Applications in Plant Sciences 2013 1(4): 1200295

Primer Note

Polymorphic microsatellite loci for *Virola sebifera* (Myristicaceae) derived from shotgun 454 pyrosequencing[1]

Na Wei[2,8], Christopher W. Dick[2,3], Andrew J. Lowe[4,5], and Michael G. Gardner[4,6,7]

[2]Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, Michigan 48109-1048 USA; [3]Smithsonian Tropical Research Institute, P.O. Box 0843-03092, Balboa, Ancón, Republic of Panama; [4]Australian Centre for Evolutionary Biology and Biodiversity, School of Earth and Environmental Science and Environment Institute, University of Adelaide, North Terrace, Adelaide, South Australia 5001, Australia; [5]Science Resource Centre, Department of Environment and Natural Resources, Adelaide, South Australia 5005, Australia; [6]School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia; and [7]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, South Australia 5001, Australia

- *Premise of the study:* Polymorphic microsatellite loci were characterized in the dioecious neotropical rainforest tree *Virola sebifera*. The markers will be used to study ecological and genetic impacts of hunting and landscape change in this vertebrate-dispersed, insect-pollinated tree species.
- *Methods and Results:* Simple sequence repeat (SSR) markers were screened from genomic libraries of South American *V. sebifera* obtained by shotgun 454 pyrosequencing. Primer pairs were tested on Panamanian samples (N = 42). Approximately 52% of the 61 tested SSR markers amplified, and 16% were polymorphic. Ten selected polymorphic SSR loci contained seven to 15 alleles per locus, and polymorphic information content averaged 0.694. Observed heterozygosity ranged from 0.465 to 0.905, and expected heterozygosity was between 0.477 and 0.876.
- *Conclusions:* The 10 polymorphic loci will be useful in studying gene flow and genetic structure at local and regional spatial scales in *V. sebifera*.

Key words: microsatellite loci; Myristicaceae; shotgun 454 pyrosequencing; *Virola sebifera*.

Neotropical nutmeg (*Virola sebifera* Aubl.; Myristicaceae) is a wide-ranging canopy tree found in mature tropical forests from Central America to the Amazon Basin and Guiana Shield. Like other species in its genus, *V. sebifera* is dioecious, pollinated by small insects, and dispersed by vertebrates (primarily large birds) that consume the nutrient-rich red aril covering its seeds (Howe, 1981). Given the high mobility and considerable seed loads of large avian dispersers, seed-mediated gene flow in *V. sebifera* may play an important role in maintaining genetic variation within and among populations. However, as increasing anthropogenic activities (e.g., hunting and landscape change) adversely impact the abundance and/or habitat of frugivores (Wright, 2003; Vetter et al., 2011), it is important to investigate how changing vertebrate densities may impact gene flow and population structure in *V. sebifera* and other tropical forest tree species.

To address these and other questions, we developed a set of polymorphic microsatellite DNA markers for *V. sebifera*, based on genomic DNA libraries obtained from French Guiana samples by shotgun 454 pyrosequencing (Gardner et al., 2011a, b).

METHODS AND RESULTS

Previously developed genomic libraries of *V. sebifera* (Gardner et al., 2011a, b) were obtained using the combined genomic DNA of six French Guiana individuals, sampled from tagged trees in trails or permanent forest inventory plots in three localities: Sentier la Mirande (4°51′N, 52°20′W; tag no. S35, S31), Sentier Rorota (4°52′N, 52°15′W; S104, S110), and Iracoubo (5°25′N, 53°5′W; S230, S235). Genomic DNA was isolated from each individual using NucleoSpin Plant II (Macherey-Nagel, Düren, Germany), then pooled with equal concentrations (~0.8 µg/individual) for subsequent 454 pyrosequencing. Standard GS-FLX Titanium library preparation was adopted. After DNA nebulization, small fragments of length <350 bp were removed. Fragmented DNA was then ligated with MID-tagged (MID5, ACGAGTAGACT) adapters. This barcoded *V. sebifera* DNA library was multiplexed with seven other species in a single run of GS-FLX Titanium, which rendered *V. sebifera* 12.5% of the picotiter plate.

We used the program QDD version 2 (Meglécz et al., 2010), set at default parameters, to search for simple sequence repeat (SSR) loci with ≥5 uninterrupted motif repeats from 90,164 read sequences (mean read length = 367 bp) (Gardner et al., 2011a, b). The SSR marker output was further restricted to A and B primer designs in QDD version 2, so as to exclude loci with complex flanking regions (i.e., containing repeat units). We obtained a total of 526 SSR loci, of which 315 contained dinucleotide motifs, followed by 182 tri-, 21 tetra-, six penta-, and two hexanucleotide motifs. Following the suggestions of Gardner et al. (2011a), we first focused on loci containing at least 10 pure repeat units of di-, tetra-, and pentanucleotide SSR motifs, which were expected to be more polymorphic than other motifs. However, because of an unexpected low rate of amplification success and polymorphism, we also included compound

[1]Manuscript received 14 June 2012; revision accepted 28 August 2012.

The authors thank C. Scotti-Saintagne and I. Scotti for contributing *Virola sebifera* DNA samples and collection information. The work was supported by a Rackham Graduate Student Research Grant from the University of Michigan.

[8]Author for correspondence: weina@umich.edu

doi:10.3732/apps.1200295

Table 1. Characteristics of 10 polymorphic SSR markers developed in *Virola sebifera*.

Locus	Primer sequences (5′–3′)[a]	Repeat motif	Size range (bp)	T_a (°C)	GenBank accession no.
VSE02	F: CGGTAGTCCATTGATTGGCA R: GCTGTCATTGTGCATCTTCCT	$(AG)_{12}$	266–296	55	JX415276
VSE11	F: TATAGATGCCTGCCATTGGA R: TCGTGCGAAATTCCCTTCTA	$(AG)_{10}$	237–267	55	JX415277
VSE30	F: CATGCATGCTGGTCCATA R: TTCAGCATATTCTCATGTTCCA	$(AGT)_{10}$	159–186	55	JX415278
VSE31	F: AACTAGGGCTCTCGCAGCTT R: CCAAAGAAGTGCTCCTCAGC	$(AAT)_{12}$	183–210	55	JX415279
VSE32	F: TGCCCAAGTGGGTTTCTCTA R: CCAGTGTTTCTTCTCTTGCATC	$(AAT)_{15}$	197–221	55	JX415280
VSE36	F: AGACGGATTGAGGAGAAGCC R: CGGAGCACAGGAATGAAATC	$(ACC)_{10}$	222–243	55	JX415281
VSE38	F: CCATTTGCTCTAAGCAATTCATC R: TCACATGCGAATTGTTCACAC	$(ACT)_{14}$	214–253	55	JX415282
VSE42	F: CACCGCTACTGTTTCCTGGT R: GTGGGATGTGCCATAGAAGC	$(AG)_3G(AG)_3G(AG)_{14}$	283–306	55	JX415283
VSE45	F: TGAAATTTGTTCCCTTCTGAGG R: TGATCCATTATTCAGATGAGGC	$(TCA)_5(TCGTCA)_{14}(TCA)_3$	132–163	55	JX415284
VSE55	F: GTTGGAGACTGTCCTCGGTG R: TGCTTAACAGCATGGAATGG	$(AGT)_9$	162–192	55	JX415285

Note: T_a = annealing temperature.
[a]M13 tail (TGTAAAACGACGGCCAGT) added to the 5′ end of each forward primer.

motifs, and tri- and hexanucleotide microsatellite loci of ≥9 repeats. The final testing array contained 61 candidate SSR markers (57% in di-, 36% in tri-, 3% in tetra-, 2% in penta-, and 2% in hexanucleotide motifs).

We checked the amplification rate and polymorphism of the 61 SSR primer pairs in 42 *V. sebifera* adult trees (diameter at breast height [dbh] ≥20 cm; voucher no. *Pérez 1806* and *Pérez 1930*, STRI herbarium, Panama), which were randomly collected from the 50-ha Forest Dynamics Plot in the plateau of Barro Colorado Island (9°10′N, 79°51′W), Panama. Genomic DNA was isolated from silica-dried leaves using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA), quantified using NanoDrop 2000 (Thermo Scientific, Wilmington, Delaware, USA), and diluted to 1.5 ng/μL for subsequent PCR. The 6-μL PCR cocktail contained 1.5 ng of DNA template, 0.05 μM of M13-tagged (5′-TGTAAAACGACGGCCAGT-3′) forward primer, 0.4 μM reverse primer, 0.017 μM 6FAM-labeled M13 primer (5′-TGTAAAACGACGGCCAGT-3′), 4 mM $MgCl_2$, 3 μL GoTaq Colorless Master Mix (Promega Corporation, Madison, Wisconsin, USA) with buffer (pH 8.5), 200 μM of each dNTP, and 1 U *Taq* DNA polymerase. PCRs were carried out in a Mastercycler ep thermocycler (Eppendorf, Hamburg, Germany) following an initial denaturation at 94°C for 4 min; 28 cycles of 94°C for 30 s, 55°C for 40 s, and 72°C for 60 s; 10 cycles of 94°C for 30 s, 52°C for 40 s, and 72°C for 60 s; and a final extension at 72°C for 10 min. PCR product of 1.5 μL was added to 12 μL Hi-Di formamide (Applied Biosystems, Carlsbad, California, USA) and 0.05 μL GeneScan 500 Rox Standard (Applied Biosystems) for subsequent fragment sizing in an ABI 3730 DNA Analyzer (Applied Biosystems) by the DNA Sequencing Core Laboratory at the University of Michigan. Alleles were visualized and scored using GeneMarker version 3.7 (SoftGenetics, State College, Pennsylvania, USA). Marker polymorphism, including the number of alleles per locus, observed and expected heterozygosity, exclusion probability with one parent known, and Hardy–Weinberg equilibrium (HWE), was estimated in GenAlEx version 6.4 (Peakall and Smouse, 2006). Significance levels for multiple tests of HWE (α-level = 0.05) were adjusted by sequential Bonferroni procedure (Rice, 1989). In addition, polymorphism information content of each locus was measured using PowerMarker version 3.0 (Liu and Muse, 2005). We tested for the presence of null alleles, allelic dropout, and scoring errors (due to stuttering) using MICROCHECKER version 2.2.3 (Van Oosterhout et al., 2004).

Our results showed that 17 (49%) di-, 13 (59%) tri-, 1 (50%) tetra-, 0 penta-, and 1 (100%) hexanucleotide markers were amplifiable; but 3 (9%) di-, 6 (27%) tri-, 0 tetra-, 0 penta-, and 1 (100%) hexanucleotide SSRs were considered as polymorphic (≥6 alleles per locus) in the current study. These 10 polymorphic markers (Table 1) had mean allelic richness of 10.3 alleles per locus (Table 2). Observed heterozygosity ranged from 0.465 to 0.905, and expected heterozygosity was between 0.477 and 0.876. PIC per locus averaged 0.694 (Table 2). No allelic dropout or scoring errors were detected, but one locus (VSE02) appeared to contain null alleles. Two (VSE02 and VSE36) of the 10 loci showed deviation from Hardy–Weinberg proportions after sequential Bonferroni correction ($P < 0.006$). The overall exclusion probability with one parent known was 0.992.

CONCLUSIONS

We found that trinucleotide SSR loci exhibited better marker properties, such as higher probability of polymorphism and less stuttering, than the other motifs, particularly dinucleotide SSRs. Although the 454 genomic libraries were obtained from French Guiana samples, the markers were developed for Panamanian individuals, despite the probable high levels of genomic divergence between populations located east and west of the Andean cordilleras. Genomic divergence may partly explain the unexpected low rate of amplification (52%) and polymorphism

Table 2. Summary statistics of SSR marker polymorphism screened in 42 *Virola sebifera* individuals located in the 50-ha Forest Dynamics Plot on Barro Colorado Island, Panama.

Locus	*A*	H_o	H_e	PE	PIC
VSE02	15	0.487	0.876*	0.604	0.864
VSE11	12	0.810	0.834	0.511	0.816
VSE30	9	0.767	0.695	0.306	0.666
VSE31	10	0.762	0.764	0.384	0.734
VSE32	8	0.756	0.732	0.347	0.703
VSE36	7	0.465	0.477*	0.129	0.456
VSE38	12	0.905	0.827	0.491	0.806
VSE42	11	0.571	0.566	0.197	0.549
VSE45	10	0.561	0.544	0.178	0.524
VSE55	9	0.854	0.844	0.519	0.825
Mean	10.3	0.694	0.716	0.992¶	0.694

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

*Significant deviation from Hardy–Weinberg expectations after sequential Bonferroni correction ($P < 0.006$).

¶Probability of exclusion over all loci.

(16%) of the markers. Although one marker (VSE02) showed evidence of null alleles, and one other marker showed deviation from HWE, these markers may perform well in the South American populations. The 10 polymorphic loci characterized here will be useful for studies of gene flow and population structure in this widespread, vertebrate-dispersed, dioecious tree species.

LITERATURE CITED

Gardner, M. G., A. J. Fitch, T. Bertozzi, and A. J. Lowe. 2011a. Rise of the machines: Recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* 11: 1093–1101.

Gardner, M. G., A. J. Fitch, T. Bertozzi, and A. J. Lowe. 2011b. Data from: Rise of the machines: Recommendations for ecologists when using next generation sequencing for microsatellite development. Dryad Digital Repository. doi:10.5061/dryad.f1cb2.

Howe, H. F. 1981. Dispersal of a Neotropical nutmeg (*Virola sebifera*) by birds. *Auk* 98: 88–98.

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. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution; International Journal of Organic Evolution* 43: 223–225.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.

Vetter, D., M. M. Hansbauer, Z. Vegvari, and I. Storch. 2011. Predictors of forest fragmentation sensitivity in Neotropical vertebrates: A quantitative review. *Ecography* 34: 1–8.

Wright, S. J. 2003. The myriad consequences of hunting for vertebrates and plants in tropical forests. *Perspectives in Plant Ecology, Evolution and Systematics* 6: 73–86.