



**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2017

DEVELOPMENT OF GENE-BASED MICROSATELLITE MARKERS IN *Acer saccharum* MARSH.

Monica Harmon

Michigan Technological University, harmon@mtu.edu

Copyright 2017 Monica Harmon

Recommended Citation

Harmon, Monica, "DEVELOPMENT OF GENE-BASED MICROSATELLITE MARKERS IN *Acer saccharum* MARSH.", Open Access Master's Report, Michigan Technological University, 2017.
<http://digitalcommons.mtu.edu/etdr/434>

Follow this and additional works at: <http://digitalcommons.mtu.edu/etdr>



Part of the [Biotechnology Commons](#), [Genetics Commons](#), and the [Other Forestry and Forest Sciences Commons](#)

**DEVELOPMENT OF GENE-BASED MICROSATELLITE MARKERS IN *Acer
saccharum* MARSH.**

**By
Monica Harmon**

A REPORT

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Forest Molecular Genetics and Biotechnology

MICHIGAN TECHNOLOGICAL UNIVERSITY

2017

© 2017 Monica Harmon

This report has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Forest Molecular Genetics and Biotechnology.

School of Forest Resources and Environmental Science

Report Advisor: *Oliver Gailing*

Committee Member: *Victor Busov*

Committee Member: *Blair Orr*

School Dean: *Terry Sharik*

Table of Contents

Abstract	5
Chapter I: Introduction.....	6
Importance of sugar maple	6
Susceptibility to climate change.....	6
Genetic marker analyses in sugar maple	7
Microsatellite markers.....	7
Applications of genetic markers in sugar maple management.....	8
Objective	9
Background	9
Genetic variation analyses in sugar maple.....	9
Methods.....	11
PCR Touchdown method.....	11
Results/Discussion	12
Outlook.....	13
Acknowledgements	13
References	15
Chapter 2.....	17
Development of novel genic microsatellite markers from transcriptome sequencing in sugar maple (<i>Acer sacharrum</i> Marsh.)	17
Abstract	18
Preface.....	19
Findings.....	19
Background	19
Methods.....	21
Plant materials for microsatellite analysis	21
Library preparation, transcriptome sequencing and assembly.....	21
Functional annotation and microsatellite discovery	22
Marker development and genetic variation study.....	23
Results and discussion.....	25
Transcriptome assembly and validation.....	25
Functional annotation and microsatellite discovery	26
Microsatellite marker characterization	26

Authors' contributions.....	28
Author details	29
Acknowledgements	29
Competing Interests.....	30
Availability of data and materials	30
Funding.....	30
References	34

Abstract

Sugar maple (*Acer saccharum* Marsh) is an important ecological and economic resource. Despite its significance, there are few molecular genetic resources available to date. This study used previously developed EST libraries generated for sugar maple and other hardwood species to develop and characterize EST-SSRs in sugar maple and test their transferability to other *Acer* species. Out of 58 markers tested, 22 showed amplification in the expected size range. From these, 16 polymorphic markers were selected to test transferability and 13 of these showed amplification in at least one other *Acer* species.

Chapter I: Introduction

Importance of sugar maple

Sugar maple (*Acer saccharum* Marsh) is a valuable hardwood tree species for North America and brings in millions of dollars in revenue in the timber industry and millions more in profit from maple syrup derived from its sap (USDA/NASS 2015). Its timber is of high economic importance for example, use in basketball courts(ESPN 2017). It is also an ornamental tree: its foliage exhibits vibrant colors during autumn. Sugar maple is a deciduous tree with a native range in eastern North America from southern Tennessee to southern Canada reaching its limits when temperate forest mixes with boreal forest (Graignic et al. 2013). It produces a double winged samara and is a mainly outcrossing species (Geburek 1993). The tree is a valuable part of the ecosystems it occupies. Sugar maples play an important role in their ecosystems for example they exhibit hydraulic redistribution. This phenomenon occurs in sugar maples via taproots up-taking water from deep in the soil and depositing it in shallower soils (Dawson 1993). Furthermore, some plants in the vicinity seem to prefer this water specifically (Emerman et al. 1996). Thus, sugar maple may play an important role in sustaining plant-life in their surroundings.

Susceptibility to climate change

As the risks of climate change come to the forefront of science, there is a great need anticipate what effects will harm or alter a tree species' range and/or ability to produce timber. Studies indicate that sugar maple will move northward with climatic change but that the species may have a discontinuous distribution (Graignic et al. 2014). With a

better understanding of what is happening at a population genetics level we can better predict the species' response to a changing climate and associated stresses.

Genetic marker analyses in sugar maple

Genetic resources such as variable microsatellites for sugar maple are currently largely lacking (Graignic et al. 2013). These highly variable nuclear DNA markers can help us enhance studies of gene flow. For example, a limited number of highly variable microsatellite markers have been developed in sugar maple for paternity and gene flow analyses (Graignic et al 2013, Khodwekar et al. 2015). However, gene-based markers such as expressed sequence tag simple sequence repeats (EST-SSRs) have not been developed in sugar maple. Because EST-SSRs were developed from stress-induced Expressed Sequence TAG (EST) libraries, we know these markers are located in genes that potentially respond to stresses based on research conducted under the Hardwood Genomics Project (www.hardwoodgenomics.org). Specifically this will prove valuable as the change in climate may indeed exert abiotic stresses on the trees especially at the southern boundaries of the range. It is important to understand gene flow patterns at the fringes and their effect on genetic variation for the species' adaptation to climate change. For example, decreased gene flow between populations can result in inbreeding depression. This project is the first to use these libraries to develop genetic markers in sugar maple.

Microsatellite markers

Microsatellite markers are highly variable and locus-specific PCR-based markers. These markers are short repeats often called simple sequence repeats (SSRs) and are widespread in eukaryotic genomes (Tautz 1989). These SSRs vary from 2-6 repeated nucleotides up

to 64 bp in length (Conner and Hartl 2004). Due to errors during DNA replication, the microsatellite length will vary in number of repeats (Selkoe and Toonen 2006). So-called slipped strand mispairing is a mechanism by which the DNA misaligns after denaturation during DNA replication. The repeat regions may align themselves on a downstream or upstream repeat in the strand (Ellegren 2000). This causes differences in the repeat number and this allelic diversity is what proves valuable to study genetic variation. Microsatellites are distributed across the genome and primers are designed complementary to regions which flank the microsatellite to generate locus-specific amplification products. Some SSRs are located in expressed genes. The primer binding sites located in expressed genes (ESTs) will remain similar even across species and these loci can thus be applied across species boundaries (Ellis and Burke 2007). While genomic SSRs are hypervariable and present throughout the genome, EST-SSRs are not as variable because of their location in expressed genes (Ellis and Burke 2007). The ESTs in the present thesis come from already developed EST libraries based on research conducted under the Hardwood Genomics Project (www.hardwoodgenomics.org). In this project 51 sugar maple EST libraries were developed and sequenced through Illumina sequencing, each based on different molecular responses to stressors. Three tissue types (root, leaf and petiole) were exposed to ozone, heat, drought and would stress with some control tissue that was unstressed. RNA was extracted from these tissues and sequencing was performed to generate the libraries.

Applications of genetic markers in sugar maple management

The ability to glean valuable information about genetic variation within and among populations will benefit those involved in maintaining and managing populations of sugar

maple trees. For example, these markers can be used to assess gene flow patterns with comparisons made between unmanaged and managed populations of sugar maple. Management strategies can then be applied to mitigate effects of limited gene flow among populations. Finally, EST-SSRs could be characterized along an environmental gradient to identify gene markers with a role in local adaptation. Both genic and non-genic markers could be used on populations of sugar maple with parent and seedling samples from different parts of the species' range.

Objective

The objective of the study was to develop and characterize genic microsatellite markers (EST-SSRs) in sugar maple (*Acer saccharum* Marsh.) and to test their transferability to other *Acer* species.

Background

Genetic variation analyses in sugar maple

Studies in sugar maple show that genetic variation is high and that with emerging genetic technology, analyses of genetic variation in sugar maple have changed over time. Perry and Knowles (1991) used allozyme frequencies to assess genetic variation and specifically analyzed genetic variation within stands. Allozymes were a precursor to PCR-based markers that are soluble enzymes tested in an electrophoresis apparatus to determine genetic variation. They are codominant markers and became very important in population genetic studies (Conner and Hartl 2004). Perry and Knowles (1991) indicated moderate structuring within three natural sugar maple populations in northwestern Ontario. These data are useful but allozymes are less variable than microsatellite markers. Allozymes will not reveal anything about non-coding regions of the genome nor is it easy

to stain for certain structural proteins and making them impossible to examine (Conner and Hartl 2004).

Randomly amplified polymorphic DNA (RAPD) markers can be useful in population genetic analyses. RAPDs are generated from randomly developed primers usually 10bp in length. These primers bind randomly throughout the genome. RAPDs were used in Gunter et al. (2000) and indicated that genetic differentiation was highest among populations in the southern regions of the species' range while northern populations showed much lower variation based on genetic distance, fixation coefficients and differences in allele frequencies. These results indicate that trees in the southern range may be most affected by climate change but also the most adaptable trees. It is hypothesized that the high level of genetic variation in sugar maple could imply a genetic mechanism for adaptation. Studies using RAPDs are difficult to reproduce, interpret since RAPDs are dominant markers leaving no indication of whether the locus is hetero- or homozygous. RAPDs prove valuable but cannot differentiate between coding and non-coding and only dominant traits can be visualized (Eriksson et al. 2006). The hypothesis in Gunter et al. (2000) could be tested with microsatellite markers with higher variability and better reproducibility.

Graignic et al. (2013) developed 23 genomic microsatellite markers in sugar maple and tested them on two populations in Quebec. Additionally, Khodwekar et al. (2015) developed 7 new genomic SSRs for sugar maple which were not transferable to other *Acer* species. These microsatellite markers are useful in population genetics studies and have been used for geographic variation studies and harvesting technique studies (Graignic et al. 2014, Graignic et al. 2016). Genomic microsatellite markers have been

used for gene flow analyses and to assess fine-scale genetic structuring within a population (Khodwekar et al. 2015). However, nuclear microsatellite markers are not located in coding DNA like markers derived from EST libraries. The current study is the first report of EST-SSRs in sugar maple. EST-SSRs are locus-specific and are located in genes with putative function in stress response. They are codominant and tend to be transferable to other species of the same genus.

With the lack of genetic and genomic resources in sugar maple, the present study aims to add these markers for use in sugar and other related *Acer* species.

Methods

PCR Touchdown method

The touchdown method as described in Bodénès et al. (2012) for polymerase chain reaction (PCR) in combination with fluorescent labeling was employed in this study. This label is effective: added to the forward primer, it will attach to the PCR products. With the amount of forward primer less than half the concentration of the reverse primer, the temperature is adjusted to accommodate the reverse primer and continue to be amplified with the subsequent products leading to an easily readable labeled final product (Schuelke 2000). In adding the sequence, GTTTCTTT to the 5' reverse primer as described by Brownstein, Carpten and Smith (1996) the adenylation of the 3' end increases greatly. This method was utilized in this study to ensure better genotyping results.

Once the results of the electrophoretic separation were received, alleles were assigned with careful visual inspection of the electrophoretogram. Genetic variation was assessed using the software GeneA1Ex (Peakall and Smouse 2012) as observed and expected

heterozygosity (H_o , H_e) and number of alleles per locus (N_a). Additionally the inbreeding coefficient (F) was calculated for each marker as potential indication for null alleles.

Another software package (GENEPOP version 4.2, Raymond and Rousset 1995) was used to calculate significant differences from Hardy-Weinberg proportions and linkage disequilibrium. Linkage disequilibrium is the non-random association of alleles and can cause problems in population genetics. Bonferroni corrections were also applied to Hardy-Weinberg proportions and null allele frequencies were tested using FreeNA (Chapuis and Estoup 2007).

See Harmon et al. for exact methods (submitted Nov. 2016 to BMC Research Notes).

Results/Discussion

Out of 58 markers tested, 22 amplified loci in the expected size range. Next, all polymorphic markers (16 markers) were tested on four other *Acer* species for transferability and 13 showed amplification in at least one other *Acer* species. Overall, the markers did not show deviation from HWE or linkage disequilibrium. H_o and H_e were lower than for non-genic markers (Khodwekar et al. 2015). This high level of transferability in EST-SSRs was expected (Ellis and Burke 2007). These are the first EST-SSRs developed in sugar maple but EST-SSRs developed in other hardwood tree species show similar transferability (Lind and Gailing 2013). The lower H_o and H_e in EST-SSRs as compared to genomic SSRs is due to the location of these markers in conserved regions of the genome (i.e. in expressed genes). Two markers showed high inbreeding coefficient values and high frequencies of null alleles. Detailed results are presented in Chapter II.

These markers are the first EST-SSRs to be developed in sugar maple and are suitable for population genetics studies. Furthermore they can be used in the other *Acer* species in which they amplified polymorphic loci.

Outlook

With the availability of both neutral genomic SSRs and genic SSRs with potential role in adaptation, patterns of genetic variation and their relation to environmental and climatic variables can be studied. Also, gene flow could be studied <continue> Because these markers are gene-based, studies of latitudinal variables could be undertaken studying the different adaptations along the gradient. A range-wide analysis would be beneficial to discover what changes in genetic variation occur along environmental gradients. Looking at other populations at the same latitude as this study, including seedling samples, would allow studies of gene flow and the mating system. With this information a genetic map could be generated. Finally, with information about natural populations, those in charge of management populations might assess the diversity of their stands and make choices for best practice based on genetic diversity from natural populations.

Acknowledgements

This project was supported by the Ecosystem Science Center at Michigan Technological University and through the support of the Coverdell Fellowship. The work would not have been possible without the help of Sudhir Khodwekar who isolated DNA and helped interpret the results. The great support of Blair Orr through Monica's time in AmeriCorps VISTA.

References

Bodénès, C., Chancerel, E., Gailing, O., Vendramin, G. G., Bagnoli, F., Durand, J., Goicoechea, P. G., Soliani, C., Villani, F. & Mattioni, C. 2012. Comparative mapping in the Fagaceae and beyond with EST-SSRs. *BMC plant biology*, 12, 153.

Brownstein, M.J., Carpten, J.D. and Smith, J.R. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques*, 20, 1004-6.

Chapuis, M.-P. and Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24, 621-631.

Conner, J. K. & Hartl, D. L. 2004. A primer of ecological genetics, Sinauer Associates Incorporated.

Dawson, T.E. 1993. Hydraulic lift and water use by plants: implications for water balance, performance and plant-plant interactions. *Oecologia*, 95, 565-574.

Ellegren, H., 2000. Microsatellite mutations in the germline:: implications for evolutionary inference. *Trends in genetics*, 16, 551-558.

Ellis, J. & Burke, J. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity*, 99, 125-132.

Emerman, S.H. and Dawson, T.E. 1996. Hydraulic lift and its influence on the water content of the rhizosphere: an example from sugar maple, *Acer saccharum*. *Oecologia*, 108, 273-278.

Eriksson, G., Ekberg, I. and Clapham, D. 2006. An introduction to forest genetics. Genetic Center, Department of Plant Biology and Forest Genetics, SLU.

ESPN. 2017. *From Tree to Tipoff* [Online]. espn.com. Available: http://www.espn.com/espn/feature/story/_/id/18985386/from-tree-tipoff [Accessed April 6 2017].

Geburek, T. (1993). Are genes randomly distributed over space in mature populations of sugar maple (*Acer saccharum* Marsh.)? *Annals of Botany*, 71, 217-222.

Graignic, N., Tremblay, F. & Bergeron, Y. 2013. Development of polymorphic nuclear microsatellite markers in sugar maple (*Acer saccharum* Marsh.) using cross-species transfer and SSR-enriched shotgun pyrosequencing. *Conservation Genetics Resources*, 5, 845-848.

- Graignic, N., Tremblay, F. & Bergeron, Y. 2014. Geographical variation in reproductive capacity of sugar maple (*Acer saccharum* Marshall) northern peripheral populations. *Journal of Biogeography*, 41, 145-157.
- Gunter, L., Tuskan, G. A., Gunderson, C. A. & Norby, R. 2000. Genetic variation and spatial structure in sugar maple (*Acer saccharum* Marsh.) and implications for predicted global-scale environmental change. *Global Change Biology*, 6, 335-344.
- Lind, J.F. and Gailing, O. 2013. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* EJ Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes*, 9, 707-722.
- Khodwekar, S., Staton, M., Coggeshall, M.V., Carlson, J.E. and Gailing, O. 2015. Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). *Annals of Forest Research*, 58, 193-204.
- PE, P.R.S. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28, 2537-2539.
- Perry, D.J. and Knowles, P. 1991. Spatial genetic structure within three sugar maple (*Acer saccharum* Marsh.) stands. *Heredity*, 66, 137-142.
- Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248-249.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18, 233-234.
- Selkoe, K. A. & Toonen, R. J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, 9, 615-629.
- Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research*, 17, 6463-6471.
- USDA/NASS QuickStats Ad-hoc Query Tool. (n.d.). Retrieved February 19, 2017
- Weising, K., Nybom, H., Pfenninger, M., Wolff, K. & Kahl, G. 2005. DNA fingerprinting in plants: principles, methods, and applications, CRC press.

Chapter 2

BMC Research Notes

Development of novel genic microsatellite markers from transcriptome sequencing in sugar maple (*Acer sacharrum* Marsh.)

--Manuscript Draft--

Monica Harmon¹, Thomas Lane², Margaret Staton², Mark V. Coggeshall³, Teodora Best⁴, Chien-Chih Chen^{5,6}, Haiying Liang⁵, Nicole Zembower⁴, Daniela I. Drautz-Moses⁷, Yap Zhei Hwee⁷, Stephan C. Schuster⁷, Scott E. Schlarbaum⁸, John E. Carlson⁴, Oliver Gailing^{1*}

¹School of Forest Resources and Environmental Science, Michigan Technological University, 1400

Townsend Drive, Houghton, MI 49931

²Department of Entomology and Plant Pathology, The University of Tennessee, Knoxville, TN 37996

³Department of Forestry, University of Missouri, Columbia, MO 65274

⁴Department of Ecosystem Science and Management and Department of Plant Science, The Pennsylvania State University, University Park, PA 16802

⁵Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29630.

⁶Current affiliation: College of Agriculture, Fisheries and Forestry, Fiji National University, Nausori, Fiji.

⁷Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551, Singapore

⁸Department of Forestry, Wildlife & Fisheries, The University of Tennessee, Knoxville, TN 37996-4563

*Corresponding author: ogailing@mtu.edu

Abstract

Background: Sugar maple (*Acer saccharum* Marsh.) is a hardwood tree species native to Northeastern North America and economically valued for its wood and sugar syrup. Yet, few genetic resources have been developed for this species to date. Microsatellite markers have been a useful tool in population genetics, e.g. to monitor genetic variation and to analyze gene flow patterns. The aim of this study is to make available a reference transcriptome and microsatellite markers in sugar maple.

Findings: From 29.2Gb of RNA sequencing data derived from different tissues and stress treatments, a set of 117,861 putative unique transcripts were assembled. From this set of sequences a total of 1514 microsatellite motifs were identified. Out of 58 genic microsatellite markers tested on a population of 47 sugar maple trees in Upper Michigan, 21 amplified well, of which 15 were polymorphic and 6 were monomorphic. Values for expected heterozygosity varied from 0.224 to 0.726 for individual loci. Of the 15 polymorphic markers, 13 exhibited transferability to four other *Acer* species.

Conclusions: Genic microsatellite markers can be applied to analyze genetic variation in potentially adaptive genes relative to genomic reference markers as a basis for the management of the species genetic resources in the face of climate change.

Keywords: *Acer saccharum*, EST-SSRs, transferability, Next-generation sequencing

Preface

Lab work involved in the characterization of the genic microsatellite markers in the present study was done by Monica Harmon. MH ran PCR on the 58 primers, checked these products after agarose electrophoresis, prepared amplified products, sent these samples to Cornell University for fragment analysis on a capillary sequencer and scored the results. MH also calculated genetic variation, null allele frequencies and linkage disequilibrium between all marker pairs. The following is an updated version of the paper. Since the paper was submitted one more marker (As_di9) was tested on the whole population of sugar maple. Two more markers (As_di34; As_di41) were tested for transferability on the four other *Acer* species.

Findings

Background

Sugar maple (*Acer saccharum* Marsh.) is a hardwood species native to the Northeastern United States and Southern Canada. It is valued for its wood and syrup, supports a large industry (USDA/NASS, 2015) and is an important part of the ecosystems in which it resides (Emerman and Dawson, 1996). While microsatellite markers have many applications in population genetic analyses, few resources for sugar maple have been developed to date (Graignic et al., 2013, Khodwekar et al., 2015).

Microsatellite markers are short sequence repeats (SSRs) which are often highly variable and prone to mutations (Tautz, 1989). Genic microsatellite markers can be derived from EST-libraries and can be located in the coding regions or in the 5'- and 3' untranslated regions of mRNAs (Ellis and Burke, 2007). Due to their location in expressed genes, they often show a lower genetic variation than genomic (non-genic) microsatellites and a

higher transferability between related species (Ellis and Burke, 2007). Especially stress-responsive genes might be under selection and show clinal variation along environmental gradients. Heat and drought stress and population fragmentation are especially expected at the southern distribution edge of the species (Iverson et al., 2008), as the result of warming climates. The new genic microsatellites will be useful to analyze patterns of genetic variation and provide the basis for the management and conservation of the species' genetic resources in a changing climate.

Methods

Plant materials for microsatellite analysis

Leaf samples were collected from 47 georeferenced adult trees of one population in Houghton, MI (47°07'07.46''N, 88°35'16.41''W) (Khodwekar et al., 2015). Total genomic DNA was isolated for the 47 samples of *Acer saccharum* and for four samples of each *Acer rubrum* L. (section *Rubra*), *Acer saccharinum* L. (section *Rubra*), *Acer platanoides* L. (section *Platanoidea*) and *Acer ginnala* Maxim. (section *Ginnala*) from ~1 cm² of silica gel dried leaf samples using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany).

Library preparation, transcriptome sequencing and assembly

A total of 51 transcriptome libraries were derived from tissues (leaves, petioles, roots) of ozone, cold, heat, drought, wound stressed and unstressed half-sib seedlings. Tissue types and treatments are described in Additional file 1. Details on ozone and wound stress treatments are given in (Lane et al., 2016). Tissues were flash frozen and stored at -80 °C until RNA was extracted. Approximately 1g of frozen tissue was used for isolation of total RNA with a modified CTAB method that uses lithium chloride precipitation (Chang et al., 1993). After RNA quality assessment with an Agilent Bioanalyzer (Agilent Technologies), RNA was converted to cDNA with the Illumina TruSeq kit. Shearing, size selection, barcoding and library validation was performed as described in (Lane et al., 2016).

Paired-end sequencing of 46 sugar maple libraries was performed using the Illumina HiSeq 2000 (San Diego, CA). A further set of 5 libraries were sequenced on an Illumina MiSeq (San Diego, CA) (Additional file 1). Raw reads can be found in the NCBI Short

Read Archive (SRA) with the bioproject accession PRJNA273272. Read statistics for each library are provided in the supplemental materials.

Trimmomatic version 0.35 was used to trim raw sequencing reads of adapter sequences and low quality bases (Bolger et al., 2014). Kmer-based error correction of RNASeq reads was performed in Rcorrector (kmer length of 31) (Song and Florea, 2015).

Trimmed and corrected reads were assembled with the Trinity pipeline version r20131110 (Grabherr et al., 2011). Isoforms from the Trinity pipeline were collapsed and further assembled with cd-hit version 4.6.1 (Fu et al., 2012). Open reading frame (ORF) prediction was executed with TransDecoder version 2.0.1 (Haas et al., 2013).

TransDecoder predictions were further improved by performing searches using BLAST version 2.2.29 versus the Uniprot SwissProt protein databases (cut-off of e-value < 1e-4) and by HMMER version 3.1b2 against the pfam database (Camacho et al., 2009, Magrane and UniProt, 2011, Finn et al., 2011, Finn et al., 2016). Two methods were employed to determine assembly completeness of the sugar maple transcriptome. First, transcripts were assessed by BUSCO (Benchmarking Universal Single-Copy Orthologs) version 1.1b1 to determine the presence of universal single-copy orthologs using the early release plant database (Simao et al., 2015). Second, TransRate v1.0.3 was used to calculate quality scores for contigs based on the read alignment evidence (Smith-Unna et al., 2016).

Functional annotation and microsatellite discovery

BLAST version 2.2.29 queries were performed against the plant taxonomic division of the TrEMBL protein database and Swiss-Prot protein database (Camacho et al., 2009, Magrane and UniProt, 2011). Simple sequence repeats (SSRs) were found by searching

putative unique transcripts (PUTs) using a perl script (https://github.com/mestato/lab_code/tree/master/hwg_gssr_scripts). SSRs needed to meet specific requirements for inclusion in the final output, requiring: 2 bp repeats to have 8-200 copies, 3 bp repeats to have 7-133 copies, and 4 bp repeats to have 6-100 copies. Compound SSRs, defined as SSRs occurring with less than 15 bases of separation, were removed. Dustmasker was used to mask low complexity regions to aid primer design in Primer3 v2.3.6 (Camacho et al., 2009, Untergasser et al., 2012). The following Primer3 parameters were used: primer_product_size_range = 100–450, primer_min_tm = 55.0, primer_max_tm = 65.0, primer_min_gc = 40, primer_max_gc = 60, primer_max_poly_x = 3, primer_gc_clamp = 2.

Marker development and genetic variation study

The DNA was diluted to ~ 1ng/μl with double deionized water (Ultra Pure Water, Molecular Grade, Phenix Research Products). A 15μl PCR reaction was prepared consisting of 6μl double deionized water (Ultra Pure Water, Molecular Grade from Phenix Research Products), 5μl HotFIREPol (Solis Biodyne, containing 2mM dNTPs, 1U Taq polymerase, 10mM MgCl₂), 0.2 μl of 5 μM forward primer with the M13 tail (5'-CACGACGTTGTAAACGAC-3'), 1.5μl of 5 μM 5' dye labelled (6-FAM) M13 primer, 0.5μl of 5 μM pig-tailed (5'-GTTTCTT-3') reverse primer (Kubisiak et al., 2013, Schuelke, 2000) (Sigma Aldrich Inc., St. Louis, MO) and 2μl DNA (~ 2ng). The PCR touchdown reaction was performed in an Applied Biosystems 2720 Thermal Cycler (Foster City, CA) consisting of an initial denaturation at 95 °C for 15 minutes, 10 touchdown cycles of 1 minute at 94 °C, 1 minute at 60 °C (decreasing 1 °C each cycle), and 1 minute at 72 °C, followed by 25 cycles of 94 °C for 1 minute, 50 °C for 1 minute and 72

°C for 1 minute. The final extension was at 72 °C for 20 minutes. PCR products were checked on 1.5% agarose gels and stained with 2µl GelRed (10,000X in water; Biotium, Hayward, CA) and exact fragment sizes were determined after electrophoretic separation on an ABI Prism® Genetic Analyzer 3730 with Gene-Scan™ LIZ-500 as internal size standard. Alleles were assigned to bins after careful visual inspection using GeneMarker® V2.6.7 (SoftGenetics). F

Primer pairs (Sigma Aldrich Inc., St. Louis, MO) were tested for amplification initially in four randomly selected samples. Markers that amplified polymorphic products in the expected size range were amplified in all 47 samples. Genetic variation assessment was conducted using GenAlEx 6.502 (Peakall and Smouse, 2006). Specifically, observed and expected heterozygosity (H_o , H_e , respectively) (Nei, 1973) and number of alleles (N_a) were calculated. Further, the inbreeding coefficient (F) was estimated for each locus in the population using GENEPOP version 4.2 (Raymond and Rousset, 1995, Rousset, 2008). Significant deviations from Hardy-Weinberg proportions were tested using Fischer's exact test in GENEPOP. Additionally pairwise linkage disequilibrium was tested for all marker pairs in GENEPOP. Bonferroni corrections were applied to adjust for multiple testing (Rice, 1989).

Results and discussion

Transcriptome assembly and validation

Illumina sequencing of 51 transcriptome sugar maple libraries of varying tissue types and treatments produced 282 million reads (29.9 Gb), available at NCBI SRA bioproject accession PRJNA273272 (Additional file 1). *De novo* assembly of reads yielded 117,861 putative unique transcripts (PUTs) with an average length of 945 bases and an N50 length of 1,667 bases. TransDecoder version 2.0.1 was used to identify 67,537 open read frames (ORFs) derived from 51,390 PUTs (43.6%). The peptide sequences calculated from the ORFs average 284 amino acids in length. Assembly accuracy was assessed as quality metrics based on examination of read mapping using Transrate (Smith-Unna et al., 2016) and identified an overall mapping rate of 72.8%, and over 85% of those mapped reads supported the assembly (i.e. both pairs aligned to the same transcript in the correct orientation without overlapping either transcript end). Out of a total of 956 BUSCO ortho-groups searched, 914 were found in the BUSCO database. These were further classified as 434 complete single-copy genes, 446 complete but duplicated genes, and 34 fragmented copies of genes. This represents the first transcriptome reference for sugar maple and will be a valuable genomic resource for future studies, for example, to generate markers, to use as a reference for gene expression sequencing or to identify candidate stress-response genes.

Functional annotation and microsatellite discovery

Sugar maple PUTs were used as queries for BLAST searches against the proteins in the Swiss-Prot and plant TrEMBL (Magrane and UniProt, 2011) databases yielding 64,458 (54.7%) PUTs with matches to TrEMBL plant proteins and 49,760 (42.2%) PUTs with matches to Swiss-Prot. Simple Sequence Repeat (SSR) extraction found 6,279 SSRs in 5,965 PUTs (5.1%). SSRs were identified for 2 bp, 3 bp, and 4 bp motifs. The 4 bp motifs were the rarest, making up 0.97% of total SSRs found. The shorter 2 bp and 3 bp motifs were much more common, making up 78.7% and 20.4% of total SSRs found, respectively. Primers were designed for 1,514 SSRs using the flanking regions of the repeats (Additional file 2). Of these 1,149 were 2 bp repeats ranging from 8 to 30 repeats, 346 were 3 bp repeats ranging from 7 to 20 repeats, and 19 were 4 bp repeats ranging from 6 to 15 repeats. Additional file 2 contains a report of the SSR analysis along with primers designed for SSRs meeting the appropriate requirements and functional annotations for SSR markers.

Microsatellite marker characterization

Out of the 58 primers tested, 15 amplified polymorphic loci and 6 amplified monomorphic loci (Table 1, Additional file 2). For polymorphic loci H_e ranged from 0.224 to 0.726, H_o from 0.143 to 0.722 and N_a from 2 to 10. A total of 11 out of the markers 15 markers showed no significant deviation from Hardy-Weinberg proportions after Bonferroni correction, while markers As_di12 and As_di38 ($p < 0.05$) and markers As_di34, As_di48 ($p < 0.01$) showed significant and positive F values. No significant linkage disequilibrium was found between marker pairs after Bonferroni correction ($p < 0.05$).

A total of 16 markers were tested for transferability in three DNA samples from each of the four *Acer* species which represent three taxonomic sections: *A. saccharinum* (section *Rubra*), *A. rubrum* (section *Rubra*), *A. platanoides* (section *Platanoidea*) and *A. ginnala* (section *Ginnala*) using the same PCR protocol as described above. The results showed that most primers amplified a multi-banding pattern across species. However, 13 markers amplified single loci in the expected size range in at least one other *Acer* species (Table 2). Nine markers were monomorphic in some or all species (Table 2). Only one of the markers, As_tetra1, was polymorphic in more than one species.

The level of genetic variation for individual polymorphic markers varied considerably. Mean variation at polymorphic markers was much lower for genic (H_o : 0.424; H_e : 0.543) than for non-genic markers in the same sugar maple population (H_o : 0.708; H_e : 0.822 (Khodwekar et al., 2015)). Similarly, lower genetic variation was observed at genic EST-SSRs than at genomic SSRs in several *Quercus rubra* and *Q. ellipsoidalis* populations (Lind and Gailing, 2013). The generally lower genetic variation at genic EST-SSRs suggests that these markers are not selectively neutral, but subject to positive selection (Ellis and Burke, 2007). Genetic variation observed at genic EST-SSRs in sugar maple was lower than estimates obtained in other wind-pollinated tree species such as North American oak species *Quercus rubra* and *Quercus ellipsoidalis* from the same geographic region ($H_e = 0.700$, range: 0.690 to 0.740, (Lind and Gailing, 2013)). F values are slightly higher in these genic microsatellite markers as compared to non-genic markers in the same population (Khodwekar et al., 2015) but most F values are not significantly different from zero. High F values at individual markers could be due to the presence of null alleles. Marker As_di34 with high F values and significant deviation

from Hardy-Weinberg proportions ($p < 0.001$) showed no amplification in nearly half of the samples suggesting high null allele frequencies. The other marker, As_di48, with a very high F value showed high amplification success comparable to the other markers. Overall high variation was observed in F values ranging from -0.081 for As_di7 to 0.705 for As_di48 which may suggest balancing or positive selection at some markers.

The overall level of genetic variation, the absence of linkage disequilibrium and low to moderate deviations from HWE for most markers suggest the suitability of these markers for population genetic studies in sugar maple. In addition, markers showed high transferability to other *Acer* species providing at least two new polymorphic markers for the four other *Acer* species, and even six new polymorphic markers for one of the species, *A. platanoides*. High transferability of markers across *Acer* sections suggests their potential usefulness for population genetic analyses in other maple species. In contrast, only two out of 13 genomic SSRs amplified polymorphic loci in one other species, *Acer ginnala* (Khodwekar et al., 2015). This high transferability is typical of EST-SSRs and has been shown repeatedly in crop plants and is more prevalent than in non-EST-SSR markers (Ellis and Burke, 2007). With the availability of both neutral genomic SSRs and genic SSRs with potential role in adaptation, patterns of genetic variation and their relation to environmental and climatic variables can be studied.

Authors' contributions

OG, MS, MC, SS, JEC conceived and designed the experiment. MC and SS provided plant materials. MH, TL, TB, HL, NZ, and DM performed experiments. OG, MH, TL, and MS analyzed the data. OG, MH, MS and TL wrote the paper. All authors read, edited, and approved the final manuscript.

Author details

¹School of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931. ²Department of Entomology and Plant Pathology, The University of Tennessee, Knoxville, TN 37996. ³Department of Forestry, University of Missouri, Columbia, MO 65274. ⁴Department of Ecosystem Science and Management and Department of Plant Science, The Pennsylvania State University, University Park, PA 16802. ⁵Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29630. ⁶SciVentures Inc., 4900 Hopyard Rd, Suite 100, Pleasanton, CA 94588. ⁷Department of Forestry, Wildlife & Fisheries, The University of Tennessee, Knoxville, TN 37996-4563.

Acknowledgements

This project was supported by the National Science Foundation Plant Genome Research Program (TRPGRA2 IOS-1025974) as part of the Hardwood Genomics Project. Additional funding was provided by the Ecosystem Science Center and the Life Science and Technology Institute (LSTI) at Michigan Technological University. This work would not have been possible without the contributions of Sudhir Khodwekar at Michigan Technological University who isolated DNA and helped with interpreting the results. Monica Harmon was partly supported by the OSM/VISTA Master of Science degree program. We also acknowledge the Genomics Core Facility of the Huck Institutes of the Life Sciences at Penn State University for MiSeq data production.

Competing Interests

The authors declare that they have no competing interests.

Availability of data and materials

Raw reads can be found in the NCBI Short Read Archive (SRA) with the bioproject accession PRJNA273272.

Funding

The study was funded the National Science Foundation Plant Genome Research Program (TRPGRA2 IOS-1025974) as part of the Hardwood Genomics Project.

Table 1 Primer sequences and descriptions of 21 microsatellite markers developed in *Acer saccharum*

Locus	Primer Sequences (5' - 3')	Repeat Motif	Size Range (bp)	**Expected Fragment Size	A	H _o	H _e	F	p
As_di1	F: TCCCAGGCATGAACAAGGTT R: TGCAGTAAGTTGACAGCTCT	(AC) ₉	251 - 273	230	9	0.581	0.658	0.117	0.0034
As_di7	F: GGGTCTGTCTCTGTTTCTGCA R: ACAGGGTTCAGTCTGAGCTGTG	(AC) ₉	151 - 161	134	6	0.651	0.602	-0.081	0.2578
^m As_di11	F: AGAGAACCACCAAGGATGCA R: CAGGGAGCCATTTCACTCTGA	(GA) ₈	187	167	1	0	0	-	-
As_di12	F: AAGACATCTTGAGGGCGGTG R: TGTAAGTGCATAACGGGCCA	(TC) ₉	447 - 458	425	4	0.194	0.271	0.283	0.0014*
^m As_di13	F: TCAAGAAATACTGGCTCAGGTCA R: ACATGCATGTTGAGCGATTGT	(TC) ₈	264	238	1	0	0	-	-
As_di15	F: GGGCAGAGAGGGAATTCGAG R: TGGGGAGACAGAACTTGTGC	(TC) ₈	259 - 265	235	4	0.375	0.429	0.126	0.0236
As_di16	F: AATTGCCTGTGGTGGGAACT R: ACTCCACTCTCTTTCTTGCTCA	(GT) ₈	211 - 218	190	4	0.675	0.680	0.007	0.8696
^m As_di19	F: GACCTGACCACCTCCTCCTA R: AGGGCAACATACACGGGATCG	(CT) ₉	228	205	1	0	0	-	-
As_di21	F: TGTCAGCAGCCCTACAGTTG R: ACAGGTCACGATCTCTCCCT	(GT) ₈	135 - 142	113	4	0.450	0.609	0.261	0.0455
^m As_di27	F: TCATGACCATGACCCAACACT R: TCCTGTGGATTCTTTTGTATCTGGT	(TC) ₈	386	363	1	0	0	-	-
^m As_di30	F: GATCCCCTTCGTTGCTGACA R: GGACTGCCCGATTGATACTCA	(TG) ₉	457	432	1	0	0	-	-
^m As_di31	F: CTCCACCACCATCCAACCAA R: TCCTCAGCTTCTTGGGCTTG	(AC) ₈	187 - 189	167	1	0	0	-	-
As_di34	F: AACGGATGGCAAGCTAGCTT R: CAGGCCTGCCAGAACTAA	(AC) ₈	225 - 240	218	4	0.217	0.726	0.701	<0.0001 **
As_di35	F: TGTTAGTCTCCTCCACACGT R: CAGCAGCAGCAGCAAACT	(TC) ₈	151 - 153	129	2	0.143	0.224	0.362	0.0751
As_di36	F: ATGTGAGTCCGTGAGTCCGT R: ATAAGCAGCAGCAGCAGACA	(TG) ₈	237 - 245	212	5	0.475	0.606	0.216	0.0322
As_di37	F: TGGTGGGTAGCAGCAAAAGA R: TGCACATGGGATGATGATCAGT	(TG) ₁₁	167 - 184	151	8	0.722	0.704	-0.025	0.9324
As_di38	F: ACAGAGAGAGAGAGAGCTTGT R: TACCGTCATGGCCGATGATG	(AC) ₈	175 - 179	154	3	0.184	0.362	0.491	0.0012*

As_di41	F: AAGCTGAGAAACCCAAAGCA R: CACCACCCAACCCTTTTCCT	(TA) ₈	259 - 271	236	6	0.526	0.625	0.157	0.2675
As_di48	F: AGGTTTCGGGTTTTGAATCTTCA R: CAAGGACTTTGGCTCTGCTG	(TA) ₈	173 - 179	155	4	0.200	0.678	0.705	<0.0001 **
As_di49	F: TGCAACTGTTGAGTGGTGGA R: ACAAGTCGAACAACCCGTTG	(TG) ₁₀	159 - 183	133	10	0.625	0.636	0.017	0.0528
As_tetra1	F: TTGACGGAGAGCTTGGTTCC R: AAAACCCAATTCGCCACGTG	(TGCT) ₆	257 - 273	241	5	0.341	0.337	-0.012	0.4674
[†] Mean					6	0.424	0.543	0.222	-
Mean					4	0.308	0.395	0.222	

A: number of alleles per locus, H_o: observed heterozygosity, H_e: expected heterozygosity, F: inbreeding coefficient, *: significantly different from Hardy-Weinberg proportions ($\alpha = 0.05$) after Bonferroni corrections, **: significantly different from Hardy-Weinberg proportions ($\alpha = 0.01$) after Bonferroni corrections ^m: these markers amplified monomorphic loci and were not tested on the whole population. **: expected fragment size based on EST contigs. Actual fragments were longer as tailed primers were used for amplification. [†]Mean variation was calculated for polymorphic markers only to allow for comparisons with other related studies on sugar maple (Khodwekar et al., 2015).

Table 2 Transferability of microsatellite loci to other *Acer* species

<i>A. rubrum</i>	<i>A. saccharinum</i>	<i>A. platanoides</i>	<i>A. ginnala</i>
		As_di1	As_di1*
As_di7	As_di7*	As_di9	As_di9*
	As_di12	As_di15	As_di15*
			As_di16
	As_di21*	As_di21	
As_di35	As_di35*	As_di35*	
		As_di36*	
	As_di37*	As_di37*	
	As_di38		
As_di48	As_di48*	As_di48*	As_di48*
		As_di49	As_di49*
		As_tetra1	As_tetra1

*: Monomorphic in species as indicated by column

References

1. USDA/NASS. QuickStats Ad-hoc Query Tool. In. Quickstats.nass.usda.gov; 2015.
2. Emerman SH, Dawson TE. Hydraulic lift and its influence on the water content of the rhizosphere: An example from sugar maple, *Acer saccharum*. *Oecologia*. 1996;108(2):273-278.
3. Graignic N, Tremblay F, Bergeron Y. Development of polymorphic nuclear microsatellite markers in sugar maple (*Acer saccharum* Marsh.) using cross-species transfer and SSR-enriched shotgun pyrosequencing. *Conservation Genetics Resources*. 2013;5(3):845-848.
4. Khodwekar S, Staton M, Coggeshall MV, Carlson JE, Gailing O. Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). *Annals of Forest Research*. 2015;58:193-204.
5. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res*. 1989;17:6463-6471.
6. Ellis JR, Burke JM. EST-SSRs as a resource for population genetic analyses. *Heredity*. 2007;99(2):125-132.
7. Iverson LR, Prasad AM, Matthews SN, Peters M. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *For Ecol Manage*. 2008;254(3):390-406.
8. Lane T, Best T, Zembower N, Davitt J, Henry N, Xu Y, Koch J, Liang H, McGraw J, Schuster S *et al*. The green ash transcriptome and identification of genes responding to abiotic and biotic stresses. *BMC genomics*. 2016;17:702.
9. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*. 1993;11(2):113-116.
10. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
11. Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*. 2015;4.
12. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644-U130.
13. Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28(23):3150-3152.
14. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494-1512.
15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST plus : architecture and applications. *BMC Bioinformatics*. 2009;10.
16. Magrane M, UniProt C. UniProt Knowledgebase: a hub of integrated protein data. *Database*. 2011.
17. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*. 2011;39:W29-W37.

18. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*. 2016;44(D1):D279-D285.
19. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-3212.
20. Smith-Unna R, Bournnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research*. 2016;26(8):1134-1144.
21. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3-new capabilities and interfaces. *Nucleic Acids Research*. 2012;40(15).
22. Kubisiak TL, Nelson CD, Staton ME, Zhebentyayeva T, Smith C, Olukolu BA, Fang GC, Hebard FV, Anagnostakis S, Wheeler N *et al.* A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*). *Tree Genet Genomes*. 2013;9(2):557-571.
23. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol*. 2000;18(2):233-234.
24. Peakall R, Smouse PE. GENEALX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 2006;6:288-295.
25. Nei M. Analysis of gene diversity in subdivided populations. *Proc Nat Acad Sci USA*. 1973;70(12):3321-3323.
26. Raymond M, Rousset F. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered*. 1995;86(3):248-249.
27. Rousset F. GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour*. 2008;8(1):103-106.
28. Rice WR. Analyzing tables of statistical tests. *Evolution*. 1989;43(1):223-225.
29. Lind J, Gailing O. Genetic structure of *Quercus rubra* L. and *Q. ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genet Genomes*. 2013;9:707-722.