Controlled Conditions). *Floresta e Ambiente* 24, e20150129. (in Spanish). doi: <http://doi.org/10.1590/2179-8087.012915>
- Wee, A. K. S., Li, C., Dvorak, W. S., and Hong, Y. (2012). Genetic diversity in natural populations of *Gmelina arborea*: implications for breeding and conservation. *New Forests* 43(4), 411–428. doi: <https://doi.org/10.1007/s11056-011-9288-2>