Microsatellite primers for eastern North American *Celtis* (Cannabaceae) species

Andrew Hayes Song Wang Tyler William Smith

Introduction

Celtis pumila Pursh [sensu Whittemore (2013); synonym *C. tenuifolia* Nutt.; Cannabaceae] is a shrub or small tree native to eastern North America, ranging from Missouri to Pennsylvania, south to Georgia and Alabama. Disjunct populations occur in Michigan, Indiana, and Southern Ontario. It is a Threatened Species in Canada (COSEWIC, 2003). However, conservation management has been hampered by difficulty distinguishing *C. pumila* from *C. occidentalis* L., with which it co-occurs. Both of these species are sympatric with *C. laevigata* Willd in the USA. These three taxa are characterized by intergrading morphology (Fernald, 1950), polyploidy (Whittemore, 2013)(Whittemore 2013), and need taxonomic revision (Sherman-Broyles et al., 1997).

We required highly-informative molecular markers to assess the genetic structure and taxonomy of Celtis in eastern North America. Previous studies of Celtis have employed AFLPs (Whittemore, 2005; Whittemore and Townsend, 2007). However, we elected to develop microsatellite primers, which have several advantages: co-dominant expression; multi-allelic loci; and easier scoring.

Methods and Results

We extracted DNA from silica-dried leaf tissue of 1 individual of *C. pumila*, collected from Trenton, Ontario in 2016, using a Nucleospin Plant II kit (Machery-Nagel, Bethlehem, Pennsylvania, U.S.A.). The sample was submitted to the Georgia Genomic Facility at the University of Georgia (Athens, Georgia, U.S.A) for genome skimming. DNA was fragmented using the Bioruptor UCD-300 sonication device (Diagenode, Denville, New Jersey, U.S.A). Illumina TruSeqHT libraries were prepared using the Kapa Library Preparation Kit (KR0453-v2.13; Kapa Biosystems, Wilmington, Massachusetts, U.S.A.) with custom indexes from Faircloth and Glenn (2012). Libraries were sequenced using an Illumina HiSeq with PE150 reads (Illumina, San Diego, California, U.S.A.). The sequence reads are available in the NCBI Sequence Read Archive, BioSample Accession SAMN13198269. A total of 1.99 x106 HiSeq reads were imported and paired in Geneious 7.0.6 (Biomatters, Auckland, New Zealand). Illumina TruSeq adapters and bases with an error probability limit above 0.05 were trimmed. A de novo assembly was performed and consensus sequences were exported from Geneious as FASTA files and imported into PAL_finder v0.02.04 (Castoe et al., 2012). 13,069 loci with 2-6 bp repeats were designed at default minimum lengths. We selected 40 primer pairs for further study, based on the criteria: both the forward and reverse primers were found at only one locus; the motif length was 3-6 base pairs; there were more than 10 repeats.

We applied all 40 primer pairs to eight samples: two *Celtis occidentalis* and six *C. pumila* collected from populations across the range of the two species: *C. occidentalis* : COHA1 (Point Pelee, Ontario); VARC1 (Virginia); *C. pumila*: DWHA12-2 (Point Pelee, Ontario); ID-ID1 (Indiana); KYHR1 (Kentucky); PASH1 (Pennsylvania); VAWR2 (Virginia); and GC1 (Trenton, Ontario). 10 of the primer pairs did not show any product on agarose gels after electrophoresis, and were excluded from further testing. We scored the remaining 30 loci for polymorphisms among the eight samples using a 3500xl Genetic Analyzer (Applied Biosystems). We identified a primary set of 11 loci that produced clear, polymorphic fragments in all samples (Table 1). We

also identified a set of 11 secondary loci, which had potential utility but would require further development to validate, due to reduced amplification success or absence of polymorphisms (Table S1). We amplified the 11 primary loci in 113 samples from five *C. pumila* populations and one population of *C. occidentalis* and *C. laevigata*. Population and voucher data are presented in Appendix 1. The ploidy of all samples was determined via flow cytometry (Hayes, 2019). Three loci produced inconsistent results in the full sample, including genotypes with more alleles than expected for the individual (i.e., more than 2 alleles for diploids, more than 3 alleles for triploids), and were removed. The remaining 8 loci are summarized in Table 2. Samples were genotyped and manually scored using Geneious 11.1.4 (Biomatters Ltd.).

PCR reactions were carried out with 8 uL master mixes containing 0.24 uL DMSO, 4.0 uL 2X Phusion Master Mix w/ High Fidelity Buffer (New England Biolabs), 0.416 uL mix of Left and Right Primers (Sigma-Aldrich, Oakville, ON) 0.192 uL of FAM or VIC-labelled CAG Tag (10uM), 2.152 uL ddH 2 O and 1.0 uL DNA, using a T-100 Thermal Cycler (Bio-Rad, Hercules, California, U.S.A.). Thermal cycling began with 5 min denaturation at 95°C, followed by the touchdown phase with 15 cycles of 30 s denaturation at 95°C, 30 s annealing from 72° to 57°C (−1°C per cycle), and 30 s elongation at 72°C, followed by a generic amplification stage of 20 cycles of 30 s denaturation at 95°C, 30 s annealing at 55°C and 30 s elongation at 72° followed by a 5 min final elongation at 72°C and a holding temperature of 15°C. Amplicons were incorporated with fluorescent labels FAM (Sigma-Aldrich, Oakville, ON) and VIC (Life Technologies) and pooled into four groups of three which were then sequenced by capillary electrophoresis using a 3500xl Genetic Analyzer (Applied Biosystems).

All 8 primer pairs were polymorphic, and produced consistent amplification. We recorded 47 alleles, 3-14 per locus. The *Celtis pumila* samples, all of which are triploid, exhibited very low diversity: 31 of 32 individuals at Pinery Provincial Park shared the same multi-locus genotype (MLG); 16 of 19 individuals from Point Pelee National park shared a different MLG; and 8 of 9 individuals from Sideling Barrens shared a third MLG. All five *C. pumila* populations had MLGs shared by multiple individuals (Table 2). This suggests reproduction in these populations is largely apomictic.

Diversity was higher in the diploid populations. Two of the 26 individuals from the Point Pelee National Park population of *C. occidentalis* shared the same MLG, and all 10 *C. laevigata* individuals had unique MLGs. We tested each locus in each species for departure from Hardy-Weinberg Equilibrium (HWE) using the hw.test function in the pegas R package (Paradis, 2010). Only locus CSSR28 for *Celtis occidentalis* differed significantly from HWE ($p = 0.032$). All other loci in *C. occidentalis* were in HWE, and all loci were in HWE in the *C. laevigata* population.

Conclusions

We have validated a set of eight microsatellite markers that cross-amplify in three *Celtis* species in eastern North America, and show variation within and among populations and species. This provides a necessary tool for exploring taxonomic relationships and conservation genetics in this group.

Tables

Table 1. Characterization of 11 polymorphic microsatellite loci based on 2 *C. occidentalis* and 6 *C. pumila* samples.

Locus	Motif	Primer sequences (5'-3')	Allele size	Ta	Genbank Acc
CSSR13	AATT(28)	F: AAAGTGACAAAGAGTTTTAAATGGG R: AAATCAACTCGAATAAAGAGGGC	190-220	58.2	MN611935
CSSR20	ACT(39)	F: TTGGGCAAACCAAAATTGC R: GCCAGAGGACAACAATTTGC	232-290	59.5	MN611937
CSSR22	TTCGTC(30)	F: CAAACTCAGCAGCTTCTTCTCC R: CGCCTTTGAGGATCTTCTCC	360-385	60	MN611938
CSSR16	TTG(33)	F: GGAGCCATAAATTTCGTGAAGG R: TTCCGCCATTGTCACTTCC	264-297	61.1	MN611936
CSSR28	AGGCTG(30)	F: GTGAGGCCAACCATGAGAGG R: GATCAGGTGGTGAATGTGGC	190-232	60.5	MN611940
CSSR31	ACTACC(30)	F: TGCCTTCTTTACAACTTTGAGTGC R: CAGAATCGCTTTACGACCCC	450-480	60.6	MN611942
CSSR30	ACTCAG(24)	F: TTGCTTTCTCCACCATTCCC R: CAAAGAGAGAAATTACGGAGACCC	212-235	59.8	MN611941
CSSR36	ATACAC(30)	F: TTTTGAGGTGCCCTTTAATGC R: GATGCTGGTCACATGGTTCC	440-461	59.8	MN611943
CSSR24	ATCTTC(30)	F: TTCGTATGATGAAATTTGGTTGC R: GTGGGAAAGACAACAATGCC	385-445	59.5	MN611939
CSSR37	TTGGGC(36)	F: TGGGTCCATCAGAAGTTACCG R: CCACATTTAGTCCCACATTGC	275-348	59	MN611944
CSSR38	ATAGGG(30)	F: GGTAAGAAGGCCAAGTTTCAAGC R: TCATCATCCTTGTCCCTACGC	306-370	61.1	MN611945

Table 2. Secondary SSR loci

Locus	Motif	Primer sequences (5'-3')	Allele size (bp)
CSSR5	AGGTGG	F: TATCTTCCCTTGTTGGCACG	240-280
		R: GGAGTTGAGGGTGTTGGAGG	
CSSR7	CGGGG	F: AACTAGGGATGCAAATGGGG	560*
		R: CGAATTGGTCCCCAGAGC	
CSSR15	TTTG	F: GTGTGAAAACCCCAAATTGC	530-550
		R: CTAATTTCGCAAACCCAGC	
CSSR18	ATG	F: CATGGAGAACGGGATTTTGG	600-640
		R: GAGCCCACTCAGATAAGCCC	
CSSR25	ATATAG	F: AAAAGTCGCAAAAGAGAGATAGC	310-330
		R: TTTCTCACTATTTGTCCTATACATAGCC	
CSSR27	AGGTCG	F: AAAATCAACGTGGCTGAGGC	195*
		R: TCCTCTTCCACGACCTCTCG	
CSSR29	AGCTTG	F: TTTTCTTCTGCATCCATCAACC	100-120
		R: CCTGAGTCCATGCTTGTGAGC	
CSSR32	ACAGGG	F: GAATCAGGGTTGGGGTGG	75-110
		R: GTACAAGGAAGGGGACAGGG	
CSSR33	ACAGAG	F: CCAAAATCAAAACGAATCTCTGC	160-200
		R: TGGATTCCAAAATGGGTTCG	
CSSR34	AAGTTC	F: CACCATCTTTGGGGACAAGG	327*
		R: AAATTTGCCAACAAGGTCAGG	
CSSR39	ATACAC	F:TTTTGAGGTGCCCTTTAATGC	420-460
		R:GATGCTGGTCACATGGTTCC	

*monomorphic loci

Table 3. Allele and genotype richness, and heterozygosity for 8 SSR loci in four triploid *C. pumila* populations

	Pinery PP (n=32)				Point Pelee NP (n=19)				Sideling Barrens (n=9)				Hermit Ridge (n=9)				St. Joe SP (n=8)			
	A	G	He	Ho	A	G	He	Ho	A	G	He	Ho	A	G	He	Ho	A	G	He	Ho
CSSR16	2	1	0.75	1	2	1	0.75	1	5	2	0.82	1.00	4	3	0.89	1.00	4	2	0.81	1.00
CSSR20	3	1	0.89	1	3	1	0.89	1	6	2	0.92	1.00	6	3	0.96	1.00	9	5	0.98	1.00
CSSR22	4	2	0.89	1	4	3	0.89	1	3	2	0.89	1.00	3	3	0.87	0.78	5	4	0.91	0.86
CSSR28	2	1	0.75	1	2	1	0.75	1	3	2	0.79	1.00	4	3	0.92	0.67	4	4	0.90	1.00
CSSR30	1	1	0.00	0	1	1	0.00	0	1	1	0.00	0.00	2	2	0.30	0.22	2	2	0.33	0.25
CSSR31	1	1	0.00	0	2	2	0.15	0	2	2	0.74	0.89	3	3	0.69	0.56	3	3	0.51	0.43
CSSR36	2	1	0.75	1	2	1	0.75	1	1	1	0.00	0.00	3	2	0.53	0.33	3	3	0.60	0.50
CSSR38	2	1	0.75	1	2	1	0.75	1	3	1	0.89	1.00	2	1	0.75	1.00	4	4	0.83	0.25
MLG	2				4				2				3				4			

A: number of alleles; G: number of unique genotypes; He: expected heterozygosity; Ho observed heterozygosity; MLG: number of unique multi-locus genotypes. He was calculated as $1 - \sum p^3$, where p is the frequency of each allele present in the population. Triploid allele frequencies were calculated using the simfreq function of the R package polysat (Clark and Jasieniuk 2001), which accounts for allele dosage ambiguity.

Table 4. Allele and genotype richness, and heterozygosity for 8 SSR loci in two diploid *Celtis* populations

	C. occidentalis			C. laevigata		
	Point Pelee NP			Sumter NF		
	A	He	Ho	A	He	Ho
CSSR16	6	0.69	0.88	1	0.00	0.00
CSSR20	7	0.76	0.85	6	0.71	0.90
CSSR22	2	0.17	0.19	2	0.18	0.20
CSSR28	3	0.56*	0.62*	4	0.60	0.38
CSSR30	1	0	0	2	0.10	0.10
CSSR31	3	0.48	0.42	1	0.00	0.00
CSSR36	2	0.42	0.42	2	0.32	0.40
CSSR38	2	0.31	0.23	4	0.48	0.60
MLG	25			10		

Columns as for Table 3. * indicates this locus differs signficantly from Hardy-Weinberg equilibrium for this population.

Table 5. Allele richness for 8 microsatellite primers in three Celtis species

		C. pumila (n = 77)		*C. occidentalis (n = 26)*		*C. laevigata (n = 10)*	
	Total	Species	Private	Species	Private	Species	Private
CSSR16	9	8	3	6	1	1	0
CSSR20	14	11	2	7	2	6	1
CSSR22	6	5	2	2	0	2	1
CSSR28	4	4	0	3	0	4	0
CSSR30	3	2	1	1	0	2	1
CSSR31	4	3	1	3	1	1	0
CSSR36	3	3	0	2	0	2	0
CSSR38	4	4	0	2	0	4	0

Total: total number of alleles over all 113 samples; Species: alleles present in each species; Private: number of alleles unique to each species.

Table 6. Voucher information for Celtis populations used in this study

Location	Latitude	Longitude	Collection Data
Point Pelee National Park, Ont	41.9522	-82.5152	*
Pinery Provincial Park, Ont	43.2575	-81.8347	*
PASH: Sideling Barrens, PA	39.7327	-78.3477	T. W. Smith & V. J. Nowell PASH1, 6 June 2016
KYHR: Hermit Ridge Road, KY	37.8221	-84.1623	T. W. Smith & V. J. Nowell KYHR1, 10 June 2016
MOSJ: St. Joe State Park, MO	37.8193	-90.5306	T.W. Smith, A. Hayes, T. Garant MOSJ, 2 June 2017
SCSF: Sumter National Forest, SC	34.5427	-81.7618	T.W. Smith, A. Hayes, T. Garant SCSF1, 8 June 2017
VARC: New River, VA	37.3891	-80.8672	T.W. Smith & V. J. Nowell TWS 16-007, 8 June 2016
IDID: Indiana Dunes National Lakeshore, IN	41.6737	-87.0192	T.W. Smith & V. J. Nowell IDID1, 11 June 2016
VAWR: Wilderness Road State Park, VA	36.6349	-83.5217	T.W. Smith & V. J. Nowell VAWR2, 9 June 2016
GC: Trenton, Ont	44.2231	-77.5867	T.W. Smith & P. Catling, GC1, 29 June 2016

* All vouchers are deposited at DAO. No *Celtis pumila* vouchers were collected from Ontario populations, as it is a listed species (Threatened) in Canada.

Acknowledgements

T. Garant, V. Nowell, D. VanExan and P. Catling for assistance in the field, S. Grund and W. W. Overbeck provided locality information. This work was funded by Point Pelee National Park.

References

Castoe, T. A., A. W. Poole, A. P. J. de Koning, K. L. Jones, D. F. Tomback, S. J. Oyler-McCance, J. A. Fike, et al. 2012. Rapid Microsatellite Identification from Illumina Paired-End Genomic Sequencing in Two Birds and a Snake. *PLOS ONE* 7: e30953.

COSEWIC. 2003. COSEWIC assessment and update status report on the dwarf hackberry *Celtis tenuifolia* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa, Ontario, Canada.

Faircloth, B. C., and T. C. Glenn. 2012. Not All Sequence Tags Are Created Equal: Designing and Validating Sequence Identification Tags Robust to Indels S.-H. Shiu [ed.], *PLoS ONE* 7: e42543.

Fernald, M. L. 1950. Gray's manual of botany. Van Nostrand Company, New York.

Hayes, A. 2019. The genetic structure of *Celtis tenuifolia* and comparisons to the related species *C. occidentalis*, and *C. laevigata*: implications for the conservation management of threatened populations in Southern Ontario. MSc Thesis. Carleton University, Ottawa, Ontario, Canada.

Paradis, E. 2010. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 26: 419–420.

Sherman-Broyles, S. L., W. T. Barker, and L. M. Schulz. 1997. Ulmaceae Mirbel. *In* Flora of North America Editorial Committee [ed.], Flora of North America Volume 3: Magnoliophyta: Magnoliidae and Hamamelidae, Oxford University Press, New York.

Whittemore, A. T. 2005. Genetic Structure, Lack of Introgression, and Taxonomic Status in the *Celtis laevigata—C. reticulata* Complex (Cannabaceae). *Systematic Botany* 30: 809–817.

Whittemore, A. T. 2013. Ulmaceae (Elm Family). *In* G. Yatskievych [ed.], Steyermark's Flora of Missouri: Volume 3, Missouri Botanical Garden and the Conservation Commission of the State of Missouri, St. Louis, Missouri.

Whittemore, A. T., and A. M. Townsend. 2007. Hybridization and Self-compatibility in *Celtis*: AFLP Analysis of Controlled Crosses. *Journal of the American Society for Horticultural Science* 132: 368–373.