

The genetic diversity of triploid *Celtis pumila* and its diploid relatives *C. occidentalis* and *C. laevigata*

Andrew Hayes, Song Wang, Alan Whittemore, Tyler Smith

Abstract

Celtis pumila is a shrub or small tree, native to eastern North America. Its taxonomic status is uncertain, as it frequently co-occurs with plants morphologically intermediate to its sympatric congeners, *C. occidentalis* in the north, and *C. laevigata* in the south. While these three taxa are generally treated as distinct taxa, many authors have suggested hybridization may be blurring the boundaries among them. We collected tissue samples from populations of all three taxa across the full range of *C. pumila*. Flow cytometry data confirm the co-occurrence of diploids and triploids across this range, and, at two locations, tetraploids. Microsatellite data confirms that there are two diploid species, *C. occidentalis* and *C. laevigata*, and that they do hybridize where they co-occur in southern Missouri and Illinois. The triploid samples formed two geographically-oriented clusters. The northern group showed some subtle genetic and morphological similarity to *C. occidentalis*, while the southern group was more similar to *C. laevigata*. However, while there may be rare gene-flow from the diploids to the triploid, most triploid populations are dominated by one or a few apomictic clonal genotypes.

Keywords

apomixis, *Celtis*, hybridization, microsatellite, polyploidy, triploid

Introduction

(remove heading from paper)

Celtis pumila Pursh (sensu Whittemore (2013); *C. tenuifolia* Nutt. in Barker (1997) and earlier authors; Cannabaceae; Dwarf Hackberry) is a shrub or small tree, native to eastern North America. Its range consists primarily of the southeastern United States, from Missouri, Kentucky, and Vermont south to Florida, and from Kansas and Texas east to the Atlantic Coast (Barker 1997). There are also smaller, disjunct populations in the Great Lakes regions of Ohio, Indiana, Michigan, and southern Ontario (Wagner 1974). *Celtis pumila* forms part of a poorly-understood complex of apomictic shrubs and small trees extending across the continent, referred to as the “small tree complex” by Whittemore (2008).

Within this range, *C. pumila* is sympatric with two diploid species of *Celtis* L. subg. *Celtis*: *C. occidentalis* L. in the north, and *C. laevigata* Willd. in the south. The three species share several distinctive characteristics: corky ridges on larger stems and branches; polymorphic ovate or lanceolate leaves with more-or-less oblique bases; andromonoecious, with male and bisexual flowers on the same twigs, with pollen catapulted explosively into the air exactly as described for *Morus* L. by Taylor et al. (2006); and spherical drupes 5–10 mm in diameter, with a thin mesocarp over a hard stone (Barker 1997).

Celtis laevigata and *C. occidentalis* are both canopy trees, typically with one strong vertical trunk (occasionally 2 or 3 parallel upright trunks), reaching 30 m in height at maturity, if small then with erect leaders, and bark

with corky warts or ridges over the whole trunk and major branches. *Celtis pumila* is limited to the subcanopy, less than 8 m in height, with leaning trunks and horizontal or arching leaders or (with age) multiple spreading trunks, and bark with large corky warts at base of trunk and around old wounds, otherwise smooth; in harsh habitats (open dunes and alvars, serpentine barrens etc.) it forms stunted shrubs less than 2 m tall. *Celtis laevigata* generally has narrow, lanceolate, lance-triangular leaves that are generally entire (but var. *smallii* has 2-15 irregular teeth on each margin). *Celtis occidentalis* has broader, ovate or triangular leaves with conspicuously serrate margins [(6)-9-39 teeth on each margin]; in contrast, *C. pumila* leaves are shorter, entire or with up to 17 teeth per margin. *Celtis occidentalis* and *C. pumila* can also be distinguished by the number of major areoles on the leaves: *C. occidentalis* has 5-8, compared to 3-5 in *C. pumila* (Wagner 1974).

However, while typical forms of the three species appear quite distinctive, intermediates abound, displaying combinations of characters that defy ready identification. Small, vegetative plants are especially challenging to identify. This variation has resulted in conflicting opinions on the appropriate taxonomic classification of the group. Fernald (1950) recognized three species (*C. occidentalis*, *C. laevigata* and *C. tenuifolia*), which he divided into eight subspecies based on slight morphological differences and growth substrate. Gleason and Cronquist (1991) maintained the same three species, but eliminated the subspecies. Furthermore, plants which Fernald (1950) placed in *C. occidentalis* (as *C. occidentalis* var. *pumila*) were subsumed within *C. tenuifolia* by Gleason and Cronquist (1991). A related species, *Celtis reticulata* Torrey, occurs from Texas and Kansas, where it is sympatric with *C. pumila* and *C. laevigata*, ranging west to the Pacific coast. *Celtis reticulata* has been treated as a variety of both *C. occidentalis* and *C. laevigata* (Barker 1997).

Many authors have suggested hybridization may be contributing to the confusion (e.g., Buck and Bidlack 1998, COSEWIC 2003). Indeed, Fernald (1950) noted that the three species were often “seemingly confluent”. However, recent research has suggested that hybridization is not as prevalent as these references suggest. Whittemore (2005) found no evidence for introgression between *Celtis laevigata* and *C. reticulata*, two species formerly claimed to introgress frequently, although one of these species, *C. reticulata*, has since been found to show frequent apomixis. Whittemore and Townsend (2007) found that *Celtis* species are not easy to hybridize: extensive controlled pollination experiments involving 23 potential parents belonging to seven *Celtis* species yielded only two hybrid progeny.

The main questions addressed by this project are: 1) How many genetic groups of *Celtis* exist in eastern North America, and do they relate directly to their cytotypes (distinct diploid and triploid groups)? 2) Is there any evidence of hybridization or gene flow between these groups?

Answering these questions is necessary to resolve the taxonomic relationships in this group. It is also important to inform conservation planning. *Celtis pumila* is a Threatened species in Canada, but uncertainty around its identification and suspicions of hybridization with *C. occidentalis* present challenges to effective population management.

Methods

Field sampling

In May of 2016 and May and June of 2017, we sampled 38 *Celtis* sites across the eastern United States and southern Canada (Table 1). Wild *Celtis* were sampled from representative sites across the range of the genus in the eastern US and Canada. In order to test for interspecific hybridization and potential gene flow among the species, concentrated sampling was carried out in the area around the lower Ohio River and its junction with the Mississippi, where the ranges of all three native *Celtis* species overlap. In addition, one population of *C. reticulata* Torr., a western member of the “small tree complex” similar in morphology to *C. pumila*, was collected in north-central Texas, along with a sympatric population of *C. laevigata*. Whenever possible, we sampled 20 individuals from a location, and included the full range of morphological variability present. A minimum distance of 2 m between plants was used to reduce the probability of sampling from the same clone or sibling groups. In total, 381 samples were collected, with sample sizes ranging from 4-30 and averaging 15

samples per population (mean = 15.2). Vouchers are deposited at DAO (herbarium acronyms follow Thiers 2020).

Table 1. *Celtis* populations sampled.

Site	Location (State/Prov, County)	Latitude	Longitude
Canada			
ONPE	Ontario, Essex	41.952	-82.515
ONLO	Ontario, Hastings	44.256	-77.125
ONGC	Ontario, Hastings	44.223	-77.587
ONPA	Ontario, Hastings	44.147	-77.300
ONPP	Ontario, Lambton	43.257	-81.835
United States			
ALCC	Alabama, Butler	31.956	-86.853
ALHR	Alabama, Butler	31.725	-86.472
ALAC	Alabama, Crenshaw	31.869	-86.261
ARCC	Arkansas, Izard	35.934	-91.911
ARWR	Arkansas, Izard	35.911	-91.928
ARGR	Arkansas, Randolph	36.284	-91.186
ILGG	Illinois, Hardin	37.598	-88.381
ILLG	Illinois, Jackson	37.683	-89.396
ILWB	Illinois, Johnson	37.373	-88.925
ILHE	Illinois, Pope	37.583	-88.441
INCB	Indiana, Steuben	41.630	-85.083
INEM	Indiana, Knox	38.786	-87.469
INID	Indiana, Porter	41.745	-87.087
KYBL	Kentucky, Robertson	38.451	-83.993
KYHR	Kentucky, Madison	37.822	-84.162
MIWA	Michigan, Washtenaw	42.390	-84.054
MOCC	Missouri, Shannon	37.169	-91.126
MOGR	Missouri, St. Louis	38.503	-90.702
MOKC	Missouri, St. Louis	38.500	-90.689
MOMA	Missouri, Jefferson	38.379	-90.394
MOSJ	Missouri, St. Francois	37.819	-90.530
MSCS	Mississippi, Clarke	32.105	-88.697
MSSM	Mississippi, Kemper	32.870	-88.720
MSWD	Mississippi, Marshall	34.662	-89.463
OHCQ	Ohio, Erie	41.388	-82.830
OHES	Ohio, Erie	41.425	-82.643
PABE	Pennsylvania, Bedford	40.013	-78.432
PASH	Pennsylvania, Fulton	39.732	-78.347
SCCC	South Carolina, Laurens	34.335	-82.028
SCSF	South Carolina, Laurens	34.542	-81.761
TXFW	Texas, Tarrant	32.837	-97.476
VADS	Virginia, Bath	37.896	-79.799
VARC	Virginia, New River	37.389	-80.867
VAWR	Virginia, Claiborne	36.634	-83.521

For each sample, three leaves were removed and dried with silica gel. These were stored at room temperature until extraction occurred. Eight samples were taken from each population for genetic analysis. All samples from populations with 8 or fewer individuals were included in the genetic analysis (OHES, INCB, ILHE, MOCC, MSWD).

Flow Cytometry

We used flow cytometry to assess the ploidy of all of our samples. For each sample, a 0.7 cm x 0.7 cm piece of leaf tissue was chopped into fine powder in 750 μ l Tris MgCl₂ buffer (Pfossner et al. 1995) and let stain on ice in complete darkness for 30 min with 50 μ l RNAse and 250 μ l Propodium Iodide mixed with Galbraith buffer (Galbraith et al. 1983). Samples were then analyzed with a Gallios Flow Cytometer (Beckman Coulter, Brea, California, U.S.A.) and compared to petunia and soy standards. The flow cytometry histograms were analyzed with the `flowPloidy` R package (Smith et al. 2018).

An individual of *Celtis sinensis* at the U. S. National Arboretum (living plant accession number NA 69704-J) yielded a diploid meiotic chromosome number ($2n = 20$), and a nuclear DNA content of $2C = 1.037$ pg, confirming a genome size of $C = 0.5185$ pg for a haploid set of ten chromosomes. Surveys of multiple *Celtis* spp. (data not shown) indicate that this C value is constant across species of subg. *Celtis*, so ploidy levels were calculated using this figure.

DNA Extraction and Microsatellite Assays

For each leaf sample, 15 mg of silica was ground with 5 mm stainless steel beads in a 2 mL Eppendorf tube at 30 Hz for 2 min using a Retsch MM 300 Tissue Lyser (Retsch, Haan, Germany). From the ground tissue, genomic DNA was extracted using the Plant Nucleospin II Genomic DNA Kit (Machery-Nagel, Düren, Germany). Concentration and quality of obtained DNA was verified using a Nanodrop 2000 UV-Vis Spectrophotometer (ThermoScientif, Walktham, Massachusetts, U.S.A.).

We used genome-skimming to develop eight nuclear polymorphic microsatellite markers for use in this study (see Appendix XX). Polymerase Chain Reactions were carried out with 8 μ L master mixes containing 0.24 μ L DMSO, 4.0 μ L 2X Phusion Master Mix w/ High Fidelity Buffer (New England Biolabs), 0.416 μ L mix of Left and Right Primers (Sigma-Aldrich, Oakville, ON), 0.192 μ L of FAM or VIC-labelled CAG Tag (10 μ M), 2.152 μ L ddH₂O and 1.0 μ L DNA, using a T-100 Thermal Cycler (Bio-Rad, Hercules, California, U.S.A.). The thermal cycling profile of touchdown PCR (TD-PCR) was used to prevent mispairings of primers and achieve higher quality amplification (Don et al. 1991). PCR products were run and observed on a 1% agarose gel stained with GelRed (Biotium, Hayward, California, U.S.A.) and viewed with a High Performance UV Transiluminator (UVP, Upland, California, U.S.A.) with a 100 bp DNA ladder (New England Biolabs) to confirm the presence and size of the amplicons, along with negative controls for all primer pairs to confirm absence of contaminants prior to genotyping.

Amplicons were then pooled into groups and visualized by capillary electrophoresis using a 3500xl Genetic Analyzer (Applied Biosystems) with the GeneScan 500 and 1200 LIZ Size Standard (Applied Biosystems). Individual samples were genotyped with Geneious v11.1.4 software (Biomatters, Auckland New Zealand) and manually scored. Samples that failed for three or more SSR loci were dropped from the analysis. A total of 332 samples were genotyped and cytotyped at at least 6 loci.

Population structure

STRUCTURE analysis (Falush et al. 2003) was used to assign the diploid samples into genetic clusters, and identify possible inter-cluster hybrids. We excluded triploids and tetraploids from this analysis, as they appear to be predominantly apomictic (based on displaying limited or no diversity within populations, see below), and so violate the assumptions of the STRUCTURE model. As we expected our samples to include at least two reproductively-isolated species (genetic clusters), we set the model to leave allele frequencies uncorrelated among clusters. We ran 20 replicates for each value of K (1-10). Each replicate started with 20,000 burn-in cycles, followed by 50,000 simulations. We calculated Evanno's ΔK (Evanno et al. 2005) to select the most informative number of groups.

We classified individuals with an inferred proportion of admixture greater than 70% for one of the clusters as members of that species. Individuals which were not assigned to any cluster with more than 70% probability

were identified as possible hybrids. This threshold was selected arbitrarily. The results were identical using any threshold between 66 and 72%. Setting the threshold to 75% results in two additional potential hybrids being identified, both from populations already identified with the lower threshold.

We complemented the STRUCTURE analysis with principal coordinate analysis (PCoA). PCoA is less informative than STRUCTURE, but makes no assumptions about reproductive biology, so can accommodate samples with a mixture of sexual diploids and apomictic polyploids. The PCoA was based on Bruvo's distance (Bruvo et al. 2004), as computed by the R package **polysat** (Clark and Jasieniuk 2011). Bruvo's distance incorporates genotypic distances between individuals of different ploidies, and uses a stepwise mutation model, as appropriate for microsatellites (Kimura and Ota 1975).

Genetic diversity

The R packages **polysat**, **adegenet** (Jombart 2008) **poppr** (Kamvar et al. 2014) and **hierfstat** (Goudet and Jombart 2015), were used to calculate population-level genetic diversity statistics for the populations sampled. Observed heterozygosity (H_e), gene diversity (H_s), and the inbreeding coefficient (F_{is}) were calculated only for the diploids, as the triploid populations appear to be mostly apomictic and so violate the assumptions underlying F statistics.

Species-level genetic diversity measures were calculated for diploid clusters from the STRUCTURE analysis (i.e., *C. occidentalis* and *C. laevigata*) were calculated, including number of private alleles (A_p), observed heterozygosity (H_e), gene diversity (H_s), the inbreeding coefficient (F_{is}), and differentiation among populations (F_{st}).

We used a distance threshold to identify triploid clones, following the approach of Meirmans and Tienderen (2004). PCR artifacts and somatic mutations may introduce small differences between individuals from the same clonal lineage. We set the distance threshold based on visual inspection of a histogram of triploid inter-individual distances.

We calculated pairwise- F_{st} among diploid clusters and triploid groups to better characterize their relationships.

Morphology

Following completion of the molecular analysis, herbarium specimens were examined to identify morphological characters corresponding to the genetic groups revealed. We included specimens from A, CAS, F, MO, NA, NY, TEX, UC, and USCH.

Results

Our samples included 171 diploids (DNA-ploidy mean 1.22 pg, sd 0.03) at 26 locations, 159 triploids (DNA-ploidy mean 1.78 pg, sd 0.13) at 24 locations, and 2 tetraploids (DNA-ploidy mean 2.32 pg, sd 0.15) at two locations (Table 2). 12 of the 26 triploid populations co-occurred with diploids. One of the tetraploids was collected in an otherwise triploid population, and one was collected from a location with diploids and triploids.

Diploid Population Structure

STRUCTURE analysis of the diploid samples indicated they formed two distinct clusters; the ΔK value peaked at 2 groups (Fig. 1). Based on field examination of leaf morphology (unpublished data), and geographic distribution, these clusters correspond to the species *C. occidentalis* and *C. laevigata*.

Table 2.

Site	species	N	MLG	Ho	Hs	Fis
Canada						
ONGC	pumila	6	1			
ONLA	occidentalis	2	1			
	pumila	3	1			
ONPA	pumila	1	1			
ONPE	occidentalis	26	25	0.458	0.433	-0.029
	pumila	19	1			
ONPP	pumila	32	1			
USA						
ALAC	laevigata	3	3	0.542	0.594	0.084
	pumila	5	2			
ALCC	laevigata	8	8	0.444	0.593	0.287
ALHR	laevigata	6	6	0.427	0.521	0.126
	pumila	2	1			
ARCC	laevigata	8	8	0.561	0.520	-0.015
	pumila	1	1			
ARGR	laevigata	6	6	0.479	0.544	0.095
ARWR	laevigata	7	7	0.518	0.528	-0.031
ILGG	pumila	9	1			
ILHE	occidentalis	1	1	0.714		
	pumila	1	1			
	hybrid	2	2	0.375	0.750	0.267
ILLG	laevigata	8	8	0.511	0.613	0.147
	hybrid	1	1	0.429		
ILWB	laevigata	3	3	0.625	0.583	-0.152
	hybrid	3	3	0.417	0.542	0.183
	pumila	4	1			
INCB	occidentalis	4	4	0.438	0.490	0.064
INEM	laevigata	4	4	0.500	0.620	0.172
	occidentalis	5	5	0.500	0.562	0.080
	hybrid	1	1	0.500		
INID	pumila	5	1			
KYBL	pumila	1	1			
KYHR	pumila	9	3			
MIWA	pumila	8	1			
MOCC	occidentalis	2	2	0.625	0.500	-0.361
	pumila	6	5			
MOGR	occidentalis	2	2	0.750	0.571	-0.278
	pumila	6	2			
MOKC	occidentalis	8	8	0.556	0.534	-0.045
MOMA	laevigata	5	5	0.475	0.456	-0.045
	occidentalis	1	1	0.500		
	hybrid	4	4	0.594	0.589	-0.031
MOSJ	pumila	8	4			
	tetraploid	1	1			
MSCS	laevigata	12	12	0.354	0.426	0.260
MSSM	laevigata	8	8	0.344	0.432	0.147
MSWD	laevigata	6	6	0.433	0.510	0.123
OHCQ	occidentalis	2	2	0.438	0.719	0.238
	pumila	6	1			
OHES	occidentalis	8	8	0.433	0.452	0.023

Site	species	N	MLG	Ho	Hs	Fis
PABE	pumila	2	2			
PASH	pumila	9	2			
SCCC	laevigata	2	2	0.438	0.429	-0.133
	pumila	5	1			
	tetraploid	1	1			
SCSF	laevigata	10	10	0.317	0.313	-0.035
TXFW	laevigata	1	1	0.625		
	reticulata	3	2			
VARC	occidentalis	2	2	0.500	0.469	-0.056
VAWR	pumila	8	3			

Note: Site codes as in Table 1. The first two letters indicate the state or province. N is the number of individuals sampled from each species at a location. MLG is the number of unique multi-locus genotypes detected. Ho is observed heterozygosity, Hs is gene diversity, and Fis is the within-population fixation index. Gene diversity and Fis were not calculated for triploids (clonal apomicts) or populations where only a single individual was genotyped.

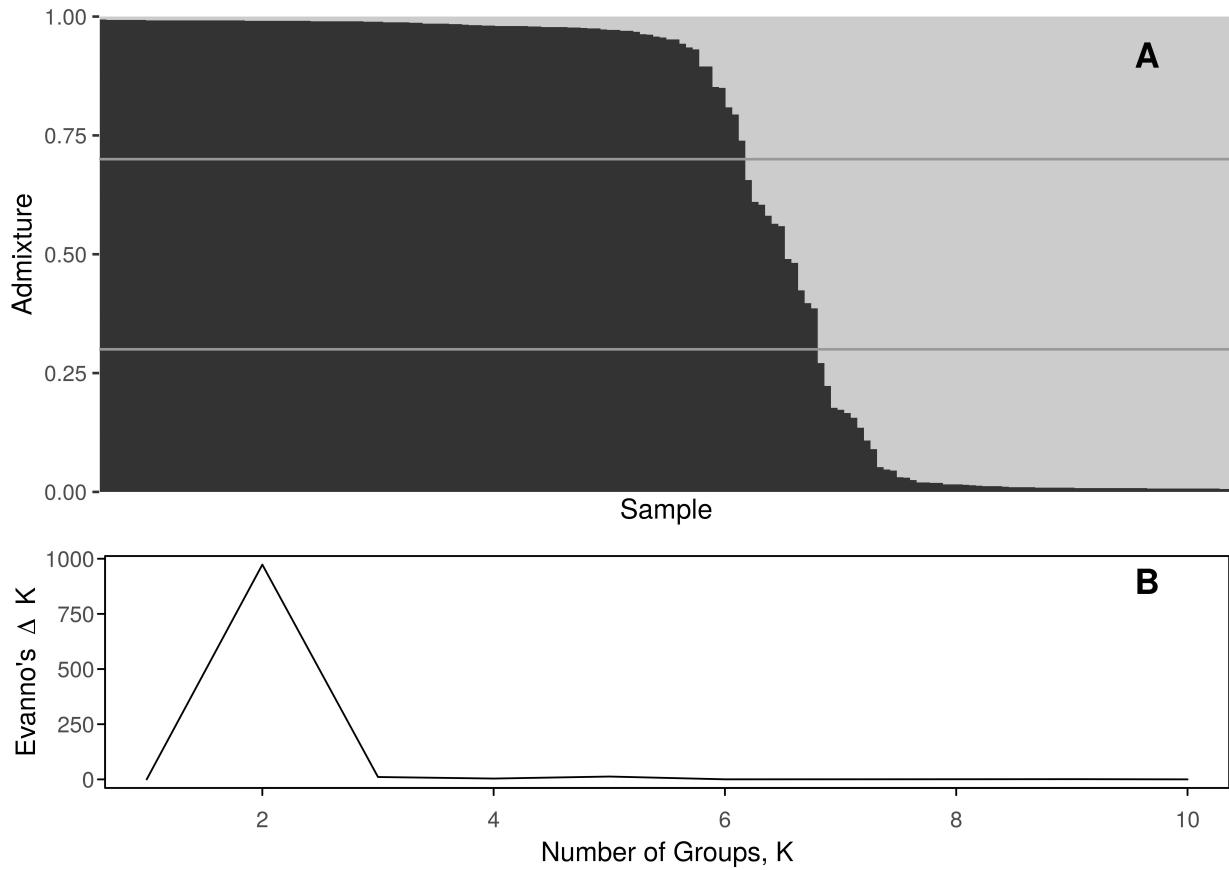


Figure 1: STRUCTURE analysis of diploid samples. A. STRUCTURE plot. Samples are sorted in order of proportion of inferred admixture from the two identified clusters, *C. laevigata* (dark) and *C. occidentalis* (light). This highlights the 11 putative hybrids, distinguished as having less than 0.7/more than 0.3 inferred contribution from each cluster. The horizontal lines show these thresholds. B. Evanno's delta K values for STRUCTURE analyses, showing the clear peak at $K = 2$.

There were 97 individuals in the *C. laevigata* cluster, and 63 in the *C. occidentalis* cluster. Additionally, 11 individuals had an inferred proportion of ancestry < 70% from both species. We interpreted this as evidence of potential inter-species hybridization (Fig. 1). These 11 individuals were found in ILHE, ILLG, ILWB (southern Illinois), INEM (southern Indiana), and MOMA (southern eastern Missouri). Both *C. laevigata* and *C. occidentalis* were common throughout this area, and co-occurred together with the hybrids at INEM and MOMA (Fig. 2).

Principal Coordinates analysis (PCoA) of the diploids revealed the intermediate genetic position of the putative hybrids between the two diploid species (Fig. 2). There is no evidence of a third diploid plant group corresponding to *C. pumila*.

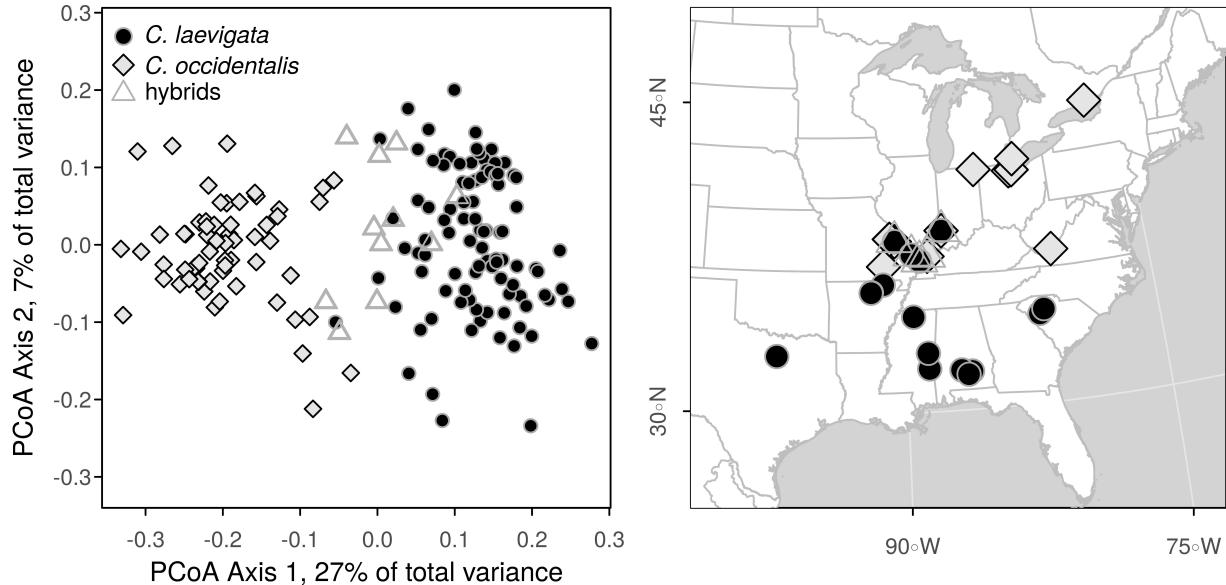


Figure 2: Diploid *Celtis*. Left: PCoA of the Bruvo distances among 171 diploid *Celtis* sample microsatellite genotypes. Right: Geographic locations of the samples from the PCoA plot.

Triploid Population Structure

The PCoA revealed two geographic clusters of triploids (Fig. 3). The northern samples, including Ontario, northern Ohio, southeastern Michigan, and southern Pennsylvania formed a relatively tight cluster; the remainder of the samples, including locations from Indiana and Virginia to Arkansas and Mississippi, formed a looser cluster. The single population of *C. reticulata* was placed within the southern triploid cluster.

Note to Alan: The outlying group the upper right corner of Fig. 3 is two populations from Illinois, ILGG (9 individuals) and ILHE (1 individual). These are all the triploids from these two locations. The other outlying cluster, on the right side of the plot, is population ILWB (4 individuals). There are additional Illinois populations that plot in the main Southern cloud. Not having a good explanation, I didn't comment on this in the text. It's very suggestive though!

The two “southern” points that are closest to the “northern” cluster are one individual from Alabama, and another from Missouri.

When the diploids, triploids and tetraploids are combined in a single PCoA, the diploid hybrids are again placed between the two diploid species (Fig. 4). The southern triploids form a cloud between the two diploid species, but with more overlap with the *C. laevigata* cluster. The northern triploids are contiguous with the

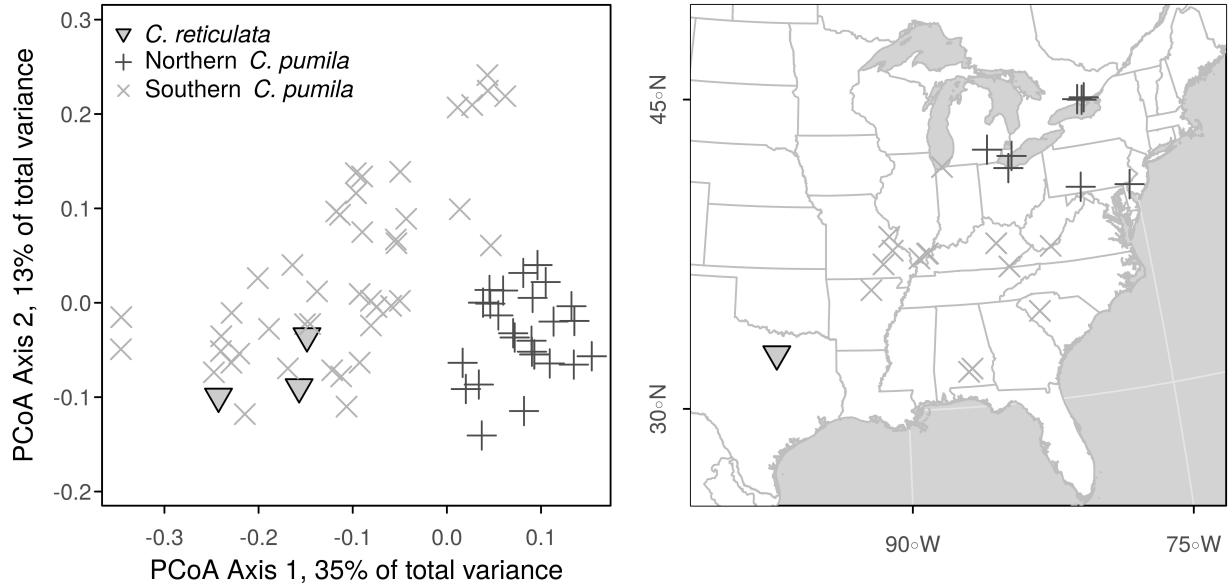


Figure 3: Triploid *Celtis*. Left: PCoA of Bruvo distances among 159 triploid *Celtis* sample microsatellite genotypes. Right: Geographic locations of the samples from the PCoA plot.

Table 3: Population diversity summaries for diploid **Celtis** samples

	Ap	Ho	Hs	Fis	Fst
<i>occidentalis</i>	7	0.517	0.501	-0.030	0.054
<i>laevigata</i>	17	0.474	0.517	0.082	0.069

southern triploids, but are more removed from the diploids. The two tetraploids are also in an intermediate position between the diploids.

Genetic diversity

When the two diploid clusters (i.e., *C. occidentalis* and *C. laevigata*) were compared across all 8 loci, *C. occidentalis* had seven private alleles, and *C. laevigata* had seventeen (Table 3).

Visual inspection of the triploid inter-individual distances showed a bimodal distribution (Fig. 5). We interpreted the lower mode to represent distances within clonal lineages. In apomictic *Celtis*, the embryo is formed by meiotic diplospory (unpublished data), so limited genetic variation may be created by crossing-over during the incomplete meiosis that forms the unreduced megasporangium, and plants in the same lineage are not expected to be genetically identical (van Baarlen et al. 2000, van Dijk and Bakx-Schotman 2004). Accordingly, we set a threshold distance of 0.1 in identifying apomictic clones.

After applying this threshold, we found a total of 32 triploid multi-locus genotypes (MLG). 28 of the MLGs were found in a single population each. The remaining four MLG were found in 2-6 populations each. The same MLG was shared by all individuals at three Ontario locations (ONGC, 6 individuals; ONPE, 19 individuals; PPP, 32 individuals), one Michigan location (MIWA, 8 individuals), one Ohio location (OHCQ, 6 individuals); and by one of two individuals sampled at one Pennsylvania location (PABE). Each of these locations are part of the northern cluster of *C. pumila*.

Another MLG is shared between two Illinois locations (ILGG, 9 individuals, and ILHI, 1 individual). A third is shared between an Illinois location (ILWB, 4 individuals) and a Missouri location (MOCC, 1 individual).

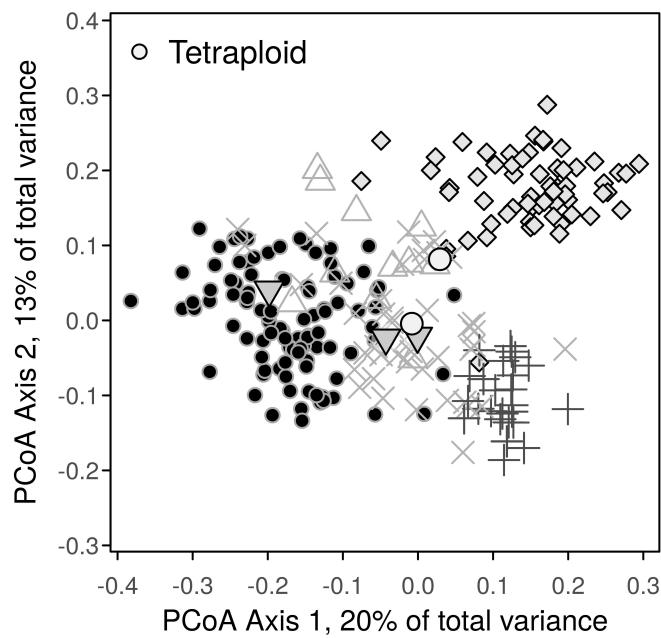


Figure 4: PCoA of Bruvo distances among 332 *Celtis* sample microsatellite genotypes, including diploids, triploids and tetraploids. Symbols are the same as in previous plots, with the addition of white circles for the tetraploid samples.

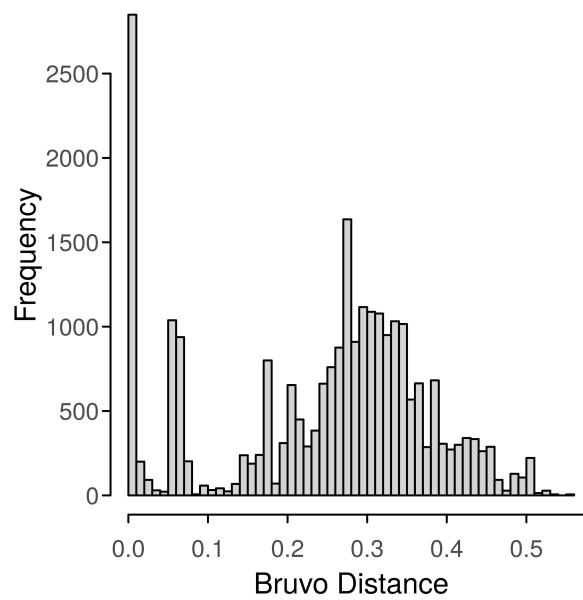


Figure 5: Bruvo Distance among triploid samples

Table 4: Pairwise Fst among **Celtis** genetic clusters

	laevigata	occidentalis	North	South	C. reticulata
laevigata	0.000	0.169	0.138	0.045	0.017
occidentalis	0.169	0.000	0.135	0.104	0.060
North	0.138	0.135	0.000	0.077	0.029
South	0.045	0.104	0.077	0.000	0.011
C. reticulata	0.017	0.060	0.029	0.011	0.000

The fourth multi-location MLG is shared between two locations in Missouri (MOSJ, 4 individuals; and MOGR, two individuals).

The northern triploid cluster, which spans nearly 700 km, has only six multi-locus genotypes, and there are only three in Ontario. Furthermore, most of Ontario consists of only one genotype (57 of 61 individuals sampled). This is in stark contrast to the southern cluster, which has 26 multi-locus genotypes in total, and 12 within 400 km of each other in Illinois, Missouri and Arkansas. Additionally, with the exception of Pennsylvania, all of the northern populations have only one MLG per population, whereas the southern populations have up to five. Thus, there is greater genetic diversity in the southern region than the northern region, and greater diversity within individual southern populations than individual northern populations.

Fst values were calculated between the various species (Table 4). The differentiation between the triploid groups and each of the diploids is less than the differentiation between the diploids. The northern triploids are nearly equidistant from both diploids, while the southern triploids are approximately half the distance from *C. laevigata* as they are from *C. occidentalis*.

Morphology

There was considerable morphological variation within the triploids, but little of this variation is aligned with the two geographic clusters revealed by our analysis. The northeastern and southwestern plants are similar in almost all their morphological characters, differing only in the range of variation in leaf serration and pedicel morphology. The northeastern plants have more serrate leaves (the leaf subtending a pedicel is almost always toothed, often with as many as 20-40 teeth per leaf, while corresponding leaves on southwestern *C. pumila* are usually entire, seldom with more than ten teeth per leaf), and pedicels of the northeastern plants can be longer (4-19 mm long, vs. 5-11 mm in southwestern plants) and sometimes sparsely pubescent or glabrous (southwestern plants have the pedicels always hispid). Both morphologies are found in the Carolinas and Georgia, where we have few SSR samples.

Discussion

Our analysis confirms two distinct diploid *Celtis* species in eastern North America, *C. occidentalis* and *C. laevigata*. Typical mature plants are readily distinguishable based on leaf morphology, and they have largely distinct ranges. However, where they co-occur, particularly in southern Illinois, Indiana, and Missouri, we could not reliably identify immature plants. In fact, our provisional field identifications of samples in this area were frequently wrong, and were corrected with our SSR fingerprint and ploidy data.

While hybridization among *Celtis* taxa has long been suspected in eastern North America (Fernald 1950), our data is the strongest evidence to date supporting this hypothesis. Buck and Bidlack (1998) invoked hybridization to explain inconsistencies in clusters defined by morphological and isozyme data in *C. laevigata*, *C. occidentalis*, and *C. reticulata*. However, patterns they observed could also have resulted from plasticity, or lack of taxonomically informative variation in the markers they employed. Whittemore and Townsend (2007) were unable to produce *C. occidentalis* x *C. laevigata* hybrids in controlled crosses, but only two attempts were made, using the same two individual parent trees.

The sampling we report here is more extensive than previous studies. Our structure analysis shows two clear, geographically coherent genetic clusters corresponding to *C. occidentalis* and *C. laevigata*, with a small number of intermediates. These intermediate individuals were all found in locations where both species co-occur in close proximity, further supporting our view that these are true hybrids.

In contrast, SSR variation in the triploids does not correlate well with morphological variation, or with any published taxonomy. These minor morphological differences between the southern and northern *C. pumila* clusters could be due to genetic influence from *C. occidentalis* (with strongly toothed leaves and longer, glabrous pedicels) in the north and *C. laevigata* (with generally entire leaves and shorter pedicels, though its pedicels also are usually glabrous) in the southern regions. However, even these few characters overlap broadly when the two triploid clusters are compared. There is no sharp division into two clusters, as seen with the SSR data, and the triploids as a whole share characters that are not present in either of the diploid species (see above).

The inclusion of our single population of *C. reticulata* in the southwestern cluster was unexpected, but it makes sense on morphological grounds, as the two taxa have many characters in common. Both are large shrubs or small trees to 12 m tall, with leaning trunks and horizontal or arching leaders or (with age) multiple spreading trunks. They are generally triploid and apomictic, with pollen largely or entirely aborted, with the stainable grains (if present) very variable in size and number of pores, or else anthers remaining small and indehiscent. They also show the same complex positional leaf polymorphism. Juvenile leaves are thin, green, ovate or elliptical with shorter, blunter apices, and always strongly toothed. Leaves of fertile laterals are triangular to triangular-ovate, with margins entire or variously toothed, and apex acute or acuminate. Leaves of vigorous leaders are much larger (up to twice as long), ovate, and the marginal serration and apex often differ strikingly from leaves on lateral branches of the same plant. It is not possible to draw taxonomic conclusions from a three *C. reticulata* individuals, but it is clear from our results that future work must consider the whole *C. pumila* - *C. reticulata* complex (the “small tree complex” of Whittemore (2008)) as a whole.

Notably absent in our diploid data was any indication of a third group corresponding to *C. pumila*. It’s possible that diploid *C. pumila* exists, but it is clearly uncommon. However, other surveys we are conducting have revealed rare plants of *C. reticulata*, which our data indicates is part of the same genetic complex.

Gametophytic apomicts are commonly polyploid (Bicknell and Koltunow 2004, Asker and Jerling 2017). Several different explanations have been suggested for this correlation (Whitton et al. 2008), and some authors have found a connection with hybridization between species or distinct genotypes (Hojsgaard and Hörandl 2019). It is tempting, given our data, to suggest that *C. pumila* may be a triploid derived from both *C. occidentalis* and *C. laevigata*. We have established that these two species do, at least occasionally, hybridize. Further, the triploids show some overlap with diploid hybrids in our PCoA (Fig. 4), though the triploid clusters are centered well below and to the right of the diploid hybrids in our diagram; and the genetic distance (pairwise Fst) between the triploids and the diploids is smaller than the distance between the diploid species (Table 4).

However, *C. pumila* differs strikingly from putative hybrids between *C. laevigata* and *C. occidentalis* both morphologically and ecologically. It has characters not found in either of the diploid species, but found in diploid *C. reticulata*, especially the distinctive growth form (small, with leaning or spreading trunks, becoming multitrunked with age; leaders ± horizontal or arching) and their strong adaptation to edaphically dry sites, such as sand deposits and thin soil over rock outcrops. By contrast, *C. laevigata* and *C. occidentalis* are both excurrent trees that can reach 30 m in height, typically with one strong vertical trunk (occasionally 2 or 3 parallel upright trunks) and erect leaders, and they typically grow in well-watered lowland (often bottomland) sites. This suggests that diploids belonging to the “small tree complex,” perhaps now occurring only farther west, need to be investigated as possible contributors to the extensive triploid populations found across the continent.

The differing relationships between the southern and northern triploids and the diploid species may indicate recent or ongoing gene flow from *C. laevigata* and/or *C. occidentalis* to adjacent triploid lineages. Recurrent formation of triploids could be providing a one-way transfer of alleles from the diploid species to regional populations of *C. pumila*. It has already been noted above that the minor morphological differences between

the northeastern and southern clusters of *C. pumila* follow the morphology of the diploid species in these respective areas. Gene flow from diploid sexual plants into triploid apomictic congeners is well known in many apomictic groups, but the mechanism is not well understood. Verduijn et al. (2004) have found that tetraploid crosses may be important in this process. We have found that triploid *Celtis* do produce occasional tetraploid progeny in the vicinity of sexual diploids, and analysis of the genetic makeup of these tetraploids will be important in understanding genetic processes in triploid *Celtis* populations. Note that even low levels of outcrossing are sufficient to maintain significant genetic diversity in apomictic populations (Hodac et al. 2019).

The fact that triploid *C. pumila* occurs beyond the southern limit of *C. occidentalis*, and also north of the northern limit of *C. laevigata*, indicates that regardless of its origin, it has now established stable populations independent of its diploid congeners. Triploid species are rare, but not unknown, and include widespread species such as *Rubus canadensis* Linnaeus (Craig 1960) and *Hieracium lepidulum* Stenstroem (Chapman et al. 2004). The conventional viewpoint is that these species, with predominantly apomictic reproduction, are an evolutionary dead-end (Stebbins Jr 1950, Arrigo and Barker 2012) and their persistence depends on recurrent formation from diploid parents. This does not seem to be the case with either of the aforementioned species, nor *Celtis pumila*. All three of these species maintain widespread and abundant populations, apparently independent of their diploid progenitors.

Conservation Implications

The most recent assessment of *C. pumila* in Canada (COSEWIC 2003), suggested that the tree-sized plants (up to 6m tall) growing in Eastern Ontario forests may be hybrids of *C. occidentalis* and *C. pumila*, and as such would not warrant legal protection. We have confirmed that these plants are in fact part of the same *C. pumila* cluster as in southwestern Ontario. In fact, most of the plants in eastern Ontario, which are subcanopy trees, share the same MLG as plants in southwestern Ontario, which includes a mix of dwarf forms (< 2 m tall) and taller trees.

Acknowledgments

Louise Riedel completed the chromosome count of the *Celtis sinensis* sample used to establish the genome size for our flow cytometry analysis. Victoria Nowell and Tim Garant assisted with field work, and Tracey James assisted with flow cytometry. Permission and assistance in sampling populations was provided by Elizabeth Shimp (Shawnee National Forest), Mike Moomey and Jill Fox (Trail of Tears State Park), Circle B Campground (Indiana), Glen Palmgreen and Alicia Ihnken (Pinckney Recreation Area), KOA Missouri, Christopher Crabtree (St. Joe State Park), Tony Flemin (Clarkco State Park), M. S. Whittington (Holly Springs National Forest), Malissa Briggler (Current River Conservation Area), and Dan VanExan (Lower Trent Fish and Game Club).

This project was funded by Point Pelee National Park, with assistance from Tammie Dobbie. This study was not preregistered with an institutional registry.

Author Contributions

AH designed and implemented the study, under the supervision of TWS. AH and TWS conducted the fieldwork. AH and SW completed the molecular labwork. AH and TWS analyzed the data. AW provided herbarium data. AH was the primary author of the manuscript, with contributions from TWS and AW.

Literature Cited

- Arrigo, N., and M. S. Barker. 2012. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* 15: 140–146.
- Asker, S. E., and L. Jerling. 2017. *Apomixis in Plants*. Boca Raton, Florida: CRC Press.
- Barker, W. T. 1997. *Celtis*. Pp. 376–379 in *Flora of North America North of Mexico*, vol. 3,. New York: Oxford University Press.
- Bicknell, R. A., and A. M. Koltunow. 2004. Understanding apomixis: recent advances and remaining conundrums. *The Plant Cell* 16: S228–S245.
- Bruvo, R., N. K. Michiels, T. G. D’Souza, and H. Schulenburg. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* 13: 2101–2106.
- Buck, G., and J. Bidlack. 1998. Identification of *Quercus* and *Celtis* Species using morphological and electrophoretic data. *Proceedings of the Oklahoma Academy of Science* 78: 23–33.
- Chapman, H., B. Robson, and M. L. Pearson. 2004. Population genetic structure of a colonising, triploid weed, *Hieracium lepidulum*. *Heredity* 92: 182–188.
- Clark, L. V., and M. Jasieniuk. 2011. polysat: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources* 11: 562–566.
- COSEWIC. 2003. *COSEWIC Assessment and Update Status Report on the Dwarf Hackberry Celtis Tenuifolia in Canada*. Ottawa, Ontario, Canada: Committee on the Status of Endangered Wildlife in Canada.
- Craig, D. L. 1960. Studies on the cytology and the breeding behaviour of *Rubus canadensis* L. *Canadian Journal of Genetics and Cytology* 2: 96–102.
- Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick. 1991. ‘Touchdown’ PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 19: 4008.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164: 1567–1587.
- Fernald, M. L. 1950. *Gray’s Manual of Botany*. New York: Van Nostrand Company.
- Galbraith, D. W., K. R. Harkins, J. M. Maddox, N. M. Ayres, D. P. Sharma, and E. Firoozabady. 1983. Rapid Flow Cytometric Analysis of the Cell Cycle in Intact Plant Tissues. *Science* 220: 1049–1051.
- Gleason, H. A., and A. Cronquist. 1991. *Manual of Vascular Plants of Northeastern United States and Adjacent Canada*. 2nd ed. Bronx, New York: New York Botanical Garden.
- Goudet, J., and T. Jombart. 2015. *Hierfstat: Estimation and Tests of Hierarchical F-Statistics*. R Package Version 0.04-22.
- Hodac, L., S. Klatt, D. Hojsgaard, T. F. Sharbel, and E. Hörandl. 2019. A little bit of sex prevents mutation accumulation even in apomictic polyploid plants. *BMC Evolutionary Biology* 19: 170.
- Hojsgaard, D., and E. Hörandl. 2019. The rise of apomixis in natural plant populations. *Frontiers in Plant Science* 10: 358.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
- Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

- Kimura, M., and T. Ota. 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proceedings of the National Academy of Sciences* 72: 2761–2764.
- Meirmans, P. G., and P. H. V. Tienderen. 2004. genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4: 792–794.
- Pfosser, M., E. Heberle-Bors, A. Amon, and T. Lelley. 1995. Evaluation of sensitivity of flow cytometry in detecting aneuploidy in wheat using disomic and ditelosomic wheat–rye addition lines. *Cytometry* 21: 387–393.
- Smith, T. W., P. Kron, and S. L. Martin. 2018. flowPloidy: An R package for genome size and ploidy assessment of flow cytometry data. *Applications in Plant Sciences* 6: e01164.
- Stebbins Jr, C. L. 1950. Variation and evolution in plants. *Variation and evolution in plants*.
- Taylor, P. E., G. Card, J. House, M. H. Dickinson, and R. C. Flagan. 2006. High-speed pollen release in the white mulberry tree, *Morus alba* L. *Sexual Plant Reproduction* 19: 19–24.
- Thiers, B. 2020. Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden. Website <http://sweetgum.nybg.org/science/ih/>.
- van Baarlen, P., P. J. van Dijk, R. F. Hoekstra, and J. H. de Jong. 2000. Meiotic recombination in sexual diploid and apomictic triploid dandelions (*Taraxacum officinale* L.). *Genome* 43: 827–835.
- van Dijk, P. J., and J. M. T. Bakx-Schotman. 2004. Formation of unreduced megasporangia (diplospory) in apomictic dandelions (*Taraxacum officinale*, s.l.) is controlled by a sex-specific dominant locus. *Genetics* 166: 483–492.
- Verdijnen, M. H., P. J. van Dijk, and J. M. M. Van Damme. 2004. The role of tetraploids in the sexual–asexual cycle in dandelions (*Taraxacum*). *Heredity* 93: 390–398.
- Wagner, W. H. 1974. Dwarf hackberry (Ulmaceae: *Celtis tenuifolia*) in the great lakes region. *The Michigan Botanist* 13: 73–99.
- Whittemore, A. T. 2008. Exotic species of *Celtis* (Cannabaceae) in the flora of North America. *Journal of the Botanical Research Institute of Texas* 2: 627–632.
- 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 *Steyermark's Flora of Missouri*, vol. 3, ed. G. Yatskievych. St. Louis, Missouri: Missouri Botanical Garden and the Conservation Commission of the State of 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.
- Whitton, J., C. J. Sears, E. J. Baack, and S. P. Otto. 2008. The dynamic nature of apomixis in the angiosperms. *International Journal of Plant Sciences* 169: 169–182.