Conserv Genet
DOI 10.1007/s10592-006-9274-y

ORIGINAL PAPER

Microsatellite markers for *Caesalpinia echinata* Lam. (Brazilwood), a tree that named a country

Sônia Cristina Oliveira Melo · Fernanda Amato Gaiotto · Fernanda Barbosa Cupertino · Ronan Xavier Corrêa · Alessandra Maria Moreira Reis · Dário Grattapaglia · Rosana Pereira Vianello Brondani

Received: 21 November 2006 / Accepted: 8 December 2006

Abstract *Caesalpinia echinata*, commonly known as Pau-brasil (Brazilwood), the famous tree that named Brazil is native to the Atlantic forest. Men extensively exploited it ever since discovery and colonial times due to its value as a source of red dye. As a consequence, Brazilwood is a threatened species with populations reduced to small forest fragments. Ten polymorphic microsatellite loci were developed from an enriched genomic library. Using fluorescently-labeled primers, a total of 83 alleles were found after analyzing a sample of 44 trees. These high genetic information content markers should allow detailed investigations of mating systems, gene flow, population structure and paternity in natural populations.

Keywords SSR · Atlantic forest · Natural population · Allelic diversity · Conservation

S. C. O. Melo · F. A. Gaiotto (✉) · F. B. Cupertino · R. X. Corrêa
Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Ilheus, BA, Brazil
e-mail: gaiotto@uesc.br

A. M. M. Reis · D. Grattapaglia
Universidade Católica de Brasília, Brasilia, DF, Brazil

D. Grattapaglia
Embrapa Recursos Genéticos e Biotecnologia, Brasilia, DF, Brazil

R. P. V. Brondani
Laboratório de Biotecnologia, Embrapa Arroz e Feijão, Goiania, GO, Brazil

Main text

Caesalpinia echinata is a species of the Leguminosae family native to the Atlantic forest. This red-hued tree received the common name ''Pau-brasil'' by the Portuguese explorers as ''Pau'' is portuguese for ''wood'', and ''brasil'', derived from ''brasa'', means ''ember''. This tree soon became the most valuable source of red dye, an important colonial export item, to a point that it gave Brazil its name. Because of intensive exploitation natural populations of *C. echinata* were reduced to small forest fragments (Lima et al. 2002; Lira et al. 2003). Remnant populations have apparently suffered the constant loss of genetic variability due to genetic drift, inbreeding and reduced gene flow (Lima et al. 2002; Lira et al. 2003). *C. echinata* is currently included in the IUCN Red List of Threatened Species as ''endangered'' (Varty 1998).

Although Brazilwood is considered an important national tree and valued as the best wood for musical instrument fabrication, especially violins (Bueno 2002), very little is known about several aspects of its ecological and genetic features that could aid in devising efficient strategies for its conservation. RAPD markers have been used for a preliminary look at the genetic structure of Brazilwood forest fragments in Brazil (Cardoso et al. 1998) revealing a significant amount of variation among populations in the state of Rio de Janeiro. Maternally inherited chloroplast microsatellites also revealed a major structured pattern with most of the genetic variation found between geographical regions and between populations within regions (Lira et al. 2003). To date, however, no clear picture of the overall genetic variation for the recombining nuclear DNA has been generated due to the lack of adequate, co-dominantly inherited markers. Here

we report the development of a battery of ten highly informative nuclear microsatellite markers for *C. echinata*.

The enriched genomic library was obtained using a protocol described by Brondani et al. (1998). Genomic DNA from an individual tree of *C. echinata* was extracted from leaves according to Ferreira and Grattapaglia (1998). Approximately 50 μg of genomic DNA were digested with *Sau*3AI, and fragments in the desired range of 280–600 bp were recovered by DEAE-cellulose NA-45 membrane (Schleicher and Schuell, NY) via electrophoresis on 2% agarose gel. The fragments were attached to adapters of the *Sau*3AI restriction site. Fragments containing AG/TC repeats were selected by hybridization to biotinylated oligonucleotides complementary to the repetitive sequence, and were recovered by magnetic beads linked to streptavidine. Fragments enriched for simple sequence repeats were amplified by PCR, cloned in plasmid vector pGEM-T (Stratagene, CA) and transformed into chemically competent *E. coli* cells. All recombinant colonies were picked, grown overnight at 37°C, followed by plasmid DNA isolation from individual colonies through an alkaline lysis miniprep procedure. Using universal primers, DNA inserts were sequenced on a 377 automated DNA sequencer (Applied Biosystems, CA) using dye-terminator fluorescent chemistry. Primers pairs flanking the repetitive sequences were designed with the software ''Primer 3.0'' (Lincoln et al. 1991). Twenty-six pairs of primers were designed and screened by silver staining detection in polyacrylamide gels. Ten loci were selected based on allelic hypervariability, low stuttering and ease of interpretation in a sample of eight individual plants. The forward primers for these ten microsatellite markers were 5′ end labeled with HEX, 6-FAM or TET dyes.

Microsatellites were characterized with a sample of 44 adult individuals of *C. echinata* from a natural Atlantic Rain Forest fragment of 200 ha (Serra do Teimoso Reserve: 15°12′S, 39°29′W—Bahia State, Brazil), where were found about 727 different vascular plant species (Amorin et al. 2005). The PCR cocktail (13 μl) contained 30 ng of genomic DNA, 250 μM dNTPs, PCR buffer (10 mM Tris–HCl, 50 mM KCl, 2.25 mM $MgCl_2$ pH 8.3), 2.5 μg/μl BSA (New England Biolabs), 0.2 μM of each primer and 1 unit of *Taq* DNA polymerase (Invitrogen). Amplifications were performed using the following protocol: 96°C for 2 min followed by 30 cycles of 94°C for 1 min, the primer specific annealing temperature T_a (see Table 1) for 1 min, 72°C for 1 min, and a final extension cycle at 72°C for 7 min. Alleles were genotyped on 377

Table 1 Information of 10 loci to *C. echinata*

Locus	Primer pair sequence (5′–3′)	Repeat motif	T_a (°C)	Size range (bp)	A	H_E	H_O	Q	I	GenBank accession no.	HW P-value
CE02	ggCAAggTgAACAgCAAC CATACACACggCCTggTT	$(GA)_{17}$	56	250–290	13	0.834	0.750	0.67661	0.05075	BV682861	0.0208
CE07	ACTAgTCggAACTCgCA CTgTgATTCgTCAATCTT	$(CT)_{25}$	59	90–144	16	0.775	0.818	0.61977	0.06690	BV682862	0.2064
CE09	CTCCTCCTTCTAAgTTgC gTCATgCTCCAAgTTCTC	$(CT)_{24}$	59	140–160	8	0.675	0.985	0.44220	0.16692	BV682863	0.0000
CE11	TCTCTCTCCTCCAACTCC CCTCTTCgTACgTTCATC	$(TC)_{15}$	59	100–140	10	0.714	0.341	0.52099	0.17988	BV682864	0.0000
CE14	CTTTATgggCCAgATCATgT gAgCCTTgAATCCgTCCA	$(CT)_3AT(CT)_3$	59	190–200	4	0.244	0.045	0.12226	0.59579	BV682865	0.0000
CE18	CATgCAAAgCTAATCTgTg gCCATggAgAAgTgTTAC	$(TA)_1CTGTTG\ (TA)_2$	59	178–200	3	0.574	0.568	0.32709	0.25455	BV682866	0.0021
CE19	CAAAgACCAAATTgAgAA AgTgAgAgCATggACTgT	$(AG)_{18}$	59	220–250	8	0.602	0.364	0.41355	0.18554	BV682867	0.0000
CE23	gCTCCgATTATCTCCTCC AAgACCAAAgACgTgAgC	$(AG)_{11}$	63	100–140	5	0.692	0.386	0.45504	0.15442	BV682868	0.0000
CE25	CAggATCCAACgTTAACT ggCCACAATgTATCAgAA	$(AG)_{24}$	59	170–200	9	0.705	0.523	0.50972	0.12043	BV682869	0.0003
CE26	gATCACACACACCACCTCTCT TgCCAAggTggAgTCAAg	$(GA)_{19}$	63	170–190	7	0.476	0.568	0.27012	0.32646	BV682870	0.0745

Primer pair sequence, repeat sequence, annealing temperature (T_a), size ranges, number of alleles/locus (A), expected heterozygosity (H_E), observed heterozygosity (H_O), paternity exclusion probability (Q), probability of genetic identify (I) based on 44 adult individuals genotypes, and Hardy-Weinberg equilibrium probability (HW P-value)

automated DNA sequencer (Applied Biosystems), using Genescan® v.3.1 and Genotyper® v2.5 softwares and ROX-500 Size Standard (Applied Biosystems).

Genetic parameters, as numbers of alleles/locus (A), expected and observed heterozygosity (H_E and H_O, respectively) were estimated using Genetic Data Analysis (GDA) (Lewis and Zaykin 2000). The probability of genetic identity (I) and the paternity exclusion probability (Q) (Weir 1996) were also estimated. Genetic data for all 10 markers were submitted to a Fisher's exact test (10000 bootstraps) to test for Hardy-Weinberg proportions and linkage disequilibrium.

All 10 microsatellite markers were highly polymorphic, producing 83 alleles in total. Number of alleles/locus varied from 3 to 16, with an average of 8.3 (Table 1). Almost all loci, but CE07 and CE26, did not conform to HW expected genotypic frequencies with H_O usually lower than H_E in most loci. However, three loci displayed an excess of heterozygous individuals (CE07, CE09, CE26) (Table 1). As several two-locus linkage disequilibrium tests were deemed significant, combined values of I and Q could not be estimated for all loci (Weir 1996). However, a battery of eight microsatellite markers (CE09, CE11, CE14, CE18, CE19, CE23, CE25, and CE26), found to be in linkage equilibrium, allowed us to estimate partial combined values of I and Q at 0.5×10^{-6} and 97%, respectively.

In conclusion, this first set of microsatellite markers developed for *C. echinata* is able to discriminate individuals and is adequate for parentage testing and genetic population studies. It should significantly increase the possibilities of in-depth investigation of the genetic status of current forest fragments as well as remnant populations in preserved areas. These markers are currently being used to investigate genetic structure and gene flow patterns in natural populations of *Caesalpinia echinata* in southern Bahia (Brazil), in order to elaborate sound strategies for conserving this important icon tree. Furthermore it will be interesting to verify the picture emerging from the analysis of this set of co-dominant markers with that revealed by previous studies (Cardoso et al. 1998; Lira et al. 2003) that suggested a strongly structured pattern of variation among populations based on RAPD and cpSSR markers.

Acknowledgements Research supported by Brazilian National Research council—CNPq, Bahia State Foundation for Research Support—FAPESB, and Biodiversitas Foundation. S.C.O.M and F.B.C. held CNPq fellowships.

References

Amorin AM, Fiaschi P, Jardim JG, Thomas WW, Clifton BC, Carvalho AMV (2005) The vascular plants of a forest fragment in southern Bahia, Brazil. SIDA 21(3):1727–1752

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theor Appl Genet 97:816–827

Bueno E (2002) Nova viagem à terra do pau-brasil. In: Bueno E (ed) Pau-brasil, Axis Mundi, São Paulo, pp 19–38

Cardoso MA, Provan J, Powell W, Ferreira PCG, de Oliveira DE (1998) High genetic differentiation among remnant populations of the endangered *Caesalpinia echinata* Lam. (Leguminosae Caesalpinioideae). Mol Ecol 7:601–608

Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética, 3rd. Embrapa-Cenargen, Brasília 220 p

Lewis PO, Zaykin D (2000) Genetic data analysis: computer program for the analysis of allelic data. http:// alleyn.eeb.uconn.edu/gda. Cited 20 Sept 2006

Lima HC, Lewis GP, Bueno E (2002) Pau-brasil: uma biografia. In: Bueno E (ed) Pau-brasil, Axis Mundi, São Paulo, pp 39–76

Lincoln SE, Daly MJ, Lander ES (1991) '*Primer*' software. Available from E. Lander. Whitehead Institute, Cambridge, MA

Lira CF, Cardoso SRS, Ferreira PCG et al (2003) Long-term population isolation in the endangered tropical tree species *Caesalpinia echinata* Lam. revealed by chloroplast microsatellites. Mol Ecol 12:3219–3225

Varty N (1998) *Caesalpinia echinata*. In: IUCN2006. 2006 IUCN red list of threatened species. Available via http://www.iucnredlist.org of subordinate document. Cited on July15th 2006

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, MA