American Journal of Botany

American Journal of Botany: e268–e270. 2011.

AJB Primer Notes & Protocols in the Plant Sciences

Development and characterization of microsatellite loci in *Pericopsis elata* (Fabaceae) using a cost-efficient approach[1]

Claire Micheneau[2,4], Gilles Dauby[2], Nils Bourland[3], Jean-Louis Doucet[3], and Olivier J. Hardy[2]

[2]Evolutionary Biology and Ecology Unit CP 160/12, Faculté des Sciences, Université Libre de Bruxelles, Av. F. D. Roosevelt 50, B-1050 Brussels, Belgium; and [3]Université de Liège, Gembloux Agro-Bio Tech, Unité de Gestion des Ressources forestières et des Milieux naturels, Laboratoire de Foresterie des Régions tropicales et subtropicales, B-5030 Gembloux, Belgium

- *Premise of the study*: Microsatellite loci were developed in the endangered *Pericopsis elata* using a combination of low-cost procedures.
- *Methods and Results*: Microsatellite isolation was performed simultaneously on three distinct species through a newly available procedure that associates multiplex microsatellite enrichment and next-generation sequencing, allowing the rapid and low-cost development of microsatellite-enriched libraries through the use of a 1/32nd GS-FLX plate. Genotyping using M13-like labeling in multiplexed reactions allowed additional cost savings. From 72 primers selected for initial screening, 21 positively amplified *P. elata*, and 11 showed polymorphism with two to 11 alleles per locus and a mean value of 5.4 alleles per locus.
- *Conclusions*: These microsatellite loci will be useful to further investigate the level of genetic variation within and between natural populations of *P. elata* in Africa.

Key words: Fabaceae; microsatellites; next-generation sequencing; *Pericopsis elata*; tropical timber tree; universal fluorescent-labeled primer.

Pericopsis elata (Harms) Meeuwen, also known as afrormosia, African teak, or assamela, is a deciduous tree that produces one of the most valued tropical timbers of Africa. However, habitat degradation, together with increasing pressures for international exports and regeneration problems, has led to important declines in some of its populations (Dickson et al., 2005). Consequently, the species is now considered endangered and is included in both the IUCN Red List of Threatened Species and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) Appendix II (IUCN, 2010). Ongoing efforts, including several research programs, are currently developed to improve the management of the species in exploited forests (Bourland et al., 2010), and the acquisition of relevant information about genetic diversity and population genetic structure is highly awaited.

In the current study, we described how the combination of low-cost procedures may lead to significant cost savings for the development of 11 polymorphic microsatellite loci in *P. elata*. These markers were further used to describe the level of genetic variation in two populations in Central Africa.

[1]Manuscript received 10 February 2011; revision accepted 12 May 2011.

This study was financed by the National Fund for Scientific Research of Belgium (F.R.S.-FNRS, grant MIS F.4.519.10.F and grant 2.45.77.10) and by Gembloux Agro-Bio Tech (GxABT-ULg, TROPDIV-Project PPR 10.000). We acknowledge Thibaut Malausa for allowing us to benefit from the project 'EcoMicro' (AIP BioResources grant, INRA) and the R&D budget of Genoscreen (Lille, France). For their help with sampling, we also thank the forest company Pallisco (Douala, Cameroon) and J. P. Mate of the University of Kisangani (Kisangani, Democratic Republic of Congo).

[4]Author for correspondence: cmichene@ulb.ac.be

doi:10.3732/ajb.1100070

METHODS AND RESULTS

Microsatellite primers were isolated in *P. elata* at the 454 GS-FLX titanium platform of Genoscreen (Lille, France) through a newly described procedure that associates multiplex microsatellite enrichment and pyrosequencing (Malausa et al., 2011). This isolation procedure is particularly cost-efficient as it allows the generation of 32 microsatellite-enriched libraries through a single 454 GS-FLX sequencing run (with either four samples per 1/8 plate, or eight samples per 1/4 plate), where each library typically allows defining primers for 100 to 1000 microsatellite markers (Malausa et al., 2011). We further optimized this step by mixing genomic DNA of two additional species with *P. elata* (Fabaceae), namely *Greenwayodendron suaveolens* (Engl. & Diels) Verdc. (Annonaceae) and *Scorodophloeus zenkeri* Harms (Fabaceae) [results not shown for these two latter species]. This approach can reduce the cost for obtaining a microsatellite library through a commercial service to ca. $600 per species.

Total DNA was extracted from silica gel–dried leaves using NucleoSpin Plant kits (Macherey-Nagel, Düren, Germany). A mixture of ca. 5 μg of DNA of the three species (including a single sample per species) was sent to Genoscreen (Lille, France), and was used to isolate microsatellite loci following the protocol of Malausa et al. (2011). Briefly, enriched libraries were first constructed using the following eight motifs to enrich total DNA: $(AG)_{10}$, $(AC)_{10}$, $(AAC)_8$, $(AGG)_8$, $(ACG)_8$, $(AAG)_8$, $(ACAT)_6$, and $(ATCT)_6$. The resulting enriched DNA was sequenced by Genoscreen (see Malausa et al., 2011 for additional details). The sequencing run generated 23 901 reads for the three species, which were analyzed using the bioinformatic QDD pipeline (Meglécz et al., 2010): sequences shorter than 80 bp and sequences containing microsatellite motifs shorter than five repeats were discarded. Sequences with significant BLAST hits but with flanking region identity levels below 90% were also discarded. Sequences displaying only BLAST hits for which pairwise similarity between the complete overlapping part of the flanking regions was greater than 90% were grouped into contigs aligned by ClustalW and a two-thirds majority rule was used to build a consensus

 © 2011 Botanical Society of America

sequence (see Malausa et al., 2011). Primer pairs were designed automatically by Primer3 (Rozen and Skaletsky, 2000) within QDD with the criteria described in Malausa et al. (2011): i.e., primers were designed only if the resulting PCR product was between 80 and 500 bp long; the flanking region must contain, at most, a five-base mononucleotide stretch, or two repeats of any di-hexa base-pair motif; the annealing temperature of primers was between 50°C and 64°C; and the difference in annealing temperature between the forward and the reverse primer was not greater than 4°C.

From a total of 7760 primer pairs generated from Primer3 for 846 loci, we selected 72 primer pairs among loci representing the longest di-, tri, and tetra-nucleotide repeats, and having more than 10 flanking nucleotides between microsatellite motifs and designed primers. To reduce the cost of dye-labeled primers, we opted for a M13-like labeling technique (see below). Hence, the forward sequence of each selected locus was redesigned by incorporating to the 5′ end one of the four Q1–Q4 unique universal sequences (see Table 1), after having checked that these newly designed sequences were suitable for primer mix and did not present any hairpin structures or possible primer-dimers. The 72 designed primer pairs were first tested individually to verify amplification in *P. elata*, with the following PCR conditions: 1 μL buffer (10×), 0.4 μL $MgCl_2$ (25 mM), 0.3 μL dNPTs (10 mM each), 0.2 μL of each primer (0.01 mM), 0.05 μL *Taq* polymerase (TopTaq DNA Polymerase, 5 U/μL, QIAGEN, Venlo, Netherlands), 1 μL of template DNA (of ca. 10–50 ng/μL), and H_2O to make a final volume of 10 μL. Amplifications were performed as follows: 94°C (4 min), followed by 40 cycles of 94°C (30 s), 56°C (45 s), 72°C (1 min), and a final extension at 72°C for 10 min. PCR products were run out on a 1% agarose gel and stained with SYBR Safe (Invitrogen, Merelbeke, Belgium).

A total of 21 primer pairs generated products in *P. elata*, and were reamplified to quantify levels of polymorphism on eight individuals chosen over the species distribution in Cameroon, Congo, and the Democratic Republic of Congo (DRC). PCRs were conducted individually using the modified protocol of Schuelke (2000), which incorporates three primers (M13-like protocol): an unlabeled reverse primer, a Q-tailed forward primer (incorporating the unique sequence added to the 5′ end, i.e., either Q1, Q2, Q3, or Q4), and a third primer composed of this same unique universal sequence but with a fluorescent dye attached to the 5′ end: either 6-FAM (for Q1), NED (for Q2), VIC (for Q3), or PET (for Q4) (Table 1). We estimated that this approach allowed reducing the costs of primers by ca. five times because only four labeled primers have to be ordered (and they can be further used for other species as well). PCR conditions were identical to those mentioned before, but the Q-tailed forward primer was used at one-third the amount (0.7 μM) of the reverse and fluorescently labeled primers (2 μM). PCR amplifications were as follows: 94°C (4 min), followed by 20 to 30 cycles each of 94°C (30 s), 56°C (45 s), 72°C (1 min), plus 10 cycles each of 94°C (30 s), 53°C (45 s), 72°C (45 s), and a final extension at 72°C for 10 min.

Preliminary population genetic analyses were carried out on a total of 154 individuals from two different locations (127 individuals from eastern Cameroon: 3°09′55″N, 14°18′53″E; and 27 from northern DRC: 0°17′36″N, 25°18′52″E). After excluding loci that did not amplify or were monomorphic, we selected 13 primer pairs that were amplified in two multiplexed reactions (Table 1) using the QIAGEN Multiplex kit (QIAGEN) in a final 15 μL reaction volume. PCRs were carried out as follows: 7.5 μL Multiplex PCR Master Mix, 1.5 μL primer mix (with Q-tailed primers at 0.7 μM and reverse primers at 2 μM), 0.15 μL of each fluorescent Q1–Q4 primers (10 μM), 1.5 μL DNA, and 3.9 μL H_2O. Multiplex PCR programs consisted of 95°C for 15 min, followed by 23 to 30 cycles each of 94°C (30 s), 57°C (90 s), 72°C (90 s), followed by 10 cycles each of 94°C (30 s), 53°C (45 s), 72°C (45 s), and a final extension at 60°C for 30 min. One microliter of PCR products was directly added to 12 μL HiDi formamide and 0.2 μL GeneScan 500 LIZ size standard, and run on an ABI 3730 sequencer (Applied Biosystems, Lennik, Netherlands). Results were analyzed using GeneMapper Software Version 3.7 (Applied Biosystems). The number of alleles per locus, observed and expected heterozygosities, and tests for deviation from the Hardy–Weinberg equilibrium (HWE) were calculated using SPAGeDi 1.3 (Hardy and Vekemans, 2002). Results were adjusted for multiple comparisons using a sequential Bonferroni correction.

When combining together all loci and both populations, 11 of 13 loci were polymorphic, with the number of alleles per locus ranging from two to 11 and a mean value of 5.4 alleles per locus (Table 1). At the population level, only five of 11 loci were polymorphic in Cameroon, with a number of alleles ranging from two to four, while all loci were polymorphic in DRC, with the number of alleles ranging from two to 10. Expected heterozygosity (H_e) ranged from 0 to 0.578 with a mean value of 0.204 in Cameroon, and from 0.107 and 0.841 with a mean value of 0.522 in DRC. In total, all polymorphic loci significantly deviated from HWE in Cameroon because of heterozygote deficiency, while in DRC only three out of 11 loci deviated from HWE (Table 2). We checked whether departure from HWE at a given locus might be explained by the presence of null alleles by using the software INEst (Chybicki and Burczyk, 2009),

TABLE 1. Characterization of 11 microsatellite markers isolated from *Pericopsis elata*. Forward and reverse primer sequences, labeled primer, repeat motif, annealing temperature (T_a), range of allele sizes (bp), number of alleles (N_a), and GenBank accession number are shown for each primer pair.

Locus	Primer sequence (5′–3′)*	Labeled primer*	Repeat motif	T_a (°C)	Allele size (bp)	N_a	GenBank Accession No.
P5†	F: Q1-GGATATTAGGGCTTTAAGTAACCGT	Q1-6-FAM	$(CA)_{10}$	56	132–136	3	JF508137
	R: TGGGAGCATTCCTAAGTCCA						
P12†	F: Q4-GAAAAAGGGATGCCATCAAA	Q4-PET	$(TC)_{12}$	56	140–156	3	JF508138
	R: CATGGTTTTGTCTCTGCACAA						
P13‡	F: Q1-AATTGCCAGCTGGTTTTCAT	Q1-6-FAM	$(TC)_{13}$	56	145–154	2	JF508139
	R: AATGGCAGGCACGAATGTAT						
P14†	F: Q2-TCTTCGGGATACCAGTGTCA	Q2-NED	$(CT)_{12}$	56	154–168	4	JF508140
	R: TTTTTGACATACATGGCACCTT						
P19‡	F: Q3-CCAATATTTGGTTAAAGCTGAACAC	Q3-VIC	$(GA)_{10}$	56	148–162	5	JF508141
	R: CACTTTGGAAGGTCAGGTCC						
P24‡	F: Q4-CCATAATCCTTTGCTGTTCTCA	Q4-PET	$(CT)_{11}$	56	168–174	4	JF508142
	R: GAGCAGTTAGGGTTTCCCAA						
P32†	F: Q4-GTAATCGGGAGGCACGTAAA	Q4-PET	$(GA)_{12}$	56	196–228	11	JF508143
	R: GATGAAATATGCAAGGAGCGA						
P50‡	F: Q2-GGCTTGTGGGACCTATGATG	Q2-NED	$(AC)_{13}$	56	230–244	3	JF508144
	R: GAATGACACCATTACAATGAGGA						
P52‡	F: Q4-CTTTTCCTCCTCGCACATTC	Q4-PET	$(CTT)_{12}$	56	237–258	6	JF508145
	R: GATGAGACCCGTTTGGAAGA						
P55‡	F: Q3-TGGCAAGATGAATGAGTCCA	Q3-VIC	$(AC)_{19}$	56	240–264	10	JF508146
	R: GGTTTCCTCTTTGTATGCAAGAAC						
P66‡	F: Q2-CCCTTGAGCGAACATTCATC	Q2-NED	$(AC)_{13}$	56	294–320	8	JF508147
	R: CATTTGGTGATAAGAGATGTGAAA						

* Q1 = TGTAAAACGACGGCCAGT (Schuelke, 2000); Q2 = TAGGAGTGCAGCAAGCAT; Q3 = CACTGCTTAGAGCGATGC; Q4 = CTAGTTATTGCTCAGCGGT (Q2–Q4, after Culley et al., 2008).

† Locus included in multiplex 1 (multiplex 1 also included two other monomorphic loci that are not described here, P33 and P35).

‡ Locus included in multiplex 2.

TABLE 2. Characterization of 11 loci from two populations of *Pericopsis elata* (Cameroon and DRC). Sample size (N), number of alleles (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), and tests of Hardy–Weinberg equilibrium (HWE) are shown for each microsatellite locus.

	Cameroon					DRC				
Locus	N	N_a	H_o	H_e	HWE*	N	N_a	H_o	H_e	HWE*
P5	127	1	0	0		27	3	0.519	0.590	
P12	127	1	0	0		27	2	0.037	0.107	
P13	127	1	0	0		27	2	0.185	0.440	**
P14	118	1	0	0		25	4	0.280	0.352	
P19	127	3	0.260	0.492	****	27	4	0.407	0.579	
P24	127	1	0	0		27	3	0.296	0.425	
P32	127	1	0	0		27	10	0.704	0.827	
P50	127	2	0.339	0.447	**	27	3	0.111	0.108	
P52	126	3	0.135	0.353	****	27	4	0.407	0.637	****
P55	127	4	0.205	0.578	****	27	8	0.926	0.840	
P66	124	4	0.226	0.378	****	25	7	0.360	0.841	****

* Empty cells indicate no deviation from HWE values.
** ($P < 0.01$) and **** ($P < 0.0001$) indicate significant deviation from HWE after sequential Bonferroni correction.

which jointly estimates inbreeding and null allele frequencies to account for deviation from HWE. Under the individual inbreeding model (IIM), null allele frequency estimates were significantly different from zero at all loci that deviated from HWE in DRC, but no evidence for the presence of null alleles was detected in Cameroon. Seven of these loci (i.e., P13, P50, P19, P24, P52, P14, P12) also amplified one sample of *P. laxiflora* (Benth. ex Baker) Meeuwen from Benin. Although it was not possible to test intraspecific polymorphism for this species, it is worthwhile to note that P19 and P35 were heterozygote, and that alleles found in P12, P13, P14, P24, and P52 were distinct from those recorded in *P. elata*.

CONCLUSIONS

In this study, we developed 11 microsatellite loci in the endangered timber tree *P. elata*. Our approach combines existing protocols but led to a notable decrease of expenses and a significant time savings: (1) we used a powerful procedure for the development of a microsatellite-enriched library that associates multiplex microsatellite enrichment and next-generation sequencing, (2) this low-cost procedure was implemented for three species simultaneously, reducing by three the cost per species for a microsatellite library, and (3) the use of M13-like labeled primers notably reduced expenses related to primer tests. The microsatellite primers described in the current study are currently used to further investigate levels of genetic variation and spatial genetic structure in *P. elata* across its natural range in Central and West Africa.

LITERATURE CITED

BOURLAND, N., Y. L. KOUADIO, G. COLINET, AND J.-L. DOUCET. 2010. *Pericopsis elata* (Harms) van Meeuwen in the south-eastern part of Cameroon: Ecological and pedological approaches to improve the management of an endangered commercial timber species. *The International Forestry Review* 12: 111.

CHYBICKI, I. J., AND J. BURCZYK. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *The Journal of Heredity* 100: 106–113.

CULLEY, T. M., S. G. WELLER, A. K. SAKAI, AND K. A. PUTNAM. 2008. Characterization of microsatellite loci in the Hawaiian endemic shrub *Schiedea adamantis* (Caryophyllaceae) and amplification in related species and genera. *Molecular Ecology Resources* 8: 1081–1084.

DICKSON, B., P. MATHEW, S. MICKLEBURGH, S. OLDFIELD, D. POUAKOUYOU, AND J. SUTER. 2005. An assessment of the conservation status, management and regulation of the trade in *Pericopsis elata*. Fauna & Flora International, Cambridge, United Kingdom.

HARDY, O. J., AND X. VEKEMANS. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.

IUCN. 2010. IUCN Red List of Threatened Species. Version 2010.4. Website http://www.iucnredlist.org [accessed 29 August 2011].

MALAUSA, T., A. GILLES, E. MEGLÉCZ, H. BLANQUART, S. DUTHOY, C. COSTEDOAT, V. DUBUT, ET AL. 2011. High-throughput microsatellite isolation through 454 GS-FLX titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644.

MEGLÉCZ, E., C. COSTEDOAT, V. DUBUT, A. GILLES, T. MALAUSA, N. PECH, AND J. F. MARTIN. 2010. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics (Oxford, England)* 26: 403–404.

ROZEN, S., AND H. SKALETSKY. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.)* 132: 365–386.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.