Article (refereed)

Hernandez, G.; Buonamici, A.; Walker, K.; Vendramin, G. G.; Navarro, C.; Cavers, S.. 2008 Isolation and characterization of microsatellite markers for Cedrela odorata L. (Meliaceae), a high value neotropical tree. Conservation Genetics, 9, 2. 457-459. doi:10.1007/s10592-007-9334-y

Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree.

Running title (45 chars): Microsatellites for *Cedrela odorata*

Hernandez G[1,2], Buonamici A[3], Walker K[2], Vendramin GG[3], Navarro C[1] and Cavers S[2,*]

[1] Centro Agrónomico Tropical de Investigación y Enseñanza, Cartago, Turrialba 7170, Costa Rica

[2] NERC Centre for Ecology and Hydrology, CEH Edinburgh, Bush Estate, Penicuik, Midlothian EH26 0QB, Scotland, UK

[3] Plant Genetics Institute, Consiglio Nazionale delle Ricerche (CNR), 50019 Sesto Fiorentino (Florence), Italy.

[*]To whom correspondence should be addressed: Dr S Cavers, Centre for Ecology and Hydrology Edinburgh, Bush Estate, Penicuik, Midlothian EH26 0QB, Scotland, UK. Tel. +44 131 445 4343, Fax. +44 131 445 3943, Email scav@ceh.ac.uk

Keywords (5): *Cedrela odorata*, Cedar, Meliaceae, microsatellites

Abstract (82 words)

We describe 9 primers for amplification of microsatellite loci for the Neotropical tree *Cedrela odorata* L. (Meliaceae). Loci were isolated from an enriched library derived from a single DNA sample from a tree in Costa Rica. Levels of polymorphism were determined using samples from a large progeny trial. Across loci, the number of alleles ranged from 14 to 30. Observed heterozygosity levels ranged from 0.61 to 0.88. No linkage disequilibria were detected although some departures from HWE were found, probably due to a Wahlund effect.

Cedrela odorata (Meliaceae), known as Cedro Amargo or Spanish Cedar, is a high-value species of the Mahogany family, widely distributed in the neotropics (Holdridge et al. 1997). Internationally valued for its high quality wood, *C. odorata* has been used for construction, furniture and boat building amongst other things (Cordero and Boshier 2003). As a result, heavy selective logging has severely reduced wild populations, to the point that the species is now threatened at a provenance level (Patiño 1997).

Previous studies have assessed genetic structure using AFLP and universal chloroplast DNA markers (Cavers et al 2003a,b) and quantitative traits (Navarro et al. 2005), thoroughly describing genetic variation in the species at a landscape scale in Mesoamerica. Now, attention is turning to restoration efforts and best practice for sourcing seed. A major progeny trial is now underway in Costa Rica to examine the effects of landscape context on progeny fitness, as recent studies have shown that alteration of the forest surrounding seed trees changes gene flow patterns (e.g. Rocha and Aguilar, 2001), with potentially significant consequences for inbreeding rates in progeny arrays. To enable analysis of mating system variation in these progeny arrays, a set of microsatellites were isolated for *C. odorata*. Loci were optimised for PCR and screened for polymorphism using 487 individuals in 68 families, from 12 populations distributed across Mesoamerica.

DNA was extracted from silica-gel-dried leaf material (DNeasy Plant mini kit, QIAGEN) from a single adult tree of *C. odorata* from Costa Rica. A microsatellite library enriched for di- (AG, GT, AT, GC) and trinucleotide (CAA, ATT, GCC) repeats was constructed following Edwards *et al*. (1996). The enriched DNA was cloned into pGEM-T vector and transformed into JM109 cells. A total of 200 clones were sequenced with an Amersham

MegaBACE 1000 automated sequencer using DYEnamic ET Terminator Sequencing Kit (Amersham Biosciences).

Screening for polymorphism and optimisation of PCR conditions was carried out using leaf material from families from different, widely-separated provenances (Tulum and Xpujil, Mexico; Cedros, Honduras; Hojancha, Costa Rica) to maximise the potential for detection of polymorphism. All microsatellites were amplified using 25.0 ul PCR reactions consisting of the following: 2.0 ul template DNA, 15.4 ul H_2O, 2.0 ul primers, 2.5 ul 10X buffer, 0.5 ul dNTPs, 0.4 ul BSA, 0.2 ul Taq DNA polymerase (New England Biolabs). Reactions were run on a Hybaid MBS thermocycler to the following protocol: 1 min at 94 °C, then 30 cycles of 60 s at 94 °C, 60 s at 55 °C, 60 s at 72 °C and finally 5 min at 72 °C. Optimal PCR reaction conditions for each of the polymorphic pairs were determined by testing a range of annealing temperatures (55.0 -70.3 °C) and different template DNA concentrations (1:5, 1:10, 1:20 and 1:1). In all cases the forward primer was labelled with either IRD 700 or 800 fluorescent label (MWG Biotech). PCR products were then separated on 6% polyacrylamide gel (25 cm), and visualized using a LI-COR 4200 IR2 automated genotyper. PCR products were run out alongside a microSTEP DNA size standard (Microzone Limited) and fragment sizes were scored using SAGA™ software. Numbers of alleles were calculated, and observed and expected heterozygosity, exclusion probability (Cervus v3.0, Marshall et al 1998) and null allele rate were estimated (Microchecker v2.2.3, van Oosterhout et al. 2004). Tests for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium, corrected for multiple comparisons and considering 1 individual per family, were carried out using Genepop v3.3 (Raymond & Rousset 1995). Results for effective numbers of alleles ($N_{A)}$, observed (H_O) and expected heterozygosity (H_E) are reported (Table 1) for the collection as a whole and for a single population (Tikal) to demonstrate the within-population utility of the markers.

In 80% of the sequences a microsatellite motif was detected. However, about 130 sequences were discarded because of unfavourable properties for primer design. In 30 cases primers were designed, using PRIMER 3 software (Rozen & Skalestky 2000). Of these, 9 gave clear, interpretable band patterns and were polymorphic (Table 1). The numbers of alleles per locus ranged from 14 to 30 with levels of observed heterozygosity from 0.61 to 0.88. Null alleles were detected at rates of >0.05 at ced41, ced61a, ced65 and ced131. In all cases, these are most likely due to scoring errors due to stuttering, as indicated by low frequencies of heterozygote genotypes with size differences of a single repeat unit (van Oosterhout et al. 2004). Across the whole dataset, only loci ced2 and ced44 showed no departure from HWE. For some loci, this was contributed to by the presence of null alleles (highest null allele frequency estimated at locus ced61a with 18.1%, Table 1), but most probably reflects a heterozygote deficiency due to combination of samples from several, widely-distributed populations (Wahlund effect). No linkage disequilibria between loci ($P > 0.05$) were observed, suggesting that these loci should be valuable markers for population genetics and parentage analysis for *Cedrela odorata*.

References

Cavers S, Navarro C, Lowe AJ (2003)a A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. Conserv Genet 4:571-580

Cavers S, Navarro C, Lowe AJ (2003)b Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L. (Meliaceae) in Mesoamerica. Molec Ecol 12:1451-1460

Cordero J, Boshier DH (2003) Árboles de Centroamérica: un manual para extesionistas. Oxford Forestry Institute, UK. CATIE, Turrialba, CR.

Holdridge LR, Poveda LJ, Jiménez Q (1997) Árboles de Costa Rica. San José, CR. Centro Científico Tropical. v. 1, p. 269

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molec Ecol 7: 639–655

Navarro C, Cavers S, Pappinen A, Tigerstedt P, Lowe AJ, Merilä J (2005) Contrasting Quantitative Traits and Neutral Genetic Markers for Genetic Resource Assessment of Mesoamerican *Cedrela odorata*. Silvae Genet 54(6): 281-292

Patiño F (1997) Genetic resources of *Swietenia* and *Cedrela* in the Neotropics. United Nations Food and Agriculture Organization. Rome, Italy.

Raymond M, Rousset F (1995) genepop (Version 1.2): Population genetics software for exact tests and ecumenicism. J Hered 86: 248–249

Rocha OJ, Aguilar G (2001). Variaton in the breeding behaviour of the dry forest tree *Enterolobium cyclocarpum* (Guanacaste) in Costa Rica. Am J Bot 88: 1600-1606

Rozen S, Skaletsky H (2000) *Primer3 on the WWW for general users and for biologist programmers.* In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker : software for identifying and correcting genotyping errors in microsatellite data. Molec Ecol Notes 4: 535-538

1 Acknowledgements

2 This work was part of the project SEEDSOURCE, funded by the EC under FP6 INCO

3 (contract 003708). Library development, cloning and sequencing was done by CNR and

4 optimisation at CEH by GH during his M.Sc., sponsored by CATIE and Ministerio de Ciencia

5 y Tecnología.

Table 1. Primer sequences, characterisation and basic descriptive statistics of 9 microsatellite markers isolated from *Cedrela odorata*.

Locus	Array	Primer Sequences (5' - 3')	Allele Size / bp	T (°C)	N	N_A	H_O	H_E	Pr (Ex_1)	Pr (Ex_2)	Null rate	GenBank Accession no.
Ced2	$(GA)_{20}$	**F:**TTTGCTTTGAGAAACCTTGT*	130-170	55	400	19	0.850	0.882	0.614	0.762	0.017	EF413962
		R:AACTTTCGAATTGGTTAAGG			(63)	(12)	(0.841)	(0.842)				
Ced18	$(GA)_{23}$	**F:**CAAAGACCAAGATTTGATGC*	130-150	55	403	19	0.797	0.844	0.545	0.708	0.029	EF413963
		R:ACTATGGGTGGCACAACTAC			(69)	(16)	(0.826)	(0.875)				
Ced41	$(TC)_{18}$	**F:**TCATTCTTGGATCCTGCTAT*	120-160	55.5	451	21	0.745	0.910	0.693	0.819	0.100	EF413964
		R:GTGGGAAAGATTGTGAAGAA			(72)	(12)	(0.792)	(0.839)				
Ced44	$(TG)_{14}(AG)_{17}$	**F:**ACTCCATTAACTGCCATGAA*	180-240	55.5	456	30	0.882	0.931	0.755	0.860	0.028	EF413965
		R:ATTTTCATTCCCTTTTAGCC			(74)	(17)	(0.959)	(0.920)				
Ced54	$(GA)_{15}(AG)_6G(GA)_5$	**F:**GATCTCACCCACTTGAAAAA*	120-160	55	408	28	0.843	0.936	0.772	0.871	0.053	EF413966
		R:GCTCATATTTGAGAGGCATT			(70)	(21)	(0.886)	(0.931)				
Ced61a	$(TG)_{10}$	**F:**CAATCAAACCAAAAATGGAT*	240-270	55.5	428	14	0.605	0.868	0.581	0.737	0.181	EF413967
		R:GCAAATTAACCAGAAAAACG			(71)	(10)	(0.535)	(0.843)				
Ced65	$(GA)_7(CA)_{14}$	**F:**GAGTGAGAAGAAGAATCGTGATAGC*	160-200	55.5	468	17	0.618	0.802	0.456	0.632	0.136	EF413968
		R:GAGGTTCGATCAGGTCTTGG			(76)	(12)	(0.605)	(0.798)				
Ced95	$(CT)_{17}(AC)_{13}$	**F:**ATTTTCATTCCCTTTTAGCC*	190-250	55	389	28	0.918	0.829	0.715	0.834	0.043	EF413969
		R:TTATCATCTCCCTCACTCCA			(63)	(18)	(0.905)	(0.913)				
Ced131	$(CT)_{16}$	**F:**CTCGTAATAATCCCATTCCA*	80-120	55	419	14	0.745	0.829	0.507	0.677	0.053	EF413970
		R:GGAGATATTTTTGGGGTTTT			(66)	(10)	(0.773)	(0.823)				
			Total cumulative exclusion probabilities						0.999	0.999		

* Indicates fluorescently labelled primer. Abbreviations are: number of individuals (*N*), number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities and exclusion probability Pr(Ex - for first parent $_1$ and second parent $_2$). Values in brackets for *N*, N_A, H_O, H_E are single-population estimates from the population Tikal.