

POLYMORPHIC MICROSATELLITE MARKERS FOR A WIND-DISPERSED TROPICAL TREE SPECIES, *TRIPLARIS CUMINGIANA* (POLYGONACEAE)¹

NA WEI^{2,4} AND CHRISTOPHER W. DICK^{2,3}

²Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, Michigan 48109-1048 USA; and ³Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Balboa, Ancón, Republic of Panama

- *Premise of the study:* Novel microsatellite markers were characterized in the wind-dispersed and dioecious neotropical tree *Triplaris cumingiana* (Polygonaceae) for use in understanding the ecological processes and genetic impacts of pollen- and seed-mediated gene flow in tropical forests.
- *Methods and Results:* Sixty-two microsatellite primer pairs were screened, from which 12 markers showing five or more alleles per locus (range 5–17) were tested on 47 individuals. Observed and expected heterozygosities averaged 0.692 and 0.731, respectively. Polymorphism information content was between 0.417 and 0.874. Linkage disequilibrium was observed in one of the 66 pairwise comparisons between loci. Two loci showed deviation from Hardy–Weinberg equilibrium. An additional 14 markers exhibiting lower polymorphism were characterized on a smaller number of individuals.
- *Conclusions:* These microsatellite markers have high levels of polymorphism and reproducibility and will be useful in studying gene flow and population structure in *T. cumingiana*.

Key words: gene flow; microsatellite marker; PacBio sequencing platform; single-molecule real-time sequencing; *Triplaris cumingiana*; wind dispersal.

Triplaris cumingiana Fisch. & C. A. Mey. ex C. A. Mey. (Polygonaceae) is a wind-dispersed, dioecious tree species found in humid forests of lower Central America and western South America (Croat, 1978). It forms an obligate mutualistic relationship with the stinging ant *Pseudomyrmex triplarinus* (Croat, 1978) as an antiherbivore defense. Unlike most dioecious tree species that have inconspicuous unisexual flowers (Bawa and Opler, 1975), flower sexual dimorphism is pronounced in *T. cumingiana*, which produces bright red bracts signaling flowers on female trees during the dry season in Panama. The dioecious mating system, which permits unambiguous identification of the maternal and paternal contribution to seedlings, combined with the ease of sexual determination, make *T. cumingiana* of particular interest for studies of pollen- and seed-mediated gene flow in tropical tree species. To investigate the ecological and genetic impacts of pollen and seed dispersal in *T. cumingiana* as compared to tropical trees of alternative pollination and dispersal syndromes, we developed polymorphic microsatellite markers in *T. cumingiana*. We used single-molecule real-time sequencing (SMRT) implemented in the PacBio RS platform (Pacific Biosciences, Menlo Park, California, USA) because it is capable of generating long reads (Wei et al., 2014).

METHODS AND RESULTS

Genome shotgun sequences were obtained using PacBio's high accuracy mode of circular consensus sequencing (Wei et al., 2014). In brief, genomic DNA from one reproductive-sized tree of *T. cumingiana* that grows in the 50-ha Forest Dynamics Plot (FDP) on Barro Colorado Island, Panama (9°10'N, 79°51'W; tag no. 199017), was used for PacBio 500-bp SMRTbell library preparation. Following sonication, DNA fragments averaging 500 bp were ligated with two 55-nucleotide hairpin adapters, and then sequenced on the PacBio RS platform using C2 chemistry. Four SMRT cells generated a total of 178,122 circular consensus reads. Quality control was performed to remove homopolymer-rich sequences and poor-quality portions of individual reads (for details, see Wei et al., 2014). The resulting high-quality sequences were used for microsatellite searching and primer design in QDD version 2.1 (Megléc et al., 2010). In total, 795 microsatellite loci were obtained, in which 686 loci contained pure repeat motifs (524 di-, 143 tri-, 15 tetra-, three penta-, and one hexanucleotide repeat motifs). From these pure microsatellites, loci of ≥ 9 repeats with dinucleotide motifs and loci of ≥ 7 repeats with other repeat motifs were retained for marker validation. This test array was composed of 62 microsatellite loci (32 di-, 25 tri-, three tetra-, one penta-, and one hexanucleotide motifs).

To test these primers, we isolated genomic DNA using a modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987) from lyophilized leaves of 47 *T. cumingiana* adult trees growing in the 50-ha FDP (voucher no. Pérez 1862, Smithsonian Tropical Research Institute herbarium [STRI], Panama). After the initial check of primer amplification on three individuals, we found 39 primer pairs generated easily interpretable allelic patterns, of which 11 loci were monomorphic. We then tested the remaining 28 polymorphic markers on another nine individuals, for which two loci showing apparent null alleles were excluded from further analyses. Microsatellite loci of ≥ 5 alleles based on these 12 samples were screened on an additional 35 individuals. The 8- μ L PCR reactions contained 4 ng DNA, 6.25 nM HEX-labeled or 9.40 nM FAM-labeled M13 primer (TGTAACACGACGCCAGT), 0.075 μ M M13-tagged forward primer, 0.3 μ M reverse primer, 4 mM MgCl₂, 4 μ L GoTaq Colorless Master Mix (Promega Corporation, Madison, Wisconsin, USA) including 200 μ M of each dNTP and 1 unit *Taq* DNA polymerase, and H₂O. PCRs were carried out using two different thermocycling conditions. For most of the tested primers, we used a touchdown protocol (a; Table 1): 94°C for 4 min; 28 cycles of

¹Manuscript received 20 June 2014; revision accepted 22 July 2014.

This work was supported by a CTFS-ForestGEO grant from the Smithsonian Tropical Research Institute and the Center for Tropical Forest Science.

⁴Author for correspondence: weina@umich.edu

doi:10.3732/apps.1400051

TABLE 1. Characteristics of 12 polymorphic microsatellite markers developed in *Triplaris cumingiana*.

Locus	Primer sequences (5'–3') ^a	Repeat motif	Allele size range (bp)	T _a (°C) ^b	GenBank accession no.
TRI_01	F: GGCTTTAATTCACCATTAGCC R: TTGCATCCACACCTAGCAAC	(AAT) ₈	337–418	b	KF680412
TRI_07	F: GCCTGACATGATCAAATCCTC R: TTTCAATTGTTGACGGGATG	(ACAT) ₈	220–364	b	KF680415
TRI_09	F: GAAGTTGGCAGTCGAGGTTTC R: CAAGCTCCAAACTCCCTCAG	(AAAG) ₈	194–242	a	KF680417
TRI_20	F: ATTTGCCACTCCGCTACTTG R: CTCATCATACGATGGCGTTC	(AAG) ₉	196–217	a	KF680422
TRI_26	F: ATAGCCTCTAGCCCGACCTG R: GGGCTCTTCTGCTAGGGTTC	(ACATAT) ₇	196–238	a	KF680426
TRI_27	F: TCCCTCAGACTGTCCAAAGC R: AGCCAATTGATTGGTTTCAAG	(AAG) ₁₇	154–238	a	KF680427
TRI_31	F: GCAAATCATAATTGGGCTTACC R: CTGCCCTAAACGATCTCACC	(AT) ₉	200–224	b	KF680430
TRI_38	F: TGGCTTGACTTGTCGATGTG R: CCACAATTACAAACCACAAAG	(AT) ₁₂	109–127	b	KF680432
TRI_40	F: TACACGGGAGCTTGATTTC R: ATAAACCTAGGCACGGAGGC	(AG) ₁₀	232–254	a	KF680433
TRI_45	F: TCATGAGGGAAGATGAGTTCCG R: AAATAAATGGGCACGATAGC	(AG) ₂₆	106–122	a	KF680437
TRI_49	F: GTCGGCCTGCTTCTTTCTC R: TGCGACTTGTAACGCAACG	(AG) ₁₉	123–149	a	KF680440
TRI_55	F: AACCCCTGACGATCATTGC R: CAATTTGAAGCAAGCTGAGTG	(AG) ₁₇	288–304	a	KF680444

Note: T_a = annealing temperature.

^aM13 tail (TGTA AACGACGGCCAGT) added to the 5' end of each forward primer.

^ba = 59°C (decreasing 0.2°C per cycle) in a touchdown PCR protocol; b = 54.5°C in a nontouchdown PCR protocol (see text for details).

94°C for 30 s, 59°C (decreasing 0.2°C per cycle) for 40 s, and 72°C for 60 s; 10 cycles of 94°C for 30 s, 53°C for 40 s, and 72°C for 60 s; and a final extension at 72°C for 10 min. When the above protocol produced weak PCR amplicons, we followed a nontouchdown protocol (b; Table 1): 94°C for 4 min; 28 cycles of 94°C for 30 s, 54.5°C for 40 s, and 72°C for 60 s; 10 cycles of 94°C for 30 s, 51.5°C for 40 s, and 72°C for 60 s; and a final extension at 72°C for 10 min. PCR amplicons were multiplexed by combining one HEX-labeled locus of 1.6 μL and one FAM-labeled locus of 1.4 μL, with 11.5 μL Hi-Di formamide (Life Technologies, Carlsbad, California, USA) and 0.05 μL GeneScan 500 ROX Size Standard (Life Technologies), before loading to a single lane on an ABI 3730 DNA Analyzer (Life Technologies). Alleles were called using GeneMarker version 2.4.1 (SoftGenetics, State College, Pennsylvania, USA).

We examined marker characteristics including allelic richness, observed and expected heterozygosity, and probability of exclusion (PE₂, when one parent was known; PE₃, of a parent pair) using GenAlEx version 6.5 (Peakall and Smouse, 2012). Polymorphism information content (PIC) was estimated using PowerMarker version 3.25 (Liu and Muse, 2005). Exact tests of Hardy–Weinberg equilibrium (HWE) and locus pairwise linkage disequilibrium (LD) were conducted in GENEPOP version 4.2.2 (Rousset, 2008). The *P* values of HWE and LD tests were adjusted for multiple comparisons using Holm's correction (Holm, 1979).

We first focused on 12 markers with an average allelic richness of nine (range 5–17) (Table 2). Observed heterozygosity ranged from 0.426 to 0.936 (mean = 0.692), and expected heterozygosity was between 0.475 and 0.885 (mean = 0.731). Locus PIC averaged 0.705, and overall exclusion probability was 0.998 (with one parent known) and 1.000 (of a parent pair). We observed LD between the loci TRI_27 and TRI_31 (Holm's adjusted *P* = 0.038). Two loci (TRI_26 and TRI_38) showed HWE deviation (Holm's adjusted *P* < 0.007). These 12 microsatellite markers should provide resolution for studying gene flow and genetic structure in *T. cumingiana*. In addition, we provide information on 14 less-polymorphic loci (2–4 alleles per locus tested on 12 individuals) (Appendix 1), which can be potential candidate markers if more genetic information is required.

CONCLUSIONS

We characterized novel microsatellite markers in the dioecious, insect-pollinated, wind-dispersed tropical tree *T. cumingiana* to

improve understanding of the processes of pollen- and seed-mediated gene flow in tropical forests. This will be done in parallel with studies of tree species with alternative pollination and dispersal syndromes. These markers will also be useful for studying the ecological responses of *T. cumingiana* (e.g., dispersal, recruitment) to rapid changes in temperature and rainfall patterns, as the distribution of this species is associated with

TABLE 2. Summary statistics of microsatellite marker polymorphism tested on 47 reproductively mature trees of *Triplaris cumingiana*, growing in the 50-ha Forest Dynamics Plot on Barro Colorado Island, Panama.

Locus	A	H _o	H _e	PE ₂	PE ₃	PIC
TRI_01	13	0.915	0.877	0.605	0.909	0.866
TRI_07	17	0.800	0.847	0.554	0.890	0.835
TRI_09	8	0.766	0.740	0.366	0.759	0.717
TRI_20	7	0.830	0.775	0.398	0.769	0.747
TRI_26	8	0.511*	0.691	0.303	0.700	0.664
TRI_27	15	0.936	0.885	0.626	0.920	0.874
TRI_31	7	0.652	0.790	0.419	0.786	0.762
TRI_38	9	0.511*	0.825	0.484	0.834	0.804
TRI_40	8	0.660	0.655	0.269	0.670	0.631
TRI_45	6	0.489	0.481	0.127	0.466	0.454
TRI_49	5	0.809	0.736	0.317	0.666	0.688
TRI_55	5	0.426	0.475	0.116	0.384	0.417
Mean	9.0	0.692	0.731	0.998 [§]	1.000 [§]	0.705

Note: A = number of alleles per locus; H_e = expected heterozygosity; H_o = observed heterozygosity; PE₂ = probability of exclusion with one parent known; PE₃ = probability of exclusion of a parent pair; PIC = polymorphism information content.

* Significant deviation from Hardy–Weinberg equilibrium after Holm's correction (adjusted *P* < 0.007).

[§] Cumulative probability of exclusion over multiallelic loci.

high soil phosphorous content and high dry-season intensity (Condit et al., 2013).

LITERATURE CITED

BAWA, K. S., AND P. A. OPLER. 1975. Dioecism in tropical forest trees. *Evolution* 29: 167–179.

CONDIT, R., B. M. J. ENGELBRECHT, D. PINO, R. PÉREZ, AND B. L. TURNER. 2013. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. *Proceedings of the National Academy of Sciences, USA* 110: 5064–5068.

CROAT, T. B. 1978. Flora of Barro Colorado Island. Stanford University Press, Stanford, California, USA.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65–70.

LIU, K. J., AND S. V. MUSE. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics (Oxford, England)* 21: 2128–2129.

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.

PEAKALL, R., AND P. E. SMOUSE. 2012. GenAIEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics (Oxford, England)* 28: 2537–2539.

ROUSSET, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

WEI, N., J. B. BEMMELS, AND C. W. DICK. 2014. The effects of read length, quality and quantity on microsatellite discovery and primer development: From Illumina to PacBio. *Molecular Ecology Resources* 14: 953–965.

APPENDIX 1. Fourteen additional polymorphic microsatellite markers of *Triplaris cumingiana* screened on 12 individuals sampled from the 50-ha Forest Dynamics Plot on Barro Colorado Island, Panama. PCRs follow a touchdown protocol (see text for details).

Locus	Primer sequences (5'–3')	Repeat motif	Allele size range (bp)	A	H _o	H _e	PIC	GenBank accession no.
TRI_06	F: CCTTCCAAACAAGGCTTACC R: GGTCTTGGATCAGCTGAAGG	(ACT) ₇	272–281	2	0.083	0.080	0.077	KF680414
TRI_13	F: TGTGTATACCACAAAGCCGAAG R: TCTTCAATCGTTCTGCCTCC	(AGG) ₁₅	115–118	2	0.667	0.444	0.346	KF680419
TRI_28	F: TCAAACGATACATTCCATTCTG R: TTGGAATGTTAGGATTGGCG	(AAT) ₁₄	108–114	2	0.167	0.444	0.346	KF680428
TRI_30	F: AAAGGGAGGAGAAGAATGGTG R: TCTGCATGGTGTCTCATAAAC	(AAT) ₁₁	125–212	4	0.250	0.358	0.338	KF680429
TRI_36	F: GGAGTTGACTTGCAATTGGG R: TCATACCCAGTTAACCCATGC	(AG) ₉	163–177	2	0.222	0.444	0.555	KF680431
TRI_44	F: TTTAGCCACAATTGCTCAAGAC R: AAAGATCGTCGTTCTCCAC	(AT) ₂₈	148–166	4	0.250	0.705	0.651	KF680436
TRI_51	F: CATGTACCAACTGAACCTGTC R: CTCTTGACCGACCGACGAG	(AC) ₂₀	147–153	3	0.455	0.368	0.425	KF680441
TRI_52	F: TTTCTTGGGTAATTAGTGAGGG R: TAATCCCTGTAGCGTAATCCC	(AG) ₂₃	115–121	4	0.182	0.380	0.446	KF680442
TRI_54	F: GTTTGACCAAGGTTGACCAG R: GGGAAAGAACAAGAAGGAAGG	(AG) ₂₆	121–123	2	0.083	0.080	0.077	KF680443
TRI_56	F: CTAATCGATTGAGGTTCGTGG R: TTGGCAGCAATCTAAGTCCC	(AG) ₁₅	138–142	3	0.818	0.661	0.652	KF680445
TRI_57	F: CAGCTGCTATTGCTCTCAGC R: TATTTCCAACCAATCTCCCG	(AG) ₁₄	147–155	3	0.833	0.517	0.420	KF680446
TRI_58	F: GAACATCCCAACAACATCCC R: TAGTGGTCGGCAAGCTAGTG	(AG) ₁₄	266–268	2	0.636	0.483	0.471	KF680447
TRI_59	F: GGTGGATGTGGCAGTGTAG R: GATCCGAAATTTGCCGTTAC	(AG) ₁₃	190–192	2	0.667	0.444	0.346	KF680448
TRI_62	F: TAGCGACGGATAAGCTAGGG R: TTATTCTGCCATCACCGCTC	(AG) ₁₁	175–187	3	0.091	0.169	0.281	KF680450

Note: A = number of alleles per locus; H_e = expected heterozygosity; H_o = observed heterozygosity; PIC = polymorphism information content.