Identification and Characterization of Microsatellite Loci in the Tuliptree, Liriodendron tulipifera (Magnoliaceae)

Authors: Ricardo Gutiérrez-Ozuna, and Matthew B. Hamilton
Source: Applications in Plant Sciences, 5(8)
Published By: Botanical Society of America
URL: https://doi.org/10.3732/apps.1700032
**Key words:** Liriodendron tulipifera; Magnoliaceae; microsatellites; next-generation sequencing; population genetic studies; simple sequence repeats (SSRs); tuliptree.

Liriodendron tulipifera L., commonly known as tuliptree, tulip poplar, or yellow poplar, is a pioneer tree in the family Magnoliaceae native to eastern North America. It has a wide geographic distribution in the southeastern and mid-Atlantic United States and occurs in diverse habitats. To facilitate population genetic analyses of effective population size and population structure, we developed genomic microsatellite (simple sequence repeat [SSR]) markers without the potential limitations of previously reported SSRs. *Liriodendron tulipifera SSRs have been developed from expressed sequence tags (ESTs; Xu et al., 2006, 2010; Yang et al., 2012; Zhang et al., 2015) located in or near functional genes, and consequently, they are more likely to be affected by natural selection (Ellis and Burke, 2007). Liriodendron chinense (Hems.) Sarg. genomic (noncoding, nontranscribed) microsatellite loci have been cross-amplified in *L. tulipifera* (Yao et al., 2008). Cross-species amplification of microsatellite loci might result in ascertainment bias, where polymorphism is reduced when loci are transferred to related species (Ellegren et al., 1995).

Preliminary tests of loci from Yao et al. (2008) carried out with 10 *L. tulipifera* individuals showed low polymorphism (results not shown). Nonneutral evolution or ascertainment bias can potentially impact the estimation of population genetic parameters. Therefore, we identified and characterized polymorphic genomic microsatellite loci in *L. tulipifera* using Illumina next-generation sequencing and a bioinformatics pipeline.

**METHODS AND RESULTS**

**Microsatellite development**—Total DNA from leaves of one *L. tulipifera* individual collected on the main campus of Georgetown University in Washington, D.C., USA, was extracted using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA). A genomic DNA library for Illumina paired-end sequencing was prepared from 4 μg of DNA following the PCR-free library prep kit from Illumina (Illumina, San Diego, California, USA). DNA was sheared to 550 bp and sequenced as 150 bp paired-end reads on an Illumina HiSeq 2500 at the Bioscience Institute of Virginia Tech (Blacksburg, Virginia, USA). We used PAL_FINDER_v0.02.04 (Castoe et al., 2012) to extract reads containing perfect microsatellites (unreadable and identical repeats). The reads were imported to PAL_FINDER and analyzed in two different ways: (1) as Illumina paired-end reads filtered to include ≥12 tri-, ≥10 tetra-, ≥8 penta-, and ≥6 hexanucleotide repeats, and (2) as FASTQ sequence files converted to FASTA format, treated as ≥34 single-end reads, and filtered to include ≥15 di-, ≥10 tri-, ≥8 tetra-, ≥6 penta-, and ≥4 hexanucleotide repeats. One potential advantage of using both methods is the development of loci with a broader range of amplification fragment sizes. In both cases, we identified microsatellite loci with flanking sequences suitable for PCR primer design or potentially amplifiable loci (PALs). Raw reads were deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA; BioProject no. PRJNA331147, BioSample no. SAMN05417503). Summaries of reads containing microsatellite repeats and PALs (with primer sequences) detected using both methods are available in Appendices S1 and S2.

1 Manuscript received 6 April 2017; revision accepted 23 May 2017.

The authors thank the Smithsonian Environmental Research Center (Edgewater, Maryland, USA), James Madison’s Montpelier (Orange, Virginia, USA), and Saddler’s Woods Conservation Association (Haddon Township, New Jersey, USA) for allowing sampling of specimens. R.G.O. was supported by a fellowship from the Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico). The authors are grateful for funding from the Georgetown Environment Initiative, the Center for the Environment, and the Department of Biology at Georgetown University.

1 Author for correspondence: rg453@georgetown.edu

doi:10.3732/apps.1700032

Applications in Plant Sciences 2017 5(8): 1700032; http://www.bioone.org/loi/apps © 2017 Gutiérrez-Ozuna and Hamilton. Published by the Botanical Society of America. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC-SA 4.0), which permits unrestricted noncommercial use and redistribution provided that the original author and source are credited and the new work is distributed under the same license as the original.
We selected a set of 77 PALs to empirically assess amplification using three individuals. We amplified each locus in 25-μL PCR reactions (1× OneTaq Standard reaction buffer, 160 μM dNTPs, 0.2 μM forward primer, 0.2 μM reverse primer, 0.625 units OneTaq DNA polymerase [New England Biolabs, Ipswich, Massachusetts, USA], 1 μL template DNA [concentration was not determined], and ddH₂O to 25 μL). Thermocycling conditions were 94°C (30 s); followed by 30 cycles of denaturation at 94°C (30 s), annealing at 50–61°C (30 s, Table 1), and extension at 68°C (30 s); and a final extension of 68°C (5 min). Fifty-one primer pairs yielded products of the expected size without nonspecific amplification and were then tested for polymorphism in seven individuals, by visualizing PCR products on 3% agarose gels. Of these 51 loci, 23 were polymorphic and used to genotype 30 to 52 total individuals collected from three old-growth locations in the native range of *L. tulipifera* (Appendix 1). Because *L. chinense*, the other single species in *Liriodendron*, has a restricted geographic distribution in the native range of *L. tulipifera*, we selected a set of 23 polymorphic genomic microsatellite loci isolated from *Liriodendron tulipifera*.

**Table 1. Characteristics of 23 polymorphic genomic microsatellite loci isolated from *Liriodendron tulipifera***

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer sequences (5′–3′)</th>
<th>Repeat motif</th>
<th>Observed amplicon size (bp)</th>
<th>Fluorescent labeling method</th>
<th>Identifier</th>
</tr>
</thead>
</table>
| Lt006 | F: AGTTGGGAATTGGGACAAGG  
R: TGGACTTTCCATCGAGTTGACG | (AAG)₁₂ | 416–437 | M13-labeled | KX869968 |
| Lt011 | F: GCATGGACATGGTGTAACCC  
R: TTTGATGTCGCCCCTACTGCG | (AAC)₁₂ | 113–140 | 6-FAM¹ | KX869967 |
| Lt014 | F: CTTCATGAAACAAAGGAAACC  
| Lt023 | F: CCGTGCCCGACATTTGAAGC  
| Lt025 | F: AGTTGGAATTTGGGAACG  
| Lt032 | F: GCTTCCGACACAAACATGCG  
| Lt035 | F: CACAGACTTGCGTTCCATTAC  
| Lt036 | F: TTAGAGTGGATACCTGTTAACC  
| Lt043 | F: TCCATCTCTTTTCTTGGC  
| Lt052 | F: TGTTCCGAGATTTGTTGCC  
| Lt054 | F: CCGCTGATGGTGGATCTTTG  
| Lt059 | F: CTGCCCTTCCTAATCTTTGAG  
| Lt060 | F: CCTCTCTCTGGACCTTGGC  
| Lt061 | F: TTGGGAGATTTGAGGCAGGC  
| Lt064 | F: AAGAGCTCATTCCTCAGGAG  
| Lt066 | F: CCTGCGCCTGTTGATCTGCG  
| Lt068 | F: AAATCCCTTAAAGGCGTTCC  
| Lt070 | F: TTGCTGCGTCTATCTTACC  
| Lt075 | F: CATCGGAACATTGCCTTCCTC  
| Lt077 | F: TATCCACAGGCCCTACAA  
| Lt079 | F: GGAGACCTGCGTTTAACTGCG  
| Lt080 | F: GGCTCGAATTTCTGGTCTCC  
| Lt081 | F: ATGATTCCGCCAAGTCTCTC  

Note: Tₐ is annealing temperature.

¹PCR multiplex sets are indicated as 1, 2, or 3.
²Numbers are either GenBank accession numbers or Illumina sequence identifiers associated with NCBI’s Short Read Archive (BioProject no. PRJNA331147, BioSample no. SAMN05417503).
Table 2. Genetic properties by individual and pooled sampled locations of 23 polymorphic microsatellite markers developed in *Liriodendron tulipifera*.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Montpelier (n = 20)</th>
<th>Frog Canyon (n = 20)</th>
<th>Saddler's Woods (n = 12)</th>
<th>Pooled locations (n = 52)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>A</td>
<td>H_e</td>
<td>H_o</td>
</tr>
<tr>
<td>Lt006</td>
<td>10</td>
<td>6</td>
<td>0.789</td>
<td>0.800</td>
</tr>
<tr>
<td>Lt011</td>
<td>20</td>
<td>6</td>
<td>0.726</td>
<td>0.600</td>
</tr>
<tr>
<td>Lt014</td>
<td>60</td>
<td>5</td>
<td>0.868</td>
<td>0.800</td>
</tr>
<tr>
<td>Lt023</td>
<td>10</td>
<td>5</td>
<td>0.505</td>
<td>0.500</td>
</tr>
<tr>
<td>Lt025</td>
<td>10</td>
<td>4</td>
<td>0.658</td>
<td>0.900</td>
</tr>
<tr>
<td>Lt032</td>
<td>10</td>
<td>5</td>
<td>0.700</td>
<td>0.700</td>
</tr>
<tr>
<td>Lt035</td>
<td>19</td>
<td>7</td>
<td>0.750</td>
<td>0.263***</td>
</tr>
<tr>
<td>Lt036</td>
<td>20</td>
<td>8</td>
<td>0.846</td>
<td>0.850</td>
</tr>
<tr>
<td>Lt043</td>
<td>20</td>
<td>6</td>
<td>0.805</td>
<td>0.750</td>
</tr>
<tr>
<td>Lt052</td>
<td>9</td>
<td>4</td>
<td>0.699</td>
<td>0.667</td>
</tr>
<tr>
<td>Lt054</td>
<td>19</td>
<td>7</td>
<td>0.764</td>
<td>0.579</td>
</tr>
<tr>
<td>Lt059</td>
<td>20</td>
<td>5</td>
<td>0.853</td>
<td>0.650</td>
</tr>
<tr>
<td>Lt060</td>
<td>10</td>
<td>5</td>
<td>0.774</td>
<td>0.600</td>
</tr>
<tr>
<td>Lt061</td>
<td>20</td>
<td>4</td>
<td>0.458</td>
<td>0.450</td>
</tr>
<tr>
<td>Lt064</td>
<td>20</td>
<td>5</td>
<td>0.750</td>
<td>0.600</td>
</tr>
<tr>
<td>Lt066</td>
<td>10</td>
<td>4</td>
<td>0.695</td>
<td>0.600</td>
</tr>
<tr>
<td>Lt068</td>
<td>20</td>
<td>12</td>
<td>0.837</td>
<td>0.900</td>
</tr>
<tr>
<td>Lt070</td>
<td>20</td>
<td>3</td>
<td>0.578</td>
<td>0.450</td>
</tr>
<tr>
<td>Lt075</td>
<td>10</td>
<td>2</td>
<td>0.395</td>
<td>0.500</td>
</tr>
<tr>
<td>Lt077</td>
<td>20</td>
<td>4</td>
<td>0.742</td>
<td>0.700</td>
</tr>
<tr>
<td>Lt079</td>
<td>20</td>
<td>4</td>
<td>0.391</td>
<td>0.400</td>
</tr>
<tr>
<td>Lt080</td>
<td>10</td>
<td>3</td>
<td>0.542</td>
<td>0.200*</td>
</tr>
<tr>
<td>Lt081</td>
<td>10</td>
<td>6</td>
<td>0.811</td>
<td>0.600</td>
</tr>
</tbody>
</table>

*Note: A = number of alleles; F = fixation index; H_e = expected heterozygosity under random mating; H_o = observed heterozygosity; n = number of individuals sampled; N = number of individuals genotyped; NullIIM = estimate of null allele frequency given the individual inbreeding model (IIM); PIC = polymorphism information content.

See Appendix 1 for locality and voucher information.

Significant deviation from Hardy–Weinberg expected genotype frequencies: *P < 0.05, **P < 0.01, ***P < 0.001.
† 95% highest posterior density interval does not include zero.
China and Vietnam, we were not able to test for cross-species amplification. As cross-amplification of genomic SSRs has limited success in plants (Merritt et al., 2015) and success declines as genetic divergence increases (Barbara et al., 2007), we did not test for cross-amplification in other Magnoliaceae.

For fragment analyses, PCR products were fluorescently labeled either using primers tailored with a 5′ M13(−21) sequence following Schuelke (2000) or using primers with a 5′ fluorophore and amplified in multiplex (Table 1). In the tailored labeling method, two PCR reactions were carried out using the same reverse primer. The first PCR used an M13(−21)-tailed locus-specific forward primer, while the second used a universal fluorescently labeled M13(−21) as a forward primer. The products of the first PCR were purified using StrataPrep PCR Purification Kit (Agilent Technologies, Santa Clara, California, USA) and then used as the template for the second PCR. Fluorescent products were electrophoresed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, California, USA), and amplicons sizes were estimated with either orange or red DNA size standard (MCLAB, San Francisco, California, USA) and GeneMapper software 3.7 (Applied Biosystems) using the Local Southern sizing algorithm.

Microsatellite data analysis—Genotypes appeared diplodiploid, displaying at most two alleles per locus per individual. Data were analyzed by sampled location and as a pooled population (Table 2). For each locus, number of alleles, expected heterozygosity (HE), expected heterozygosity under random mating (Hr), and polymorphism information content (PIC = 1 − (∑(f(x)−f)^2)) were calculated for each population (Table 2). For the pooled population, were estimated using CERVUS 3.0.3 (Kalinoski et al., 2007). We used GENEPOP 4.2 (Rousset, 2008) to test deviation from Hardy–Weinberg expected heterozygote frequency (HWE) using default values for Markov chain parameters, and to estimate the fixation index (F = |Hr − Hr/Hr|; Hamilton, 2009) for the pooled population.

Population genetic parameters are listed for each sampled location showing loci exhibited two to 12 alleles, with almost all alleles common to each location (Table 2). Lack of population differentiation (FST = 0.077 estimated using GENEPOP) justified pooling genotypes from the three locations. In the pooled population, observed and expected heterozygosities ranged from 0.233 to 0.865 and 0.272 to 0.876, respectively. Six loci showed significant deviations from HWE (Table 2) with deficits of heterozygotes that could be attributed to numerous causes. One hypothesis is nonrandom mating, which we tested using INEST (Schuelke, 2000) or using CozaN, Jim. J. H. et al. 2015. Microsatellite ‘evolution’: Directionality or bias? Nature Genetics 11: 360–362. Ellis, J. R., and J. M. Burke, 2007. EST-SSRs as a resource for population genetic analyses. Heredity 99: 125–132. Hamilton, M. B, 2009. Population genetics. Wiley-Blackwell, Hoboken, New Jersey, USA.


Taken together with thousands of PALs, this study provides useful resources for population genetic studies of L. tulipifera.

CONCLUSIONS

The 23 microsatellite markers developed here do not suffer from ascertainment bias and show high levels of polymorphism.

LITERATURE CITED


Gutiérrez-Ozuna and Hamilton—Liriodendron tulipifera microsatellites

Appendix 1. Locality and voucher information for the Liriodendron tulipifera samples used in this study.

<table>
<thead>
<tr>
<th>Location</th>
<th>County, State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Voucher no.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frog Canyon, Smithsonian</td>
<td>Anne Arundel, Maryland</td>
<td>38.884286</td>
<td>−76.552695</td>
<td>MARY1021991</td>
<td>20</td>
</tr>
<tr>
<td>Environmental Research Center</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>James Madison’s Montpeher</td>
<td>Orange, Virginia</td>
<td>38.226667</td>
<td>−78.179444</td>
<td>MARY1021990</td>
<td>20</td>
</tr>
<tr>
<td>Sadder’s Woods</td>
<td>Camden, New Jersey</td>
<td>39.007022</td>
<td>−75.087750</td>
<td>MARY1021989</td>
<td>12</td>
</tr>
</tbody>
</table>

Note: n = number of individuals sampled.

* A voucher sample for each location was deposited in the Norton-Brown Herbarium (MARY), University of Maryland, College Park, Maryland, USA.

http://www.bioone.org/loi/apps

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 14 Aug 2019
Terms of Use: https://bioone.org/terms-of-use