Applications in Plant Sciences

Applications in Plant Sciences 2013 1(12): 1300050

Primer Note

Development of genomic microsatellites in *Gleditsia triacanthos* (Fabaceae) using Illumina sequencing[1]

Sandra A. Owusu[2], Margaret Staton[3], Tara N. Jennings[4], Scott Schlarbaum[5], Mark V. Coggeshall[6], Jeanne Romero-Severson[7], John E. Carlson[8], and Oliver Gailing[2,9]

[2]School of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, Michigan 49931 USA; [3]Clemson University Genomics Institute, Clemson, South Carolina 29634 USA; [4]Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon 97331 USA; [5]Department of Forestry, Wildlife and Fisheries, University of Tennessee, Knoxville, Tennessee 37996-4563 USA; [6]School of Natural Resources, University of Missouri, Columbia, Missouri 65211 USA; [7]Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556 USA; and [8]Department of Ecosystem Science and Management and Department of Plant Science, Pennsylvania State University, University Park, Pennsylvania 16802 USA

- *Premise of the study:* Fourteen genomic microsatellite markers were developed and characterized in honey locust, *Gleditsia triacanthos*, using Illumina sequencing. Due to their high variability, these markers can be applied in analyses of genetic diversity and structure, and in mating system and gene flow studies.
- *Methods and Results:* Thirty-six individuals from across the species range were included in a genetic diversity analysis and yielded three to 20 alleles per locus. Observed heterozygosity and expected heterozygosity ranged from 0.214 to 0.944 and from 0.400 to 0.934, respectively, with minimal occurrence of null alleles. Regular segregation of maternal alleles was observed at seven loci and moderate segregation distortion at four of 11 loci that were heterozygous in the seed parent.
- *Conclusions:* Honey locust is an important agroforestry tree capable of very fast growth and tolerance of poor site conditions. This is the first report of genomic microsatellites for this species.

Key words: agroforestry; Fabaceae; *Gleditsia triacanthos*; microsatellite; next-generation sequencing.

Honey locust (*Gleditsia triacanthos* L.), a common leguminous tree native to the eastern and central United States, occurs on rich bottomlands and rocky upland slopes and is a frequent invader of abandoned fields (Schnabel et al., 1998). It is used in land reclamation efforts due to its fast growth and tolerance of poor site conditions (Preston and Braham, 2002). Honey locust populations are characterized by wide genetic variation in adaptive traits such as winter hardiness in northern races and more nutritious fruits in the south. Application of molecular genetic techniques such as marker-assisted selection can be used in enhancing germplasm selection efficiently compared to traditional breeding procedures. Development of genetic resources in this important species will aid studies of genetic diversity among its populations, and will help identify genes underlying desirable traits of interest in developing sustainable management strategies for this important but underutilized species. Identification of genetic diversity within and among regions in honey locust through assessment of genetic variation in different environments is also important for efficient selection of genotypes for land restoration purposes.

To date, no microsatellite resources have been developed in honey locust for characterizing its genetic resources. Gene-based microsatellite markers (expressed sequence tag–simple sequence repeats [EST-SSRs]) developed for related species such as *Medicago truncatula* Gaertn., *Ceratonia siliqua* L., and *Copaifera officinalis* (Jacq.) L. show low transferability and are not polymorphic in honey locust (data not shown). Next-generation sequencing is now frequently used for easy and rapid development of microsatellite markers across many different taxa (Jennings et al., 2011), reducing time and costs for sample processing and sequencing. Using low-coverage, paired-end Illumina genome sequencing of *G. triacanthos*, we characterized 14 nuclear microsatellite markers and assessed their variability in 36 samples from a provenance trial.

METHODS AND RESULTS

Low-coverage whole genome sequencing (Jennings et al., 2011) was used to produce an initial set of genomic resources for 10 hardwood tree species, including honey locust (Staton et al., in prep.). Illumina libraries were created from sonicated genomic DNA extracted from leaflets of one individual (seed parent of 88 single-tree progeny, Appendix 1) of *G. triacanthos* per the manufacturer's protocol (QIAGEN DNeasy96 Plant Kit; QIAGEN, Hilden, Germany). Libraries were constructed using Illumina TruSeq version 2 index sequencing adapters (Illumina, San Diego, California, USA), and then pooled in an equimolar mixture and sequenced using 101-bp paired-end chemistry on an Illumina HiSeq 2000 at the Oregon State University Center for Gene Research and Biocomputing. After sequencing, reads for individual libraries were sorted by index. Because the input DNA was sheared to a

[1]Manuscript received 6 June 2013; revision accepted 23 July 2013.

The authors thank Nick Wheeler for his guidance and contributions, and Richard Cronn, Brian Knaus, and Kimberly Hansen for assistance with Illumina library construction and sequencing. This research was supported by the National Science Foundation Plant Genome Research Program (grant no. TRPGRA2 IOS-1025974). Additional funding was provided by the Ecosystem Science Center and the Biotech Research Center of Michigan Technological University.

[9]Author for correspondence: ogailing@mtu.edu

doi:10.3732/apps.1300050

TABLE 1. Characteristics of 14 novel genomic microsatellite markers developed in *Gleditsia triacanthos*.[a]

Locus	Primer sequences (5′–3′)[b]	Amplicon size (bp)	Repeat motif	T_a (°C)	Size range (bp)
GLT002	F: NED-TAAAAAGTAACCTTAAAGG	104	$(AT)_9$	56	103–147
	R: AGTAAAGAGGTAACGATTT				
GLT021	F: 6-FAM-ATATCACCAATTTAAGACC	100	$(AG)_{11}$	56	94–98
	R: GTACACAAAACTTCGAGAG				
GLT026	F: VIC-AAGCTTGATTAGAGAAATT	127	$(AT)_{14}$	56	113–143
	R: AGATAGTTCCTTTCAGTTG				
GTT057	F: PET-CAGGTAAAACATGAGATTGATGC	121	$(TA)_9$	56	127–157
	R: TTCCATAAAATCAGTCATGCAA				
GTT063	F: NED-CTCTTGCGCACACTAAAACG	116	$(AC)_{12}$	56	147–187
	R: CGTACGGTGACACTTGTGC				
GTT073	F: VIC-CATGATTTAGAGAGAGAAATGTTTTGG	109	$(GA)_{13}$	56	134–164
	R: AACCAAGCCCTTCATTTATGG				
GTT114	F: 6-FAM-TCAAGCTAGTTAGCCTTCCTGC	121	$(TC)_{20}$	56	102–138
	R: AAATATGGGAGCAATGAACC				
GTT116	F: NED-CTAAAGCTTGACTTCTGAATCC	134	$(CT)_8$	56	131–143
	R: CGCTATATCGGAATCCCTGC				
GTT117	F: PET-GGTGGTATGTGCAAGCAAGC	120	$(TA)_8$	56	110–122
	R: CTTGAGCCACCCATTACCC				
GTT118	F: 6-FAM-CAGTCCCACCTTCACTAGCC	119	$(CT)_8$	56	110–130
	R: TGCGTGTAATCTGAGCTTGG				
GTT126	F: PET-TGGATTAAGTTGTAAAGCGAGG	109	$(AT)_8$	56	98–146
	R: CCGTCAAACTTAAGACCCACC				
GTT131	F: 6-FAM-CTTTGAACTCTAATACTCTGGTTGC	100	$(AC)_9$	56	91–153
	R: TCAACCACCTTAAGACATCCC				
GTT132	F: VIC-CAGTCCTCATGTCTAGTCTAGTGC	105	$(AT)_{11}$	56	90–130
	R: CAATCTCTGGTGCAAGATGC				
GLT4027	F: 6-FAM-AGGAATTATTCTCTACCAA	107	$(TCCA)_6$	56	91–107
	R: CGAATCTCATTTTATACAA				

Note: T_a = annealing temperature.

[a] Geographic coordinates for the provenances are given in Appendix 1.

[b] The fluorescent label is shown with the forward primer.

modal length of ~160 bp, paired sequences were joined into overlapping extended contigs using FLASH (Magoč and Salzberg, 2011) with default settings. Using an input of 14,888,028 paired-end sequences, FLASH constructed 13,775,803 contigs that ranged in size from 93 to 191 bp. An SSR finder script (Staton et al., in prep.) for di-, tri-, and tetranucleotide repeats identified 61,086 microsatellite motifs. Microsatellites were defined as a 2-bp motif repeated eight to 40 times, a 3-bp motif repeated seven to 30 times, or a 4-bp motif repeated six to 20 times. Using the program CAP3 (Huang and Madan, 1999), redundant sequences were filtered from the SSR-containing sequences (identity ≥95%), leaving only putatively unique loci. The filtered reads were assessed with Primer3 (Rozen and Skaletsky, 2000) to identify primers using default program settings with slight modifications: melting temperature = 54°C minimum, 58°C optimum, and 62°C maximum; amplicon size = 100 bp minimum, 200 bp maximum; and primer length = 17 bp minimum, 19 bp optimum, and 25 bp maximum. A total of 4715 primer pairs flanking microsatellite motifs were identified (4084 di-, 544 tri-, and 87 tetranucleotide motifs). The genomic SSR data are publicly available through the National Center for Biotechnology Information (NCBI) Short Read Archive.

Amplification and polymorphism of primers for 108 dinucleotide and 36 tetranucleotide repeat motifs were assessed in a panel of seven unrelated individuals (seed parent and six potential pollen donors of the 88 single-tree progeny; Appendix 1) after electrophoretic separation on the QIAxcel Fast Analysis System using the QIAxcel DNA High Resolution Kit for microsatellite analysis (QIAGEN). Polymorphic loci were amplified in 36 samples from a provenance experiment (Kellogg Forest, Michigan, 28 provenances, latitudinal range: 30°11′N–42°45′N, longitudinal range: 76°19′W–106°37′W; Appendix 2) and in 88 single-tree progeny using fluorescent-labeled forward primers (6-FAM, PET, NED, and VIC). Amplification products were separated on an ABI Prism Genetic Analyzer 3730 (Applied Biosystems, Foster City, California, USA) and scored with GeneMapper version 4.0 (Applied Biosystems). PCRs were performed in a 15-μL reaction mix that contained 3 μL of 5× HOT FIREPol Blend Master Mix Ready to Load (contains 10 mM $MgCl_2$, 0.6 units of HOT FIREPol *Taq* polymerase, and 2 mM dNTPs; Solis BioDyne, Tartu, Estonia), 2 μL each of 5 μM fluorescent-labeled forward (Applied Biosystems) and reverse primers (Sigma-Aldrich, St. Louis, Missouri, USA), 6 μL double deionized water (DNase- and RNase-free), and 2 μL DNA (~1.8 ng/μL). Amplification was carried out in a Peltier Thermal Cycler (GeneAmp PCR system 2700, Applied Biosystems). The PCR profile was as follows: 15 min denaturation at 95°C, followed by 35 cycles of 45 s denaturation at 94°C, a 45 s annealing step at the annealing temperature (Table 1), a 45 s elongation at 72°C, and a final extension step at 72°C for 20 min. Observed (H_o) and expected (H_e) heterozygosities (Nei, 1973) and number of alleles (A) were calculated in GENEPOP version 4.0.10 (Raymond and Rousset, 1995). Pairwise linkage disequilibrium for all loci was also calculated in GENEPOP.

All 144 primer pairs amplified products in the expected size range and 14 were polymorphic in the set of seven unrelated individuals after electrophoretic separation on the QIAxcel Fast Analysis System (QIAGEN). Using the diversity panel of 36 individuals from the species distribution range (Appendix 2),

TABLE 2. Genetic properties of the 14 novel microsatellite markers in *Gleditsia triacanthos*.

Locus	A	H_o	H_e
GLT002	19	0.886	0.914
GLT021*	3	0.333	0.549
GLT026*	11	0.735	0.864
GTT057*	11	0.778	0.829
GTT063*	12	0.667	0.793
GTT073*	12	0.944	0.825
GTT114*	17	0.889	0.877
GTT116	6	0.214	0.716
GTT117*	6	0.257	0.400
GTT118*	7	0.639	0.705
GTT126	20	0.853	0.934
GTT131*	9	0.778	0.821
GTT132*	17	0.639	0.895
GLT4027*	5	0.676	0.729

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity.

* Primers characterized in the seed parent and 88 progeny.

the 14 microsatellite markers showed relatively high levels of polymorphism with number of alleles per locus ranging between three and 20 (Table 2). Genomic microsatellites are associated with high levels of polymorphism due to their occurrence in the less conserved untranscribed regions of DNA. H_o ranged from 0.214 to 0.944 and H_e from 0.400 to 0.934 (Table 2). H_o and H_e were similar in the samples for each locus, except for GTT116 (Table 2), which had a high number of missing data, indicating low incidence of null alleles for most of the 14 microsatellite markers. No significant linkage disequilibrium was detected between markers ($P < 0.05$) after Bonferroni correction. Out of the 14 loci, 11 were heterozygous in the seed parent of 88 progeny, and regular segregation of the maternal alleles was assessed in the progeny using a χ^2 test. Segregation distortion was observed for GLT026 and GTT131 ($P < 0.05$), and for GTT117 and GTT4027 ($P < 0.01$). Distorted segregation of alleles could be attributed to a variety of both genetic and physiological factors, including pollen-tube competition, pollen lethals, preferential fertilization, and elimination of zygotes (Lu et al., 2002).

CONCLUSIONS

This is the first report of genomic microsatellites for *G. triacanthos*. The high levels of polymorphism at the 14 loci are especially useful for gene flow and mating system analyses in a species that is functionally dioecious. The microsatellites will also facilitate the study of the effect of isolation and fragmentation on genetic variation and structure in *G. triacanthos* populations.

LITERATURE CITED

Huang, X., and A. Madan. 1999. CAP3: A DNA sequence assembly program. *Genome Research* 9: 868–877.

Jennings, T. N., B. J. Knaus, T. D. Mullins, S. M. Haig, and R. C. Cronn. 2011. Multiplexed microsatellite recovery using massively parallel sequencing. *Molecular Ecology Resources* 11: 1060–1067.

Lu, H., J. Romero-Severson, and R. Bernardo. 2002. Chromosomal regions associated with segregation distortion in maize. *Theoretical and Applied Genetics* 105: 622–628.

Magoč, T., and S. L. Salzberg. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)* 27: 2957–2963.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA* 70: 3321–3323.

Preston Jr., R. J., and R. R. Braham. 2002. North American trees. Iowa State Press, Ames, Iowa, USA.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

Schnabel, A., J. D. Nason, and J. L. Hamrick. 1998. Understanding the population genetic structure of *Gleditsia triacanthos* L.: Seed dispersal and variation in female reproductive success. *Molecular Ecology* 7: 819–832.

Appendix 1. Geographic coordinates of *Gleditsia triacanthos* seed parent (784) and potential pollen parents (Butternut Valley, Memphis, Tennessee, USA) used for microsatellite marker screening.

Accession no.	Locality	Latitude	Longitude
784*	DeKalb, Tennessee, USA	35°54′46.607″N	85°54′32.083″W
780	DeKalb, Tennessee, USA	35°54′47.090″N	85°54′31.678″W
781	DeKalb, Tennessee, USA	35°54′47.090″N	85°54′31.678″W
782	DeKalb, Tennessee, USA	35°54′47.256″N	85°54′32.114″W
783	DeKalb, Tennessee, USA	35°54′46.697″N	85°54′31.244″W
785	DeKalb, Tennessee, USA	35°54′51.779″N	85°54′29.778″W
786	DeKalb, Tennessee, USA	35°54′50.392″N	85°54′30.755″W

* Individual used to make the Illumina library.

Appendix 2. Geographic coordinates of *Gleditsia triacanthos* provenances from a provenance experiment, Kellogg Forest, Michigan, USA.

Accession no.	Locality	Latitude	Longitude
60	Bulloch, Georgia, USA	32°28′N	81°46′W
90	Franklin, Kentucky, USA	38°10′N	84°52′W
91	Franklin, Kentucky, USA	38°10′N	84°52′W
109	Hickman, Kentucky, USA	36°45′N	89°05′W
110	Hickman, Kentucky, USA	36°45′N	89°05′W
154	Washington, Texas, USA	30°11′N	96°37′W
160	Washington, D.C., USA	38°55′N	77°00′W
161	Lancaster, Pennsylvania, USA	40°01′N	76°19′W
166	Huntington, Pennsylvania, USA	40°22′N	77°54′W
169	East Carroll, Louisiana, USA	32°47′N	91°10′W
218	Chester, South Carolina, USA	34°43′N	81°13′W
234	Aiken, South Carolina, USA	33°34′N	81°43′W
247	Franklin, Kansas, USA	37°45′N	95°10′W
258	Bienville, Louisiana, USA	32°32′N	92°55′W
261	Bienville, Louisiana, USA	32°32′N	92°55′W
278	Warren, Ohio, USA	39°25′N	84°12′W
281	Delaware, Ohio, USA	40°22′N	82°57′W
300	Fairfax, Virginia, USA	38°51′N	77°19′W
305	Ogle, Illinois, USA	42°01′N	89°20′W
327	Monroe, Arkansas, USA	34°40′N	91°19′W
337	Texas, Missouri, USA	37°31′N	91°50′W
340	Phelps, Missouri, USA	37°55′N	91°55′W
341	Phelps, Missouri, USA	37°55′N	91°55′W
342	Larimer, Colorado, USA	40°34′N	105°04′W
366	Story, Iowa, USA	42°01′N	93°32′W
367	Story, Iowa, USA	42°01′N	93°32′W
370	Erie, Ohio, USA	41°27′N	82°42′W
387	Burke, North Carolina, USA	35°45′N	81°46′W
422	Bernalillo, New Mexico	35°04′N	106°37′W
444	Polk, Iowa, USA	41°43′N	93°35′W
445	Polk, Iowa, USA	41°43′N	93°35′W
447	Ingham, Michigan, USA	42°45′N	84°30′W
457	Monongalia, West Virginia, USA	39°37′N	79°57′W
461	Piatt, Illinois, USA	39°47′N	88°37′W
465	Piatt, Illinois, USA	39°47′N	88°37′W