Characterization of Microsatellites Identified by Next-Generation Sequencing in the Neotropical Tree Handroanthus billbergii (Bignoniaceae)

Authors: Morillo, Eduardo, Buitron, Johanna, Limongi, Ricardo, Vignes, Helene, and Argout, Xavier

Source: Applications in Plant Sciences, 4(5)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1500135

Applications in Plant Sciences 2016 4(5): 1500135

Primer Note

Characterization of microsatellites identified by next-generation sequencing in the Neotropical tree *Handroanthus billbergii* (Bignoniaceae)[1]

Eduardo Morillo[2,4], Johanna Buitron[2], Ricardo Limongi[2], Helene Vignes[3], and Xavier Argout[3]

[2]Instituto Nacional de Investigaciones Agropecuarias (INIAP), Casilla Postal 17-01-340, Quito, Ecuador; and
[3]Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR 1334 AGAP, Montpellier, France

- *Premise of the study:* We developed microsatellite (simple sequence repeat [SSR]) markers in the Neotropical tree *Handroanthus billbergii* (Bignoniaceae), to be applied in assessment of genetic diversity in this species as a reference for inferring the impact of dry forest fragmentation in Ecuador.
- *Methods and Results:* Using next-generation sequencing, we detected a total of 26,893 putative SSRs reported here. Using an ABI 3500xl sequencer, we identified and characterized a set of polymorphic markers in 23 individuals belonging to three populations of *H. billbergii*.
- *Conclusions:* We report a set of 30 useful SSR markers for *H. billbergii* and a large list of potential microsatellites for developing new markers for this or related species.

Key words: Bignoniaceae; *Handroanthus billbergii*; microsatellites; next-generation sequencing (NGS).

Handroanthus billbergii (Bureau & K. Schum.) S. O. Grose (syn. *Tabebuia billbergii* (Bureau & K. Schum.) Standl.) (Bignoniaceae), known as "guayacan negro" in Spanish, is a characteristic deciduous tree species of tropical dry forests in the Tumbesian region. This region, recognized as one of the most endangered biodiversity hotspots, is included among the most threatened ecosystems in Ecuador (Espinosa et al., 2011). According to the most recent taxonomic revision, *H. billbergii* belongs to the genus *Handroanthus* Mattos, a group of trees known for wood that is among the heaviest and hardest known in the world (Grose and Olmstead, 2007). Because this species produces very hard, durable, and high-quality wood, it has been overexploited by the local people and dominates the market, and therefore its diversity is declining rapidly (Détienne and Vernay, 2011).

To study the impact of forest fragmentation on *H. billbergii* populations in Ecuador, a survey of genetic diversity using highly polymorphic and codominant markers is proposed. Microsatellite markers (also known as simple sequence repeats [SSRs]) are available in the closely related species *Tabebuia aurea* (Silva Manso) Benth. & Hook. f. (Braga et al., 2007), but transferability of these markers to *H. billbergii* was almost null. Furthermore, a genetic survey with inter-simple sequence repeats (ISSRs) revealed low polymorphism rates, primarily at the intra-population level (Rueda, 2015). In this study, we report the development of SSR markers in *H. billbergii* by applying next-generation sequencing (NGS). We also report an estimation of the genetic information content of a set of 30 polymorphic markers surveyed in a representative group of individuals from three different populations.

METHODS AND RESULTS

As biological material for the DNA library, we used samples from two representative trees referenced in populations 1 and 2 (Appendix 1) during field prospecting expeditions in Ecuador. Genomic DNA was isolated from dried leaf samples using a procedure reported in Rueda (2015). The DNA library was prepared using the Nextera DNA Sample Kit (Ref. GAO9115; Illumina, San Diego, California, USA). DNA fragmentation started with 50 ng of purified DNA, followed by end-polishing and sequencing adapter ligation to prepare di-tagged DNA fragment libraries. The quality of the libraries was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA), and the concentration was quantified using the KAPA Library Quantification Kit for Illumina Sequencing (KR0390; Kapa Biosystems, Woburn, Massachusetts, USA). The sequencing was performed on a MiSeq Sequencer (Illumina) using the 2×300-bp read mode. To reduce the raw data set and maximize the length of the sequences, an assembly of the 2,169,901 reads was performed with the ABYSS assembler (Simpson et al., 2009). To search for microsatellite markers among the assembled contigs (856,149), we used the MIcroSAtellite identification tool (MISA; Thiel et al., 2003). The search used specific criteria for the design of primers; these included motifs 2–6 bp in size, with a minimum repeat number of four repetitions, and with a maximum difference between two SSRs of 100 bp. The primers would thus amplify one or more repeats, effectively encompassing them as the same repeat motif if these were 100 bp or less from each other.

Among the 61,074 SSRs identified, primers were designed for 26,893 SSRs, of which 11,438 were dinucleotide, 8119 were trinucleotide, 3545 were

[1]Manuscript received 2 December 2015; revision accepted 27 January 2016.

This work was funded by the Instituto Nacional de Investigaciones Agropecuarias (INIAP) in scientific collaboration with the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Genetic Improvement and Adaptation of Mediterranean and Tropical Plants Research Unit (UMR AGAP).

[4]Author for correspondence: eduardo.morillo@iniap.gob.ec

doi:10.3732/apps.1500135

TABLE 1. Description of 30 new microsatellite markers developed in *Handroanthus billbergii*.

Locus[a]	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)	GenBank accession no.
mTb002	F: GGAACGTGCTAGTGTGTGTG	$(GT)_{15}$	235–297	KT715664
	R: AGAGAGTGAGTTGCAACAAAAGT			
mTb003	F: GACCTTCTGCTTGTGTTCC	$(TG)_{15}$	142–157	KT715665
	R: CTGTAAGTGTTAATTCTGCTGCT			
mTb005	F: GCAAGGGTGGGGTAGAG	$(TC)_{16}$	186–218	KT715666
	R: GGGCAACGCATCTTGT			
mTb006	F: GGGCTTCACATTGGTTG	$(TC)_{18}$	162–172	KT715667
	R: GGCATTTCCCAAGAACA			
mTb011	F: TAATTTCCGGGTGCAGA	$(AG)_{17}$	167–193	KT715668
	R: CACTGGTCTCTCACATTTCAC			
mTb013	F: CTTCTCATTCATTTTGGTGG	$(TC)_{15}$	354–368	KT715669
	R: GCTTCAACACTTTCACACATC			
mTb014	F: TCAGTGCAACTCCATTCC	$(CT)_{15}$	247–259	KT715670
	R: AATGAACGGCATCATCTTT			
mTb015	F: AGCAACACAAGGAGCATTT	$(TC)_{21}$	275–301	KT715671
	R: GTCAGACCCAATAACTTACCTTC			
mTb016	F: CCAGCCTCAGTTTCAGTTC	$(AG)_{15}$	126–156	KT715672
	R: CCATTGGGATCTCTGCTT			
mTb017	F: TGAACATGGAACAGAGCAA	$(AG)_{16}$	216–274	KT715673
	R: AGCCCAAGCGGATACA			
mTb018	F: GTGGTGCAGCGACTTCT	$(CT)_{15}$	230–238	KT715674
	R: ACATCATCGTCATCCTCATC			
mTb019	F: CAAATAAAAGTCATAGCAGAGG	$(AG)_{19}$	272–312	KT715675
	R: TGGCATTGAACACAACTC			
mTb020	F: CATTGACTCGTTGTCCC	$(AG)_{15}$	305–341	KT715676
	R: GATCCTACAGTCTCACATAGAAG			
mTb021	F: ATTGTTGATGAAGGGCAAA	$(TC)_{18}$	269–277	KT715677
	R: GGGCAAGGCTAAAGGAA			
mTb022	F: AGATCCACGAACCCAAAA	$(AG)_{15}$	202–240	KT715678
	R: GAACGCCGAAGTGTGAG			
mTb023	F: AGCGCAATGTGATAAGAGCT	$(AC)_{14}$	252–262	KT715679
	R: GCCCTTCATTCTTGGTGAGC			
mTb024	F: CCATTTGCTTGCCTTACCCA	$(TA)_{12}$	199–209	KT715680
	R: AAGCAAACAACCACTCTGCA			
mTb025	F: CAAAGTGAGAGGAACTGAAAA	$(AT)_{11}$	354–368	KT715681
	R: GGACACGAGCCAAGAAG			
mTb027	F: TTTTCACAACCAGTAACTTC	$(ATT)_{15}$	165–201	KT715682
	R: GGTGTTTGGCATTACTTT			
mTb028	F: TGGCAAGGACATAATCTTCAAGA	$(TAA)_{17}$	185–215	KT715683
	R: AAAACCCCAAATTCACTCCCT			
mTb029	F: AACGAAAGAGGCGAGGT	$(TTA)_{13}$	273–282	KT715684
	R: TCCACCCATGTCCAATC			
mTb031	F: TGCAAGTCCTGGGAAGCATA	$(CTT)_{12}$	194–215	KT715685
	R: GCACGAACAGAATGTCCAGG			
mTb032	F: CGTCGAATATCTAGTGTGGG	$(AAT)_{13}$	189–213	KT715686
	R: ACAGATGAAGAGAAAACCAAAG			
mTb033	F: ACAAGGAAGTAAATTGCAACTCG	$(ATT)_{15}$	171–205	KT715687
	R: ACCAGACTCCAAACACGACT			
mTb035	F: TCCTAATTCACCAACTTCC	$(ATT)_{15}$	151–183	KT715688
	R: GTCTGTAAGCCACATAGACTG			
mTb036	F: CGACTTCCACCATCCAA	$(TTA)_{14}$	149–170	KT715689
	R: CCTTTCTTTTGCAGCCC			
mTb037	F: TCTTGTTGGGAATAATTGGA	$(AAT)_{15}$	235–265	KT715690
	R: GCATTAGGCAAAATTCGAG			
mTb041	F: CGACATTCTTGCTCCCAATCA	$(TAA)_{16}$	221–260	KT715691
	R: AAACAGCGGCAAGAAAGGTT			
mTb043	F: GGTCTAGCACGTGACTAACC	$(ATT)_{16}$	232–275	KT715692
	R: CCCAATACGAGGCATATGTGA			
mTb052	F: TGACAGTGAAAAGTTGCCACA	$(AAT)_{18}$	147–171	KT715693
	R: TCATCGCAATATGTACACGATTG			

[a]Annealing temperature for all loci was 55°C.

tetranucleotide, 941 were pentanucleotide, 336 were hexanucleotide, and 2514 contained complex SSR motifs (Appendix S1).

For the SSR screening, we selected primer pairs flanking di- and trinucleotide SSR motifs with a minimum of 10 repetitions and amplifying fragments between 400 and 1000 bp in length. We screened a set of 55 primers: 25 primer pairs corresponding to dinucleotide SSR motifs and 30 primer pairs corresponding to trinucleotide SSR motifs. Amplification of the expected SSR fragments was first observed using agarose gel electrophoresis. Eight primer pairs that did not yield a single, well-amplified PCR product were excluded from the polymorphism survey. For screening the 47 remaining primer pairs, 23 individuals of three *H. billbergii* populations located in Ecuador (Appendix 1) were analyzed on an ABI 3500xl sequencer (Applied Biosystems, Waltham, Massachusetts,

TABLE 2. Results of initial SSR primer screening in three populations of *Handroanthus billbergii*.[a]

	Population 1 (Mangahurco) (N = 9)			Population 2 (Puna) (N = 7)			Population 3 (Arenillas) (N = 7)			Total		
Locus	A	H_e	H_o	A	H_e	H_o	A	H_e	H_o	A	PIC	CA[b]
mTb002	6	0.76	0.80	5	0.75	1.00	6	0.78	1.00	8	0.80	+
mTb003	2	0.18	0.20	2	0.50	0.60	2	0.42	0.60	2	0.33	+
mTb005	3	0.62	0.20	3	0.61	0.66	3	0.65	0.25	8	0.84	+
mTb006	3	0.53	0.75	3	0.64	0.60	4	0.72	0.80	4	0.65	+
mTb011	3	0.59	0.50	3	0.66	0.80	4	0.48	0.40	5	0.58	+
mTb013	1	0	0	2	0.32	0	3	0.62	0	4	0.42	+
mTb014	2	0.48	0.40	3	0.66	0.40	3	0.62	1.00	4	0.66	—
mTb015	3	0.56	0.40	4	0.58	0.80	3	0.54	0.60	6	0.56	+
mTb016	3	0.53	0.75	5	0.68	0.40	8	0.84	1.00	9	0.79	+
mTb017	5	0.78	0.75	8	0.86	1.00	10	0.90	1.00	17	0.91	+
mTb018	3	0.46	0.40	2	0.42	0.60	2	0.48	0.80	3	0.42	+
mTb019	8	0.86	0.60	7	0.78	0.80	7	0.84	0.80	14	0.87	+
mTb020	7	0.82	0.60	5	0.75	1.00	7	0.82	0.80	12	0.89	+
mTb021	4	0.72	0.60	3	0.64	1.00	4	0.72	0.40	5	0.73	+
mTb022	6	0.80	0.80	6	0.80	0.80	5	0.72	1.00	9	0.77	+
mTb023	2	0.32	0.40	2	0.18	0.20	1	0	0	3	0.17	+
mTb024	2	0.42	0.60	4	0.71	1.00	4	0.64	1.00	5	0.58	+
mTb025	2	0.44	0	6	0.82	0.40	4	0.66	0.20	6	0.69	+
mTb027	2	0.37	0.50	6	0.80	1.00	5	0.77	1.00	8	0.84	+
mTb028	3	0.62	1.00	8	0.84	1.00	3	0.50	0.66	8	0.77	+
mTb029	2	0.50	1.00	3	0.58	1.00	2	0.27	0.33	3	0.48	+
mTb031	2	0.18	0.20	3	0.54	0.40	2	0.32	0	3	0.32	+
mTb032	1	0	0	4	0.65	0.50	2	0.18	0.20	5	0.34	+
mTb033	4	0.70	0.40	4	0.72	0.80	4	0.64	0.60	6	0.68	+
mTb035	3	0.56	0.80	3	0.58	0.40	4	0.66	0.60	5	0.58	+
mTb036	2	0.32	0	3	0.46	0.20	2	0.32	0	4	0.36	+
mTb037	6	0.78	0.60	6	0.81	0.75	3	0.62	0.80	8	0.82	+
mTb041	5	0.68	0.80	3	0.56	0.60	5	0.72	0.60	6	0.69	+
mTb043	4	0.70	0.40	4	0.58	0.40	5	0.74	0.80	9	0.74	+
mTb052	1	0	0	5	0.72	0.80	5	0.76	1.00	8	0.75	+
Mean	3.3	0.510	0.482	4.1	0.641	0.664	4	0.599	0.608	6.5	0.638	

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; N = number of individuals sampled; PIC = polymorphism information content.

[a] Locality and voucher information are provided in Appendix 1.

[b] Cross-amplification in the sympatric species *H. chrysanthus*: + = successful amplification; — = no amplification.

USA). The forward primers were end-labeled with an M13 extension (5′-CACGACGTTGTAAAACGAC-3′), and the SSR amplifications were multiplexed in a 10-μL volume reaction using FAM, NED, PET, and VIC chemistries. We used 20 ng of DNA template, 1× buffer (including $MgCl_2$ with a final concentration of 1.5 mM), 0.08 μM of the M13-labeled primer, 0.1 μM of the reverse primer, 0.1 μM of M13 fluorescent primer, 0.5 mM of $MgCl_2$, 200 μM dNTP, and 0.05 U/mL of *Taq* DNA polymerase. The following program was used in a 384 multiplex PCR arrangement: 94°C for 5 min, 35 cycles at 94°C for 45 s, 55°C for 60 s, 72°C for 75 s, including a touchdown of 10 cycles at 94°C for 45 s, 60°C decreasing 0.5°C per cycle, and a final elongation step of 72°C for 30 min. Electropherograms were analyzed with GeneMapper version 4.1 using GeneScan 600 LIZ as a size standard (Applied Biosystems). Allele calling was obtained by checking for each data point in the amplification peaks. The scoring edition of the SSR profiles allowed us to obtain a good genotype assignment and the identification of null alleles and possible scoring errors. Statistical parameters as number of alleles per locus, observed (H_o) and expected heterozygosities (H_e), polymorphism information content (PIC) of each locus, and the presence of linkage disequilibrium (LD) were calculated with PowerMarker version 3.25 (Liu and Muse, 2005).

Thirty microsatellite loci showing good genetic profiles and clear allelic size variability were characterized as polymorphic markers (Table 1). Among the 23 genotyped individuals, a total of 197 alleles were scored with a mean of 6.5 alleles per locus (Table 2). Primers mTb017, mTb019, and mTb020 showed the greatest number of alleles and therefore have higher PIC values (0.87–0.91). Based on polymorphism interpretation of PIC values (Botstein et al., 1980), only eight markers with PIC values lower than 0.5 showed moderate genetic diversity and the remaining 24 markers were very informative, with PIC values higher than 0.5. H_e and H_o mean values ranged from 0.51 to 0.641 and 0.482 to 0.664, respectively, in the three screened populations, and evidence of significant LD was found for 24 of the 435 possible SSR pairwise combinations after Bonferroni corrections. However, a further analysis with more individuals will be useful for confirming possible genetic associations between the reported SSR markers. Finally, primer transferability was also tested in the sympatric related species *H. chrysanthus* (Jacq.) S. O. Grose, showing good results of cross-amplification in all but one of the 30 markers (Table 2).

CONCLUSIONS

The set of 30 SSR markers reported here will be used for surveying genetic diversity in a larger sample size of *H. billbergii* populations in the dry forests in Ecuador. In addition, the list of 26,893 SSRs published here (Appendix S1) is available for developing other markers in *H. billbergii* or related species.

LITERATURE CITED

BOTSTEIN, D., R. L. WHITE, M. SKOLNICK, AND R. W. DAVIS. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314–341.

BRAGA, A. C., A. M. M. REIS, L. T. LEOI, R. W. PEREIRA, AND R. G. COLLEVATTI. 2007. Development and characterization of microsatellite markers for the tropical tree species *Tabebuia aurea* (Bignoniaceae). *Molecular Ecology Notes* 7: 53–56.

DETIENNE, P., AND M. VERNAY. 2011. Les espèces du genre *Tabebuia* susceptibles de fournir le bois d'ipé. *Bois et Forêts des Tropiques* 307: 69–77.

Espinosa, C. I., O. Cabrera, A. L. Luzuriaga, and A. Escudero. 2011. What factors affect diversity and species composition of endangered Tumbesian dry forests in Southern Ecuador? *Biotropica* 43: 15–22.

Grose, S. O., and R. G. Olmstead. 2007. Taxonomic revisions in the polyphyletic genus *Tabebuia* s.l. (Bignoniaceae). *Systematic Botany* 32: 660–670.

Liu, K., and S. V. Muse. 2005. PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics* 21: 2128–2129.

Rueda, A. 2015. Estudio de la diversidad genética de poblaciones de guayacán sabanero (*Tabebuia billbergii*) de bosques secos del Ecuador. Tesis Ing. Biotecnología, ESPE, Sangolquí, Ecuador.

Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones, and I. Birol. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research* 19: 1117–1123.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 10: 411–422.

Appendix 1. Voucher information for *Handroanthus billbergii* samples used in this study.[a]

Species	Collection locality	Geographic coordinates	*N*
H. billbergii (Bureau & K. Schum.) S. O. Grose	Mangahurco, Loja, Ecuador (Population 1)	4°8′6.108″S, 80°26′22.812″W	9
H. billbergii	Puna, Guayas, Ecuador (Population 2)	2°42′57.204″S, 80°5′27.204″W	7
H. billbergii	Arenillas, El Oro, Ecuador (Population 3)	3°29′2.112″S, 80°6′28.583″W	7
H. chrysanthus (Jacq.) S. O. Grose	Mangahurco, Loja, Ecuador	4°8′6.108″S, 80°26′22.812″W	3

Note: *N* = number of analyzed individuals.
[a]Representative trees (codes Tb24 and Tb37) used for developing the markers were sampled from population 1.