# 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

by

Andrew Hayes

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements of the degree of

Master of Science

in

Biology

Carleton University
Ottawa, Ontario

# Abstract

*Celtis tenuifolia* is a threatened shrub species in southern Ontario which co-occurs with *Celtis occidentalis* and *Celtis laevigata* across the eastern United States. Although it is quite common further south, only six disjunct sites in Ontario currently exist, and these fragmented populations are susceptible to extinction due to habitat loss caused by quarrying operations, sand pit expansion, and beetle infestations. It is difficult for conservation managers to distinguish the three related *Celtis* species morphologically, and taxonomists have questioned the species limits in this group. Using microsatellite markers and flow cytometry, we assessed the relationship between genetic diversity, ploidy, and morphology in this group. We have determined that *C. tenuifolia* is an apomictic triploid with hybrid origins, and flow cytometry provides a reliable tool for distinguishing it from the sexual diploids *C. occidentalis* and *C. laevigata*. Furthermore, the same *C. tenuifolia* genotypes occur in a variety of ecological settings, demonstrating remarkable phenotypic plasticity. The results of these assessments will provide direction to the conservation management of *Celtis tenuifolia*, specifically providing field workers with reliable identification of the plant, as well as resolving concerns around hybridization and the risks of inbreeding and outbreeding depression.

# Acknowledgements

I would first like to thank my supervisor, Tyler Smith, for his support, guidance and patience throughout this project. In addition, I would like to thank my co-supervisor, Root Gorelick, and my committee members Joseph Bennett and Julian Starr for sharing their advice, perspective, and encouragement, Song Wang and Tracey James for their lab work and instruction, Victoria Nowell, Tim Garant, and Alan Whittemore for assistance in the field, and Justine McNaught for assisting in diagram generation. Funding was provided by Point Pelee National Park, with assistance from Tammie Dobbie.

I would also like to express my gratitude to the following individuals for permission and advice on sampling from their properties: Elizabeth Shimp (Shawnee National Forest), Mike Moomey and Jill A. Fox (Trail of Tears State Forest), the Park Manager at Circle B Campground Indiana, Glen Palmgreen and Alicia Ihnken (Pinckney Recreation Area), staff of KOA Campground Missouri, Christopher Crabtree (St. Joe State Park), Tony Flemin (Clarkco State Park), M.S.Whittington (Holly Springs National Forest), and Malissa Briggler (Current River Conservation Area).

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

## Plant Biodiversity and Conservation

In many ways, plants are the most important organisms on the planet. They provide the oxygen that animals need to keep their cells alive and are the base of the food chain for most animal life (including humans) on the planet. They can also influence climate, make and preserve soil, and provide habitat to many animal species. Plants are therefore essential in the sustainability of most ecosystems on the planet. From a human perspective, they are essential sources of medicine, food, and countless other useful products.

There are over 381,000 vascular plants species on Earth, and about 2000 new species are described each year. About 21 percent of all plants species are currently threatened with extinction (Willis, 2017). Some major impacts in recent years have been industrialization, over-exploitation, loss of pollinators, desertification, and climate change, mostly all caused, or exacerbated, by human activities (Willis, 2017).. In 2012, the International Union of Conservation of Nature estimated that the current plant extinction rate was 1000 to 10000 times higher than what it would be under natural conditions (Bapat, *et al*., 2012).

Global plant biodiversity is essential to keep those ecosystems functioning on which humans rely, which is why the threat of plant extinction is such an important issue. It has been shown that higher local and regional biodiversity enhances ecosystem services over time, especially during environmental change (Duffy, 2009). The combined results of hundreds of diversity-function experiments to date have shown that the loss of species

significantly reduces the ability of plant communities to produce new biomass by capturing biologically important resources (Cardinale, *et al*., 2007).

On a species level, genetic diversity is also extremely important because it allows populations to adapt to changing environmental conditions. The greater the variety of alleles present, the more likely a population will have individuals with the traits capable of surviving shifts in rainfall, temperature, soil conditions etc, to pass on their genes to the next generation (Toro and Caballero, 2005).

Endangered plant species that have experienced habitat loss and severe reductions in population sizes are even further at risk of losing genetic diversity through natural genetic processes such as inbreeding and genetic drift (Gauthier, *et al*., 2010). Inbreeding can reduce fertility and fitness by increasing homozygosity of recessive detrimental alleles, in a process known as inbreeding depression (Charlesworth, 2009).

Many factors influence the genetic diversity of species. Mutation, genetic drift, gene flow, and natural selection are four major forces that effect genetic differentiation. Mutation, genetic drift and natural selection generally produce genetic differentiation between populations, whereas gene flow between those populations will have the opposite effect, re-establishing genetic homogeneity (Slatkin, 1987). Other ecological factors affect the genetics of populations, such as size and spatial structure, methods of reproduction, and the dispersal of pollen and seeds. All of these factors in combination can produce a substantial amount of genetic variation within a species (Loveless and Hamrick, 1984).

## *Celtis* Taxonomy

The genus *Celtis* comprises a group of trees and shrubs with ca. 60 species globally and 6 species in North America (Sherman-Broyles, *et al.,* 1997). They are monecious, wind-pollinated, with bird-dispersed seeds. *Celtis* has traditionally been placed in the Ulmaceae (Elm Family), but recent molecular data indicates it more appropriately belongs in Cannabaceae (Whittemore, & Townsend, 2007). Opinions vary among taxonomists as to how many taxa occur in eastern North America. Fernald (1950) recognized three species, *C. occidentalis* Linnaeus (Common Hackberry), *C. tenuifolia* Nuttall (Dwarf Hackberry), and *C. laevigata* Willdenow (Sugarberry), which he further 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).

Whether, and how, to recognize this morphological variation taxonomically remains unresolved (Sherman-Broyles, *et al.* 1997). The common occurrence of plants morphologically intermediate to the three species has led many to suspect hybridization between *C. tenuifolia, C. occidentalis,* and *C. laevigata* is frequent (e.g., COSEWIC, 2003, Buck and Bidlack, 1998). Indeed, Fernald (1950) noted that the three species were often "seemingly confluent". However, in controlled pollination experiments, Whittemore *et al*. (2007) was unable to cross the species. Whittemore has since been able to produce allotetraploid plants by placing pollen from *C. occidentalis* or *C. laevigata* on receptive *C. tenuifolia* stigmas (Whittemore pers. comm.). More data is necessary to

determine how frequently such plants occur in nature, and if they are capable of establishing and producing seed of their own. Previous taxonomic work on this group has been limited to morphological data; molecular markers are needed to determine if there is in fact hybridization and gene flow among these putative species.

*Celtis tenuifolia*, as currently understood, is a shrub or small tree with serrate leaves sparingly toothed towards the apex, and small fruits ranging from orange to brown to cherry red. The range of *C. tenuifolia* consists primarily of the southeastern United States, from Missouri, Kentucky, and Pennsylvania south to Florida, and from Kansas and Oklahoma east to the Atlantic Coast (Sherman-Broyles, *et al.,* 1997). There are also smaller, disjunct populations in the Great Lakes regions of Ohio, Indiana, Michigan, and southern Ontario (Wagner, 1974). *Celtis occidentalis* is a large canopy tree with larger serrate leaves and fruits ranging from dark orange to purple to blue-black. It can be found in Manitoba, Ontario and Quebec and its U.S. range comprises Wyoming, Colorado and every state east of there, with the exception of Florida and Louisiana in the south and North Dakota in the north. *Celtis laevigata* has eliptic or lanceolate leaves and its range stretches from Indiana in the north, to Texas in the west, and covers the rest of the south eastern United States (Sherman-Broyles, *et al.,* 1997)

Assessing species boundaries requires a clearly defined species concept. One of the most commonly used is the Biological Species Concept (Mayr, 1942), which emphasizes reproductive isolation as a requirement for defining species. This is unsatisfactory for plants, where many accepted species maintain their genetic and morphological integrity despite occasional hybridization with other distinct species. Using the possibility of fertile hybrids as justification for merging otherwise discrete species would produce a taxonomy that under-represents the evolutionary diversity of

vascular plants.

Many alternatives to the biological species concept have been proposed to address issues like the one described above (reviewed in Stuessy 2009). De Queiroz (2007) developed a framework to reconcile these different approaches, the General Lineage Concept of species. De Queiroz defines species as "separately evolving metapopulation lineages". Metapopulation here refers to an inclusive group of connected subpopulations, and lineage indicates that they share an ancestor-descendent relationship. Crucially, De Queiroz' system accepts that the actual criteria used to define species will depend on their biological and evolutionary context. In practice, that means that a metapopulation lineage may be defined by genetic or morphological differentiation, reciprocal monophyly, reproductive isolation, ecological divergence, or some combination thereof. Some level of subjectivity remains, but the key issue is no longer deciding which criterion to use (i.e., reproductive isolation vs diagnosability), but rather deciding how much data is necessary to reach a conclusion. This flexibility means the same concept can be usefully applied to sexual and clonal species, and is not dependent on the availability of a single type of data.

## *Celtis tenuifolia* Conservation

Clarifying the taxonomic status of *C. tenuifolia* has practical implications, as it is of conservation concern in Canada. It was designated Special Concern in April of 1985, and was uplisted to Threatened[1] under the Species at Risk Act (SARA) in November

[1] Likely to become endangered if nothing is done to reverse the factors leading to their extirpation or extinction (COSEWIC, 2019)

2003. This is due to the fact that only six disjunct sites in Ontario currently exist (containing 1000 plants total), in addition to concerns about habitat loss caused by quarrying operations, sand pit expansion, and beetle infestations (COSEWIC, 2003). It is listed as secure[2] in the United States (N5) and globally secure (G5), but imperiled[3] in Ontario (S2) and Canada (N2) (NatureServe, 2019). The most serious threats (in order of decreasing priority) are altered natural disturbance regimes (coastal processes, fire, wind, and disease), bark beetles, snails, inappropriate logging activities, development (cottage, rural residential, agricultural and commercial), and aggregate extraction. Less serious threats include plant competition (alleopathic plants which produce biochemicals that affect its growth, reproduction and survival, as well as exotic or invasive plants) and recreational activity (off-road vehicles, trampling, horses, trail maintenance and firewood collection) (Parks Canada Agency, 2011).

Seedling survival is suspected to depend on forest canopy edges and openings or prairie and savanna habitats, as the plant is moderately shade intolerant (Parks Canada Agency, 2011). The populations in southern Ontario occur in three different growth environments: dry, sandy soil of the shoreline and inland sand dunes north of Lake Erie; gravel and till mounds left by glaciers on the ridge tops above the Trent River; and in limestone alvars on Pelee Island and in Hastings County (Parks Canada Agency, 2011).

*Celtis tenuifolia* has traditionally been thought to be a sexually reproducing species (Farrar, 1995, Waldron, 2003). However, recent data has suggested that it reproduces asexually, through pseudogamous apomixis (Whittemore, 2013, unpublished

[2] At very low risk of extinction or elimination due to a very extensive range, abundant populations or occurrences, and little to no concern from declines or threats (NatureServe, 2019)

[3] At high risk of extirpation in the jurisdiction due to restricted range, few populations or occurrences, steep declines, severe threats, or other factors (NatureServe, 2019)

data). Pseudogamous apomixis is the asexual formation of a seed without fertilization; pollen stimulates the egg cell to develop into an embryo, but fertilization does not take place, and no male genes are inherited (Craig, 1960; Bicknell and Koltunow, 2004; Chapman, *et al.,* 2004). Consequently, apomictic seed is genetically identical to the mother tree.

The populations in eastern Ontario (Sterling, Point Ann and Lonsdale) are quite isolated from the others, being 200 km from the nearest Ontario population near the Niagara River, and 450 km from the nearest population in Pennsylvania. Isolation from other populations may lead to increased random genetic drift and inbreeding, and reduced gene flow among populations, leading to an overall reduction in genetic diversity within those populations, and a higher genetic differentiation compared to other populations. This can in turn lead to poor recruitment and reduced fitness. (Young, *et al.,* 1996, Budd *et al*., 2015).

The *Celtis tenuifolia* populations of concern are located at the northern limits of its range, and are small and isolated in relation to the core populations further south in the United States. Because of their geographically marginal location, these sites contain plants that have potentially diverged genetically from the main range of the species through natural selection and genetic drift, and also potentially exhibit lower genetic diversity within their populations. This has been the long-standing view concerning peripheral populations (Lesica & Allendorf, 1995). However, Eckert, *et al.* (2008) showed that this is not always the case. In a review of 134 plant population genetics studies, they found that only 64.2% of the studies showed reduced genetic diversity in peripheral populations and 70.3% showed increased genetic differentiation towards the range margin. On top of this, the actual diversity between peripheral and core populations

were not significantly large (Eckert, *et al.,* 2008). This was verified by another recent study concerning the endangered species *Magnolia acuminata* (Linnaeus) Linnaeus which showed little reduction in genetic diversity from their fragmented peripheral populations in Canada to the core populations in the U.S. This was attributed to effective pollen dispersal over the long-lived history of the species (Budd, *et al.* 2015).

However, such peripheral populations have the potential to possess unique genetic characteristics and adaptations to local conditions (Lesica & Allendorf, 1995). This is especially important considering current global warming impacts potentially pushing the range of plant species northward. Considering the higher risk of extinction for small isolated populations, it is worthwhile to discover if these populations hold potential sources for adaptation to their northern environment.

Aldo Leopold, a famous forester and conservationist once said, “The first precaution to intelligent tinkering is to keep every cog and wheel.” (Leopold, 1949) This encapsulates the view that ecosystems are made up of species that are worth saving, even if it is unclear what their actual function is (Benson, 2012). It is unclear what particular function *Celtis tenuifolia* serves in its respective ecosystems, but that function could be irreplaceable, and will be lost if we allow this species to become extirpated.

## Polyploidy in plants

When an organism has more than two complete chromosome sets in its somatic cells, this is known as polyploidy (Campbell, *et al.*, 2009). Although not always the case, most animals are diploid, which means that there are two complete sets of chromosomes in every somatic cell (Griffiths*, et al.,* 2012). When there is a replication event, the

number of chromosome sets may increase to three (triploid), four (tetraploid), five (pentaploid), etc. (Ramsey and Schemske, 2002).

Polyploidy is common in the plant kingdom, and has been a key instrument in the evolution of many species. It is estimated that 70% of all angiosperms (flowering plants, including *Celtis* species) have gone through at least one episode of polyploidization (Soltis and Soltis, 1999). The traditional view of polyploid formation is that each polyploid species formed once, and thus contain only the genetic diversity provided by a single pair of parents, resulting in a genetically homogeneous species. Further studies from the 1990s, however, have shown that typically polyploid plant species form multiple times from different parental genotypes, resulting in a more diverse assortment of polyploid genotypes (Soltis and Soltis, 1999). A polyploid derived from only one species is referred to as 'autopolyploid' whereas a polyploid derived from hybridization (an organism breeding with an individual of a another species) is called an "allopolyploid" (Soltis, *et al.,* 2007).

Triploids present an unusual case of polyploidy in the plant kingdom. The historical narrative concerning triploid plants species is that they are a dead end, evolutionarily speaking, because they are unable to produce viable offspring (Rich, *et al*., 2008, Kohler, *et al.,* 2010). This is because the mating of a triploid to a lower ploidy causes incompatibility in the endosperm leading to nonviable progeny. This endosperm collapse due to the unbalanced ratio of paternal and maternal genomes is known as "triploid block" (Yamauchi, *et al*., 2004). There are also further complications for those progeny that overcome the triploid block, however, where plants undergoing meiosis will produce aneuploid gametes from improper chromosome pairings, forming sterile progeny (Kohler, *et al.,* 2010). The complications of "triploid block" can be avoided entirely by

species that undergo parthenogenesis (Yamauchi, *et al*., 2004). In the case of *C. tenuifolia* this could be why the species has evolved to reproduce through apomixis.

There are, however, many examples of triploid species in the wild, growing in mixed cytotype populations of diploid, triploid, and tetraploid. These species provide examples of what is known as a "triploid bridge", a process that actually influences the rate of tetraploid formations in a population (Husband, 2003).

Triploids are first produced through the union of a reduced diploid (n) and an unreduced diploid (2n) gamete. These unreduced diploid gametes can be created in a variety of ways (one example is illustrated in Figure 2), through mistakes in meiosis which occur after replication of sister chromatids (Brownfield and Kohler, 2011). It is not known which method is present in *Celtis tenuifolia.*

These triploids may then form unreduced triploid gametes (3n) which fuse with reduced diploid gametes (n), or form reduced triploid gametes (n) which fuse with other unreduced triploid gametes (3n) to form tetraploids (4n) (Husband, 2003). Some examples of species that grow in mixed populations of diploid, triploid, and teterploid are wandplant (*Galax urceolata* (Poiret) Brummitt; Barringer and Galloway, 2017), fireweed (*Chamerion angustifolium* (Linnaeus) Holub; Husband, 2004), orchard grass (*Dactylis glomerata (*Linnaeus) Linnaeus; Zohary and Nur, 1959), and an alpine species of buttercup (*Ranunculus kuepferi* Greuter & Burdet; Schinkel*, et al.,* 2017).

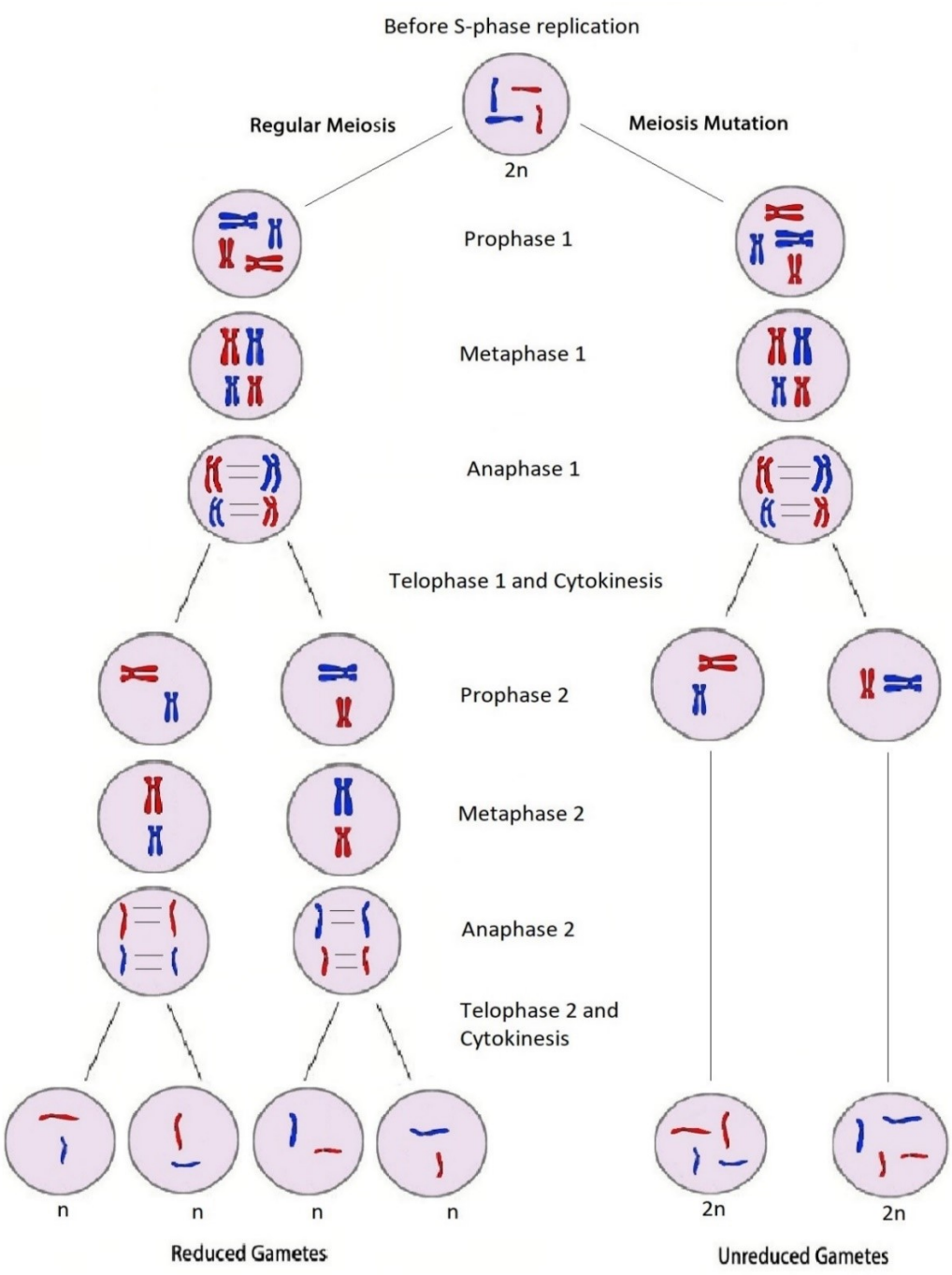


**Figure 1**: Example of regular gamete formation (left) vs unreduced gamete formation (right). During unreduced gamete formation, the second division in meiosis does not take place, leaving twice the normal number of chromosomes. Metaphase 2, Anaphase 2, Telophase 2 and Cytokinesis do not occur. After Prophase 2, sister chromatids separate but remain in the same cell.

There are also species which are predominantly triploid, without other cytotypes present in their populations. This is rare, but it does occur elsewhere in the plant kingdom. A few examples are smooth blackberry (*Rubus canadensis* Linnaeus; Craig, 1960), a species of hawkweed (*Hieracium lepidulum* Stenstroem; Chapman, *et al.,* 2004), and a naturally occurring subspecies of onion in India (*Allium cep* Linn. var. *viviparum* (Metz.) Alef.; Singh*, et al.* 1967). The lack of diploids and tetraploids in these species' populations indicates that they are not a step towards tetraploidy, via triploid bridge. The literature suggests that these species reproduce through apomixis. The traditional viewpoint is that this clonal reproduction will eventually lead to an evolutionary dead-end, (Stebbins, 1950, Arrigo and Barker, 2012)

## Microsatellites

The most widely used markers in population genetics to date are microsatellites, and they have applications in genome mapping, forensics, molecular ecology, population genetics, and conservation studies (Guichoux, *et al*., 2011). Also known as short tandem repeats (STRs), simple-sequence repeats (SSRs), or variable number tandem repeats (VNTRs), they are composed of repeating motifs one to six nucleotides long (Allendorf, *et al*., 2013). An example would be: ACTACTACTACT, which could also be written as $(ACT)^4$. Ideally, microsatellites should be 'perfect' which means that they are composed entirely of repeats of a single motif, with no interruptions or base pairs not belonging to the motif between repeats (Vieira, 2016).

The variation in the number of repeats in a microsatellite is believed to come from the slippage of the polymerase enzyme when replicating DNA, which produces a high

level of polymorphism among these markers. This in turn makes them extremely useful for assessing levels of genetic similarity among closely related species. As well, this technique is relatively inexpensive when compared with next generation sequencing (Hodel, *et al.,* 2016).

Early studies using SSRs have found that plant microsatellites specifically are highly polymorphic, somatically stable, and inherited in a co-dominant, Mendelian type fashion. (Morgante and Olivieri, 1993). This makes them ideal for plant studies which quantify diversity based on genetic distance and estimate the degree of relatedness between genotypes (Vieira, 2016).

It is important to note that microsatellites are presumed to be non-coding (Selkoe & Toonen, 2006; Allendorf *et al.* 2010; Hodel *et al*. 2016) and as such are not expected to reflect selection pressure. Rather, they are an indicator of gene flow and sexual relationships amongst different populations. Thus, we infer that there is greater gene flow between populations with similar microsatellite profiles than between those with divergent profiles. On the other hand, coding genes are subject to selection. Consequently, similarity between populations in their coding genes may reflect gene flow or convergent selection. This confounds population genetic analysis (Rowe *et al*. 2017).

## Flow Cytometry

Flow cytometry is a widespread method to estimate the nuclear DNA content of biological cells. Although first used mainly for biomedical research, the precision of this method has led to accurate estimates of thousands of genome sizes in plants (Dolezel, *et al*., 2007) This is accomplished by adding DNA-selective fluorochromic dyes to biological samples in liquid suspension and analyzing their fluorescence. These

fluorescent intensity values are compared to an internal or external standard of known genome size to get a clear estimate of the DNA content of the sample in question. (Dolezel, *et al*., 2007)

The plants involved in this study are historically categorized as two different cytotypes: diploid (two copies of the nuclear genome) or triploid (three copies). The verification of this can be easily seen in comparing the fluorescence values of the two samples, when considering that the diploid should show up at 2/3 of the value of the triploid. This method was used in this study to analyze the ploidy of the collected samples of *C. tenuifolia, C. occidentalis* and *C. laevigata*.

This allowed for evaluation of genetic relationships among cytotypes, and the clarification of Whittemore's preliminary work which indicates that *C. occidentalis* and *C. laevigata* are diploid, while *C. tenuifolia* is triploid. It also provided verification of whether or not samples with intermediate morphological forms are tetraploid, which would suggest a hybridization event between two separate species.

## Objectives

The main objective of this study is to assess genetic diversity within and among *C. tenuifolia* populations, genetic divergence among environments, and diversity levels in Ontario relative to US populations. I have also assessed the gene flow between *C. tenuifolia, C. occidentalis* and *C. laevigata,* as well as to developed markers to distinguish between the three species. The results of these assessments will help provide direction to the conservation management of *Celtis tenuifolia*, specifically providing field workers with reliable identification of the plant, as well as identify potential genetic issues that will need to be included in conservation planning.

In order to carry through with these objectives, a group of microsatellite markers needed to be developed for *Celtis tenuifolia*, the process of which is the subject of the following chapter. Samples of *C. tenuifolia, C. occidentalis,* and *C. laevigata* were collected from locations across the range of *C. tenuifolia* in the southeastern United States and were compared genetically to samples from southern Ontario, using flow cytometry and the analysis of the microsatellite data.

**Figure 2**: A typical *C. tenuifolia* plant

# Chapter 2 : Development of microsatellite primers

## Introduction

*Celtis tenuifolia* is a shrub or small tree that occurs in mid-eastern North America, with a core distribution from Missouri to Pennsylvania and south to Georgia and Alabama, and a few fragmented, disjunct populations in Southern Ontario, Michigan and Ohio. The Canadian populations have experienced significant losses from bark beetle infestations and its habitat is potentially under threat from quarrying operations and sand pit expansion in eastern Ontario, (COSEWIC, 2003), which is why it is listed as Threatened by the Species at Risk Act (SARA).

To gain an understanding of the genetic structure of *Celtis* species in eastern North America, a group of microsatellite markers were required. These allowed me to assess genetic diversity within and among *C. tenuifolia* populations, genetic divergence among environments, and diversity levels in Ontario relative to US populations. As there is some uncertainty about the taxonomic relationships between the three *Celtis* species, these markers also allowed me to assess gene flow between *C. tenuifolia, C. occidentalis* and *C. laevigata*. The results of these assessments will help provide direction to the conservation management of *Celtis tenuifolia*, specifically providing field workers with reliable identification of the plant, as well as identify potential genetic issues that will need to be included in conservation planning.

No molecular markers are currently available for this genus. Thus we have isolated and developed eight polymorphic microsatellite loci in order to analyze genetic diversity within and among populations.

## Methods and Results

Genomic DNA was extracted from silica-dried leaf tissue of 1 individual of *C. tenuifolia* collected from Trenton, Ontario in 2016 using a Nucleospin Plant II kit (Machery-Nagel, Bethlehem, Pennsylvania, U.S.A.) following the manufacturer's protocol. The sample was submitted to the Georgia Genomic Facility at the University of Georgia (Athens, Georgia, U.S.A) for isolation of microsatellite loci and primer development. They use the Illumina Next-Generation Sequencing platform to skim the genome and identify potential microsatellite loci based on the presence of repeating motifs (and flanking sequences) in individual reads. A brief description is as follows:

DNA was fragmented using the Bioruptor UCD-300 sonication device (Diagenode, Denville, New Jersey, U.S.A). Libraries compatible with Illumina TruSeqHT 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 quantified with Qubit (Life Technologies, Burlington, Ontario, Canada) and sequenced using an Illumina MiSeq vs 600 cycle kit (Illumina, San Diego, California, U.S.A.). MiSeq 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 on the first 1,000,000 sequences where both reads of any pair were ≥200 bases. Consensus sequences between 200 and 420 bp were exported from Geneious as FASTA files and imported into MSATCOMMANDER 1.0.8 beta (Faircloth, 2008). A total of 10,277 loci with perfect di-, tri-, tetra, penta, or hexa-nucleotide repeats were designed at default minimum lengths (i.e., 8 repeats for di- and tri-nucleotide motifs, 6 repeats for tetra-

nucleotide motifs) and combining loci with the greatest number of motif repeats were selected for further testing.

Of these 10,277 potential loci, a selection of 40 primer pairs with tri-, tetra-, penta-, or hexa-nucleotide repeats were chosen 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 motif repeats

We used these 40 primer pairs to amplify 6 *C. tenuifolia* and 2 *C. occidentalis* samples. 10 of the primer pairs did not show any product on agarose gels after electrophoresis, and were excluded from further testing. The 30 loci that were successfully amplified were screened for polymorphisms using a 3500xl Genetic Analyzer (Applied Biosystems) and narrowed down to 11 loci which were then used to analyze the populations in this study.

The screenings of amplification and polymorphism were done with 2 samples of *Celtis occidentalis* and 6 samples of *Celtis tenuifolia* collected from populations across the range of the two species: *C. occidentalis* : COHA1 (Point Pelee, Ontario); VARC1 (Virginia); *C. tenuifolia*: DWHA12-2 (Point Pelee, Ontario); ID-ID1 (Indiana); KYHR1 (Kentucky); PASH1 (Pennsylvania); VAWR2 (Virginia); and GC1 (Trenton, Ontario).

Polymerase Chain 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 dd$H_2O$ and 1.0 uL 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). 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 (-1C 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°C 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 samples used for screening successfully amplified at all 30 loci when tested through agarose gel electrophoresis. However, when analyzed by capillary electrophoresis for polymorphisms, only 17 showed consistent and significant differentiation between samples. Individual samples were genotyped and manually scored using Geneious 11.1.4 (Biomatters Ltd.).

11 of the most successful primers were pooled into four genotyping runs (Table 1), using FAM and VIC dyes to differentiate between fragment sizes that would potentially overlap, and these were subsequently evaluated to determine whether they were appropriate to use with the *Celtis* populations in this study.

## Conclusions

The primer pairs developed for this project successfully amplified 11 polymorphic microsatellite loci in *C. tenuifolia* and *C. occidentalis* across the range of the two species. In the next chapter, we demonstrate how we used these primers to assess genetic diversity across the *Celtis* species.

**Table 1**: Characterization of 11 polymorphic *Celtis* microsatellite loci

| Genoytyping group | Locus | Repeat Motif | Primer Name | Primer sequence (5' - 3') | Size range (bp) | $T_a$ (°C) | Fluorescent dye |
|---|---|---|---|---|---|---|---|
| Group 1 | CSSR13 | AATT(28) | test_05704 | F: AAAGTGACAAAGAGTTTTAAATGGG | 190 - 220 | 58.2 | FAM |
| | | | test_12045 | R: AAATCAACTCGAATAAAGAGGGC | | | |
| | CSSR20 | ACT(39) | test_06429 | F: TTGAGGCAAACCAAAATTGC | 232 - 290 | 59.5 | VIC |
| | | | test_16316 | R: GCCAGAGGACAACAATTTGC | | | |
| | CSSR22 | TTCGTC(30) | test_13121 | F: CAAACTCAGCAGCTTCTTCTCC | 360 - 385 | 60 | FAM |
| | | | test_00968 | R: CGCCTTTGAGGATCTTCTCC | | | |
| Group 2 | CSSR16 | TTG(33) | test_05972 | F: GGAGCCATAAATTTCGTGAAGG | 264 - 297 | 61.1 | FAM |
| | | | test_22620 | R: TTCCGCCATTGTCACTTCC | | | |
| | CSSR28 | AGGCTG(30) | test_16992 | F: GTGAGGCCAACCATGAGAGG | 190 - 232 | 60.5 | VIC |
| | | | test_14627 | R: GATCAGGTGGTGAATGTGGC | | | |
| | CSSR31 | ACTACC(30) | test_01845 | F: TGCCTTCTTTACAACTTTGAGTGC | 450 - 480 | 60.6 | FAM |
| | | | test_10401 | R: CAGAATCGCTTTACGACCCC | | | |
| Group 3 | CSSR30 | ACTCAG(24) | test_04500 | F: TTGCTTTCTCCACCATTCCC | 212 - 235 | 59.8 | FAM |
| | | | test_15226 | R: CAAAGAGAGAAATTACGGAGACCC | | | |
| | CSSR36 | ATACAC(30) | test_04288 | F: TTTTGAGGTGCCCTTTAATGC | 440 - 461 | 59.8 | FAM |
| | | | test_22991 | R: GATGCTGGTCACATGGTTCC | | | |
| Group 4 | CSSR24 | ATCTTC(30) | test_12573 | F: TTCGTATGATGAAATTTGGTTGC | 385 - 445 | 59.5 | FAM |
| | | | test_03386 | R: GTGGGAAAGACAACAATGCC | | | |
| | CSSR37 | TTGGGC(36) | test_12463 | F: TGGGTCCATCAGAAGTTACCG | 275 - 348 | 59 | FAM |
| | | | test_11261 | R: CCACATTTAGTCCCACATTGC | | | |
| | CSSR38 | ATAGGG(30) | test_22411 | F: GGTAAGAAGGCCAAGTTTCAAGC | 306 - 370 | 61.1 | VIC |
| | | | test_08658 | R: TCATCATCCTTGTCCCTACGC | | | |

Note: $T_a$ = theoretically optimal annealing temperature (touchdown PCR was carried out with a $T_a$ of 72° - 57°C for all loci)

# Chapter 3: The genetic structure of *Celtis tenuifolia* and comparisons to the related species *C. occidentalis,* and *C. laevigata:* Implications to the conservation management of threatened populations in Southern Ontario

## Introduction

*Celtis tenuifolia (*Dwarf Hackberry*)* is an endangered species of shrub growing in southern Ontario. It co-occurs with *Celtis occidentalis* (Common Hackberry), which, as its name implies, is quite abundant. There is taxonomic uncertainty regarding the two species, and there is suspected hybridization between them. In order to properly manage the remaining populations of *C. tenuifolia* in Southern Ontario, we must be able to accurately differentiate it from *C. occidentalis.* The correct identification is crucial in order to direct our efforts towards protecting the rare species in the genus.

Given the morphologically variable nature of the *Celtis* genus, molecular tools were needed for confident identification. The data I have collected provides a robust assessment of the genetic variation in *C. tenuifolia* across its range, and enables direct comparison with *C. occidentalis* and *C. laevigata* (Sugarberry*).* This has allowed for the assessment of the genetic distinctions among species, and graphical demonstration of the genetic distance in the three species through Principle Coordinate Analysis.

There is a certain amount of uncertainty concerning the taxonomic status of the *Celtis* genus. As mentioned previously, the classification has gone through numerous changes over the years, through Fernald (1950) to Gleason and Conquist (1991), as well as the Flora of North America's (Bouffard, 1997) treatment which considers five different historical designations as *Celtis tenuifolia*.

So far all of the taxonomy has been based on morphology as there have been no

molecular markers developed to help classify this species and differentiate it from related species of *C. occidentalis* and *C. laevigata.* The microsatellite markers developed for this project are the first of its kind for this genus.

There have also been recent suggestions by Whittemore (2013) that *C. tenuifolia* is triploid and reproduces apomictically. Flow cytometry was used in this study to clarify the relationships among cytotypes in this group..

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) Do these genetic groups (along with their respective cytotypes) have distinct leaf morphology? 3) Is there any evidence of hybridization or gene flow between these groups?

In terms of the populations specific to Ontario, the main questions are: 1) Which cytotypes do we have in Ontario? 2) Can we reliably identify them by genotype, cytotype and morphology? 3) How does the diversity of the Ontario populations of *C. tenuifolia* compare to the diversity of the species as a whole? Do they genetically diverge substantially from the core populations in the U.S., or lack diversity when compared?

These specific questions will need to be addressed because of their ramifications to conservation management.

## Methods

### 1) Field Sampling

In May of 2016 and May and June of 2017, we sampled 38 *Celtis* populations across the eastern United States (Table 2, Figure 12).

**Table 2:** Location of sampled populations of *Celtis* spp.

| Site code | State/Province, County, Site | Latitude | Longitude |
|---|---|---|---|
| | **CANADA** | | |
| ONPE | Ontario, Essex, Point Pelee National Park | 41.95224 | -82.515 |
| Lonsdale | Ontario, Hastings, Lonsdale | 44.25655 | -77.125 |
| GC | Ontario, Hastings, Lower Trent Valley Fish and Game Club | 44.22311 | -77.587 |
| PtA | Ontario, Hastings, Point Anne | 44.15776 | -77.3 |
| PPP | Ontario, Lambton, Pinery Provincial National Park | 43.25751 | -81.835 |
| | **UNITED STATES** | | |
| ALCC | Alabama, Butler, Cedar Creek Bridge - Creampot Rd | 31.95625 | -86.853 |
| ALHR | Alabama, Butler, Honeysuckle Rd | 31.72516 | -86.472 |
| ALAC | Alabama, Crenshaw, Arlena Church - Garnersville Rd | 31.86926 | -86.261 |
| ARCC | Arkansas, Izard, Campbell Cemetery | 35.93434 | -91.912 |
| ARWR | Arkansas, Izard, White River | 35.91122 | -91.928 |
| ARGR | Arkansas, Randolph, Galbraith Rd | 36.28403 | -91.187 |
| ILGG | Illinois, Hardin, Garden of the Gods Recreation Area | 37.59833 | -88.382 |
| ILLG | Illinois, Jackson, Little Grand Canyon | 37.68365 | -89.396 |
| ILWB | Illinois, Johnson, Wildcat Bluff | 37.37371 | -88.926 |
| ILHE | Illinois, Pope, Giibbons Creek - Herod | 37.58313 | -88.441 |
| INEM | Indiana, Knox, N Old 41 - Emison | 38.78619 | -87.469 |
| ID-ID | Indiana, Porter, Indiana Dunes State Park | 41.67371 | -87.019 |
| INCB | Indiana, Steuben, Circle B Campground | 41.63049 | -85.084 |
| KYHR | Kentucky, Madison, Hermit Ridge Rd, Madison | 37.82217 | -84.162 |
| Tws16-7 | Kentucky, Robertson, Blue Licks Battlefield State Resort Park | 38.45171 | -83.994 |
| MIWA | Michigan, Washtenaw, Waterloo-Pickney Trail | 42.39036 | -84.055 |
| MSCS | Mississippi, Clarke, Clarkco State Park | 32.10588 | -88.697 |
| MSSM | Mississippi, Kemper, Sciples Mill Rd | 32.87059 | -88.721 |
| MSWD | Mississippi, Marshall, Wall Doxey State Park | 34.66283 | -89.464 |
| MOMA | Missouri, Jefferson, Mastodon Park | 38.37909 | -90.394 |
| MOCC | Missouri, Shannon, Conservation Commision Rd 26 | 37.16991 | -91.126 |
| MOGR | Missouri, St. Louis, Green Rock Trail | 38.50395 | -90.702 |
| MOKC | Missouri, St. Louis, KOA Campground | 38.50022 | -90.69 |
| MOSJ | Missouri, St. Francois, St. Joe State Park | 37.81934 | -90.531 |
| OHCQ | Ohio, Erie, Castalia Quarry Metro Park | 41.38885 | -82.831 |
| OHES | Ohio, Erie, East Sandusky Metro Park | 41.42536 | -82.643 |
| PABE | Pennsylvania, Bedford, Lutzville Rd - Juniata River (Raystown Branch) | 40.01396 | -78.432 |
| PASH | Pennsylvania, Fulton, Sideling Hill Creek | 39.73273 | -78.348 |
| SCCC | South Carolina, Laurens, Cane Creek - Blakely Rd | 34.33523 | -82.029 |
| SCSF | South Carolina, Laurens, Sumter National Forest - Fsr 344 Rd | 34.54269 | -81.762 |
| TXFW | Texas, Tarrant, Fort Worth Nature Centre | 32.83785 | -97.477 |
| VAWR | Virginia, Claiborne, Wilderness Rd State Park | 36.63492 | -83.522 |
| VARC | Virginia, Giles, New River (Smith Branch) - Shumate Falls Rd | 37.38912 | -80.867 |

Whenever possible, we sampled 20 individuals from a location, and included the full range of morphological variability present. A minimum distance of 2 meters between plants was used to reduce the probability of sampling from the same clone. In

total, 381 samples were collected, with sample sizes ranging from 4-30 and averaging 15 samples per population (mean = 15.24).

For each sample, three leaves were removed and immediately sealed in plastic bags with 30 g of silica gel beads to dry. 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).

**2) 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 minutes 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.).

Eleven nuclear polymorphic microsatellite markers (Chapter 2) were assayed for each sample. These markers were assumed to be nuclear from the outset, and were confirmed as such by the presence of heterozygotes in at least one sample for all 11 loci. Chloroplast loci cannot be heterozygous, as the chloroplast genome is haploid. Polymerase Chain 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 dd$H_2O$ and 1.0 uL DNA, using a T-

100 Thermal Cycler (Bio-Rad, Hercules, California, U.S.A.). The thermal cycling profile of touchdown PCR (TD-PCR, see Chapter 2) 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 four groups of three 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.

SSR primer pairs that had high failure rates, yielded allele numbers inconsistent with ploidy estimates (e.g., three or more alleles for diploid samples), or produced irregular chromatogams were excluded. Additionally, samples that failed for three or more SSR loci were dropped from the analysis. A total of 332 samples were genotyped and cytotyped at 8 loci.

**3) Flow Cytometry**

To assess the ploidy of our samples, flow cytometry was performed on all of the samples that we managed to obtain SSR data for. For each sample, a 0.7 cm x 0.7 cm piece of leaf tissue was chopped into fine powder in 750 ul Tris*$MgCl_2$ buffer (Pfosser, *et al.,* 1995) and let stain on ice in complete darkness for 30 min with 50 uL RNAse and

250 ul Proprodium Iodide mixed with Galbraith buffer (Galbraith, *et al.,* 1983). Samples were then analyzed with a Gallios Flow Cytomter (Beckman Coulter, Brea, California, U.S.A.) and compared to petunia and soy standards. The flow cytometry results were exported into R ver. 3.5.0 (R Core Team, 2018) and analyzed with the 'flowploidy' package (Smith *et al.* 2018).

**4) Data analysis**

STRUCTURE is software which uses Bayesian clustering to identify genetically homogenous groups as well as individuals with mixed ancestry (Evanno, *et al.,* 2005). 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, 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 $\Delta K$ (Evanno, *et al*., 2005) to select the most informative number of groups.

We classified individuals with an inferred proportion of ancestry greater than 70% for one of the clusters as members of that population. 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 then examined the geographical distribution of the groups delineated in the PCoA and STRUCTURE analysis. This was done in R with a list of GPS coordinates gathered for each population in the field collection process.

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, 2001). Bruvo's distance incorporates genotypic distances between individuals of different ploidies . It also evaluates genetic distance using a stepwise mutation model, as appropriate for microsatellites (Kimura and Ota, 1975).

**5) Genetic diversity and Population Structure**

The R packages 'polysat', 'adegenet' (Jombart, 2008), 'poppr' (Kamvar, *et al.,* 2014) and 'hierfstat' (Goudet, 2005),were used to calculate genetic diversity and population structure statistics for the populations sampled. Observed heterozygosity ($H_e$), expected heterozygosity ($H_o$), and the inbreeding coefficient ($F_{is}$) were calculated only for the diploids, as we assume the triploid populations are at least partly apomictic and so violate the assumptions underlying F statistics. The inbreeding coefficient ($F_{is}$), is the probability that two alleles in an inbred individual are identical by descent (Hartl and Clark, 2007). $F_{is}$ values are population summaries that don't apply to groups of individuals that are not outcrossing or at Hardy Weinberg equilibrium. Since *C. tenuifolia*

is essentially cloning itself, it is not appropriate to apply an inbreeding co-efficient to something that doesn't breed at all.

The basic population parameters of *C. occidentalis* and *C. laevigata* were calculated, including number of private alleles ($A_p$), inbreeding coefficient ($F_{is}$), the level of differentiation among populations ($F_{st}$), and the total departure from HW proportions within and among populations ($F_{it}$). We used a distance threshold to identify triploid clones, following the approach of Meirmans and van Tienderen (2004). PCR artifacts and somatic mutations may introduce small differences between individuals from the same clonal lineage. Visual inspection of the triploid inter-individual distances showed a bimodal distribution (Figure 2). We assumed the lower mode (< 0.1) was formed by clones that differed only by such artefacts, and set the threshold accordingly. We calculated pairwise $F_{st}$ values for all genetic groups as a measure of relative differentiation.

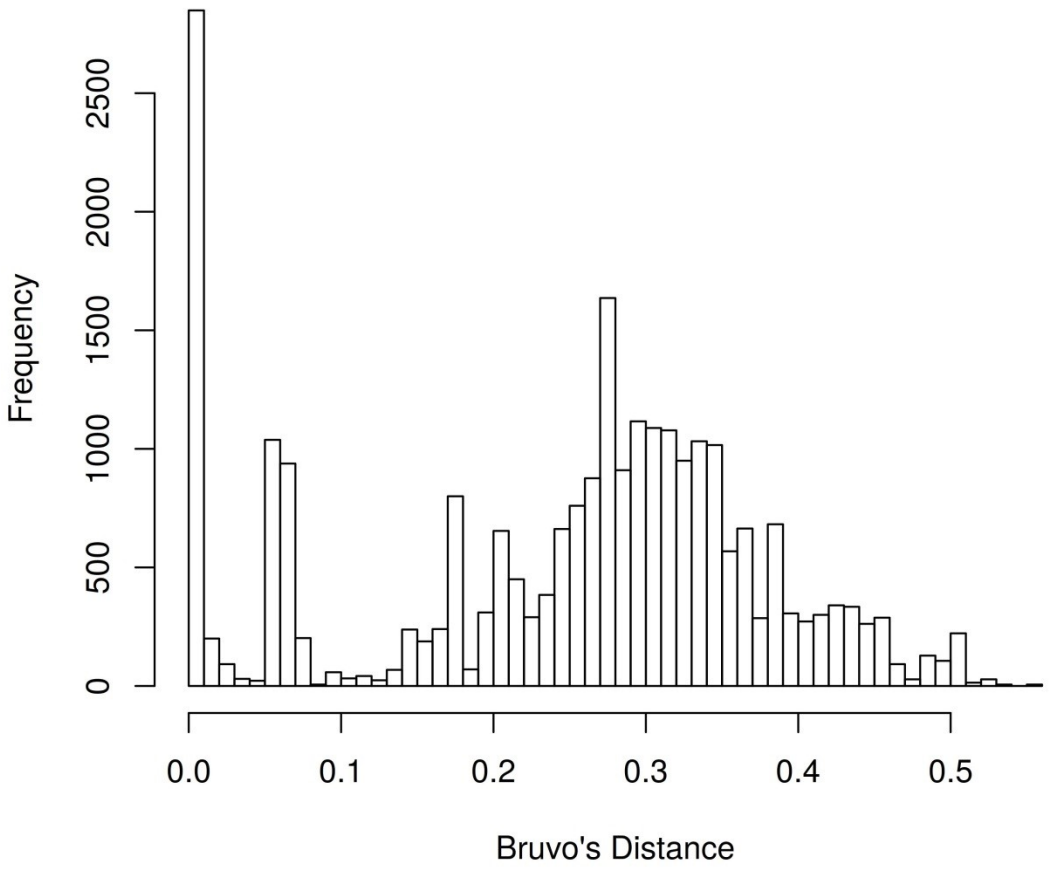


**Figure 3**: Bimodal distribution of triploid inter-individuals distance of genotypes, with a lower threshold of 0.1

**6) Morphological Differentiation**

A herbarium voucher was collected for each sample in our populations. We selected a representative leaf from each voucher for morphological analysis. We measured six variables: leaf length and width, and the number of teeth in each quadrant of the leaf. *Celtis* have oblique leaf bases, with one side rounded to the petiole, while the other is acute. The tooth density is often markedly different on the round and acute sides of the leaf, as well as in the lower and upper half. To account for this variability, we

counted the number of teeth for the lower and upper half of both sides of the leaf as discrete variables. We also calculated five ratios: the number of teeth per cm in each quadrant, and the length/width ratio of the leaf.

We calculated pair-wise correlations between all 11 variables. We limited our analysis to variables that had absolute correlation less than 0.7; this left us with three: leaf length, leaf length/width ratio, and tooth density in the upper rounded quadrant of the leaf. A permutational MANOVA (Multivariate Analysis of Variance) was done using the R package 'vegan' (Oksanen, *et al*., 2018) to analyze the morphological difference between the diploid populations based on these three non-correlated independent variables.

# Results

## 1) Flow Cytometry

Our samples included 171 diploids (Figure 3), 159 triploids (Figure 4), and 2 tetraploids (Figure 5).

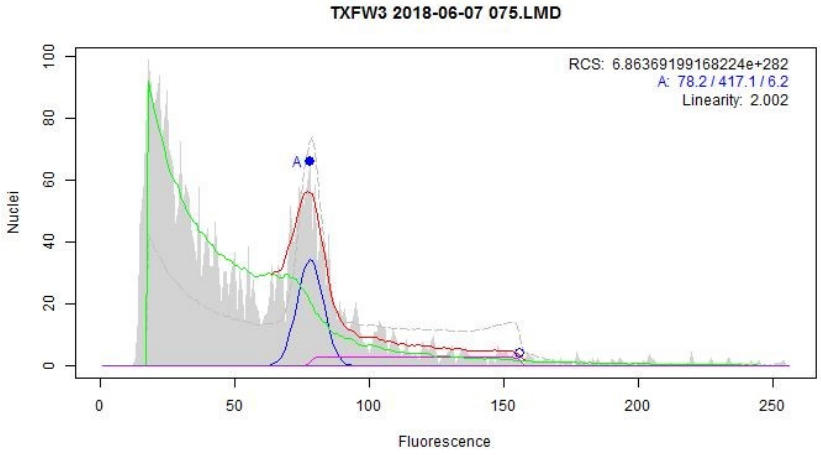


**Figure 4**: A diploid sample from the TXFW population

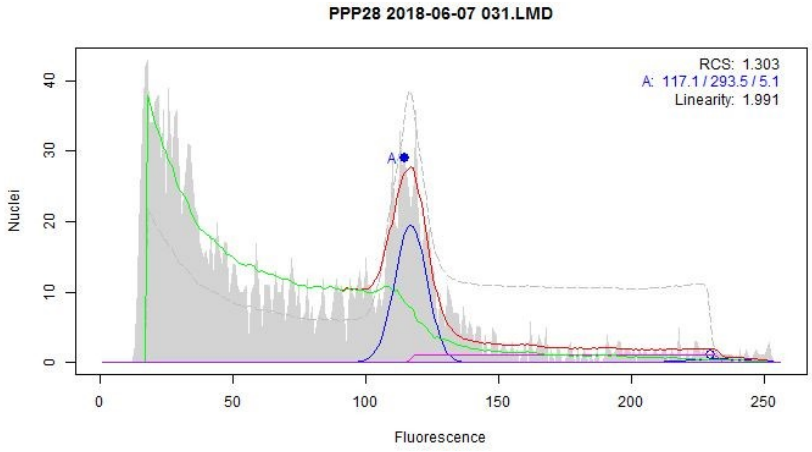


**Figure 5**: A triploid sample from the PPP population

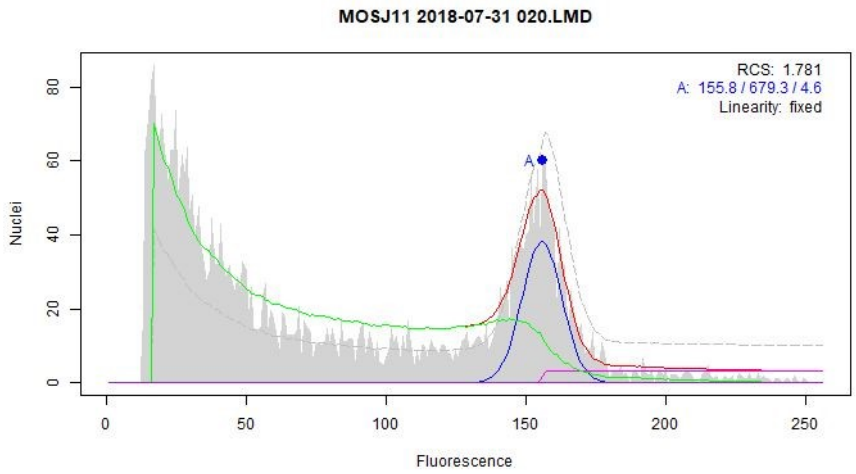


**Figure 6**: A tetraploid sample from the MOSJ population

**2) The Diploids: *C. occidentalis, C. laevigata*, and their hybrids**

The STRUCTURE analysis of the diploid samples indicated they formed two distinct clusters; the delta K value peaked at 2 groups (Figure 6). Based on leaf morphology and geographic distribution, these clusters correspond to the species *C. occidentalis* and *C. laevigata.*

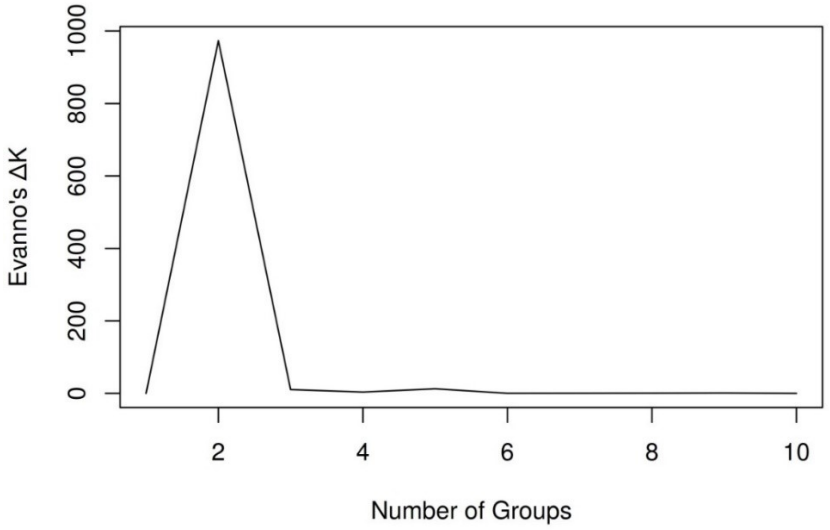


**Figure 7**: Evanno K's distance showing a clear maximum of 2 genetic groups of diploids

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 (Figure 7).

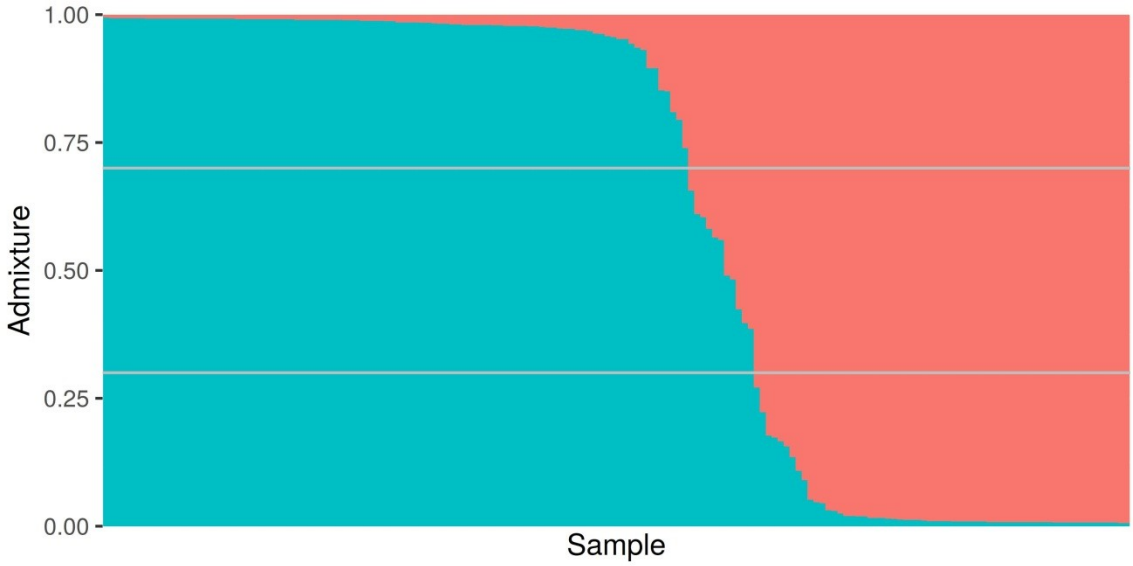


**Figure 8:** STRUCTURE analysis showing two diploid groups (*C. laevigata* in blue and *C. occidentalis* in pink) and potential inter-species hybrids

Principal Coordinates analysis (PCoA) of the diploids revealed the intermediate genetic position of the putative hybrids between the two diploid species (Figure 8). All of the putative hybrids were collected in Missouri and Illinois, where both diploid species co-occur. There is no evidence of a third diploid plant group corresponding to *C. tenuifolia*.

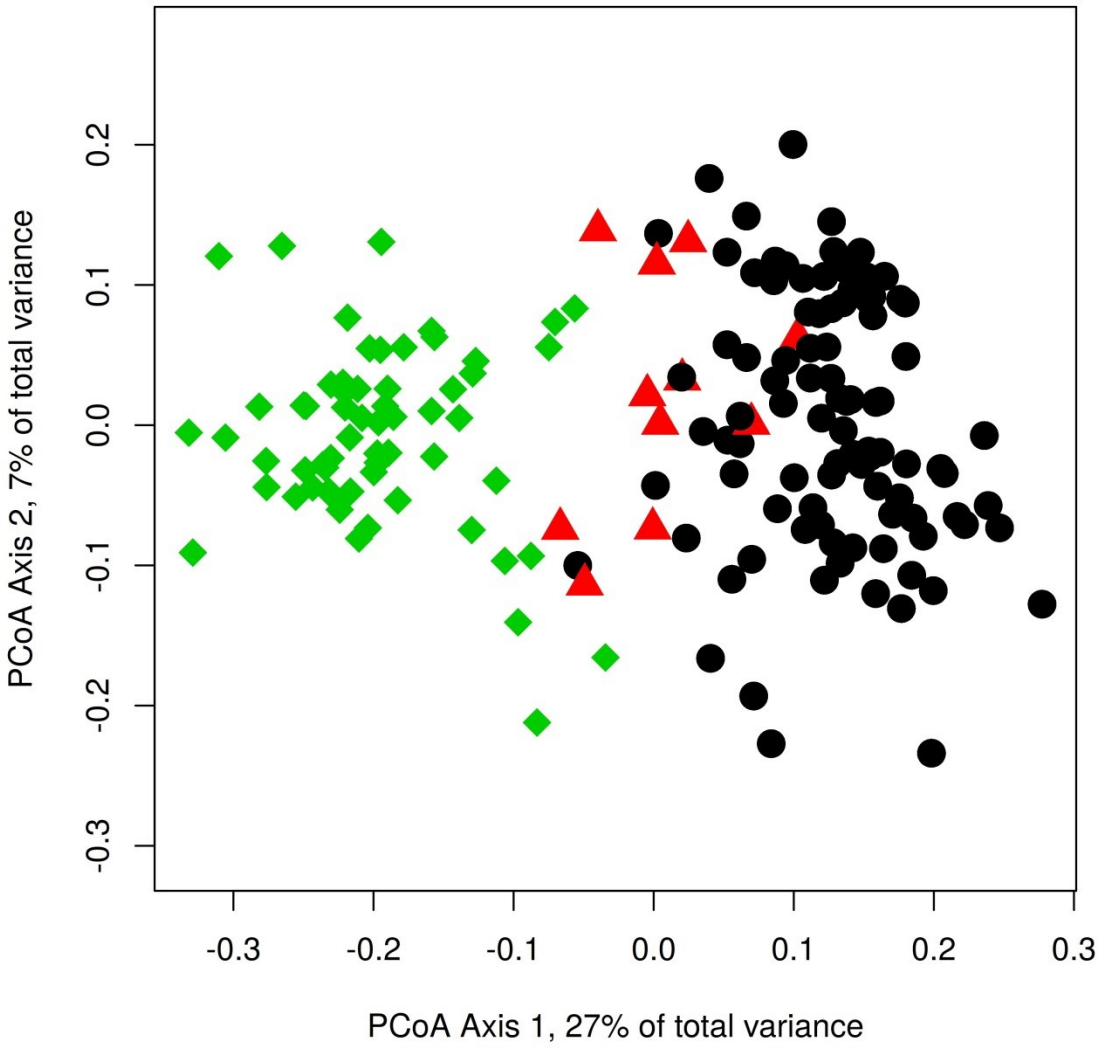


**Figure 9**: Principle Coordinate Analysis showing genetic distance between *C. laevigata* (black circles), *C. occidentalis* (green diamonds), putative diploid hybrids (red triangles)

### 3) The Triploids: *C. tenuifolia*

To clarify the ordination of the triploids, we used different symbols for six geographic groups: GL (Great Lakes: Ohio and Michigan), ON (Ontario), PA (Pennsylvania), MW (midwest, including MO, AR, ID, KY, TX), and SE (southeast, AL and SC). TX was included in the "Midwest" in order to reduce the number of symbols and simplify the plot. After applying these symbols, two groups are evident: a northern group, including the ON, GL and PA populations, and a southern group, including MW and SE (Figure 9). The two populations in Illinois formed two smaller separate satellite clusters genetically closer to the southern than the northern populations.

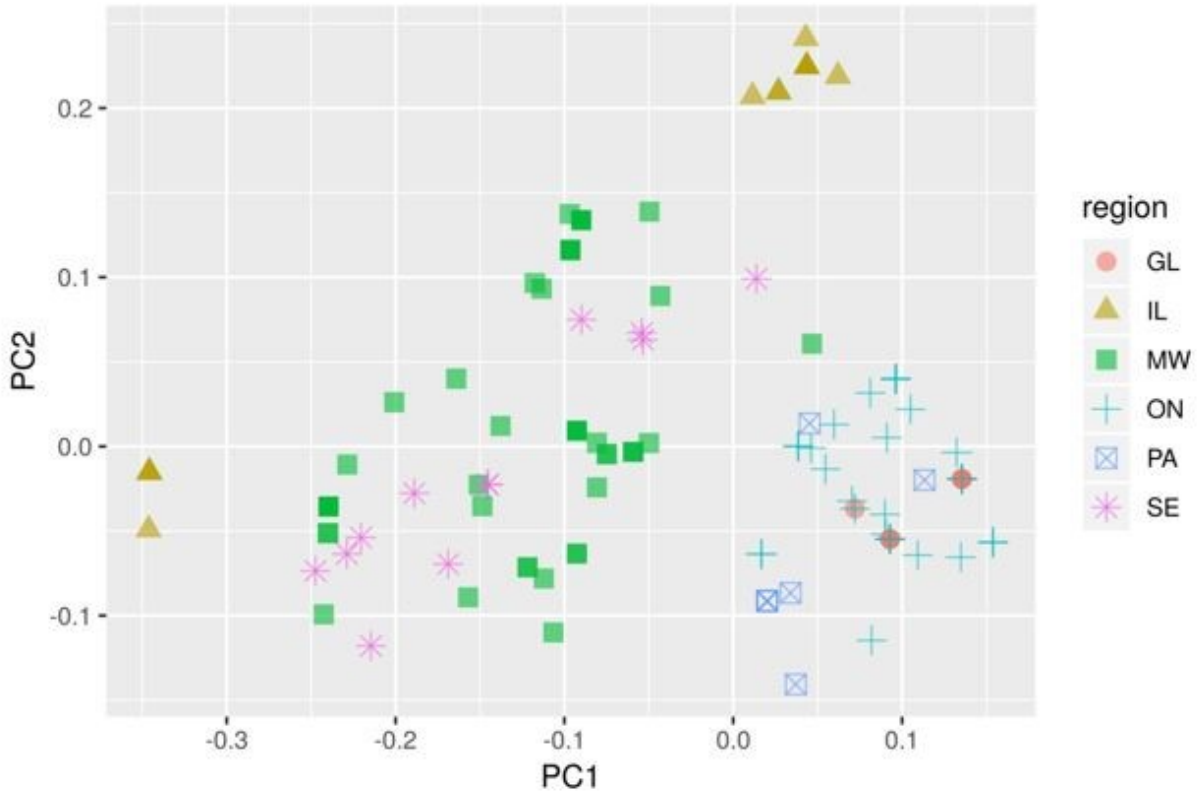


**Figure 10**: Principle Coordinate Axis showing genetic distance between the different *C. tenuifolia* populations: GL (Great Lakes: Ohio and Michigan), ON (Ontario), PA (Pennsylvania); and a southern group, including MW (Midwest: Missouri, Indiana, Arkansas, Kentucky, Texas), SE (Southeast: Alabama and South Carolina), and IL (Illinois). *Note: TX was included in the Midwest in order to reduce the number of symbols and simplify the plot.

The two main clusters can be seen as separated into a north and south geographical region, as shown in Figure 10.

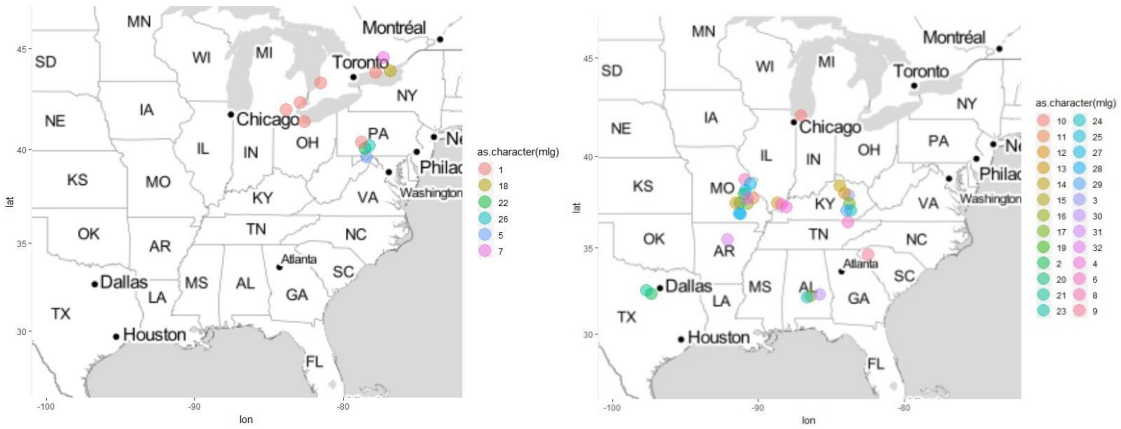


**Figure 11**: The northern (left) and the southern (right) cluster of *C. tenuifolia,* with each colour representing a separate genotype, as shown in the legends.

When the triploids were added to the diploids in a PcoA, the triploids show up genetically intermediate between the two diploid species, and overlapping somewhat with the diploid hybrids. There were also two tetraploids found, genetically intermediate to the two diploids species (Figure 11).

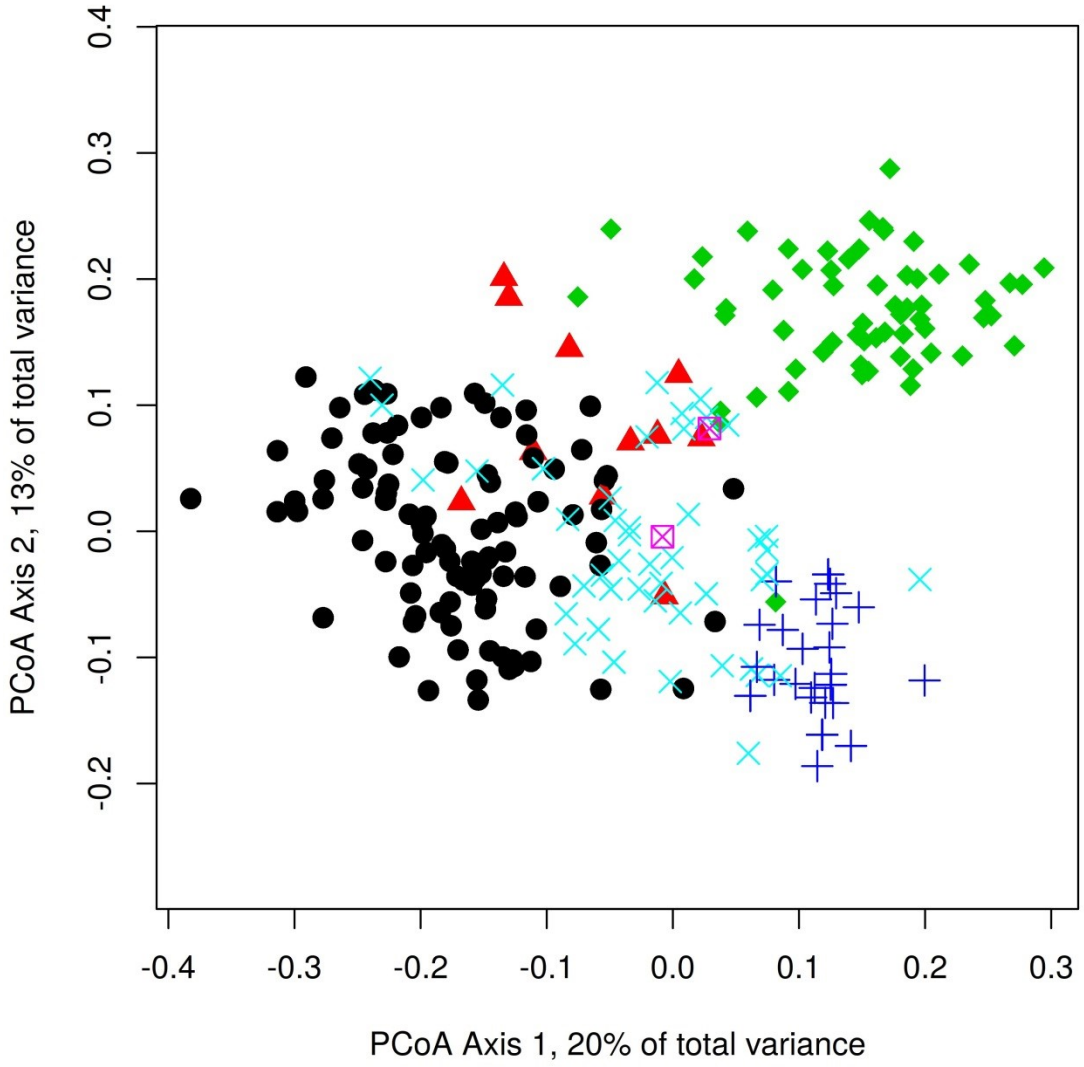


**Figure 12:** Principle Coordinate Analysis showing genetic distance between *C. laevigata* (black circles), *C. occidentalis* (green diamonds), putative diploid hybrids (red triangles), northern *C. tenuifolia* (dark blue crosses), southern *C. tenuifolia* (light blue X), and tetraploids (mauve squares).

**4) Genetic diversity**

In the 332 individuals analyzed, a total of 62 alleles were observed for 8 microsatellite loci, which is an average of 7.75 alleles per locus. (Table 3). The number of alleles per locus ranged from four (CSSR30, CSSR36) to fifteen (CSSR20).

**Table 3**: Allele sizes for 8 microsatellite loci

| Locus | Alleles size (bp) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CSSR16 | 267 | 270 | 273 | 275 | 278 | 281 | 284 | 287 | 290 | 293 | 296 | | | | |
| CSSR20 | 232 | 236 | 240 | 244 | 248 | 252 | 256 | 260 | 268 | 272 | 276 | 280 | 284 | 288 | 292 |
| CSSR22 | 364 | 366 | 369 | 373 | 376 | 379 | 381 | | | | | | | | |
| CSSR28 | 194 | 200 | 206 | 212 | 219 | 224 | 231 | | | | | | | | |
| CSSR30 | 214 | 220 | 226 | 232 | | | | | | | | | | | |
| CSSR31 | 453 | 458 | 465 | 470 | 477 | | | | | | | | | | |
| CSSR36 | 440 | 449 | 454 | 460 | | | | | | | | | | | |
| CSSR38 | 310 | 316 | 322 | 328 | 334 | 340 | 346 | 352 | 358 | | | | | | |

When the diploid populations (*C. occidentalis* and *C. laevigata)* were compared across all 8 loci, *C. occidentalis* had seven private alleles, and *C. laevigata* had seventeen (Table 4). The triploid samples had 3 alleles not present in any diploid, but all three were rare, present in less than 5% of the triploid samples.

**Table 4**: Basic population genetic parameters for diploid *Celtis* spp.

| Species | $A_p$ | $F_{IS}$ | $F_{ST}$ | $F_{IT}$ |
|---|---|---|---|---|
| *Celtis occidentalis* | 7 | -0.02656 | 0.062504 | 0.03760313 |
| *Celtis laevigata* | 17 | 0.10334 | 0.079664 | 0.1747712 |

We found a total of 32 *C. tenuifolia* multi-locus genotypes (MLG). 28 of the MLGs were found in a single population. The remaining four MLG were found in 2-6 populations each. The same MLG was shared by all individuals at GC (6 individuals; Lower Trent Valley Fish and Game Club, Ontario), ONPE (19 individuals; Point Pelee National Park, Ontario), PPP (32 individuals; Pinery Provincial Park, Ontario), MIWA (8 individuals; Waterloo-Pickney Trail, Michigan), OHCQ (6 individuals; Castalia Quarry, Ohio), and by one of two individuals sampled at PABE (Bedford County, Pennsylvania). These all correspond to the northern cluster of *C. tenuifolia.*

Another MLG is shared between 9 individuals in ILGG (Garden of the Gods, Illinois) and 1 individual in ILHE (Gibbons Creek, Herod, Illinois). The third is shared between 4 individuals in ILWB (Wildcat Bluff, Illinois) and one individual in MOCC (Conservation Committee Rd 16, Missouri). The fourth is shared between two populations in Missouri: 4 individuals in MOSJ (St. Joe Park, St. Francois County) and two individuals in MOGR (Green Rock Trail, St. Louis County).

The northern cluster, which covers 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. 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.

**Table 5:** Basic population genetic parameters for 332 individuals in 38 populations of *Celtis* spp.

| Population | Latitude | Longitude | Species | N | MLG | $H_o$ | $H_e$ | Nei's Gene Diversity | $F_{IS}$ - $(H_e - H_o)/H_e$ |
|---|---|---|---|---|---|---|---|---|---|
| ALAC | 31.86926 | -86.2607 | laevigata | 3 | 3 | 0.54167 | 0.48611 | 0.583 | -0.114285808 |
| | | | tenuifolia | 5 | 2 | | | 0.08 | |
| ALCC | 31.95625 | -86.85286 | laevigata | 8 | 8 | 0.44196 | 0.54526 | 0.583 | 0.189444609 |
| ALHR | 31.72516 | -86.47214 | laevigata | 6 | 6 | 0.42708 | 0.46484 | 0.512 | 0.081232664 |
| | | | tenuifolia | 2 | 1 | | | | |
| ARCC | 35.934342 | -91.91157 | laevigata | 8 | 8 | 0.56101 | 0.48845 | 0.523 | -0.148551198 |
| | | | tenuifolia | 1 | 1 | | | | |
| ARGR | 36.28403 | -91.18691 | laevigata | 6 | 6 | 0.47917 | 0.49306 | 0.538 | 0.028169034 |
| ARWR | 35.91122 | -91.928 | laevigata | 7 | 7 | 0.51786 | 0.4898 | 0.528 | -0.057291619 |
| GC | 44.22311 | -77.58672 | tenuifolia | 6 | 1 | | | | |
| ID-ID | 41.673707 | -87.01915 | tenuifolia | 5 | 1 | | | | |
| ILGG | 37.598331 | -88.38196 | tenuifolia | 9 | 1 | | | 0.04074 | |
| ILHE | 37.583126 | -88.44149 | occidentalis | 1 | 1 | | 0.4375 | 0.714 | |
| | | | tenuifolia | 1 | 1 | | | | |
| | | | hybrid | 2 | | | | | |
| ILLG | 37.68365 | -89.39614 | laevigata | 8 | 8 | 0.51116 | 0.56559 | 0.606 | 0.096229598 |
| | | | hybrid | 1 | | | | | |
| ILWB | 37.373705 | -88.92571 | laevigata | 3 | 3 | 0.625 | 0.49306 | 0.592 | -0.26760552 |
| | | | tenuifolia | 4 | 1 | | | 0.01111 | |
| | | | hybrid | 3 | | | | | |
| INCB | 41.63049 | -85.08357 | occidentalis | 4 | 4 | 0.4375 | 0.42188 | 0.482 | -0.037037037 |
| INEM | 38.78619 | -87.46921 | occidentalis | 5 | 5 | 0.5 | 0.5 | 0.556 | 0 |
| | | | laevigata | 4 | 4 | 0.5 | 0.52734 | 0.603 | 0.051851942 |
| | | | hybrid | 1 | | | | | |
| KYHR | 37.82217 | -84.16229 | tenuifolia | 9 | 3 | | | 0.2321 | |
| Lonsdale | 44.25655 | -77.12543 | occidentalis | 2 | 1 | 0.375 | 0.1875 | 0.25 | -1 |
| | | | tenuifolia | 3 | 1 | | | | |
| MIWA | 42.39036 | -84.05453 | tenuifolia | 8 | 1 | | | | |
| MOCC | 37.16991 | -91.12621 | occidentalis | 2 | 2 | 0.625 | 0.40625 | 0.542 | -0.538461538 |
| | | | tenuifolia | 6 | 1 | | | 0.33037 | NA |
| MOGR | 38.50395 | -90.70229 | occidentalis | 2 | 2 | 0.75 | 0.46875 | 0.667 | -0.6 |
| | | | tenuifolia | 6 | 2 | | | 0.06963 | NA |
| MOKC | 38.500219 | -90.6895 | occidentalis | 8 | 8 | 0.5558 | 0.50096 | 0.535 | -0.109484534 |
| MOMA | 38.37909 | -90.3941 | occidentalis | 1 | 1 | 0.5 | 0.25 | 0.5 | -1 |
| | | | laevigata | 5 | 5 | 0.475 | 0.4125 | 0.458 | -0.151515152 |
| | | | hybrid | 4 | | | | | |
| MOSJ | 37.81934 | -90.53063 | tenuifolia | 8 | 4 | | | 0.24524 | |
| | | | tetraploid | 1 | 1 | | | | |
| MSCS | 32.105878 | -88.697 | laevigata | 12 | 12 | 0.35417 | 0.40495 | 0.423 | 0.125401811 |
| MSSM | 32.87059 | -88.7205 | laevigata | 8 | 8 | 0.34375 | 0.39941 | 0.426 | 0.139364384 |
| MSWD | 34.662829 | -89.46396 | laevigata | 6 | 6 | 0.43333 | 0.46042 | 0.504 | 0.05882367 |
| OHCQ | 41.38885 | -82.83069 | occidentalis | 2 | 2 | 0.4375 | 0.46875 | 0.625 | 0.066666667 |
| | | | tenuifolia | 6 | 1 | | | | |
| OHES | 41.42536 | -82.64314 | occidentalis | 8 | 8 | 0.43304 | 0.42263 | 0.451 | -0.024615724 |
| ONPE | | | occidentalis | 26 | 25 | 0.45845 | 0.4247 | 0.433 | -0.079460568 |
| | | | tenuifolia | 19 | 1 | | | 0.01252 | |
| PABE | 40.01396 | -78.432 | tenuifolia | 2 | 2 | | | 0.13333 | |
| PASH | 39.732734 | -78.3477 | tenuifolia | 9 | 2 | | | 0.07407 | |
| PPP | 41.956749 | -82.5169 | tenuifolia | 32 | 1 | | | 0.00342 | |
| ptA | 44.157764 | -77.29982 | tenuifolia | 1 | 1 | | | | |
| SCCC | 34.335226 | -82.02889 | laevigata | 2 | 2 | 0.4375 | 0.29688 | 0.396 | -0.473684211 |
| | | | tenuifolia | 5 | 1 | | | 0.00889 | |
| | | | tetraploid | 1 | 1 | | | | |
| SCSF | 34.542691 | -81.76176 | laevigata | 10 | 10 | 0.31667 | 0.29672 | 0.313 | -0.067221464 |
| TXFW | 32.837848 | -97.47665 | laevigata | 1 | 1 | 0.625 | 0.3125 | 0.625 | -1 |
| | | | tenuifolia | 3 | 2 | | | 0.18519 | |
| Tws16-7 | 37.38912 | -80.86716 | tenuifolia | 1 | 1 | | | | |
| VARC | 37.38912 | -80.86716 | occidentalis | 2 | 2 | 0.5 | 0.35938 | 0.479 | -0.391304348 |
| VAWR | 36.63492 | -83.52171 | tenuifolia | 8 | 3 | | | 0.15397 | |

Note: N = number of individuals sampled, MLG = number of multi-locus genotypes,

$H_o$ = observed heterozygosity , $H_e$ = expected heterozygosity,

F = inbreeding coefficient

Fst values were calculated between the various species (Table 4). The differentiation between *C. tenuifolia* and each of the diploids is less than the differentiation between the diploids. They also show very little differentiation from the hybrid diploids (Table 6).

**Table 6:** Estimates of pairwise FST among *Celtis* spp.

| Species | *Celtis occidentalis* | *Celtis laevigata* | Mix | North *C. tenuifolia* | South C. tenuifolia |
|---|---|---|---|---|---|
| *Celtis occidentalis* | | 0.169 | 0.432 | 0.135 | 0.103 |
| *Celtis laevigata* | | | 0.026 | 0.138 | 0.044 |
| Mix | | | | 0.040 | 0.024 |
| North *C. tenuifolia* | | | | | 0.079 |
| South *C. tenuifolia* | | | | | |

The three genetic clusters were identified as three separate species, with *C. occidentalis* and *C. laevigata* occupying separate ranges, and *C. tenuifolia* spreading across the range of both of these species. The individuals that were deemed admixtures inhabited the area where the two species overlap (Figure 12).

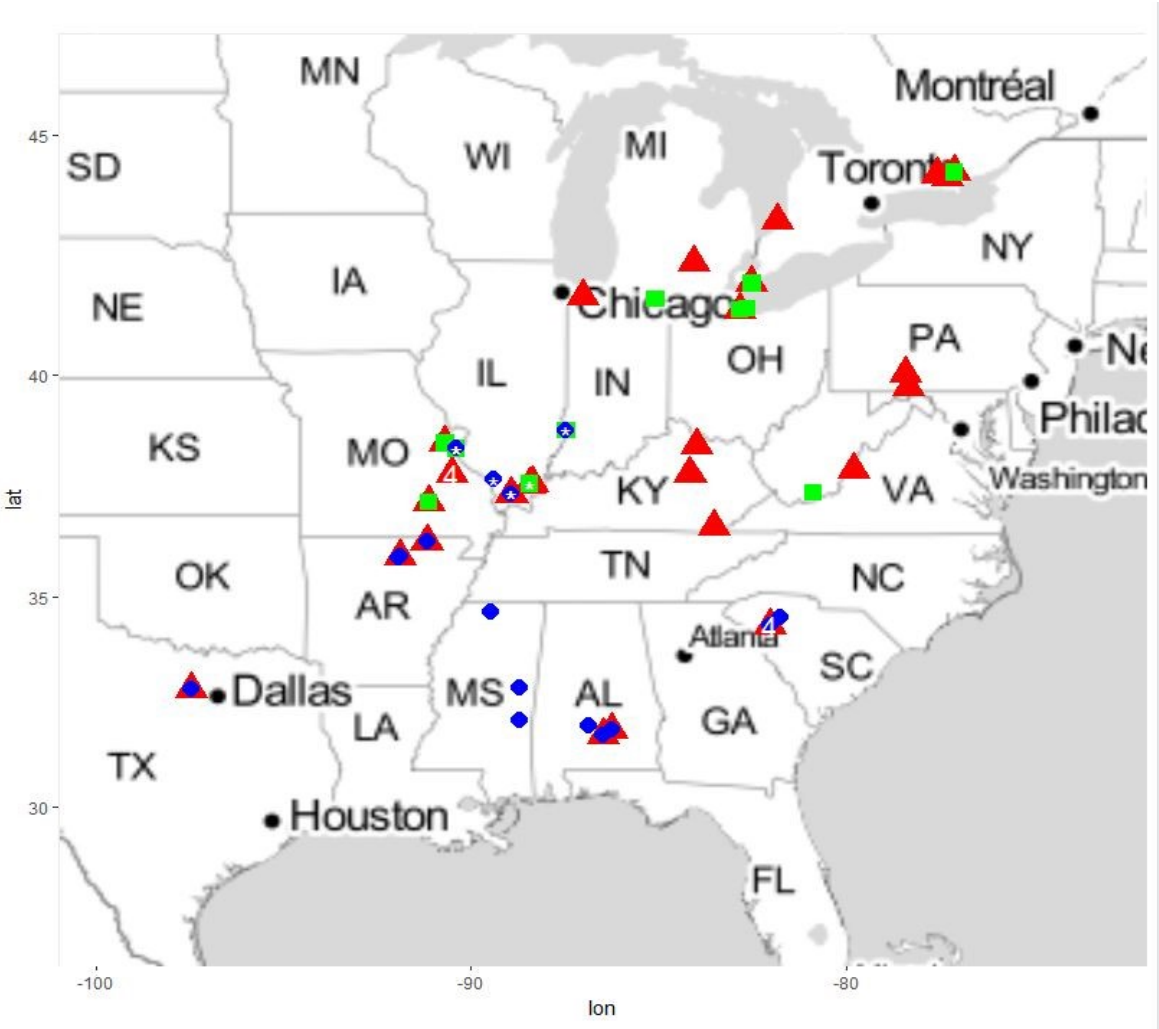


**Figure 13:** Geographical locations of all *Celtis* spp. Red triangles represent *C. tenuifolia.* Green squares represent *C. occidentalis*. Blue circles represent *C. laevigata.* Putative diploid hybrids are marked by asterisks, and the number 4 represents the two tetraploid individuals**.**

**5) Morphological Differentiation**

The MANOVA test results comparing *C. occidentalis* and C. *laevigata* based on leaf length, length/width ratio, and teeth density in the upper right quadrant of the leaf revealed a statistically significant difference between the species ($R^2 = 0.3234$, p-value = 0.0001, 9999 permutations).

When the three variables were plotted against the different species, the suspected diploid hybrids were morphologically intermediate between the two diploid species. Compared to the diploid species, *C. tenuifolia* has shorter leaves and a smaller length/width ratio overall, with intermediate teeth density (Figure 13).

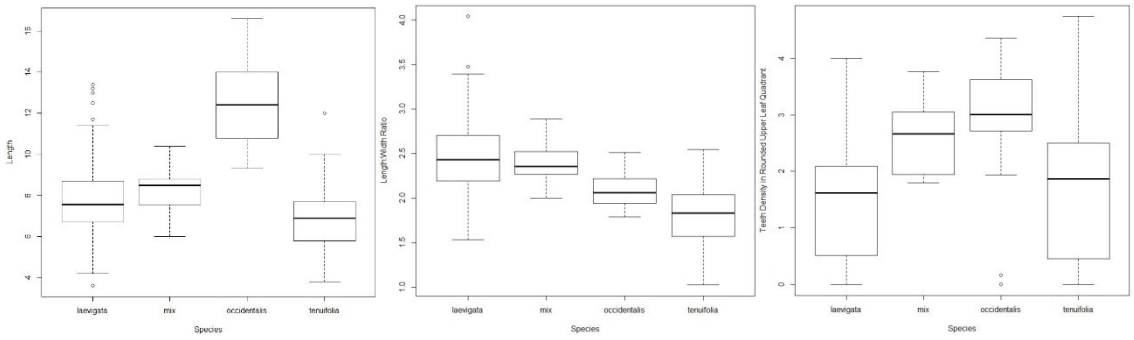


**Figure 14:** Comparisons of leaf morphologies among different *Celtis* spp. Panels (from left to right): Length, Length/Width Ratio, Teeth Density in Upper Right Quadrant. Species (from left to right in each panel): *C. laevigata*, mix (diploid hybrids), *C. occidentalis, C. tenuifolia*

## Discussion

### Three Species of *Celtis*

In the beginning of this study, we wanted to know how many genetic groups of *Celtis* exist in eastern North America, and if they related directly to their cytotypes (distinct diploid and triploid groups). Additionally, we asked if these genetic groups (along with their respective cytotypes) have distinct leaf morphology. Finally, we wanted to know if there was any evidence of hybridization or gene flow between these groups.

Through our STRUCTURE, PCoA analysis, and flow cytometry we determined that there are two distinct diploid groups (Figures 7 and 8) and that they correspond to *Celtis occidentalis* and *Celtis laevigata*. There was no diploid group that corresponded to *C. tenuifolia.* Instead, we discovered a triploid group through flow cytometry that corresponded to *Celtis tenuifolia.* This group appeared approximately genetically equidistant between the two diploid species' clusters. In total, we found three genetic groups of *Celtis* in eastern North America. Through Multivariate Analysis of Variance, we showed that these genetic groups, along with their respective cytotypes, have distinct leaf morphology. These correspond to our previous understanding of the traditional delineation of these species. (Figure 13)

Under the general lineage species concept, *Celtis occidentalis* and *Celtis laevigata* form two separate species. Despite occasional hybridization, the SSR data shows that they each form a single inter-connected genetic metapopulation, distinct from the other. *Celtis tenuifolia* can also be defined as a species under this concept. Although this triploid has evidently formed multiple times, the data is consistent with the same two diploid ancestors being involved in each case, resulting in a coherent metapopulation

lineage united by genetic and morphological characteristics.

Our analysis found strong evidence of hybridization between *Celtis occidentalis* and *Celtis laevigata*. Our STRUCTURE analysis (Figure 7) produced 11 individuals with less than 70% inferred proportion of ancestry from either species, which was the first clue that they were inter-species hybrids. Their mixed ancestry can also be clearly seen in the PCoA analysis (Figure 8), which shows their intermediate genetic position between the two species.

These admixed individuals were also restricted to areas surrounding the border of Missouri and Illinois where populations of *C. laevigata* and *C. occidentalis* were both present. This geographic distribution helps to corroborate the hybrid theory, as these potential hybrids only exist where the ranges of *C. laevigata* and *C. occidentalis* overlap.

When we compared the leaf morphology of the admixed individuals, they showed an intermediate set of traits compared to their assumed diploid parents. This can be seen by boxplots comparing length, length/width ratio, and teeth density in the upper right quadrant of the leaf (Figure 13). This is the third line of evidence that supports the theory that these individuals are in fact hybrids between *C. occidentalis* and *C. laevigata.*

In 1998, Buck and Bidlack performed an analysis of *Celtis reticulata, C. occidentalis* and *C. laevigata* to see if these species could be identified by morphological and molecular data. They used leaf and twig characteristics along with isozyme data to group the different species together. The presence of all three species in more than one cluster suggested the likelihood of hybridization among all three *Celtis* species, and of more importance to this study, between *C. occidentalis* and *C. laevigata* (Buck and Bidlack, 1998). More recently, a controlled cross-pollination experiment failed to produce hybrids of these two species (Whittemore and Townsend, 2007). Our data,

however, confirms the presence of *C. occidentalis* / *C. laevigata* hybrids in the wild. This suggests that hybridization between these two species is rare, or depends on conditions not replicated in cultivation.

### *Celtis* Species in Ontario

In Ontario specifically, there are two different species and cytotypes: *C. tenuifolia* (triploid) and *C. occidentalis* (diploid). These correspond to two distinct morphological groups according to the Multivariate Analysis of Variance tests of samples collected in the U.S. This results in the ability to reliably identify them by leaf morphology alone in some cases, and also by the fact that the majority of the *C. occidentalis* specimens observed during the sampling expedition were noticeably larger than the *C. tenuifolia*, appearing as canopy trees (25 meters and higher) as opposed to shrubs/small understory trees (1 meter to 8 meters). Compared to *C. occidentalis, C. tenuifolia* also has smaller leaves with a lower length/width ratio and less teeth around the edges (Figure 13). However, these two *Celtis* species look very similar in early stages of growth, and vegetative specimens under 8 meters tall remain difficult to distinguish. In these cases, identification must be done by determination of ploidy level through flow cytometry.

On a molecular level, identification of these two *Celtis* species through cytotype proved to be an accurate method. Having already established that there is no diploid group that corresponds to *Celtis tenuifolia*, or triploid group that corresponds to *Celtis occidentalis,* flow cytometry provides a straightforward method to identify the two species. When compared through PCoA analysis, there were overlaps in genotype, however flow cytometry unequivocally separated *C. occidentalis* and *C. tenuifolia* into diploid and triploid groups, respectively.

The diversity of the Ontario populations of *Celtis tenuifolia* is lower than the US populations. Genetic assessment showed that there are fewer multi-locus genotypes per population in Ontario (only one in each), as well as fewer multi-locus genotypes across the province, than in the United States. Missouri and Illinois had the highest number of multiple locus genotypes. To put things in perspective, these states have less geographical area than the area of Ontario occupied by *C. tenuifolia*, but have approximately 10 times the multi-locus diversity.

### *Celtis tenuifolia*: A triploid hybrid between *C. laevigata* and *C. occidentalis*

Predominantly triploid species such as smooth blackberry (*Rubus canadensis* Linnaeus, Craig, 1960) and hawkweed (*Hieracium lepidulum* Stenstroem, Chapman, *et al.,* 2004) are somewhat of an anomaly in the plant kingdom. Triploids reproduce through apomixis (See Chapter 1). The traditional viewpoint is that this will eventually lead to an evolutionary dead-end (Stebbins, 1950, Arrigo and Barker, 2012), or that the persistence of triploids depends on recurrent formation from diploid parents. This does not seem to be the case with either of the aforementioned species, nor *Celtis tenuifolia*. All three of these species maintain widespread and abundant populations, apparently independent of their diploid progenitors.

Our data suggests that *Celtis tenuifolia* is a triploid hybrid between *Celtis occidentalis* and *Celtis laevigata*. In the PCoA, the triploids show up intermediate between the two diploid species and overlapping somewhat with the diploid hybrids (Figure 11). This is consistent with the expectations for a triploid hybrid, which would contain a mixture of the alleles from both parents. This pattern is also reflected in the pairwise Fst values calculated between species (Table 6). The largest distance is between

*C. occidentalis* and *C. laevigata*, with each of these diploid species closer to *C. tenuifolia* than they are to each other.

Our documentation of genetically mixed diploid individuals demonstrates that there is occasional hybridization between *C. occidentalis* and *C. laevigata*. The geographical range of *C. tenuifolia* overlaps with both *C. occidentalis* and *C. laevigata*. In light of the evidence, it is highly likely that *C. tenuifolia* was formed in the past as a triploid hybrid of *C. occidentalis* and *C. laevigata*. A hybrid offspring is always closer genetically to both parents than each parent to each other, and our Fst values clearly show this pattern.

If *C. tenuifolia* was an autopolyploid derived from one of the diploids, we would expect low pairwise Fst values between the diploid parent and the triploid, and similar high pairwise Fst values between the two diploids and between the triploid and the non-parental diploid.

On the other hand, if *C. tenuifolia* was derived from a third (undetected) diploid, then we would expect higher pairwise Fst values between the triploid and both diploids than between the two diploids. In other words, the triploid would be more differentiated from both diploids than the diploids are from each other.

Instead, *C. tenuifolia* has a lower pairwise distance to *C. occidentalis* and *C. laevigata* than the two diploid species to each other, consistent with expectations for an allopolyploid of the two different species.

While our data indicates that *C. tenuifolia* formed from a combination of hybridization and polyploidization, we cannot distinguish between two alternative pathways. The triploids may have been formed in a single step, via the union of an unreduced gamete from one parent and a reduced gamete from the other. The resulting F1

offspring of this union would be a triploid that combines alleles from both parents. Alternatively, the triploids may have formed in two steps: the first being hybridization between the two diploid parents, producing an F1 diploid hybrid; followed by the union of reduced and unreduced gametes in the hybrids to generate the triploids in the F2 generation.

More detailed genomic data will be necessary to resolve this issue. However, the pairwise Fst values suggest that both pathways may have occurred. Triploids produced via the 1-step pathway should be genetically closer to one parent than the other, whereas the 2-step pathway would generate triploids that are equidistant to the diploid parents. Our data for the southern *C. tenuifolia* cluster show that it is more closely related to *C. laevigata* (Fst = 0.044) than it is to *C. occidentalis* (Fst = 0.103), consistent with expectations for the 1-step pathway. On the other hand, the northern *C. tenuifolia* cluster is approximately equidistant to both diploids (Fst = 0.138 and 0.135), as expected for the two-step pathway.

The limited diversity of most *C. tenuifolia* populations suggests they reproduce primarily through apomixis, as expected for triploids. It follows that each unique multi-locus triploid genotype represents a separate origin of *C. tenuifolia.* The Missouri hotspot of *C. tenuifolia* is full of different genotypes. This is indicates multiple origins of triploidy (most likely still forming) from *C. laevigata* and *C. occidentalis*, where the two species overlap. This explains the diversity all across the southern range. In the north there is a lack of diversity, because the means for creating more *C. tenuifolia* plants is lost. The *C. laevigata* genes that still exist in *C. tenuifolia* up north had to have jumped through multiple States through bird-dispersal of seeds. This is a fairly common biogeographical pattern, where there is a lot of diversity south of a glacial boundary,

around where a species will originate, and founder effects will reduce genetic diversity as only a small subset of the population is able to move north (Hewitt, 2000).

Glacial refugia tend to harbour a significant amount of the genetic diversity of a species (Petit, *et al.,* 2003). Genetic diversity tends to also be higher around the origin of a particular species, and the leading edge of migration of lower diversity, comparatively (Cruzan and Templeton, 2000). If we are to assume that the aforementioned Missouri/Illinois hotbed of diversity is the origin of *C. tenuifolia*, then according to geological data, this origin is at the base of the Laurentide Ice sheet; a retreating glacier that was covering the Northern United States and southern Canada approximately 15,000 years ago (Dyke and Prest, 1987), Slowly, over long stretches of time, populations of *C. tenuifolia* were able to move north, although the odds were against migration, as it moved further away from possible reinforcement migration from the south. There would be a filtering effect of multi-locus genotypes, as there is such a low, random chance that even one particular genotype could start a founder population. Additionally, there were likely large areas where habitat may have been unsuitable, as evidenced by the big gaps in the distribution (Figure 12). The end result here is only three *C. tenuifolia* genotypes in Ontario compared to the 25 genotypes in the southern populations; a large reduction in diversity.

**Conservation of *Celtis tenuifolia***

The most recent COSEWIC report on *Celtis tenuifolia* (2003), suggested that the tree-sized plant growing in Eastern Ontario forests may be hybrids of *C. occidentalis* and *C. tenuifolia.* We can now safely say that this is not the case, and they are in fact *C. tenuifolia*, the same species as the spindly, threadbare shrubs that can be found in the

sand dunes in Point Pelee National Park and Pinery Provincial Park. Moreover, most of the plants in Ontario share the same genotype; their morphological variation is attributable to phenotypic plasticity, rather than hybridization.

Breeding among individuals from the same population can reduce the fitness of the plants, by increasing the homozygosity of harmful recessive alleles, in a process known as inbreeding depression (Allendorf, *et al.,* 2013). Similarly, cross-breeding with other populations can reduce the fitness of the plants by introducing alleles from populations adapted to a different environment, in a process known as outbreeding depression (Allendorf, *et al.,* 2013). In a scenario in which there are threatened species in small, fragmented populations, conservation managers are typically concerned about inbreeding and outbreeding depression. However, in this scenario there is no cause for concern, since there is no breeding taking place at all. The primary concern is preserving the genotypes that are left, and keeping population numbers at a sustainable level.

Because apomictic species typically have low genetic diversity as a result of their asexual reproduction, they tend to be thought of as a lower priority compared to sexually reproducing species (Rich, *et al*., 2008). However, they still contribute to the overall local biodiversity and function like other species. *Celtis tenuifolia* has established populations across eastern North America, and well beyond the relatively small area where both of its parents co-occur. Even in areas where one or both of the diploid species co-occur with *C. tenuifolia*, they occupy different (though overlapping) habitats, and presumably serve different ecological roles. Additionally, in a case like *Celtis tenuifolia*, diversity will still exist in areas where these polyploid apomicts can be recreated from their sexually reproducing, diploid parents (Rich, *et al*., 2008).

Although *Celtis tenfuifolia* can be seen growing alongside *C. occidentalis* in the same forest environment, it can also be found growing in places where *C. occidentalis* is not, such as disturbed roadsides, sand dunes, and alvars , where it is likely filling an ecological role distinct from *C. occidentalis*. For example, the loss of *C. tenuifoli*a from alvars could lead to erosion, or lack of food and shelter for animals. We don't have specific data about the ecologies of the different *Celtis* species, but we do have enough morphological and geographical observations to show that the ecological roles of each *Celtis* species is potentially unique. The fact that *Celtis tenuifolia* has had a long, stable existence, a widespread distribution beyond the ranges of each parent, as well as a distinct habitat and form, are all arguments in favour of its conservation.

# Chapter 4: Conclusions and Implications

This study resolves long-standing confusion regarding the diversity of *Celtis* in eastern North America. The morphological variability of this group led taxonomists to question the number of species and how to delineate them. Conservation programs in Ontario were unable to reliably distinguish between the Threatened *C. tenuifolia,* and the common *C. occidentalis*. Suspicions of hybridization between these two species made it impossible for biologists to determine with confidence which plants they ought to be protecting.

As a result of our work, we now know that *C. tenuifolia* is a widespread, apomictic triploid, and it can be reliably distinguished from *C. occidentalis,* a sexual diploid, by flow cytometry.

There is very little, if any, gene flow occurring among and within its populations. For the most part, this is a strictly clonal species. The same MLG is shared by all of the individuals in many populations, and in Ontario, even between most of the separate populations.

The microsatellite data also shows that *C. tenuifolia* growing in different environments, and with dramatically different morphologies have, in some cases, identical multi-locus genotypes. This demonstrates that phenotypic plasticity is common in this species. Furthermore, if conservation managers wish to bolster declining populations, or replant *C. tenuifolia* into an area where it has been extirpated, the risks associated with inbreeding and outbreeding depression, serious concerns for sexual species, are not a concern.

It still may be the case that populations in the core range of the USA have evolved

alleles that are specific to their environment, and introducing those into the Ontario populations would make them less fit to their own habitat. A closer examination of coding DNA and further studies comparing environment to genetic profiles would have to be done to confirm this. Common gardens experiments would be useful to see if there is any genetic differentiation useful in terms of selective advantage between the different genotypes. Plants of various degrees of genotype divergence could be simultaneously grown in the same environment, mimicking soil conditions in alvar, sand dunes, and forests.

Additionally, there is another triploid *Celtis* species growing in the western United States: *Celtis reticulata*. It appears to be morphologically similar to *Celtis tenuifolia* and has the same ploidy. This suggests that it could be a western form of the same species. It would be a useful extension to this project to analyze its genetic profile to discover its relationships to *C. occidentalis* and *C. reticulata,* and if it has a similar history of multiple origins of alloploidy from these or other diploid parents.

# Literature Cited


Allendorf, F. W., Hohenlohe, P. A., & Luikart, G., 2010 Genomics and the future of conservation genetics, *Nature reviews genetics,* 11: 697 709. http://dx.doi.org/10.1038/nrg2844

Allendorf, F.W., Luikart, G., and Aitken, S.N. 2013. *Conservation and the genetics of populations. Second Edition*. Wiley-Blackwell, West Sussex., UK.

Arrigo, N. and Barker, 2012. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology*, 15: 140-146.

Bapat, V.A, Dixit, G.B., Yadav, S.R. 2012. Plant biodiversity conservation and the role of botanists. *Current Science* 102(10): 1366-1369.

Barringer, B.C., and Galloway, L. F. 2017. The Reproductive ecology of diploid and tetraploid *Galax urceolata*. *Am. Midl. Nat.* 177: 299-308.

Benson, M.H. 2012. Intelligent Tinkering: The Endangered Species Act and Resilience. *Ecology and Society*. 17(4): 28. http://dx.doi.org/10.5751/ES-05116-170428

Bicknell, R. and Koltunow A. 2004. Understanding Apomixis: Recent Advances and Remaining Conundrums. *American Society of Plant Biologists*. 16(1): 228-245.

Boufford, D. E. 1997. *Celtis tenuifolia*. In: Flora of North America Editorial Committee, eds. 1993+. Flora of North America North of Mexico. 20+ vols. New York and Oxford. Vol. 3, pp. 356-357.

Bruvo, R., Michiels, N., D'Souza, T., and Schulenburg, H. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* 13(7): 2101-2106.

Budd, C., Zimmer, E., Freeland, J.R. 2015. Conservation genetics of Magnolia acuminata, an endangered species in Canada: Can genetic diversity be maintained in fragmented, peripheral populations? *Conservation Genetics* 16: 1359-1373.

Buck, G. and Bidlack, J. 1998. Identification of *Quercus* and *Celtis* Species using morphological and electrophoretic data. *Proc. Okla. Acad. Sci.* 78: 23-33.

Brownfield, L., and Kohler, C. 2011. Unreduced gamete formation in plants: mechanisms and prospects. Journal of Experimental Botany. 62(5): 1659-1668.

Campbell, N., Reece, J., Urry, L., Cain, M., Wasserman, S., Minorsky, P., Jackson, R., 2009. *Biology. Eighth Edition*. Pearson Benjamin Cummings, San Francisco, CA.

Cardinale, B.J., Wright, P.J., Cadotte, M.W., Carroll, I.T., Hector, A., Srivastava, D.S., Loreau, M., and Weis, J.J. 2007. Impacts of plant diversity on biomass production increase through time because of species complementarity. *PNAS* 104(46): 18123-18128.

Chapman, H., Robson, B., and Pearson, ML. 2003. Population genetic structure of a colonizing triploid weed, *Hieracium lepidulum. Heredity* 92: 182-188.

Charlesworth, D., and Willis, J. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* 10: 783-796.

Clark, L., and Jasieniuk, M. 2011. Polysat: an R package for polyploidy microsatellite analysis. *Molecular Ecology Resources*, 11(3), 562-566. http://doi.org/10.1111/j.1755-0998.2011.02985.x

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. vi + 15 pp. (www.sararegistry.gc.ca/status/status_e.cfm)

COSEWIC, 2019. Committee on the Status of Endangered Wildlife in Canada. (https://www.canada.ca/en/environment-climate-change/services/committee-status-endangered-wildlife.html)

Craig, D.L. 1960. Studies on the cytology and the breeding behaviour of *Rubous Canadensis L. Can. J. Genet. Cytol.* 2: 96-102.

Cruzan, M., and Templeton, A. 2000. Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. Trends in Ecology & Evolution. 15(12): 491-496.

Dolezel, J., Greilhuber, J., and Suda, J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2(9): 2233 – 2244.

Don, R., Cox, P., Wainwright, B., Baker, K., and Mattick, J. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19(14): 4008.

Duffy, E.J. 2009. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ*. 7(8): 437-444.

Dyke, A. and Prest, V. 1987. Late Wisconsonian and Holocene History of the Laurentide Ice Sheet. *Geographie physique et Quaternaire* 41(2): 237-263.

Eckert, C., Samis, E., and Lougheed, S. 2008. Genetic variation across species' geographical ranges: the central –marginal hypothesis and beyond. *Molecular Ecology* 17: 1170-1188.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14:2611–2620

Falush, D., Stephens, M., Pritchard, J. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4): 1567-1587.

Farrar, J.L. 1995. Trees in Canada. Fitzhenry & Whiteside Limited, Markham and the Canadian Forest Service, Natural Resources Canada, Ottawa, Ontario in cooperation with Public Works and Government Services Canada. 502 pp.

Fernald, M. L. 1950. Gray's Manual of Botany : Eighth Edition. New York. NY: American Book Company.

Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science.* 220:1049–1051.

Gauthier, M., Crowe, E., Hawke, L., Emery, N., Wilson, P., Freeland, J. 2010. Conservation genetics of Pitcher's thistle (*Cirsium pitcheri*), an endangered Great Lakes endemic. *Botany*. 88: 250-257.

Gleason, H.A., and Conquist, A. 1991. Manual of Vascular Plants of Northeastern United States and Adjacent Canada: Second Edition. Bronx, NY: The New York Botanical Garden Press.

Goudet, J. 2005. Hierfstat, a package for R to compute and test variance components and F-statistics. *Molecular Ecology Notes*. 5:184-186.

Griffiths, A., Wessler, S., Carroll, S., and Doebley, J. 2012. *Introduction to Genetic Analysis. Tenth Edition*. W.H. Freeman and Company, New York, NY.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa T., Reverdel, E., Salin, F., Petit, R.J. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11(4): 591–611.

Hartl, D. and Clark, A., 2007. Principle of Population Genetics. Fourth Edition. Sinauer Associates, Inc. Publishers. Sunderland, MA.

Hewitt, G. 2000. The genetic legacy of the quaternary ice ages. *Nature* 405: 907-913.

Hodel, R.G., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Xiaoxian, L., Gitzendanner, M.A., Douglas, N.A., Germain-Aubrey, C.C., Chen, S., Soltis, D.R., and Soltis, P. 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Application in Plant Sciences*. 4(6): 1600025.

Husband, B. 2004. Chromosomal Variation in Plant Evolution. *American Journal of Botany* 91(4): 621-625.

Husband, B. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* 82: 537-546.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi: 10.1093/bioinformatics/btn129

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ2*:e281. https://doi.org/10.7717/peerj.281

Kohler, C., Scheid, O., and Erilova, A. 2010. The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* 26(3): 142-148.

Leopold, A. 1949. The Sand County Almanac, and Sketches Here and There. New York, NY: Oxford University Press.

Lesica, P., & Allendorf, F. W. 1995. When are peripheral populations valuable for conservation? *Conservation Biology*. 9(4), 753–760.

Loveless, M., and Hamrick, J. 1984. Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics* 15: 65-95.

Mayr, E., 1942. Systematics and the Origin of Species. New York, NY: Columbia University Press.

Meirmans, P., and Tienderen, P. 2004. Genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Resources*. 4(4): 792-794. https://doi.org/10.1111/j.1471-8286.2004.00770.x

Morgante, M., Olivieri, A. 1993. PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* 3(1): 175-182.

NatureServe. 2019. NatureServe Web Service. Arlington, VA. U.S.A. Available at http://services.natureserve.org.

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R., Simpson, G., Solymos, P., Stevens, M., Szoecs, E., and Wagner, H. 2018. Vegan: Community ecology package. R package version 2.4-6. http://CRAN.R-project.org/package=vegan

Parks Canada Agency. 2011. Recovery Strategy for Dwarf Hackberry (*Celtis tenuifolia*) in Canada. *Species at Risk Act* Recovery Strategy Series. Parks Canada Agency. Ottawa. vi + 43 pp.

Petit, R.J., Aguinagalde, I. ,Beulieu, J. , Bittkau, C., Brewer, S., Rachid, C., Richard, E., Silvia, F., Delphine, G. 2003. Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science.* **300**(5625): 1563-1565.

Pfosser M, Amon A, Lelley T, Heberle-Bors E. 1995. Evaluation of sensitivity of flow cytometry in detecting aneuploidy in wheat using disomic and ditelosomic wheat-rye addition lines. *Cytometry*. 21: 387–393.

Quirez, K. 2007. Species Concepts and Species Delimitation. *Systemic Biology*. 56(6): 879-886.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramsey, J. and Schemske, D. 2002. Neopolyploidy in Flowering Plants. *Annu. Rev. Ecol. Syst.* 33: 589-639.

Rich, T., Mcdonnell, E., LLedo, M. 2008. Conservation of Britain's biodiversity: the case of *Hieracium cyathis* (Asteraceae) and its relation to other apomictic taxa. *Botanical Journal of the Linnean Society*. 156 : 669-680.

Rowe, G., Sweet, M., & Beebee, T. 2017. *An Introduction to Molecular Ecology. Third Edition*. Oxford, UK: Oxford University Press.

Selkoe, K. A., & Toonen, R. J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers, *Ecology Letters*, 9(5): 615–629.

Schinkel, C., Kirchheimer, B., Dullinger, S., Geelen, D., Storme, N. and Horandl, E.2017. Pathways to polyploidy: indications of a female triploid bridge in the alpine species *Ranunculus kuepferi* (Ranunculaceae). *Plant Syst. Evol*. 303: 1093-1108.

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

Singh, F., Brat, S., and Khoshoo, T. 1967. Natural Triploidy in Viviparous Onions. *Cytologia* 32: 403-407.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236(4803):787-792.

Smith T., Kron P., Martin S. 2018. "flowPloidy: An R package for genome size and ploidy assessment of flow cytometry data." *Applications in Plant Sciences* **6**(7), e01164. doi: 10.1002/aps3.1164.

Soltis, D., Soltis, P. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* 13(9): 348-352.

Soltis, D., Soltis, P., Schemske, D., Hancock, J., Thompson, J., Husband, B., and Judd, W. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56(1): 13-30.

Stebbins G. 1950. Variation and Evolution of Plants. New York: Columbia University Press.

Stuessy, T. 2009. Chapter 11: The Species. In *Plant Taxonomy: The Systematic Evaluation of Comparative Data*. *Second Edition.* (pp:139-150) New York, NY: Columbia University Press.

Toro, M. and Caballero, A. 2005. Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B*. 360: 1367-1378. doi: 10.1098/rstb.20005.1680

Vieira, M.L., Santini, L., Diniz, A.L., and Munhoz, C. 2016. Microsatellite markers: what they mean and why they are so useful. *Genet. Mol. Biol.* 39(3): 312-328.

Waldron, G. 2003. Trees of the Carolinian Forest: a guide to species, their ecology and uses. Boston Mills Press, Ontario. 275 pp.

Wagner, W. H. 1974. Dwarf hackberry (Ulmaceae: *Celtis tenuifolia*) in the Great Lakes region. *Mich. Bot*. 13(2): 73–99.

Whittemore, A. T., & Townsend, A. M. 2007. Hybridization and Self-compatibility in *Celtis*: AFLP Analysis of Controlled Crosses. *Journal of the American Society for Horticultural Science*, 132(3): 368–373.

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

Willis, K.J. (ed.) 2017. State of the World's Plants 2017. Report. Royal Botanic Gardens, Kew.

Young, A., Boyle, T. and Brown T. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution* 11: 413-418.

Yamauchi, A., Hosokawa, A., Nagata, H. and Shimoda, M. 2004. Triploid Bridge and Role of Parthenogenesis in the Evolution of Autopolyploidy. *The American Naturalist*. 164(1) 101-112.

Zohary, D., and Nur, U. 1959. Natural triploids in the orchard grass, *Dactylis glomerata L*., polyploid complex and their significance for gene flow from diploid to tetraploid levels. *Evolution* 13: 311-317.