Development of microsatellite markers from genomic DNA of *Parashorea malaanonan* (Dipterocarpaceae) using next-generation sequencing

Crusty E. Tinio[1,3*], Saneyoshi Ueno[2], Kentaro Uchiyama[2], Lerma SJ. Maldia[3] and Nobuhiro Tomaru[1]

[1] Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601
[2] Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Tsukuba, Ibaraki, 305-8572, Japan
[3] Department of Forest Biological Sciences, College of Forestry and Natural Resources, The University of the Philippines – Los Baños, College 4031, Laguna, Philippines

*Corresponding author: Crusty E. Tinio, E-mail: cetinio@up.edu.ph

Abstract

Twenty polymorphic microsatellite markers were developed, using Next Generation Sequencing (Illumina), from genomic DNA of *Parashorea malaanonan*, a species of the Dipterocarpaceae which is ecologically and economically important in the Philippines. Thirty adult trees from a natural population were used to assess the success of PCR amplification and the degree of polymorphism. The number of alleles per locus varied from three to 13, and observed and expected heterozygosity varied from 0.200 to 0.808 and from 0.301 to 0.890 respectively. Total exclusion probabilities for the first and second parents over the 20 loci were 0.99932499 and 0.99999723 respectively. The high level of polymorphism at these loci makes it possible to obtain precise estimates of genetic parameters and thus the markers will help in studies on population genetics, conservation genetics, and molecular ecology of *P. malaanonan*.

Keywords: : *Dipterocarpaceae, microsatellite markers, Parashorea malaanonan, next generation sequencing, simple sequence repeat*

Introduction

Parashorea malaanonan (Blanco) Merr., known locally in the Philippines as 'bagtikan', belongs to the ecologically and economically important tree family the Dipterocarpaceae. It is a large, prominent tree that can reach a height of 60 m and a basal trunk diameter of up to 200 cm (Fernando et al., 2009). It is naturally distributed in Brunei Darussalam, Malaysia (Sabah, Sarawak), and the Philippines. It is categorized as critically endangered on the IUCN Red List of Threatened Species (IUCN, 2018). In the Philippines, it thrives in lowland tropical rain forest at altitudes of up to 1,300 m (Fernando et al., 2009) and it is fairly common in its native range. Although it is relatively abundant compared to other dipterocarp species in the country, its genetic resources have been rapidly reduced due to excessive logging and land use conversion (Langenberger, 2006). Gamboa-Lapitan and Hyun (2005) assessed the mating system and genetic diversity of *P. malaanonan* using isozymes. Abasolo et al. (2009) and Villarin (2015) investigated the transferability of *Shorea* microsatellite markers developed by Ujino et al., 1998, Ng et al., 2009, and Lee et al., 2004 to *P. malaanonan* and found that only a few of the markers tested were polymorphic and could therefore be used for further genetic studies in *P. malaanonan* populations. Locus specificity, amplification errors (Selkoe and Toonen, 2006) or deletion of simple sequence repeat (SSR) flanking regions among species (Lemes

DOI:10.2478/sg-2019-0004
edited by the Thünen Institute of Forest Genetics

et al., 2002) may have contributed to the limited transferability of markers even to closely related taxa within the family. This necessitates the development of new highly polymorphic microsatellite markers to aid in the generation of genetic information for investigations into genetic diversity, genetic structure, mating system, and gene dispersal via pollen and seed for this species.

For the development of microsatellite markers, our study used next generation sequencing (NGS), which has proved popular for this application. Currently, 454 (Roche) and Illumina are the NGS platforms most widely used for developing SSR markers (Vieira et al., 2016). We also used the three-primer PCR method (Blacket et al., 2012) to amplify microsatellite loci. The strategy used locus-specific forward primers with tail sequences (A, B, C or D) at the 5′ end of the primer sequences, locus-specific reverse primers, and universal forward primers (tail A, B, C or D) with fluorescent dyes. Our objective in this study was to develop and characterize polymorphic microsatellite markers for assessing the genetic diversity, genetic structure, mating system, gene dispersal, and parentage of offspring in *P. malaanonan*.

Materials and Methods

Total genomic DNA (1 µg) was extracted from inner bark tissues of an adult *P. malaanonan* tree at Mount Makiling Forest Reserve (MMFR) in the Philippines using cetyltrimethylammonium bromide (CTAB) method (Murray and Thompson, 1980) with minor modifications and purified with Genomic DNA Clean & Concentrator (Zymo Research). Voucher specimens were deposited in the herbarium of the Department of Forest Biological Sciences, College of Forestry and Natural Resources, University of the Philippines Los Baños. The purified genomic DNA was then used for library construction with a Nextera DNA Library Preparation Kit (Illumina). The library was quantified for fragment length with a MultiNA Microchip Electrophoresis System (Shimadzu Co.). Sequencing was performed on a MiSeq sequencer (Illumina) with a 600 cycle kit according to the manufacturer's instructions. Forward and reverse sequencing reads were merged with bbmerge (https://github.com/BioInfoTools/BBMap) and processed with Trimmomatic (Bolger et al., 2014) to remove adapter, low quality, and short sequences. Sequences of lengths <150 bp were removed. The remaining sequences were used as input for the CMIB pipeline (Ueno et al., 2012) to obtain unique PCR primers targeting microsatellites with the numbers of repeat units being 10, 9, 8, 7, and 6 for di-, tri-, tetra-, penta-, and hexa-SSRs, respectively. Those sequences for which primer pairs could be successfully designed were subjected to BLAST searching against the NCBI non-redundant nucleotide (nt) and non-redundant protein (nr) databases with an *E*-value cut-off of 1E-5. We selected 48 primer pairs, for which forward primers with universal primer tail sequences (Blacket et al., 2012) and reverse primers with a PIG-tail sequence (Brownstein et al., 1996) at the 5′-end of the primer sequence were synthesized.

Genomic DNA from inner bark tissues of adult trees in a *P. malaanonan* natural population in MMFR, freshly extracted using a CTAB method with minor modifications, were used in primer screening for PCR amplification and extent of polymorphism. The genomic DNA from two adult trees (dbh $\geq$ 30 cm) was used when screening for PCR amplification. Singleplex PCR was carried out with a Gene Amp PCR system 9700 using a Type-It Multiplex PCR Kit (QIAGEN) in a final volume of 6 µL containing 3.0 µL of 2× Type-it Multiplex PCR Master Mix, 0.2 µM of reverse primer, 0.2 µM of forward primer and 0.2 µM of fluorescently labeled universal primer (Blacket et al., 2012), and 1 µL genomic DNA (approximately 5 - 10 ng). The PCRs were carried out using a touchdown protocol as follows: an initial denaturation for 15 min at 95° C, followed by 35 cycles of denaturation at 94° C for 30 s, annealing at 65–60° C for 90 s, and extension at 72° C for 1 min; the annealing temperature was 65° C in the first cycle, from 64 to 61° C (-1° C/cycle) during the following four cycles, and 60° C from the 6th to 35th cycle. The PCR products were electrophoresed using a 3130-Avant Genetic Analyzer (Applied Biosystems), with a GeneScan 500 LIZ (Applied Biosystems) size standard, and fragment sizes were estimated using GeneMapper v. 4.0 software (Applied Biosystems). After screening of primers for PCR amplification, we selected the markers which were successfully amplified and constructed primer combinations for multiplex PCR using Multiplex Manager version 1.2 (Holleley and Geerts, 2009). To screen primers for polymorphism, multiplex PCR and genotyping were carried out for 30 adult trees (dbh $\geq$ 30 cm) in the same way as described above. The number of alleles per locus (*A*), observed heterozygosity (H_O), expected heterozygosity (H_E), and exclusion probability were calculated using CERVUS version 3.0.7 (Kalinowski et al., 2007). Departure from Hardy-Weinberg Equilibrium (HWE) at each locus and linkage disequilibrium for all pairs of loci were tested using probability tests implemented in Genepop version 4.2 (Rousset, 2008).

Results and Discussion

Insert size in the sequencing library ranged from 300 bp to 410 bp with a mode of 360 bp. In total, 2.69 Gb (4,476,115 pairs) were obtained, of which 3,751,025 (83.7 %) were successfully merged. The average length of the merged reads was 185 bp with a maximum of 593 bp, yielding 2,453,465 (557 Mb) sequences after cleaning with Trimmomatic. We found 2,128 (0.087 %) sequences with the microsatellites defined in the Materials and Methods section. Focusing on those microsatellites with the number of repeat units being 6, 5, 4, 3, and 2 for di-, tri-, tetra-, penta-, and hexa-SSRs respectively, 27,264 (1.11 %) of the sequences had a microsatellite motif. We designed primer pairs targeting 575 sequences, for which BLAST homology searches were performed. We found that 62 sequences

had hits against nt or nr database entries. However, 27 (43.5 %) sequences had hits against entries from non-green plants, and we avoided these sequences when synthesizing primers. Forty-eight primer pairs were synthesized and tested for PCR amplification and polymorphisms.

The initial primer screening for (singleplex) PCR amplification was done using two adult *P. malaanonan* trees. Thirty out of 48 markers were amplified to give the expected sizes and these were screened for polymorphism by means of multiplex PCR using 30 adult trees of the species. The markers were combined in five multiplex mixes, which contained between two to eight markers. Ten of the 30 markers showed monomorphic or unstable expression and 20 had stably polymorphic expression. Primer information for these 20 markers is shown in Table 1 and polymorphisms at the corresponding loci are shown in Table 2. The number of alleles per locus varied from three to 13, and observed and expected heterozygosity varied from 0.200 to 0.808 and from 0.301 to 0.890 respectively. Total exclusion probabilities for the first and second parents over the 20 loci were 0.99932499 and 0.99999723 respectively. One locus, S035, departed significantly from HWE after sequential Bonferroni correction. There was no significant linkage disequilibrium detected for any pair of the loci after sequential Bonferroni correction. The microsatellite markers developed and characterized here will be useful for use in future work on population genetics, conservation genetics and molecular ecology of the critically endangered species *P. malaanonan*.

Table 1
Primer details for the 20 microsatellite markers developed from genomic DNA of *Parashorea malaanonan*

Locus	Sequence (5'–3')[a]	Universal primer	Dye	Multiplex set	Repeat motif	T_m (°C)[b]	Product size (bp)	Accession no.[c]
S035	F: GCCTTGCCAGCCCGCACCAATGTTGCTCTCAAGGAAGAAGC R: GTTTCTTGTTGCTGGACATGAAATGTGAGC	Tail B	VIC	C	$(AG)_{12}$	63.8	123-137	LC423583, LC423584
S055	F: CGGAGAGCCGAGAGGTGCCAGAGATCAGATCGGATCGAAACCG R: GTTTCTTCCTAGCGACGGTGCTTATGCC	Tail D	PET	D	$(GA)_{14}$	64.8	114-142	LC423585, LC423586
S063	F: GCCTTGCCAGCCCGCAGGAGAATGTCAGAAAGCAAAGGC R: GTTTCTTAGCTGACTTTGAGGGTTGGTTGTC	Tail B	VIC	A	$(AG)_{11}$	61.6	125-139	LC423587, LC423588
S069	F: CGGAGAGCCGAGAGGTGCTGCCTTCGCCTTTGCCTTTCTC R: GTTTCTTGAGAGGCGAACGGAGAAGTG	Tail D	PET	A	$(TTC)_{10}$	64.6	123-143	LC423589, LC423590
S199	F: CAGGACCAGGCTACCGTGAAACAGACCCTGTGCAAATTTGC R: GTTTCTTACTCAGCAGAATCAGGAGTGGTTTG	Tail C	NED	C	$(AG)_{15}$	61.0	157-187	LC423591, LC423592
S300	F: GCCTCCCTCGCGCCAGCATGCATGTGTGTGTGTGTG R: GTTTCTTCTTCCTCTGAGCAACATTCTGAG	Tail A	6-FAM	D	$(AG)_{13}$	61.2	174-188	LC423593, LC423594
S305	F: CAGGACCAGGCTACCGTGTGTGTATTTCAGCAAGCATCCCTC R: GTTTCTTCCCGATTTCTTCAATCACCCTGCC	Tail C	NED	D	$(TTC)_{10}$	61.4	177-198	LC423595, LC423596
S312	F: CGGAGAGCCGAGAGGTGATGTAGGGACCACACATCTGGCC R: GTTTCTTACGCCAATACCACTGATGCTCCTC	Tail D	PET	C	$(TGA)_{10}$	64.2	181-196	LC423597, LC423598
S348	F: GCCTCCCTCGCGCCAGCAGACAACAAAGTCATTGGCATTGC R: GTTTCTTCCCTTTCAAACAAGTGGTTAAGAGC	Tail A	6-FAM	B	$(TC)_{10}$	64.3	193-199	LC423599, LC423600
S350	F: GCCTTGCCAGCCCGCGCGGGTGGCAAAGATACATGAACG R: GTTTCTTAGCTGACGCGTGAAACTTAGTCC	Tail B	VIC	B	$(TC)_{8}$	64.7	175-197	LC423601, LC423602
S360	F: CGGAGAGCCGAGAGGTGACTGCGGTTGAAATGATTGCTGGC R: GTTTCTTGCAGCAGGTGAACCAGAGCTC	Tail D	PET	A	$(AG)_{16}$	64.9	184-214	LC423603, LC423604
S451	F: GCCTTGCCAGCCCGCACTCTGCAGAAACTGAAAGGTGCC R: GTTTCTTGTGGGTGGAATATCAATGCACTTC	Tail B	VIC	D	$(CA)_{10}$	63.6	220-232	LC423605, LC423606
S457	F: CAGGACCAGGCTACCGTGAGGCCAGCAAGGAGATGAGTTCC R: GTTTCTTGAAGGAATCAACACAAATCGGATCC	Tail C	NED	A	$(TC)_{12}$	64.4	220-238	LC423607, LC423608
S462	F: CGGAGAGCCGAGAGGTGTGGTTGGCCAACGGAGATAGTGG R: GTTTCTTCCATGGATTGTGAGGACAGCCC	Tail D	PET	B	$(TGC)_{10}$	64.6	221-239	LC423609, LC423610
S468	F: GCCTCCCTCGCGCCAGGACTTTCCGTCTGGATCACCTGG R: GTTTCTTCTTCATCCTCGGTTCTTCTCCAG	Tail A	6-FAM	E	$(GA)_{16}$	64.7	223-255	LC423611, LC423612
S474	F: CAGGACCAGGCTACCGTGTGAAGCGAGGACTAAGGTGGAGTG R: GTTTCTTGGTCCCTCATGGTAGAAGCTTGGC	Tail C	NED	C	$(AC)_{11}$	63.9	235-245	LC423613, LC423614
S479	F: CGGAGAGCCGAGAGGTGAGACGCCAAATTTCACATTGACG R: GTTTCTTAAGCAGAGAACAAACCAAACACAG	Tail D	PET	E	$(TC)_{12}$	60.6	238-246	LC423615, LC423616
S504	F: CAGGACCAGGCTACCGTGAGCGTACCGTAGAGAGAGGAGACG R: GTTTCTTCCCACTCTCTTTCCACGATCCC	Tail C	NED	B	$(AG)_{15}$	64.8	243-270	LC423617, LC423618
S535	F: GCCTCCCTCGCGCCAGGTAAGGCATGATGGTACGCTAC R: GTTTCTTGGCAGTCCCAAGAATGACTAGGTG	Tail A	6-FAM	C	$(AG)_{11}$	62.3	259-267	LC423619, LC423620
S568	F: CAGGACCAGGCTACCGTGTTAGCAACACGGAGGATGGGTGC R: GTTTCTTGTTTGCAGCAGGTGAGGGAAGG	Tail C	NED	A	$(GCT)_{15}$	65.2	274-286	LC423623, LC423624

[a]Forward primers (F) and reverse primers (R) have universal primer tail sequences (Blacket et al., 2012) and a PIG-tail sequence (Brownstein et al., 1996) respectively.
[b]Primer melting temperature. We used annealing temperatures in the PCR touchdown protocol at 65–60 °C for all multiplex sets.
[c]Accession number in the DNA Data Bank of Japan.

Table 2
Polymorphisms at 20 microsatellite loci for *Parashorea malaanonan* (N = 30)

Locus	*A*	H_O	H_E	Test for HWE[a]
S035	8	0.519	0.827	**
S055	10	0.577	0.811	ns
S063	6	0.517	0.636	ns
S069	5	0.633	0.667	ns
S199	12	0.619	0.872	ns
S300	7	0.731	0.716	ns
S305	7	0.345	0.528	ns
S312	4	0.500	0.567	ns
S348	4	0.200	0.301	ns
S350	7	0.500	0.576	ns
S360	10	0.720	0.853	ns
S451	6	0.769	0.713	ns
S457	6	0.577	0.646	ns
S462	6	0.533	0.775	ns
S468	11	0.792	0.890	ns
S474	5	0.308	0.535	ns
S479	3	0.375	0.407	ns
S504	13	0.808	0.817	ns
S535	5	0.615	0.564	ns
S568	5	0.611	0.594	ns

A, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity.
[a]Departure from Hardy-Weinberg Equilibrium (HWE) at each locus and linkage disequilibrium for all pairs of loci were tested using probability tests implemented in Genepop version 4.2 (Rousset, 2008).
ns, not significant; ** $P < 0.01$.

References

Abasolo MA, ES Fernando, TH Borromeo and DM Hautea (2009) Cross-species amplification of Shorea microsatellite DNA markers in Parashorea malaanonan (Dipterocarpaceae). Philippine Journal of Science 138 (1):23-28

Blacket MJ, C Robin, RT Good, SF Lee and AD Miller (2012) Universal primers for fluorescent labelling of PCR fragments-an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources 12(3): 456-463. https://doi.org/10.1111/j.1755-0998.2011.03104.x

Bolger AM, M Lohse and B Usadel (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Brownstein MJ, JD Carpten and JR Smith (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20(6): 1004-1006, 1008-1010. https://www.ncbi.nlm.nih.gov/pubmed/8780871

Fernando ES, MJM Bande, RA Piollo, DD Sopot, NE Dolotina and WG Granert (2009) Dipterocarpaceae of Bohol Island, Philippines. Asia Life Sciences 18(1): 121-138

Gamboa-Lapitan P and JO Hyun (2005) Mating system of Parashorea malaanonan (M. Blanco) Merr. (Bagtikan) in Mt. Makiling, Laguna, Philippines. Philippine Agricultural Scientist 88(1): 109-121

Holleley CE and PG Geerts (2009) Multiplex Manager 1.2: a cross platform computer program that plans and optimizes multiplex PCR. BioTechniques 46(7): 511-517. https://doi.org/10.2144/000113156

IUCN (2018) The IUCN Red List of Threatened Species. Version 2018-1 [online]. Available at <http://www.iucnredlist.org> [cited 8/28/2018]

Kalinowski ST, ML Taper and TC Marshall (2007) Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16: 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Langenberger G (2006) Habitat distribution of dipterocarp species in the Leyte Cordillera: an indicator for species – site suitability in local reforestation programs. Annals of Forest Science 63: 149–156. https://doi.org/10.1051/forest:2005107

Lee SL, N Tani, KKS Ng and Y Tsumura (2004) Isolation and characterization of 20 microsatellite loci for an important tropical tree Shorea leprosula (Dipterocarpaceae) and their applicability to S. parvifolia. Molecular Ecology Notes 4: 222-225. https://doi/pdf/10.1111/j.1471-8286.2004.00623.x

Lemes MR, RPV Brondani and D Grattapaglia (2002) Multiplexed systems of microsatellite markers for genetic analysis of Mahogany, Swietenia macrophylla King (Meliaceae), a Threatened Neotropical Timber Species. Journal of Heredity 93(4): 287-290. https://doi.org/10.1093/jhered/93.4.287

Murray MG and WF Thompson (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8(19): 4321-4325. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC324241/

Ng KKS, SL Lee, Y Tsumura, S Ueno, CH Ng and CT Lee (2009) Expressed sequence tag-simple sequence repeats isolated from Shorea leprosula and their transferability to 36 species within the Dipterocarpaceae. Molecular Ecology Resources 9: 393-398. https://doi/pdf/10.1111/j.1755-0998.2008.02238.x

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources 8: 103-106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Selkoe KA and RJ Toonen (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9(5):615–629. https://doi/epdf/10.1111/j.1461-0248.2006.00889.x

Ueno S, Y Moriguchi, K Uchiyama, T Ujino-Ihara, N Futamura, T Sakurai, K Shinohara and Y Tsumura (2012) A second generation framework for the analysis of microsatellites in expressed sequence tags and the development of EST-SSR markers for a conifer, Cryptomeria japonica. BMC Genomics 13: 136. https://doi.org/10.1186/1471-2164-13-136

Ujino T, T Kawahara, Y Tsumura, T Nagamitsu, H Yoshimaru and W Ratnam (1998) Development and polymorphism of simple sequence repeat DNA markers for Shorea curtisii and other Dipterocarpaceae species. Heredity 81(4): 422-428. https://doi.org/10.1046/j.1365-2540.1998.00423.x

Vieira MLC, L Santini, AL Diniz and CF Munhoz (2016) Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology 39(3):312-328. https://doi.org/10.1590/1678-4685-gmb-2016-0027

Villarin RA (2015) Microsatellite DNA loci reveal successful cross-genus amplification in a critically endangered Parashorea malaanonan. Journal Science, Engineering and Technology 3:153-163