Ann. For. Res. 60(2): 343-350, 2017
DOI: 10.15287/afr.2017.819

ANNALS OF FOREST RESEARCH
www.afrjournal.org

Development of genic and genomic microsatellites in *Gleditsia triacanthos* L. (Fabaceae) using Illumina sequencing

Y. Wu, R. Zhang, M. Staton, S.E. Schlarbaum, M.V. Coggeshall, J. Romero-Severson, J.E. Carlson, N. Zembower, H. Liang, Y. Xu, D.I. Drautz-Moses, S.C. Schuster, O. Gailing

Wu Y., Zhang R., Staton M., Schlarbaum S.E., Coggeshall M.V., Romero-Severson J., Carlson J.E., Zembower N., Liang H., Xu Y., Drautz-Moses D.I., Schuster S.C., Gailing O., 2017. Development of genic and genomic microsatellites in *Gleditsia triacanthos* L. (Fabaceae) using Illumina sequencing. Ann. For. Res. 60(1): 343-350.

Abstract. Twenty new polymorphic genic SSRs (EST-SSRs) and 13 genomic SSRs were developed in honeylocust (*Gleditsia triacanthos*) using Illumina transcriptome and low-coverage genome sequencing. A diversity panel of 40 honeylocust samples covering large parts of the species distribution range was characterized. As expected the level of genetic variation was lower in EST-SSRs than for non-genic genomic SSRs. This is the first report of EST-SSRs for honeylocust. All markers are polymorphic and produce clear single locus amplification products and can be used for genetic diversity and gene flow analyses. The transcriptome sequencing data provide a rich resource for new marker development.

Keywords EST-SSRs, transcriptome, *Gleditsia triacanthos*, next-generation sequencing

Authors. Yawen Wu, Ruhua Zhang - Michigan Technological University, School of Forest Resources and Environmental Science, 1400 Townsend Drive, Houghton, MI 49931; Margaret Staton - Department of Entomology and Plant Pathology, The University of Tennessee, Knoxville, TN 37996; Scott E. Schlarbaum - Department of Forestry, Wildlife and Fisheries, The University of Tennessee, Knoxville, TN 37996-4563; Mark V. Coggeshall - USDA Forest Service, Northern Research Station, Purdue University, West Lafayette, IN 47907; Jeanne Romero-Severson - Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556; John E. Carlson, Nicole Zembower - The Department of Ecosystem Science and Management and Department of Plant Science, Pennsylvania State University, University Park, PA 16802; Haiying Liang - Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634; Yi Xu - Department of Plant Biology and Pathology, Rutgers University, New Brunswick, NJ 08901; Daniela I. Drautz-Moses, Stephan C. Schuster - The Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634, USA; Oliver Gailing (ogailing@mtu.edu) - University

of Göttingen, Forest Genetics and Forest Tree Breeding, Büsgenweg 2, 37077 Göttingen, Germany.
Manuscript received April 27, 2017; revised June 26, 2017; accepted June 28, 2017; online first June 29, 2017.

Introduction

Honeylocust, *Gleditsia triacanthos* L., is an outcrossing tree species of the family Fabaceae. It has a broad natural range across the central and eastern United States and occurs on a range of soil types, tolerating poor site conditions (Preston & Braham 2002). Honeylocust is also a popular fast-growing landscape plant and is suitable for planting in disturbed environments (Schnabel & Wendel 1998, Preston & Braham 2002). It is a member of the subfamily Caesalpinioideae, for which only a limited number of markers and other genomic resources are available (La Malfa et al. 2014, Gailing et al. 2017). The transfer of genic microsatellites (Expressed Sequence Tag- Simple Sequence Repeats (EST-SSRs)) developed for the related species *Ceratonia siliqua* L. (La Malfa et al. 2014) to honeylocust was not successful (Owusu et al. 2013). Earlier, we developed 14 genomic SSRs in honeylocust using low-coverage whole genome sequencing (Owusu et al. 2013). All of these markers were applicable for genetic variation assessments in a reference panel of 36 samples collected from a provenance trial (Owusu et al. 2013). However, due to signatures of null alleles in some seed parents, only six to nine of these markers were used for gene flow analyses in honeylocust population fragments (Owusu et al. 2016).

While we have already developed a limited number of non-genic microsatellite markers for honeylocust (Owusu et al. 2013), no transcriptome resources and gene-based markers such as EST-SSRs have been published for this species. The development of transcriptome resources such as stress-induced transcriptome libraries is especially important since hardwood species including honeylocust are increasingly threatened by abiotic stresses such as drought and heat, especially at their southern distribution edge. As part of the Hardwood Genomics project (http://www.hardwoodgenomics.org/) we are developing genome/transcriptome resources and genetic linkage maps in eastern North American hardwood species from different taxonomic orders. For honeylocust, we have developed genomic SSRs (Owusu et al. 2013) for genetic variation and gene flow analyses (Owusu et al. 2016) and a framework genetic linkage map for Quantitative Trait Locus (QTL) and comparative mapping based on restriction site associated sequencing (RAD-seq) markers (Gailing et al. 2017).

Here we report for the first time the development of 20 polymorphic EST-SSRs and of additional 13 polymorphic genomic SSRs for honeylocust. In addition, we describe the development of transcriptome resources for candidate gene identification and for new marker development.

Materials and methods

Root tissues of heat, cold, drought and unstressed honeylocust half-sib seedlings were used to generate five transcriptome libraries as described in Lane et al. (2016). After TruSeq Stranded mRNA library preparation we performed paired-end sequencing of the five libraries on an Illumina MiSeq (San Diego, Ca) with a read length of 151 bp and deposited raw reads in the NCBI Short Read Archive (SRA) (BioProject accession number PRJNA273269). Error correction, trimming and assembly of reads were performed as described in Lane et al. (2016). Open Reading Frames (ORFs) were predicted with TransDecoder 2.01 (Haas et al. 2013). Functional annotation was performed using BLAST version 2.229 queries against

the plant Swiss-Prot and TrEMBL protein databases at a cutoff e-value <1e-4 (Camacho et al. 2009, Magrane & UniProt 2011). Putative unique transcripts (PUTs) for SSR identification were identified using a Perl script (https://github.com/mestato/lab_code/tree/master/hwg_gssr_scripts). BUSCO (Benchmarking Universal Single-Copy Orthologs) v1.1 was used to assess transcriptome completeness (Simao et al. 2015).

For genomic SSR development, Illumina low-coverage genome sequencing was used to generate 14,888,028 paired-end sequences assembled into 13,775,803 contigs in a size range from 93-191 bp (Owusu et al. 2013, Staton et al. 2015). A total of 4715 primer pairs flanking SSR motifs were identified (Owusu et al. 2013).

Primer 3 v2.3.6 (Untergasser et al. 2012) was used to design primers with the settings described in Owusu et al. (2013).

Four honeylocust, samples from different populations were chosen to screen 25 EST-SSR and 14 nSSR primer pairs (Tables 1-2, Supporting Information). Polymorphic markers were then used to run PCR on the additional 36 samples representing the species distribution range. All 40 samples were collected from a provenance trial (Kellogg Forest, Michigan, 28 provenances, latitudinal range: 30°11′N-42°45′N, longitudinal range: 76°19′W-106°37′ W; Table 3, Supp. Info.) (Owusu et al. 2013). The PCR mix (15 µl) consisted of 6µl double deionized water (Ultra Pure Water, Molecular Grade from Phenix Research Products), 5µl HotFIREPol (Oak Biotechnologies, containing 2mM dNTPs, 1U Taq polymerase, 10mM $MgCl_2$), 0.2µl of 5 µM forward primer with M13 tail (5’-CACGACGTTGTAAAACGAC-3’), 1.5µl of 5µM 5’ dye labelled (6-FAM) M13 primer, 0.5µl of 5 µM pig-tailed (5’-GTTTCTT-3’) reverse primer (Sigma Aldrich Inc., St. Louis, MO) (Schuelke 2000, Kubisiak et al. 2013) and 2µl DNA (~ 2ng). The PCR touchdown reaction was performed in an Applied Biosystems 2720 Thermal Cycler (Foster City, CA) with the following cycling conditions: initial denaturation at 95 ºC for 15 minutes, 10 touch-down cycles of 1 minute at 94 ºC, 1 minute at 60 ºC (decreasing 1 ºC each cycle), and 1 minute at 72 ºC, followed by 25 cycles of 1 minute denaturation at 94 ºC, a 1 minute annealing step at 50 ºC, 1 minute elongation at 72 ºC, and a final extension at 72 ºC for 20 minutes. Amplification products were tested on 1.5% agarose gels stained with 2µl GelRed (10,000X in water; Biotium, Hayward, CA). Exact fragment sizes were determined by electrophoretic separation on an ABI Prism® Genetic Analyzer 3730 with Gene-ScanTM LIZ-500 as internal size standard. GeneMarker® V2.6.7 (SoftGenetics) was used to assign alleles to bins through careful visual inspection. Genetic variation parameters observed and expected heterozygosity (H_o, H_e, respectively) and number of alleles per locus (A) were calculated using GenAlEx 6.502 (Peakall and Smouse 2006, Peakall and Smouse 2012). An allele text file was exported for GENEPOP analysis. The inbreeding coefficient (F) was estimated for each locus in the population using GENEPOP version 4.2 online (http://www.genepop.curtin.edu.au/) (Raymond & Rousset 1995, Rousset 2008). Pairwise linkage disequilibrium was tested for all marker pairs in GENEPOP. Bonferroni corrections were applied to adjust for multiple testing.

Results

De novo transcriptome assembly of 7,059,396 read pairs generated 56,845 PUTs with an average length of 731 bp. We identified 30,372 ORFs from 28,642 of the PUTs. In the 56,845 PUTs we found 458 SSRs for 377 di-, 78 tri- and 3 tetra-nucleotide motifs. Primer pairs were designed for 282 SSRs, 221 for di-, 58 for tri- and 3 for tetranucleotide motifs. For genomic SSRs, a total of 4715 primer pairs flanking SSR motifs were identified and 14 polymorphic SSRs were developed earlier (Owusu et al. 2013). However, testing of these markers

in different single-tree progenies revealed null alleles for eight of these markers in some progenies limiting the application for gene flow analyses (Owusu et al. 2016). In the present study we describe the development of additional 33 polymorphic markers (20 EST-SSRs, 13 genomic SSRs).

Twenty out of the 25 EST-SSRs amplified single polymorphic loci, four did not amplify and one locus was monomorphic. For genomic SSRs, 13 out of 14 markers were polymorphic and one was monomorphic (Tables 1-2). As expected for genic markers, the genetic variation observed at EST-SSRs (H_e =0.541, H_o = 0.465, A = 5.05) was lower than for genomic SSRs (H_e = 0.605, H_o = 0.545, A = 6.50) (Table 2). The level of genetic variation differed considerably among individual markers for both genomic and EST-SSRs (Table 2). After excluding genomic SSR GTT 172 with very low variation (H_e = 0.080, H_o = 0.000, A =2), mean values for all three parameters increased for genomic SSRs (H_e = 0.646, H_o = 0.587, A = 6.846). For EST-SSRs, variation in trinucleotide repeat SSRs (H_e = 0.424, H_o = 0.370, A = 3.166) was lower than in dinucleotide repeat SSRs (H_e = 0.7155, H_o = 0.607, A = 7.875).

Significant linkage disequilibrium at the 5% level after Bonferroni correction was only detected for one marker pair (GTT143 and GT_TRI_EST45).

Discussion

Based on comparison to a plant database of 956 conserved single copy plant gene groups, the honeylocust transcriptome contains full length members of 571 groups (60%), partial length members for 229 partial groups (24%), and no representation for 155 groups (16%). This indicates the transcriptome has likely captured the majority of genes but is not completely comprehensive. Flanking primers were identified for a total of 282 SSRs according to our selection criteria (Owusu et al. 2013). The amplification success (84%) for EST-SSRs was very high indicating that these new transcriptome resources are a valuable resource for EST-SSR marker development. Transcriptome sequence data will also provide a rich resource for the identification of Single Nucleotide Polymorphism (SNP) markers. In addition to 14 markers already developed (Owusu et al. 2013), we also describe 13 new polymorphic genomic SSRs derived from low coverage genome sequencing (Staton et al. 2015). Genetic variation estimates at genomic SSRs were slightly lower but comparable to estimates (He = 0.775, Ho = 0.663, A =11.071) for 14 other genomic SSRs that were characterized in the same sample set (Owusu et al. 2013). The new genomic and EST-SSRs can be used for the assessment of genetic variation and for gene flow analyses in honeylocust (Owusu et al. 2016). Additionally, genetic characterization at these markers can be performed in populations from contrasting environments or along environmental gradients to test for selective neutrality of markers. Finally, a framework genetic linkage map was generated in honeylocust using restriction site associated DNA sequencing (RAD-seq) derived SNPs (Gailing et al. 2017). The map consists of 178 SNPs distributed across the 14 haploid chromosomes of the honeylocust genome. The new transcriptome resources are also valuable for the development and genetic mapping of SNPs and EST-SSRs in candidate genes for stress response. Quantitative trait locus (QTL) and association mapping approaches can be used to associate variation in candidate genes with adaptive phenotypes.

Conclusions

Here we report genic SSRs (EST-SSRs) for *G. triacanthos* for the first time. In addition, we describe 13 new polymorphic genomic SSRs adding additional markers for genetic diversity and gene flow analyses in *G. triacanthos* populations.

Table 1 Characteristics of 14 novel genomic (gSSRs) and 20 novel genic microsatellite markers (EST-SSRs) for *Gleditsia triacanthos*

Locus	Primer sequences (5′-3′)[b]	Amplicon size (bp)	Repeat motif	T_a (°C)	Size range (bp)
GTT 143*	F:GCATTGCAACTCTAACTGTGC R:TCTCTCTGTTGGCTTGTTGG	133	$(TC)_{12}$	60	140-164
GTT 154*	F:CATATATGGTTCGCTGCTTGG R:AGCTTAATCTCCACTTGAATCC	116	$(TA)_8$	60	130-134
GTT 156*	F:AAGTGGTATCAAAGCAGGTCG R:TTTGGAACAAATTACCCACC	142	$(TA)_{13}$	60	143-169
GTT 166*	F:CGCTCAAGAAGATAAAGGCG R:TTTGTAGTTTGCCACATCTGC	144	$(GT)_8$	60	154-186
GTT 171*	F:GGTTAGTTAGTTTCTGACCACTGC R:ACCAAAGAAGAATCTGCTTGC	134	$(AC)_{11}$	60	139-157
GTT 172*	F:ACATCATATTATGTTCGGGCTCC R:ACTCATCCTTATTACTCGTGGG	142	$(TG)_8$	60	152-158
GTT 179*	F:TCTTGTGTATTGCCTAACTTGC R:GTCAGAATTGTTCTCCGACGG	170	$(TA)_8$	60	178-190
GTT 180*	F:CCAGAAGTCAGCAAGATATTCCC R:GCAAGTTTCTGAATACTTGCG	138	$(AT)_9$	60	150-178
GTT 184*	F:TAGATACACGGCAGGGAGGG R:GTGGAGATCTAGATTCGAGTCCC	106	$(GA)_9$	60	117-121
GTT 185*	F:GCCATTCTTATGCATGTGGAGG R:AAACATGTATGGCTCGCTGC	109	$(TA)_8$	60	126-156
GTT 189*	F:GTTTCAAGGTGAGAAAGGCTCC R:AAAGAGCGCAATCCTTTACG	104	$(AG)_{12}$	60	102-122
GTT 190*	F:TGATTGAATGGCAAGCGTAAAGC R:ACCGTCATAAGCAATACAACCC	146	$(TA)_{11}$	60	151-159
GTT 193*	F:GCCTAATACCAGCCAAAGCC R:ACCCAACTACCATTCTTCGCC	158	$(AG)_9$	60	172-180
GTT 196*	F:AGAGAGGGATGATAAGATATCCTCC R:CTTGCTAAGGGATTGGGTTGC	146	$(AG)_{16}$	60	148-156
GT_TRI_EST4	F:CCAGTTGCCGTATCCAACCT R:AAAGACCCGTTCCACTCACG	152	$(AGA)_7$	60	170-179
GT_TRI_EST5	F:TGATTAGAGAGGCCCTGGCA R:ACAAACCCATCACCGGTCTC	239	$(GCA)_7$	60	250-259
GT_TRI_EST7	F:CCATGGAAGCTGCAGAGGAT R:GCAGTCTGCAATGGCTTCAG	298	$(CAG)_7$	60	308-320
GT_DI_EST21	F:TGTAGGGCACGCAGAAAGAG R:TGTTGCCACCTGATGATGCT	152	$(TC)_9$	60	170-180
GT_DI_EST23	F:GCTTGTATTGACTGCCTCATGG R:GCTACCAGATTACACGGCCA	129	$(CT)_9$	60	148-190

Table 1 (continuation)

Locus	Primer sequences (5′-3′)[b]	Amplicon size (bp)	Repeat motif	T_a (°C)	Size range (bp)
GT_DI_EST24	F:AACCCATCCCTCTCCTCCTT R:ATTGAGGCTTGGAAGGACGG	284	$(TC)_8$	60	304-330
GT_DI_EST26	F:AATCTCCCGAGTCCTACCGG R:GAGGAGTCTGTGCTGGATCAG	138	$(TC)_8$	60	159-179
GT_DI_EST27	F:CGCTTGAGCCATGAAGAGGA R:GATGTAGGTCATCCCTGCCG	155	$(GA)_{10}$	60	164-188
GT_DI_EST29	F:TCGGATGGTTCAACAGCTCC R:TGGGATTCCTAACGCCTCTT	343	$(AG)_8$	60	362-374
GT_DI_EST30	F:ACCCACATGATAGCAATCCA R:ACCACTCTTTTCTTGTCGGT	159	$(TC)_8$	60	180-192
GT_DI_EST32	F:ATTCAGAAAAGGGAGGCGGG R:CACTCCTTCACTTGTGTTTTCA	339	$(GT)_8$	60	361-369
GT_TRI_EST36	F:AGAATGGAACTGCACTCGGG R:GGTTGAGTCCCACCAGTTGT	217	$(TAC)_7$	60	238-241
GT_TRI_EST37	F:ATCTTGCTGCGTATCCTCGG R:GACGGCGAGACCATAGAGTG	357	$(CGC)_7$	60	379-382
GT_TRI_EST38	F:TGGGTGCAATTTCGGCTTTG R:CCCCTTGATGCGTAGTGGAG	328	$(CCT)_7$	60	348-351
GT_TRI_EST39	F:CAGTAGCACCTTGAGCACCA R:GAGGACGATGTTCCCGTTGT	121	$(TCG)_7$	60	145-151
GT_TRI_EST40	F:GCGTGTTGAACCAGCAAGAG R:TTCCCCAAATCACAGCCTCC	444	$(GGA)_7$	60	391-409
GT_TRI_EST41	F:GGTGGCACTTGAGAGGTCTC R:GAAAAGTGCACCAAGCCCTG	220	$(CTA)_7$	60	315-324
GT_TRI_EST42	F:TGAGGAGATGCAGAAAGGCG R:TGGGTTTTGCTTTGGTCCCA	415	$(GAA)_7$	60	429-438
GT_TRI_EST43	F:TCATGGACTTCTTGGTGCCC R:AGCCATGGGAGGAAATGCAA	175	$(TCA)_7$	60	200-206
GT_TRI_EST45	F:CTGAGCGGTGAGTGTCTTGT R:GTGTGAGAAGCAAAAGCGCA	243	$(TGG)_7$	60	267-276

Note. Abbreviations: * - genomic microsatellites, the remaining markers are genic microsatellites (EST-SSRs), T_a - annealing temperature.

Acknowledgements

The study was funded by the National Science Foundation Plant Genome Research Program (# TRPGRA2 IOS-1025974). Additional support was made available by the Ecosystem Science Center and the Biotech Research Center of Michigan Technological University. Ruhua Zhang was funded by a scholarship from Linyi University, China. We would like to thank Majid Ghafarian for help with the lab work.

Table 2 Genetic variation at genomic and genic microsatellites in 40 samples from 28 provenances

Locus	A	H_o	H_e
GTT 143	9	0.750	0.813
GTT 154	3	0.087	0.232
GTT 156	10	0.909	0.778
GTT 166	8	0.429	0.604
GTT 171	5	0.500	0.750
GTT 172	2	0.000	0.080
GTT 179	6	0.833	0.725
GTT 180	12	0.667	0.875
GTT 184	3	0.500	0.442
GTT 185	13	0.409	0.894
GTT 189	8	0.708	0.757
*GTT 190	2	1.000	0.500
GTT 193	5	0.458	0.635
GTT 196	5	0.375	0.389
Mean genomic SSRs	6.5	0.545	0.605
SE	0.92	0.278	0.237
GT_TRI_EST4	4	0.542	0.484
GT_TRI_EST5	4	0.333	0.591
GT_TRI_EST7	5	0.522	0.579
GT_DI_EST21	5	0.321	0.585
GT_DI_EST23	13	0.750	0.887
GT_DI_EST24	9	0.643	0.730
GT_DI_EST26	9	0.821	0.740
GT_DI_EST27	8	0.714	0.793
GT_DI_EST29	7	0.464	0.731
GT_DI_EST30	7	0.786	0.783
GT_DI_EST32	5	0.357	0.475
GT_TRI_EST36	2	0.107	0.101
GT_TRI_EST37	2	0.071	0.069
GT_TRI_EST38	2	0.429	0.497
GT_TRI_EST39	3	0.607	0.460
GT_TRI_EST40	2	0.296	0.252
GT_TRI_EST41	4	0.429	0.538
GT_TRI_EST42	3	0.250	0.482
GT_TRI_EST43	3	0.393	0.478
GT_TRI_EST45	4	0.464	0.558
Mean EST-SSRs	5.05	0.465	0.541
SE	0.64	0.046	0.047

Note. Abbreviations: A - number of different alleles, H_e - expected heterozygosity, H_o - observed heterozygosity, SE - Standard error, * - monomorphic markers.

References

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L., 2009. BLAST plus: architecture and applications. BMC Bioinformatics 10:421. DOI: 10.1186/1471-2105-10-421

Gailing O., Staton M.E., Lane T., Schlarbaum S.E., Nipper R., Owusu S.A., Carlson J.E., 2017. Construction of a framework genetic linkage map in *Gleditsia triacanthos* L. Plant Molecular Biology Reporter 35:177-187. DOI: 10.1007/s11105-016-1012-0

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., Leduc R.D., Friedman N., Regev A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8:1494-1512. DOI: 10.1038/nprot.2013.084

Kubisiak T.L., Nelson C.D., Staton M.E., Zhebentyayeva T., Smith C., Olukolu B.A., Fang G.C., Hebard F.V., Anagnostakis S., Wheeler N., Sisco P.H., Abbott A.G., Sederoff R.R., 2013. A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*). Tree Genetics & Genomes 9:557-571. DOI: 10.1007/s11295-012-0579-3

La Malfa S., Curro S., Douglas A.B., Brugaletta M., Caruso M., Gentile A., 2014. Genetic diversity revealed by EST-SSR markers in carob tree (*Ceratonia siliqua* L.). Biochemical Systematics and Ecology 55:205-211. DOI: 10.1016/j.bse.2014.03.022

Lane T., Best T., Zembower N., Davitt J., Henry N., Xu Y., Koch J., Liang H., McGraw J., Schuster S., Shim D., Coggeshall M.V., Carlson J.E., Staton M.E., 2016. The green ash transcriptome and identification of genes responding to abiotic and biotic stresses. BMC genomics 17:702. DOI: 10.1186/s12864-016-3052-0

Magrane M., UniProt C., 2011. UniProt Knowledgebase: a hub of integrated protein data. Database-the Journal of Biological Databases and Curation. 10.1093/database/bar009 DOI

Owusu S.A., Schlarbaum S., Carlson J.E., Gailing O., 2016. Gene flow analyses and identification of full-sib families in isolated populations of *Gleditsia triacanthos* L. Botany-Botanique 94:523-532. DOI: 10.1139/cjb-2015-0244

Owusu S.A., Staton M., Jennings T.N., Schlarbaum S., Coggeshall M.V., Romero-Severson J., Carlson J.E., Gailing O., 2013. Development of genomic microsatellites in *Gleditsia triacanthos* (Fabaceae) using Illumina Sequencing. Applications in Plant Sciences 1:1300050. DOI: 10.3732/apps.1300050

Peakall R., Smouse P.E., 2006. GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295. DOI: 10.1111/j.1471-8286.2005.01155.x

Peakall R., Smouse P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539. DOI: 10.1093/bioinformatics/bts460
Preston Jr. R.J., Braham R.R. 2002. North American trees. Iowa State Press, Ames, Iowa, USA.
Raymond M., Rousset F., 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249. DOI: 10.1093/oxfordjournals.jhered.a111573
Rousset F., 2008. GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8:103-106. DOI: 10.1111/j.1471-8286.2007.01931.x
Schnabel A., Wendel J.F., 1998. Cladistic biogeography of *Gleditsia* (Leguminosae) based on ndhF and rpl16 chloroplast gene sequences. American Journal of Botany 85:1753-1765. DOI: 10.2307/2446510
Schuelke M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233-234. DOI: 10.1038/72708
Simao F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212. DOI: 10.1093/bioinformatics/btv351
Staton M., Best T., Khodwekar S., Owusu S., Xu T., Xu Y., Jennings T., Cronn R., Arumuganathan A.K., Coggeshall M., Gailing O., Liang H., Romero-Severson J., Schlarbaum S., Carlson J.E., 2015. Preliminary genomic characterization of ten hardwood tree species from multiplexed low coverage whole genome sequencing. Plos One 10:e0145031. DOI: 10.1371/journal.pone.0145031
Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G., 2012. Primer3 - new capabilities and interfaces. Nucleic Acids Research 40: e115. DOI: 10.1093/nar/gks596

Supporting Information

The online version of the article includes Supporting Information:

Table 1. EST-SSRs, dinucleotide and trinucleotiderepeats for honeylocust (unigene version 082614)
Table 2. Genomic SSRs
Table 3. Geographic coordinates of *Gleditsia triacanthos* provenances