

**An Assessment of Genetic Diversity and Spatial Distribution of
Cercis canadensis L. (Eastern Redbud) Using Microsatellite Loci**

**A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville**

**Meher Afroze Ony
August 2019**

DEDICATION

I dedicate this manuscript to my amazing husband, Mehraj Ahmed, and my loving parents. I dearly value their unconditional support and appreciate their patience and invested time.

ACKNOWLEDGMENTS

Special thanks to my advisor Dr. Denita Hadziabdic-Guerry for giving me this opportunity to pursue my Master's degree. Her continuous support and guidance helped me to finish this project.

I would also like to thank my committee members, Dr. Robert Trigiano, Dr. William Klingeman, and Dr. John Zobel, for their assistance and encouragement throughout my project. I would also like to thank Dr. Marcin Nowicki and Ms. Sarah Boggess for their unconditional support. I also would like to give special thanks to all my lab members and peers in the Entomology and Plant Pathology Department.

I would like to express my gratitude to all collaborators and those who collected samples to complete this project. Finally, I would like to thank the University of Tennessee Institute of Agriculture and the United States Department of Agriculture for their financial support.

ABSTRACT

Cercis canadensis L. (Eastern Redbud) is a small, understory tree native to the eastern United States. This species is a popular deciduous shade tree that accounts for more than \$27 million in the annual native tree and cultivar sales in the United States. Knowledge of the genetic diversity and spatial distribution of *C. canadensis* populations in their native range is currently limited. For this study, we estimated genetic diversity and determined the spatial structure of *C. canadensis* populations at a fine scale within the Tennessee-Georgia area, and at a broad scale across the native range of the species. We hypothesized that high genetic diversity and the presence of population structure of *C. canadensis* would be evident at both fine and broad-scale perspectives, due to the species' wide range in observable morphological variation. For the fine-scale study, we utilized 15 microsatellite loci developed previously to determine genetic diversity and population structure of 174 individual, wild-type (open-pollinated) trees from 18 collection sites in Tennessee and Georgia. For the broad-scale study, we used 12 microsatellite loci to determine genetic diversity and spatial structure of 691 individual wild-type trees in 74 locations across 23 states. At both fine and broad scale, we detected two genetic clusters, high genetic diversity ($H_e = 0.60$ and $H_e = 0.67$, respectively), and moderate to high levels of genetic differentiation ($F_{ST} = 0.14$ and $F_{ST} = 0.19$, respectively) among *C. canadensis* populations. At both fine and broad scales, the majority of genetic variation were individually based (45.60%, $P < 0.001$ and 69.15%, $P < 0.001$, respectively). The high levels of genetic diversity and spatial structure of *C. canadensis* at fine and broad scales has important implications for habitat management and for future breeding programs that will target desirable horticultural traits in the species.

TABLE OF CONTENTS

INTRODUCTION.....	1
CHAPTER 1 FINE- SCALE POPULATION STRUCTURE OF <i>CERCIS CANADENSIS</i> L. (EASTERN REDBUD): THE EFFECT OF HABITAT FRAGMENTATION ON GENETIC DIVERSITY AND DIFFERENTIATION.....	4
Abstract	5
Introduction	6
Materials and Methods	10
Sample Collection.....	10
DNA Extraction	10
Microsatellite Primers and Genotyping Conditions.....	11
Genetic Diversity	12
Population Structure	13
Demography.....	15
Results	16
Microsatellite Genetic Diversity and Hierarchical Fixation Indices	16
Population Structure	17
Population Demography	20
Discussion.....	21
Appendix: Tables and Figures.....	26
CHAPTER 2 GENETIC DIVERSITY AND SPATIAL DISTRIBUTION OF <i>CERCIS</i> <i>CANADENSIS</i> IN THE UNITED STATES	40
Abstract	41
Introduction	42

Materials and Methods	43
Sample Collection.....	43
DNA Extraction	44
Microsatellite Primers and Genotyping Conditions.....	45
Genetic Diversity	46
Population Structure	47
Demography.....	49
Results.....	51
Microsatellite Genetic Diversity and Hierarchical Fixation Indices.....	51
Population Structure	52
Population Demography	55
Discussion.....	56
Appendix: Tables and Figures.....	64
CONCLUSION	80
REFERENCES.....	83
VITA	108

LIST OF TABLES

Table 1.1. Eighteen collection sites of <i>Cercis canadensis</i> collected from Tennessee and Georgia in the eastern United States.	26
Table 1.2. Genetic diversity indices across 18 collection sites of <i>Cercis canadensis</i> using 15 microsatellite loci.	27
Table 1.3. Genetic diversity indices of (i) 18 collection sites and (ii) 2 collection sites of <i>Cercis canadensis</i> using 15 microsatellite loci.	28
Table 1.4. Pairwise F_{ST} matrix of Nei's genetic distance of <i>Cercis canadensis</i> utilizing 18 collection sites.	29
Table 1.5. Analysis of molecular variance of <i>Cercis canadensis</i> across 15 microsatellite loci for (i) 18 collection sites as one hierarchical cluster, ii) 18 collection sites into two groups according to two clusters of STRUCTURE; (iii) 18 collection sites as two groups (north and south group).....	30
Table 1.6. Bottleneck determination by sign tests for <i>Cercis canadensis</i> samples utilizing 15 microsatellite loci	31
Table 2.1. Estimation of diversity indices for seventy-four <i>Cercis canadensis</i> collection sites from the Unites States using 12 microsatellite loci.....	64
Table 2.2. Genetic diversity indices of 12 microsatellite loci across 74 collection sites of <i>Cercis canadensis</i>	66
Table 2.3. Analysis of molecular variance of <i>Cercis canadensis</i> across 12 microsatellite loci for (i) 74 collection sites as one hierarchical cluster, ii) 74 collection sites into two groups according to two clusters of STRUCTURE.	67

Table 2.4. Bottleneck determination by sign tests for <i>Cercis canadensis</i> samples utilizing 12 microsatellite loci. Seventy-four collection sites were divided into two clusters according to the STRUCTURE analysis.....	68
--	----

LIST OF FIGURES

Figure 1.1. Map of location sites for 18 collection sites of <i>Cercis canadensis</i> . Red dots represent sampling sites separated by north group (east Tennessee) and south group (Georgia-Tennessee border).....	32
Figure 1.2. STRUCTURE bar graph representing two genetic clusters ($\Delta K = 2$) among 18 collection sites of <i>Cercis canadensis</i> . Each vertical bar represents an individual sample and color of the bar indicates the assignment probability of that individual to belong to one of the two identified clusters, designed in two different colors, blue or red.	33
Figure 1.3. Bayesian clustering probabilities using programs STRUCTURE and InStruct. STRUCTURE results were obtained using Delta K and graph results indicated two genetic clusters among 18 <i>Cercis canadensis</i> collection sites (A). InStruct results were obtained using deviance information criterion calculations (B). Results indicated the highest probability of two different genetic clusters ($K = 2$) across 18 <i>Cercis canadensis</i> collection sites.....	34
Figure 1.4. Neighbor-joining tree of 18 collection sites of <i>Cercis canadensis</i> was constructed and visualized using Nei's genetic distance. Numbers indicate percentage bootstrap support values using 1,000 replications.	35
Figure 1.5 (A). The Minimum Spanning Networks (MSN) of <i>Cercis canadensis</i> genotypes were calculated using Bruvo's genetic distance (considering microsatellite motif length). Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.	36

Figure 1.5 (B). The Minimum Spanning Networks (MSN) of <i>Cercis canadensis</i> genotypes were calculated using Nei's Genetic distance. Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.	37
Figure 1.6. Discriminant Analysis of Principal Components (DAPC) plot of <i>Cercis canadensis</i> individuals with 18 collection sites (A) and 12 collection sites (B). The alleles with respective locus explaining the most variance is indicated at plot axes, with percentage contributions in parenthesis.....	38
Figure 1.7. Mantel test of <i>Cercis canadensis</i> individuals across 18 collection sites in east Tennessee and around Georgia-Tennessee border. This test represents correlation between geographical and genetic distance.....	39
Figure 2.1. Map of 74 collection sites (red dots) of <i>Cercis canadensis</i> across its native distribution in the United States.....	69
Figure 2.2. Map of private alleles found in <i>Cercis canadensis</i> populations across the eastern United States. Red dots represent sampling sites that contain private alleles.	70
Figure 2.3. STRUCTURE bar graph representing two genetic clusters ($\Delta K = 2$) among 74 collection sites of <i>Cercis canadensis</i> . Each vertical bar represents an individual sample and the color of the bar indicates the assignment probability of that individual to belong to one of the two identified clusters (blue or red).	71
Figure 2.4. Map of 74 collection sites of <i>Cercis canadensis</i> across 23 states in the eastern United States. The color code represents designated cluster for each collection site according to the STRUCTURE analysis (red = cluster 1; blue = cluster 2).....	72

Figure 2.5. Bayesian clustering probabilities using programs: STRUCTURE and InStruct. STRUCTURE results were obtained using Delta K and graph results indicated two genetic clusters among 74 <i>Cercis canadensis</i> collection sites (A). InStruct were obtained using deviance information criterion calculations (B). Results indicated the highest probability of two different genetic clusters (K = 1) across 74 <i>Cercis canadensis</i> collection sites.	73
Figure 2.6. Neighbor-joining tree of 74 collection sites of <i>Cercis canadensis</i> using Nei's genetic distance. Numbers indicate percentage bootstrap support values using 1,000 replications.	74
Figure 2.7. Discriminant Analysis of Principal Components biplot of <i>Cercis canadensis</i> individuals was constructed using 1,000 permutations. The alleles with loci explaining the most variance are indicated at plot axes, with percent contributions in parentheses. Discriminant Analysis (DA) eigenvalues are also presented.	75
Figure 2.8. (A). The Minimum Spanning Networks (MSN) of <i>Cercis canadensis</i> genotypes were calculated using Bruvo's genetic distance (considering microsatellite motif length) (A). Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.	76
Figure 2.8. (B). The Minimum Spanning Networks (MSN) of <i>Cercis canadensis</i> genotypes were calculated Nei's genetic distance (B). Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.	77

Figure 2.9. Mantel test of *Cercis canadensis* individuals from 74 collection sites across 23 states in the eastern United States. The test indicated the correlation between geographical and genetic distance..... 78

Figure 2.10. Figure 2.10. Most probable DIYABC evolutionary scenario for *Cercis canadensis*. Here, populations are labeled as northeast (N1), south (N2), and northwest (N3). Scenario 1 that suggested simultaneous triple split of the originating population (N1) at t1 and occurrence of bottleneck at t1-db, provided the best support. Ne = effective population size. Model checking analyses of the closest 1% of simulated prior and posterior datasets (bottom panel) using PCA further supported scenario.

..... 79

INTRODUCTION

Cercis canadensis L. (Fabaceae) or eastern redbud is an understory temperate tree species. This deciduous species is native to regions of the mid-west to eastern United States (U.S.), as well as northeastern Mexico [1, 2]. This highly valued ornamental flowering tree presents a wide and umbrella-like crown [3], showy heart-shaped leaves, attractive and edible early spring flower, and compact form. *C. canadensis* grows well in partial shade and the species is adapted to a wide range of climate conditions and elevations [4, 5]. In addition, this commercially valued species has more than 36 cultivars [6, 7] and provides a commercial market value of \$27 million annually in the U.S. alone [7, 8].

The genus *Cercis* consists of 10 to 13 species of shrubs and small woody trees [6, 7, 9], which are indigenous to the temperate zones of Europe, Asia, and North America [5, 9, 10]. *Cercis* spp. are tolerant of high temperature and drought [11, 12], which makes these trees desirable landscape specimens due to low maintenance requirements. Besides *C. canadensis*, *C. orbiculata* Greene (California redbud or western redbud) (syn. = *C. occidentalis* Torr. ex A. Gray var. *orbiculata* (Greene) Tidestr.) [13], is the only other native species occurring in North America.

The bark of *C. canadensis* has astringent properties which are useful in treating dysentery [2, 14]. The bark and root of *C. canadensis* are also used to treat whooping cough, vomiting, fever and congestions by native American tribes in the U.S. Timber is utilized as firewood by native Americans during the winter season (USDA-PLANTS Database, 2017), but the timber has no economic value [14, 15]. The flowers are an efficient nectar source for a number of insects and hummingbirds [14, 16, 17]. The seed is a food source for several bird species such as the ring-necked pheasants (*Phasianus*

colchicus L.), bobwhite quail (*Colinus virginianus* L.), and the rose-breasted grosbeak (*Pheucticus ludovicianus* L.) [18]. White-tailed deer (*Odocoileus virginianus* Zimmermann) and grey squirrels (*Sciurus carolinensis* Gmelin) eat tree leaves, twigs, bark, and seeds [19].

Although *C. canadensis* has important ornamental, economic and environmental characteristics, very little is known about the genetic diversity and spatial structure that is present within this native tree species. Although there is a presence of highly quantitative and the continuous morphological characters [13], it is difficult to investigate the genetic diversity of this species by using solely morphological information [6]. Therefore, in this study we utilized microsatellite loci as a molecular tool to understand the genetic diversity and spatial distribution of this species across its native range in the eastern U.S. In this research, we focused on *C. canadensis* wild population distributions at both fine scale, comprising populations of trees taken around east Tennessee and Georgia-Tennessee border, and broad scale areas, comprising tree specimens from across the mid-western to eastern U.S. We hypothesized that high genetic diversity among the wild collections of *C. canadensis* would be present and wild populations would be spatially clustered across its native geographical distribution in U.S.

CHAPTER 1

FINE- SCALE POPULATION STRUCTURE OF *CERCIS CANADENSIS* L. (EASTERN REDBUD): THE EFFECT OF HABITAT FRAGMENTATION ON GENETIC DIVERSITY AND DIFFERENTIATION

Abstract

Forest fragmentation may have consequential effects on the plants through reduced genetic diversity and increased population structure due to habitat isolation, reduced population size and disturbance of pollen-seed dispersal methods. However, in case of tree species, longevity of the individuals, effective seed or pollen dispersals, mating system, and ecological dynamics may help the species to resist or overcome the negative effect of forest fragmentation. A fine scale study can shed light into the post-fragmentation genetic diversity and structure of a species. Here, we investigated the genetic diversity and population structure of *Cercis canadensis* (Eastern Redbud) wild populations at fine scale in fragmented areas around the Georgia-Tennessee border. We hypothesized high genetic diversity among the wild collections of *C.*

canadensis distributed across smaller geographical ranges. To determine the fine-scale population structure and genetic diversity around east Tennessee and the Georgia-Tennessee border, we utilized 15 previously developed microsatellite loci for genotyping 174 individuals from 18 wild collection sites. Our results indicated presence of population structure, high genetic diversity ($H_e = 0.61$), and moderate genetic differentiation ($F_{ST} = 0.14$) among the collection sites. The majority of variation was individually based (45.60%, $P < 0.001$), when compared to variation between the samples within collection sites (37.13%, $P < 0.001$) as revealed by analysis of molecular variance. Only 17.28% of the variation ($P < 0.001$) was attributed to the differences among the 2 groups. Two major genetic clusters among *C. canadensis* individuals within the smaller geographical distribution were revealed by STRUCTURE. Our data suggests that native *C. canadensis* populations in the fragmented area around Georgia-

Tennessee were able to maintain high levels of genetic diversity and spatial genetic structure. This study has important implications for habitat management efforts, and future breeding programs to improve desirable horticultural traits.

Introduction

Habitat fragmentation involves discontinuities in an organism's distribution due to geographical and/or geological barriers and/or human infrastructures [20, 21]. For example, roadways have converted forests into long, linear artificial roadside patches, that may become populated by a mix of wild-type and ornamental plant cultivars [20, 22]. Such fragmentation impacts the genetic diversity and population structure of a species in various ways depending on the species' ecology and biology [23, 24]. In a constantly changing natural setting, the dynamics of a plant population and ecological succession of plant species are shaped by the following three major challenges: how successfully the species is dispersed, established, and how long the species can persist in a specific environment [21, 25].

Habitat fragmentation may affect these processes through a number of ecological and biological modifications imposed on the fragmented area [21, 24, 26]. For instance, animal based seed dispersal and bee pollination [14] methods are affected negatively by forest fragmentation, especially at local levels [21, 27, 28]. Disruptions in dispersal processes can reduce gene flow and increase inbreeding within these spatially isolated populations [28-30]. Loss of habitat, or even its degradation caused by fragmentation, can reduce the availability of suitable habitat and therefore negatively influence species establishment [21, 26]. Finally, habitat fragmentation can lead to reductions in

population size, that result in reduced genetic variability, adaptive potential, and survival of the members in the smaller, isolated population [30, 31].

Forest fragmentation reduces the natural habitat of a species, but also creates artificial edges that differ in plant composition from the rest of the forested area [22]. Presence of forest edges may increase the habitat area for a number of broad-leaf and light-demanding tree species typically found in higher numbers in these types of ecosystems [22]. In many cases, tree species in temperate forests are less likely to be impacted by genetic consequences of forest fragmentation compared to tropical tree species [32]. This outcome is partly explained by the higher tree density and undisturbed pollen-seed dispersal. This can, in turn, ensure sufficient gene flow across these isolated populations, thus reducing potential threat of genetic declines to broad-leaf temperate tree species in North America [32-34].

Despite the apparent resilience of tree species in temperate forests, several studies provided insight about the consequences of forest fragmentation and the impact of pathogen pressure on genetic diversity and spatial structure of native North American tree species. Well-known examples include the near extinction of American chestnuts [*Castanea dentata* (Marsh.) Borkh] caused by *Cryphonectria parasitica*, (Murrill) Barr (chestnut blight) and extensive logging and fragmentation [35, 36]. High mortality of flowering dogwood (*Cornus florida* L.) as a result of dogwood anthracnose, *Discula destructiva* Redlin [37, 38] was observed a decade after the initial discovery of this fungal pathogen in the late 1970s [37-39]. Despite heavy disease pressures, this native, understory tree maintained high genetic diversity with weak genetic structure in a small geographical region (Great Smoky Mountains National Park) [37, 38].

Pinus strobus L. (eastern white pine), *Juglans nigra* L. (black walnut) and several other highly exploited tree species (due to logging and forest fragmentation) have showed little to no genetic consequences of these demographic events challenging their sustainability [40, 41].

In contrast, *Taxus baccata* L., a forest tree species in Spain, was negatively affected by chronic fragmentation and revealed strong spatial structure with a recent bottleneck history in spite of abundant seed dispersal mechanisms [42]. Fragmented systems can result in a reduction of overall species health, but also in tree mortality that can in turn restructure forest compositions in impacted communities and disrupt the ecosystem processes of tree species [36, 38, 43]. Since the European settlement of the United States (U.S.), about 220 plant species have become extinct in North America and Hawaii [44, 45]. Moreover, in the list of threatened and endangered species in the U.S., 81% were affected by anthropogenic activities [44, 46].

In geographically reduced or fragmented populations, genetic diversity can be lost through genetic drift and inbreeding, which can further reduce the ability of the affected individuals to regenerate and respond to changes of selection pressures [24, 38].

Genetic drift and inbreeding in a plant species can erode the overall population fitness and the prospects for adaptive change, thus increasing the possibility of species decline, mortality, or extinction [38, 47-49]. Due to constantly changing forest structures as a result of fragmentation, urbanization, and environmental conditions, knowledge of current genetic diversity and spatial structure of economically and/or environmentally important species is often unknown or limited. Here, we focused on the U.S. native

Cercis canadensis L. (Eastern redbud) wild populations distributed across a smaller, fragmented geographical area.

Cercis canadensis is a broad-leaf understory tree distributed in the Eastern U.S. and into the northern part of Central Mexico [1, 6, 13, 50]. This small -to -medium size ornamental tree is distinguished by its umbrella-shaped crown with various foliage colors ranging from deep purple, green to yellow [3, 4, 6]. Characteristic heart-shaped leaves, wide range of foliage and floral colors, and early spring blooms make *C. canadensis* a popular ornamental landscape tree in the temperate North America [4, 6, 14, 50]. There are more than three dozen commercial cultivars of *C. canadensis* currently available in retail and wholesale trade [6, 7]. Consequently, *Cercis* spp. cultivars have achieved an annual U.S. market value of \$27 million USD [8].

Given the ornamental popularity of this species, we utilized microsatellite loci to investigate genetic diversity and spatial structure of wild *C. canadensis* across a smaller geographical distribution in east Tennessee and around the Tennessee-Georgia border in the eastern U.S. Our hypothesis is that despite increased habitat fragmentation, wild populations of *C. canadensis* at local levels would show indication of high genetic diversity with genetic structure. To test our hypothesis, we used previously developed microsatellite loci [6] to achieve the following objectives: 1) to evaluate the fine-scale genetic diversity present within *C. canadensis* populations occurring in east Tennessee and along the Georgia-Tennessee border; and 2) to infer fine-scale patterns in the spatial distribution of *C. canadensis* along the Georgia-Tennessee border.

Materials and Methods

Sample Collection

Samples of *C. canadensis* were collected from east Tennessee and around the Georgia-Tennessee border, U.S., which is located near the center of the current native distribution range of the species (Fig 1.1). Young and newly expanded leaves from approximately 10-12 non-cultivated trees per collection site were collected at 18 locations (Table 1.1). From each tree, five to seven leaves were randomly collected from branches that were at cardinal directions around each tree. Leaves were placed in paper envelopes to dry and geographical coordinates were recorded for each sampled tree. For fine scale regional analysis, a total of 180 trees were collected; from 10 sites in Eastern Tennessee (north group) and 8 from sites around the Georgia-Tennessee border (south group). Samples collected from each location were considered a single collection site. There were 18 collection sites used for this project.

DNA Extraction

DNA was isolated from approximately 100 mg of dried leaf tissue. Samples were homogenized using a Beadmill 24 (Fisher Scientific, Pittsburgh, Pennsylvania, USA) four times with the settings of S (speed) = 6.00 m/s, T (time) = 30 s, C (cycle) = 1 and D (delay or break period) = 0.00. Between each homogenization step, samples were frozen in liquid nitrogen for a minimum 5 minutes to improve the tissue homogenization. Genomic DNA (gDNA) from each sample was isolated using the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, California, U.S.) using the manufacturer's protocol with several modifications: 2% w/v polyvinylpyrrolidone (PVP) was mixed into lysis buffer

(AP1); 8 µl of RNase was added into each sample tube; the time period of incubation at 65°C was increased to 45 minutes and inverted gently to mix the sample every two minutes; incubation step at -20°C was followed with increased time of incubation to one hour minimum. Additionally, before adding the elution buffer, ethanol was used to wash the spin columns if there was some debris left from the sample tissue. Finally, 50 µl of elution buffer (AE) pre-heated at 65°C was added and the step was repeated twice. The quality of the isolated gDNA was assessed using ND1000 visible (UV-Vis) Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, U.S.).

Microsatellite Primers and Genotyping Conditions

Primers for 68 genomic microsatellite loci [6] were synthesized by Integrated DNA Technologies (IDT, Coralville, Iowa, U.S.). Because the 68 primers were developed for a cross-species amplification study among eight species from the genus *Cercis* and another closely related species *Bauhinia faberi* Oliver [6], further screening was needed to optimize these primers for the non-cultivated *C. canadensis* samples used in this study. We used gDNA samples of five *C. canadensis* individuals from the University of Tennessee (UTK) Gardens (Knoxville, TN, U.S.) to screen all of the primers used in this study. From our initial screening results of the tested pool of 68 microsatellite loci, 15 primers pairs were selected based on polymorphic loci, as well as successful amplification rate across the five tested samples (Table 1.2).

Polymerase chain reaction (PCR) amplifications were completed, in 10 µl reaction mixture consisting of the following: 1 µl undiluted gDNA, 1 µl of 10µM of each forward and reverse primer, 0.5 µl of dimethyl sulfide, 4 µl of GoTaq G2 Hot Start Master Mix

(Promega Corp, Madison, Wisconsin U.S.) and 2.5 µl water. Both positive control (a DNA sample that amplified across all microsatellite loci) and negative control (control reaction without any DNA sample) samples were used for every primer tested to ensure validity of the data. Amplification reactions were temperature-cycled in 96 well plates using an Eppendorf Thermocycler (Eppendorf AG, Hamburg, Germany) with the following thermal profile: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and an extension at 72°C for 30 s, with a final extension of 72°C for 4 minutes. Amplified PCR products were visualized using a QIAxcel Capillary Electrophoresis System (Qiagen) and sized with a 15/600 bp internal alignment marker and a 25 to 500 bp DNA size marker. All 180 *C. canadensis* gDNA samples were tested against each of the 15 markers using the procedure described above. Failed reactions were repeated two times before declaring them missing data in the dataset. Six samples were discarded due to missing data in more than 25% loci (more than five microsatellite loci). As a result, 174 samples were included in the analyses.

Genetic Diversity

The FLEXBIN excel macro [51] was used to bin raw alleles into statistically similar allelic classes. The binned allelic data was used for all statistical analyses. Samples were divided into the following two groups based on geographical distance from each other: the 10 collection sites from Eastern Tennessee were combined in one group (north group) and the 8 collection sites from the Georgia-Tennessee border (south group) into the second group. To avoid over-representation of possible clonal samples (for example, originating from a planted cultivar, rather than an pollinated, wild type tree) in

the data set, clone-correction was completed using POPPR version 2.8.2 [52] in R Studio version 3.5.3 (R Core Team, 2019). Only unique multilocus genotypes (MLG) per collection site were used for further analysis so that unbiased allelic frequency estimates could be obtained [53].

Genetic diversity indices across 15 microsatellite loci and 18 collection sites of *C. canadensis* were calculated using package POPPR. For each microsatellite locus, we estimated number of alleles, allelic richness (A_r ; an estimation of a population's long term potential to adapt and persist in a given population), observed heterozygosity (H_o ; number of the individual heterozygotes present per locus), and expected heterozygosity (H_e ; expected heterozygosity per tested locus). Genetic fixation index (F_{ST} ; a measure of nearness to allelic fixation within a population [54-56]), genetic differentiation (F'_{ST} , a measure of relative degree of differentiation in allelic frequency [54-56]), and inbreeding coefficient (F_{IS} ; an estimator of the probability that two alleles at a random locus are from the same ancestor [57]) were also calculated using POPPR. In addition, the Shannon Weiner diversity index (H) was estimated and reflects both allele species richness and evenness of the allele distribution.

Population Structure

Population structure of the *C. canadensis* collection was analyzed with the program STRUCTURE version 2.3.4 [58] by utilizing Monte Carlo Markov Chain (MCMC) using a Bayesian method of clustering. To infer genetic clusters among *C. canadensis* individuals, we utilized the following parameters: 500,000 burn-in period with 500,000 MCMC repetitions using 20 independent chains for each K value (1 to 18). Results were

analyzed with STRUCTURE HARVESTER web version 06.94 [59]. The optimum value of K (number of inferred population clusters) was estimated using the Evanno method [60]. For STRUCTURE analysis, we used an admixture model to interpret the proportion of individuals originating from different groups or clusters. The results of the ΔK criterion attained from STRUCTURE HARVESTER were visualized using POPHELPER 2.2.6 [61] in R Studio.

However, Evanno method used by STRUCTURE HARVESTER is unable to calculate a value of ΔK below two [60]. To mitigate that issue and ensure the accuracy of our STRUCTURE results, we utilized the program InStruct, Bayesian clustering method used to infer population structure [62, 63]. This program considers the presence of clonal populations ($K = 1$) and disregards the assumption of Hardy-Weinberg equilibrium within groups of populations. The following parameters were used in this analysis: 20 independent MCMC chains for each K value (1 to 18) with burn-in period of 500,000 and 500,000 repetitions of the thinning interval using the admixture model. An admixture model was implemented to evaluate the proportion of mixed ancestry in an individual and improve the clustering [64]. Deviance Information Criterion (DIC) values were estimated to attain optimum K value [63].

Several model-free clustering approaches were implemented to delineate the population clusters. Genetic distance among different clusters was visualized creating dendrograms and minimum spanning network (MSN) using both Nei's [65] and Bruvo's genetic distances [66, 67] in POPPR. Discriminant Analysis of Principal Components (DAPC), a model-free multivariate analysis approach, is a useful tool to investigate the genetic variation among groups without any prior knowledge of clustering among the

sampled populations [68]. DAPC method was implemented and visualized with the R package ADEGENET version 2.1.1 [69], which groups genetically similar collection sites into clusters. The DAPC analysis was optimized with 31 of retained PCs (PCA eigenvalues) and cross-checked with 1,000 permutations of the dataset.

Analysis of Molecular Variance (AMOVA) [70] was estimated using POPPR with 9,999 permutations by grouping the individuals into three hierarchical groups to evaluate the molecular variance partitioning within and among the collection sites. The levels of population hierarchy were as follows: 1) 18 collection sites as one hierarchical group; 2) among groups on the basis of STRUCTURE analysis; and 3) among geographically separate north and south groups.

To estimate the isolation by distance (IBD), Mantel test [71, 72] was completed with 10,000 permutations using MASS version 7.3-51.1 [73] in RStudio. IBD discerns whether or not there is correlation between genetic and geographical distance of the studied individuals. The Mantel test was run across 18 collection sites by considering them as one population dataset as well as across two collection sites (north and south collection sites).

Demography

The program BOTTLENECK version 1.2.02 [74] was implemented to investigate the evidence of a recent bottleneck. To test whether a recent bottleneck or expansion of *C. canadensis* populations on a fine geographical scale has occurred, three mutation models were utilized: stepwise-mutation model (S.M.M), infinite allele model (I.A.M), and two-phase mutational model (T.P.M). Significance of this test under either of these

three models was evaluated by the Wilcoxon sign-ranked test with 10,000 simulations as the number of loci for this study was under 20 [74, 75].

Results

Microsatellite Genetic Diversity and Hierarchical Fixation Indices

For this data analysis, genotypic data of 174 individuals was used. After clone correction, all 174 samples resulted in unique multilocus genotypes. Therefore, results presented here include a total of 174 *C. canadensis* individuals from Tennessee and Georgia without presence of clonal samples in the dataset. An average of 7 alleles (ranged from 4 to 12) per locus were identified (Table 1.2). Mean allelic richness (A_r) was 2.88, ranging from 1.57 for locus 168a to 4.57 for locus 1057a, which demonstrates high allelic richness among *C. canadensis* individuals. The overall expected heterozygosity (H_e) across all 15 microsatellite loci was high ($H_e = 0.61$), ranging from 0.30 (locus 229a) to 0.84 (locus 1057a). Additionally, the overall Shannon-Weiner diversity index (H) was 1.22 for the 15 loci ranging from 0.58 (locus 229a) to 2.00 (locus 1057a) (Table 1.2). Furthermore, our data indicated a moderate to high population fixation value ($F_{ST} = 0.14$; ranging from 0.03 to 0.59; Table 1.2), and high population differentiation ($F'_{ST} = 0.15$; ranging from 0.03 to 0.61; Table 1.2), and presence of inbreeding among *C. canadensis* collection sites ($F_{IS} = 0.39$; Table 1.2) across the 15 microsatellite loci.

After clone correction, 174 MLGs were found, which indicated clonal samples were not present in any of the 18 collection sites. For this study, the accrued MLGs genotype dataset was analyzed in the following two settings: 1) considering all the 18 collection

sites as one population, 2) dividing the entire dataset into two groups where the first group (north group) represented collection sites from the east Tennessee area and the second group (south group) represented the collection sites near the Georgia-Tennessee border, respectively (Fig. 1.1).

The Shannon-Weiner diversity index (H) for the 18 collection sites ranged from 1.95 to 2.3 (Table 1.3), whereas the north and south groups of the two collection sites dataset showed the H index values of 4.54 and 4.38, respectively (Table 1.3). When analyzed across 18 collection sites, 22 private alleles were revealed. Private alleles were detected in 10 out of 18 collection sites, with the highest number found in the Hamilton Co.2 collection site ($n = 4$, Table 1.3). When data was partitioned into two collection sites, 29 private alleles were detected (13 and 16 across the north and south groups, respectively) (Table 1.3). For 18 collection sites, Nei's genetic diversity index (H_e), corrected for collection site sample size, ranged from 0.33 (Polk Co. 1) to 0.62 (Bradley Co.2) with an overall value of 0.61. Comparably, in the two-group analysis, the north group had lower genetic diversity when compared to the south group ($H_e = 0.53$, $H_e = 0.62$, respectively) (Table 1.3).

Population Structure

Using the Bayesian clustering analysis, STRUCTURE results indicated an optimum $\Delta K = 2$ (Fig. 1.2), suggesting that there were two major clusters among 18 collection sites of *C. canadensis*. All collection sites from the north group along with the collection sites Polk Co.1 (TN) and Polk Co.2 (TN) of the south group belonged within the first STRUCTURE-inferred cluster. The remaining six collection sites from the south group

placed in the second inferred cluster (Fig. 1.2). Also, the collection sites of the north group displayed approximately 4% of admixture (less than 80% proportion of any cluster origin). Analysis with an alternative Bayesian algorithm, InStruct (Deviance Information Criterion-based) further supported our STRUCTURE results. Both Bayesian clustering methods congruently estimated $\Delta K = 2$ (Figs. 1.3) indicating that there were two genetically distinct clusters present among the studied 18 *C. canadensis* collection sites.

Pairwise F_{ST} based on Nei's genetic distance of 18 collection sites ranged from 0.01 (between the collection sites Bradley Co.2, TN and Catoosa Co., GA) to 0.33 (between the collection sites Polk Co.1, TN and Hamilton Co.1, TN) (Table 1.4). All of these four collection sites (Bradley Co., Catoosa Co., Polk Co.1 and Hamilton Co.1) belonged to the south group. In the two collection sites level dataset, the pairwise F_{ST} between north and south groups was 0.09. The Nei's genetic distance was utilized to construct a Neighbor-Joining dendrogram and resulted in two major clades that were congruent with the STRUCTURE results. Initially, Polk Co. 1 (TN) and Polk Co. 2 (TN) collection sites were grouped into the south group based on geographical origin. However, in the dendrogram, they were placed into the first clade with all the collection sites from the north groups (TN) (Fig. 1.4). The remaining six collection sites of the south group comprised the second clade. Therefore, with exception of two collection sites (Polk Co. 1 and Polk Co. 2 from south group), collection sites with close geographic proximity belong to the same genetic clusters (Fig. 1.4).

The Minimum Spanning Network (MSN) reticulations of the complete dataset indicated the presence of the population structure with either Bruvo's (Fig. 1.5 (A)) or Nei's

genetic distances (Fig 1.5 (B)). Analysis of the MSN network using Bruvo's distances indicated the presence of population structure across the 18 collection sites of *C. canadensis* more clearly than using Nei's genetic distances.

The DAPC analysis plot for 15 microsatellite loci revealed a similar clustering pattern as observed with STRUCTURE results. The 18 collection sites were divided into two major clusters. All the collection sites from the south group except Polk Co.1 (TN, south group) and Polk Co.2 (TN, south group) were genetically close (similar to other cluster analysis) and formed a compact cluster. All 10 collection sites from the north group along with the two collection sites from south group (Polk Co.1 and Polk Co.2) were part of another cluster that displayed a wider range of genetic variation than the first cluster (Fig. 1.6). When 12 collection sites in the second cluster was analyzed independently, the DAPC plot indicated existence of a substructure among those collection sites. Also, individuals from Anderson Co.1 (TN, north group) were separated by the first axis from all other collection sites, whereas Polk Co. 1 (TN, south group) was separated by the second axis from other collection sites (Fig. 1.6).

The Analysis of Molecular Variance (AMOVA) was tested by the following groups: i) 18 collection sites as one hierarchical group; ii) 18 collection sites into two major clusters on the basis of the STRUCTURE results; and iii) 18 collection sites into two geographical groups (north and south group) (Table 1.5). The first analysis revealed that the highest amount of variance was retained within the individuals (45.60%, $P < 0.001$). Additionally, about 37.13% ($P < 0.001$) and 17.28% ($P < 0.001$) of the variance was partitioned among the individuals within the collection sites and among collection site levels, respectively. When partitioned by two clusters according to

STRUCTURE results, only 20.46% ($P < 0.001$) variation was present between two clusters, whereas 4.86% ($P < 0.001$) variations attributed to among collection sites within these two clusters (Table 1.5). Also, the greatest level of variance partitioned within individuals (41.16%, $P < 0.001$), whereas among individuals within collection sites variations was 33.52% ($P < 0.001$). When the data were divided into two geographical groups (north and south groups), 13.97% ($P < 0.001$) of observed variations occurred between the north and south groups, whereas only 8.74% ($P < 0.001$) variation was attribute to among collection sites within these two groups (Table 1.5). The majority of the variance was present at the individual level (42.60%, $P < 0.001$) and 34.68% ($P < 0.001$) variations were present among individuals within collection sites level (Table 1.5).

Isolation-by-distance analysis suggested that the geographical distance was linearly correlated ($r = 0.32$, $P < 0.001$) with genetic distance of the sampled data (Fig. 1.7). However, when analyzed separately by groups, there was no evidence of correlation between genetic and geographic distance among *C. canadensis* individuals (north group $r = -0.03$, $P = 0.70$; south group $r = -0.07$, $P = 0.14$).

Population Demography

Using the BOTTLENECK software with S.M.M, I.A.M, and T.P.M models, data was partitioned and analyzed based on STRUCTURE results. Sign test revealed no significant excess of heterozygosity among *C. canadensis* collection sites. Also, the results indicated a normal allele frequency distribution observed in both clusters, confirming that *C. canadensis* collection sites from east Tennessee and along the

Georgia-Tennessee border have not been subjected to any recent bottleneck events (Table 1.6).

Discussion

Long-term evolutionary success and ultimate species survival is influenced by population size, genetic diversity, allelic richness, fitness, and substantial gene flow, which are of fundamental importance in plant ecology, evolution, and conservation [76]. Our study revealed high levels of genetic diversity and allelic richness, moderate genetic differentiation, presence of genetic structure, and high gene flow among *C. canadensis* wild populations across the fragmented forest patches in east Tennessee and along the Georgia-Tennessee border. These results are congruent with other studies of temperate species, suggesting that populations of woody forest trees maintain high levels of genetic diversity at population level compared to herbaceous plants [40, 77-80].

Forest areas in eastern North America are affected negatively by human disturbances [81, 82], insect and pathogen infestations [83-86], drought [87-89], wildfire [90], and global climate change [85]. Human disturbances in forest ecosystems include land conversion for agriculture and settlements [81, 91], logging [92, 93], and fragmentation by building infrastructure [20]. In the 1800s and 1900s, human disturbances greatly affected the forest area in the southern Appalachian region and altered the ecosystem by logging, forest clearing, and burning, thus creating poor soil conditions and erosion [91-95]. Despite anthropogenic disturbances and habitat fragmentation, many forest trees are able to maintain high genetic diversity at population levels [40, 77-80, 96-100].

This resilience can be explained in part by species biology including, but not limited to, widespread distribution, outcrossing mating systems, high level of gene flow due to long distance pollen or seed dispersal, and presence of suitable population size [42, 96, 97].

Interestingly, *C. canadensis* populations examined in this study showed higher genetic differentiation when compared to other hardwood species [35, 41, 77, 97, 101-104].

Similar to our findings, understory temperate trees species, *Sorbus torminalis* L. (Crantz) (Wild service tree) and *Cladrastis kentukea* (Dum. Cours.) Rudd (Yellowwood), showed strong genetic differentiation, which was likely related to their patchy distributions and population structure [82, 105]. Unlike *S. torminalis* and *C. kentukea*, *C. canadensis* is widely distributed across U.S. and found in a wide range of ecological conditions. Therefore, high genetic differentiation is probably related to the tree reproductive biology, as well as a combination of several factors including dispersal method and local isolations.

Similar to many other self-incompatible species [106], *C. canadensis* depends on insect, mammal, and bird-mediated pollination for dispersal. Usually honeybees, megachilid bees, small sweat bees, butterflies and beetles are the primary pollinators of *C. canadensis* in its natural distribution [14, 16, 107]. As these insects are usually capable of flying one to several kilometers, depending on the insect species and environmental conditions, pollen-based gene flow was expected to be high across fine-scale geographical ranges [32, 108, 109]. In a large continuous forest area, a majority of the pollination that occurs by insects are within neighboring trees. In contrast, with increasing distance between isolated patches, pollen source in a fragmented area can

become rather difficult to obtain, further facilitating insufficient pollen dispersal in these fragmented patches [110-113].

Based on our study results, we reject the hypothesis of limited gene flow as a result of increased distance between the examined groups of *C. canadensis*. Limited gene flow typically leads to reduced genetic diversity, increased population structure, and inbreeding within populations [31, 34, 47], which were not consistent with our findings. Other studies focused on forest trees in fragmented landscapes found that regardless of moderate to high habitat fragmentation, many insect-pollinated tree species were able to maintain high levels of gene flow across isolated patches through increased long distance pollen dispersal [34, 114-118]. Our study is consistent with these studies as is shown by the high level of gene flow present across fragmented *C. canadensis* populations. Therefore, forest fragmentation did not negatively influence the gene flow across isolated populations of *C. canadensis* in the Georgia-Tennessee border.

Although pollen dispersal by insects is a primary medium of gene flow [119], a substantial amount of *C. canadensis* seed movement depends on other small animals. The gene flow by seed dispersal depends on the dispersal mode and behavior patterns of the animals that usually forage on the fruits and eat the seeds. Heavy *C. canadensis* fruits (pods) carry hard coated seeds that usually fall in close proximity of the tree and germinate within a year or two [14, 120]. Hence, some of the seedlings grow close to the mother tree which provides an explanation for the half-sib neighborhoods in small spatial scale [103, 119, 121-123]. Another explanation of the half-sib neighbors, is that fruits can also be ingested by small rodents (e.g., gray squirrel, eastern woodrat), white tailed deer, quails, pheasants and several other bird species which are then dispersed

in scat [14, 18, 124-126]. Related individuals (seeds that originated from the same mother tree) can be carried to a nearby location by an individual animal feeding on the fruits from the same mother tree. For example, the hoarding behavior of the rodents helps related propagules to be dispersed to a close-proximity destination [124, 127]. Birds can also occasionally ingest *C. canadensis* seeds and dispose of them in new localities across longer distances, resulting in increased gene flow among the populations [14, 18, 80]. Therefore, the direction and rate of dispersal of *C. canadensis* seeds depends on the foraging behavior of the seed carrier and the ecological conditions of their habitat.

Our results revealed the presence of spatial structure among *C. canadensis* populations in this restricted geographical area. Local spatial structure in forest tree species is profoundly influenced by seed dispersal, even if pollen-dispersal is able to maintain high gene flow [118, 128]. Usually, animals choose *C. canadensis* seed pods as an occasional or emergency food source during summer and winter months when other food resources are scarce [18, 124, 129]. This indicates that the probability of seed dispersal by animals is quite low for this species [14, 126, 130]. Moreover, *C. canadensis* seed dispersal that is mostly animal-dependent can become more limiting within isolated areas, further increasing half-sibling mating, and structuring among *C. canadensis* populations [131]. Therefore, spatial genetic structure of *C. canadensis* could be influenced by the forest fragmentation despite the presence of pollen-mediated gene flow in the small geographical location. Furthermore, our study revealed significant isolation by distance in local populations of *C. canadensis*, indicating that genetic distance between *C. canadensis* populations was significantly

correlated with the geographical distance between them. Therefore, geographical distance or barriers can be one of the reasons that influence spatial genetic structure in *C. canadensis* populations.

Appendix: Tables and Figures

Table 1.1. Eighteen collection sites of *Cercis canadensis* collected from Tennessee and Georgia in the eastern United States.

Sampling State	Collection Site	Number of Individuals	Group	Longitude	Latitude
Tennessee	Anderson Co.1	10	North Group	36.04767	-84.20664
Tennessee	Knox Co.1	10	North Group	35.95661	-84.16416
Tennessee	Loudon Co.	10	North Group	35.88551	-84.32235
Tennessee	Roane Co.	10	North Group	35.97289	-84.34798
Tennessee	Anderson Co.2	10	North Group	36.03032	-84.18986
Tennessee	Anderson Co. 3	10	North Group	36.02337	-84.18487
Tennessee	Anderson Co. 4	10	North Group	35.99935	-84.21875
Tennessee	Cocke Co.	10	North Group	35.82051	-83.14966
Tennessee	Knox Co.2	10	North Group	36.04097	-83.93403
Tennessee	Knox Co.3	10	North Group	35.94648	-83.91469
Tennessee	Polk Co. 1	10	South Group	35.15252	-84.60653
Tennessee	Polk Co. 2	10	South Group	35.18276	-84.43995
Tennessee	Bradley Co.1	10	South Group	35.12912	-85.01062
Tennessee	Hamilton Co.1	10	South Group	35.08390	-85.11900
Tennessee	Hamilton Co.2	10	South Group	35.02323	-85.37941
Tennessee	Bradley Co.2	10	South Group	35.12999	-84.73290
Georgia	Catoosa Co.	10	South Group	34.90867	-85.10970
Georgia	Whitfield Co.	10	South Group	34.73177	-85.03349

Table 1.2. Genetic diversity indices across 18 collection sites of *Cercis canadensis* using 15 microsatellite loci.

Locus	GeneBank Accession No.	Forward and Reverse Primers (5'-3')	Repeat Motif	Size Range (bp)	No. of Alleles	Ar	Ho	He	H	F _{ST}	F' _{ST}	F _{IS}
1057a	GU171393	F:TCCCTCTCAGCTTTTCATATAATCCAC R:AAAGAGAGATCGTTTAGAAGGCGG	(CCATCA) ₇	116-175	11	4.57	0.82	0.84	2.00	0.04	0.05	-0.02
127spa	GU252892	F:CCAATTCAATTCCTCTGTGTGTTG R:AACGGTGTGACTAGGAGTCAAAGG	(TC) ₄	87-112	7	3.93	0.98	0.56	1.18	0.04	0.05	-0.34
168a	GU252915	F:AACAAAAGCAAAAGCACGCTACTC R:CAGTTGCCAAAATCAGAGAAATTG	(CT) ₇	151-175	6	1.57	0.04	0.47	0.83	0.59	0.61	0.80
177b	GU252919	F:AGAAATTTAGAGACCGTGAGGTG R:TAACACACTATCCGTCATTCCAG	(GA) ₆	154-183	5	3.00	0.43	0.67	1.23	0.04	0.04	0.34
199a	GU252924	F:AATAACTCCTGGAACAATGGAGGG R:TCTATGGTTTAGACCCTTTGTCACATC	(GAGA) ₈	149-173	6	2.44	0.25	0.52	1.02	0.12	0.12	0.45
220a	GU252932	F:ACCCATTCACTACCGTTCATTGAG R:GATTCCAGATTGTCACACGTTTTG	(TATT) ₄	100-123	7	3.06	0.68	0.63	1.29	0.22	0.23	-0.26
229a	GU252937	F:CTGAGGTCCGAATGGTAATTGAAC R:CGATAATACTCGATATATGCATTGCG	(GAGAG) ₄	147-170	4	1.90	0.29	0.30	0.58	0.03	0.03	0.00
53a	GU252855	F:TCCTTTGCTCATGGTAGTCTGATG R:GACTAAAGAGTTGTGTTTCATGCC	(AAAT) ₆	128-165	10	3.50	0.25	0.75	1.66	0.07	0.07	0.65
625a	GU253092	F:TTGTGGTTCTAGCCTTTGCTTTTC R:GACTAAAGAGTTGTGTTTCATGCC	(GA) ₄	95-142	5	3.03	0.21	0.59	1.03	0.18	0.19	0.62
658a	GU253101	F:TTTTTCAGAGCGTTATCACTCAACG R:CCCTAAGTAGGAGCACTCCTTTCC	(CT) ₆	97-123	5	2.58	0.23	0.52	0.97	0.10	0.11	0.52
680a	GU253111	F:AAATTTAAAGACCCCATTCGCAAC R:ACACTCCACAAAACCTTCACTTC	(GT) ₈	144-152	5	2.07	0.00	0.59	1.00	0.32	0.33	1.00
762a	GU253134	F:TCTGTCTCACCTGCTTGCACTAAG R:GGCTCAATCTCCAAGAAAATGAAG	(TC) ₇	94-114	7	2.74	0.06	0.64	1.19	0.07	0.08	0.90
780b	GU253139	F:TAGAGCCCTATTCCCACTTGACAC R:CTTTATGAATGGTTGCTTGCTGG	(AG) ₁₂	142-175	12	4.07	0.57	0.78	1.87	0.04	0.04	0.25
871a	GU253176	F:TTCTTAAGCTAAACGGTGCATTTTG R:GATGAGGGTTGGTGTAGTGAGGAG	(CTT) ₉	112-159	12	2.99	0.15	0.73	1.61	0.11	0.11	0.77
995a	GU253208	F:GTGCTTTGTCTTTGTGTTCCATTC R:AAACTACGCGTCCCTTCCTTC	(AG) ₇	109-127	4	1.76	0.00	0.50	0.89	0.42	0.43	1.00
Overall					7	2.88	0.33	0.61	1.22	0.14	0.15	0.39

Ar = allelic richness corrected for sample size, Ho = observed heterozygosity, He = Nei's genotypic diversity, H = Shannon-Wiener index, F_{ST} = populations fixation index, F'_{ST} = population's differentiation, F_{IS} = inbreeding coefficient.

Table 1.3. Genetic diversity indices of (i) 18 collection sites and (ii) 2 collection sites of *Cercis canadensis* using 15 microsatellite loci.

(i) 18 Collection sites

Collection Sites		N	MLG	H	H_e	Pa
Name	Group					
Anderson Co.1	North Group	10	10	2.30	0.49	1
Knox Co.1	North Group	7	7	1.95	0.51	0
Loudon Co.	North Group	10	10	2.30	0.44	0
Roane Co.	North Group	10	10	2.30	0.44	0
Anderson Co.2	North Group	8	8	2.08	0.53	0
Anderson Co. 3	North Group	10	10	2.30	0.46	2
Anderson Co. 4	North Group	10	10	2.30	0.41	2
Cocke Co.	North Group	10	10	2.30	0.50	1
Knox Co.2	North Group	10	10	2.30	0.51	3
Knox Co.3	North Group	9	9	2.20	0.47	0
Polk Co. 1	South Group	10	10	2.30	0.33	2
Polk Co. 2	South Group	10	10	2.30	0.46	3
Bradley Co.1	South Group	10	10	2.30	0.54	2
Hamilton Co.1	South Group	10	10	2.30	0.49	0
Hamilton Co.2	South Group	10	10	2.30	0.58	4
Bradley Co.2	South Group	10	10	2.30	0.62	0
Catoosa Co.	South Group	10	10	2.30	0.61	2
Whitfield Co.	South Group	10	10	2.30	0.58	0
Total/Average		174	174	5.12	0.61	22

(ii) 2 collection sites

Subpopulations	N	MLG	H	H_e	Pa
North Group	94	94	4.54	0.53	13
South Group	80	80	4.38	0.62	16
Total/Average	174	174	5.12	0.61	29

N = total number of samples per collection site; MLG = number of diploid individuals multi locus genotypes after clone correction; H = Shannon-Wiener index of MLG diversity; H_e = Nei's genotypic diversity corrected for sample size; Pa = number of private alleles in each collection site.

Table 1.4. Pairwise F_{ST} matrix of Nei's genetic distance of *Cercis canadensis* utilizing 18 collection sites.

Collection Sites	Anderson Co.1	Knox Co.1	Loudon Co.	Roane Co.	Anderson Co.2	Anderson Co. 3	Anderson Co. 4	Cocke Co.	Knox Co.2	Knox Co.3	Polk Co. 1	Polk Co. 2	Bradley Co.1	Hamilton Co.1	Hamilton Co.2	Bradley Co.2	Catoosa Co.
Knox Co.1	0.11																
Loudon Co.	0.13	0.07															
Roane Co.	0.09	0.05	0.11														
Anderson Co.2	0.10	0.07	0.07	0.07													
Anderson Co. 3	0.10	0.06	0.10	0.06	0.09												
Anderson Co. 4	0.08	0.10	0.15	0.05	0.10	0.06											
Cocke Co.	0.09	0.05	0.11	0.06	0.09	0.04	0.06										
Knox Co.2	0.14	0.05	0.07	0.10	0.05	0.07	0.12	0.08									
Knox Co.3	0.14	0.08	0.05	0.09	0.06	0.09	0.12	0.09	0.05								
Polk Co. 1	0.13	0.10	0.19	0.04	0.13	0.08	0.05	0.07	0.15	0.15							
Polk Co. 2	0.13	0.08	0.06	0.11	0.05	0.10	0.14	0.11	0.04	0.05	0.17						
Bradley Co.1	0.17	0.15	0.15	0.19	0.14	0.18	0.23	0.18	0.14	0.15	0.26	0.15					
Hamilton Co.1	0.24	0.21	0.21	0.26	0.19	0.25	0.30	0.24	0.20	0.22	0.33	0.20	0.05				
Hamilton Co.2	0.19	0.15	0.17	0.20	0.16	0.19	0.24	0.18	0.16	0.18	0.26	0.18	0.05	0.06			
Bradley Co.2	0.14	0.13	0.13	0.15	0.11	0.15	0.19	0.15	0.12	0.13	0.22	0.12	0.03	0.04	0.04		
Catoosa Co.	0.14	0.12	0.13	0.15	0.12	0.15	0.20	0.15	0.11	0.13	0.22	0.12	0.03	0.05	0.03	0.01	
Whitfield Co.	0.15	0.13	0.14	0.17	0.12	0.17	0.21	0.16	0.12	0.14	0.24	0.13	0.03	0.05	0.04	0.02	0.02

Table 1.5. Analysis of molecular variance of *Cercis canadensis* across 15 microsatellite loci for (i) 18 collection sites as one hierarchical cluster, ii) 18 collection sites into two groups according to two clusters of STRUCTURE; (iii) 18 collection sites as two groups (north and south group)

(i) 18 collection sites as one group

Source of Variations	df	Sum Square	Variance	% Variation	P - value
Among collection sites	17	606.63	1.62	17.28	0.001
Among individuals within collection sites	118	1329.54	3.49	37.13	0.001
Within individuals	136	583.00	4.29	45.60	0.001
Total	271	2519.17	9.40	100.00	

$F_{IS} = 0.37$, $F_{ST} = 0.17$, $F_{IT} = 0.54$

(ii) 2 clusters (STRUCTURE)

Source of Variations	df	Sum Square	Variance	% Variation	P - value
Among clusters	1	305.78	2.13	20.46	0.001
Among collection sites within clusters	16	300.85	0.51	4.86	0.001
Among individuals within collection sites	118	1329.54	3.49	33.52	0.001
Within individuals	136	583.00	4.29	41.16	0.001
Total	271	2519.17	10.41	100.00	

$F_{CT} = 0.21$, $F_{SC} = 0.06$, $F_{IS} = 0.17$, $F_{IT} = 0.59$

(iii) 2 groups (north and south)

Source of Variations	df	Sum Square	Variance	% Variation	P - value
Among south and north groups	1	215.62	1.41	13.97	0.001
Among collection sites within two groups	16	391.01	0.88	8.74	0.001
Among individuals within collection sites	118	1329.54	3.49	34.68	0.001
Within individuals	136	583.00	4.29	42.60	0.001
Total	271	2519.17	10.06	100.00	

$F_{CT} = 0.14$, $F_{SC} = 0.10$, $F_{IT} = 0.57$

* F_{ST} = variance among collection sites relative to the total variance

* F_{IS} = inbreeding coefficient of individuals relative to population

* F_{SC} = variance among collection sites within groups

* F_{CT} = variance among groups relative to the total variance

* F_{IT} = variance in the total populations

Table 1.6. Bottleneck determination by sign tests for *Cercis canadensis* samples utilizing 15 microsatellite loci

Mutation model (excess/deficit) ^a					
Cluster	I.A.M	T.P.M	S.M.M	Mode-shift^b	P value
Cluster 1	10/5	4/11	0/15	Not shifted	NS
Cluster 2	13/3	9/6	4/11	Not shifted	NS

I.A.M. = infinite allele model; T.P.M. = two-phase mutation model; S.M.M. = stepwise mutation model; a = Excess/deficit indicates the number of loci showing excess/deficit of gene diversity under mutation-drift in populations that experienced a recent bottleneck; b = A shift in the distribution of allelic frequency classes is expected in populations that experienced a recent bottleneck; NS = not significant



Figure 1.1. Map of location sites for 18 collection sites of *Cercis canadensis*. Red dots represent sampling sites separated by north group (east Tennessee) and south group (Georgia-Tennessee border).

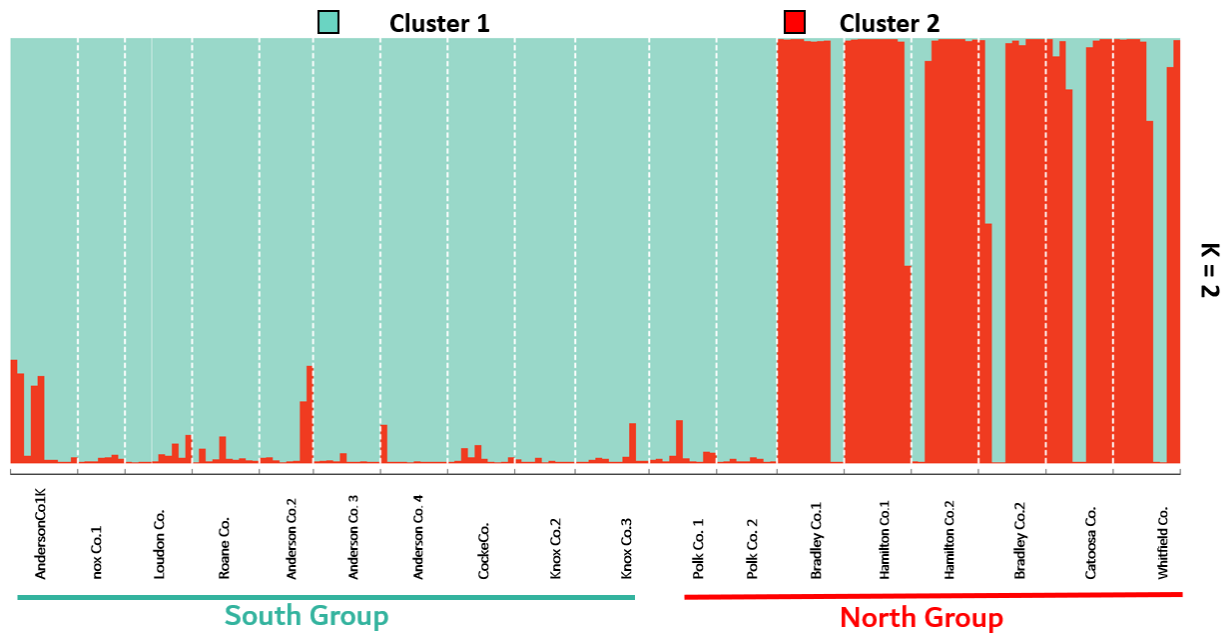
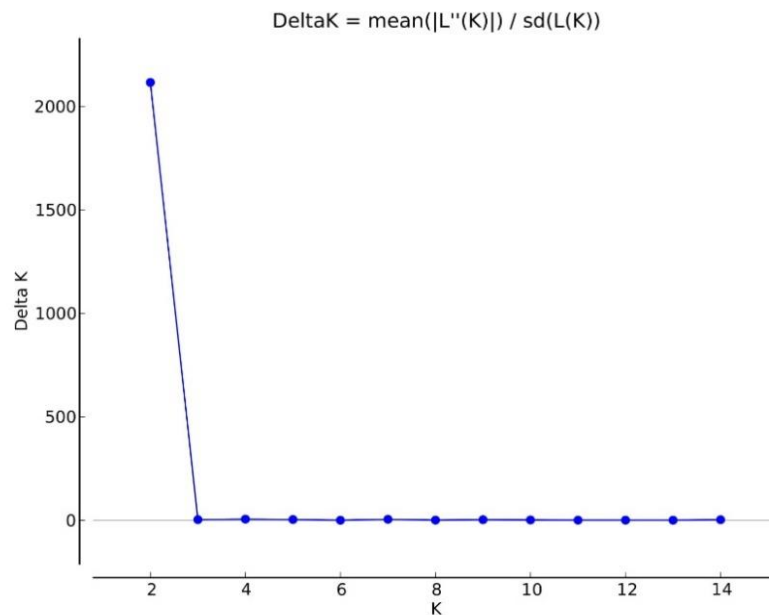


Figure 1.2. STRUCTURE bar graph representing two genetic clusters ($\Delta K = 2$) among 18 collection sites of *Cercis canadensis*. Each vertical bar represents an individual sample and color of the bar indicates the assignment probability of that individual to belong to one of the two identified clusters, designed in two different colors, blue or red.

(A)



(B)

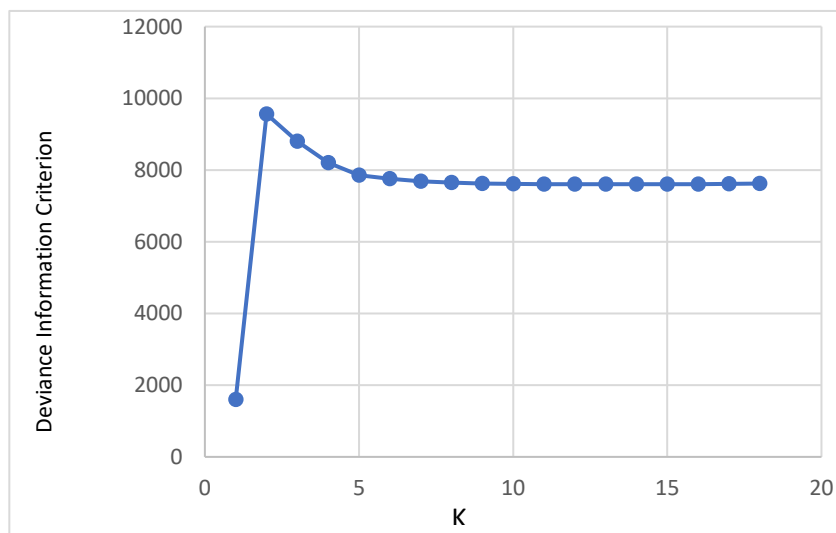


Figure 1.3. Bayesian clustering probabilities using programs STRUCTURE and InStruct. STRUCTURE results were obtained using Delta K and graph results indicated two genetic clusters among 18 *Cercis canadensis* collection sites (A). InStruct results were obtained using deviance information criterion calculations (B). Results indicated the highest probability of two different genetic clusters ($K = 2$) across 18 *Cercis canadensis* collection sites.

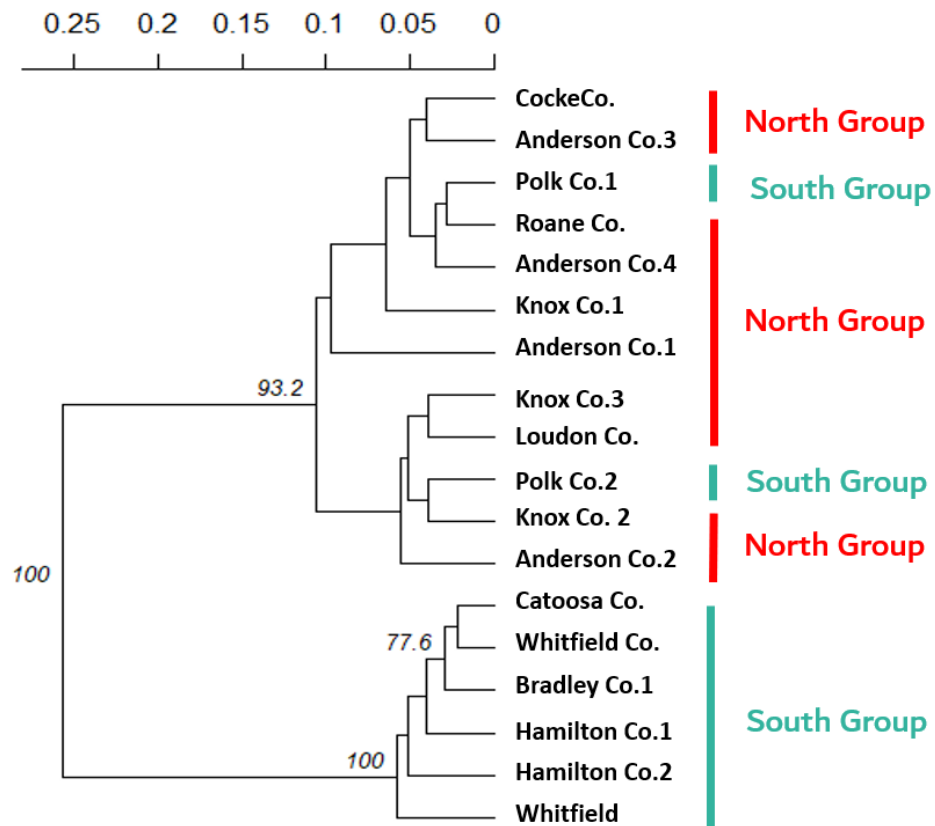


Figure 1.4. Neighbor-joining tree of 18 collection sites of *Cercis canadensis* was constructed and visualized using Nei's genetic distance. Numbers indicate percentage bootstrap support values using 1,000 replications.

(A)

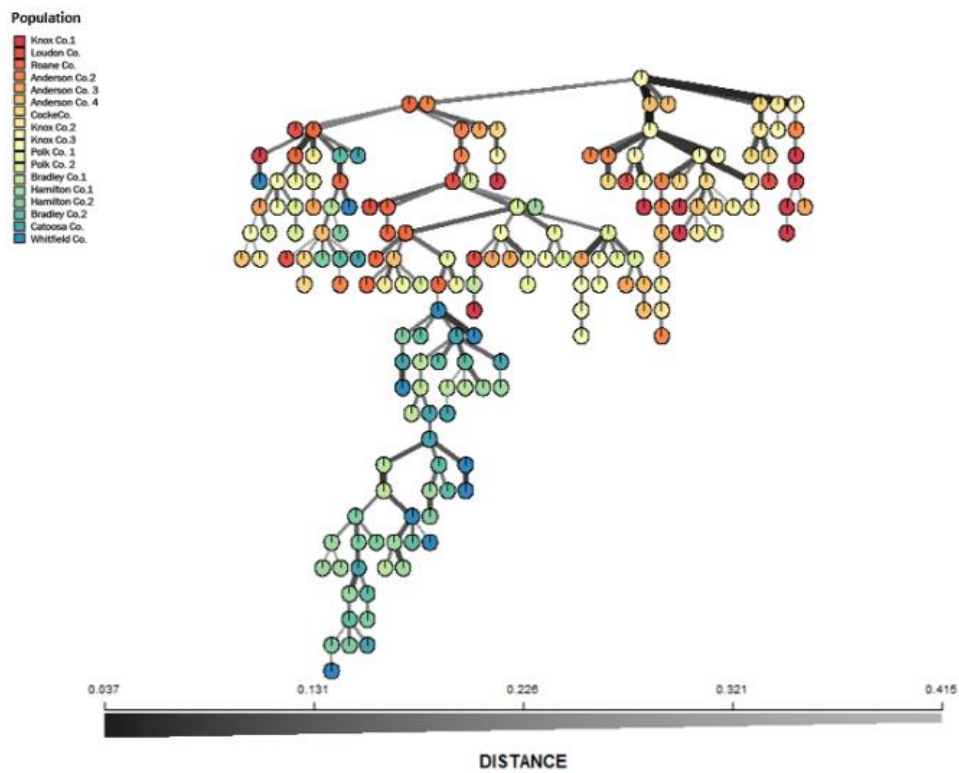


Figure 1.5 (A). The Minimum Spanning Networks (MSN) of *Cercis canadensis* genotypes were calculated using Bruvo's genetic distance (considering microsatellite motif length). Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.

(B)

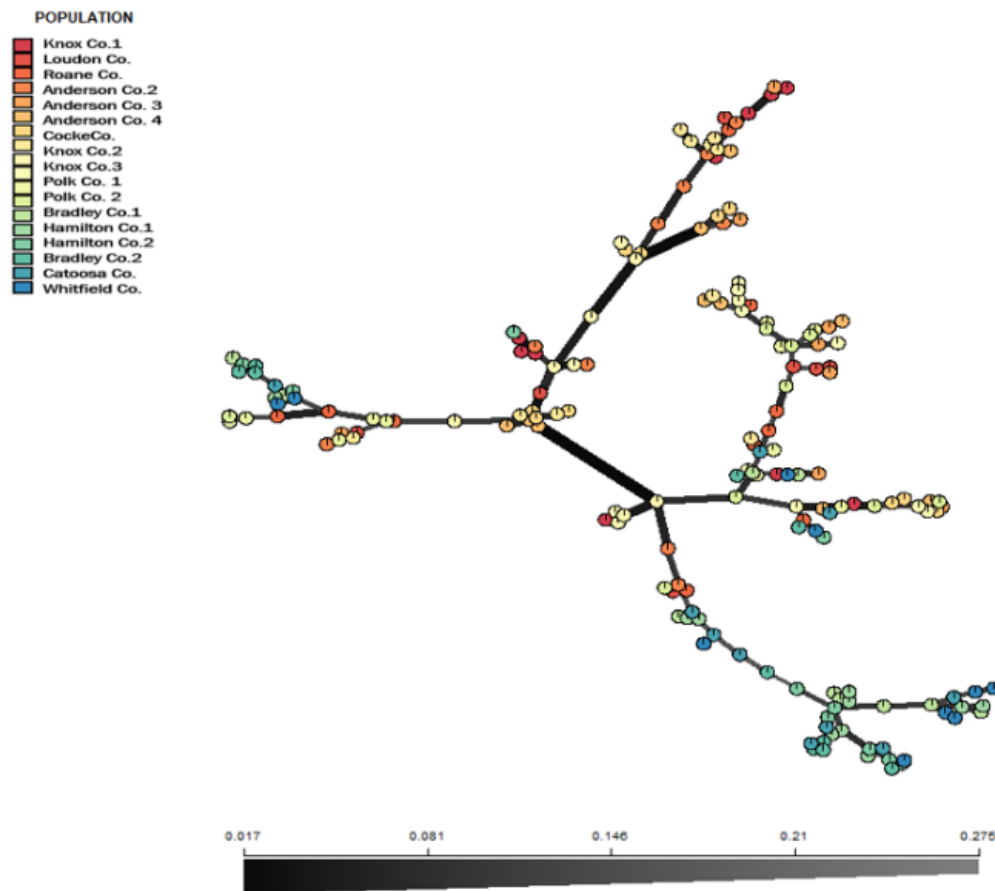
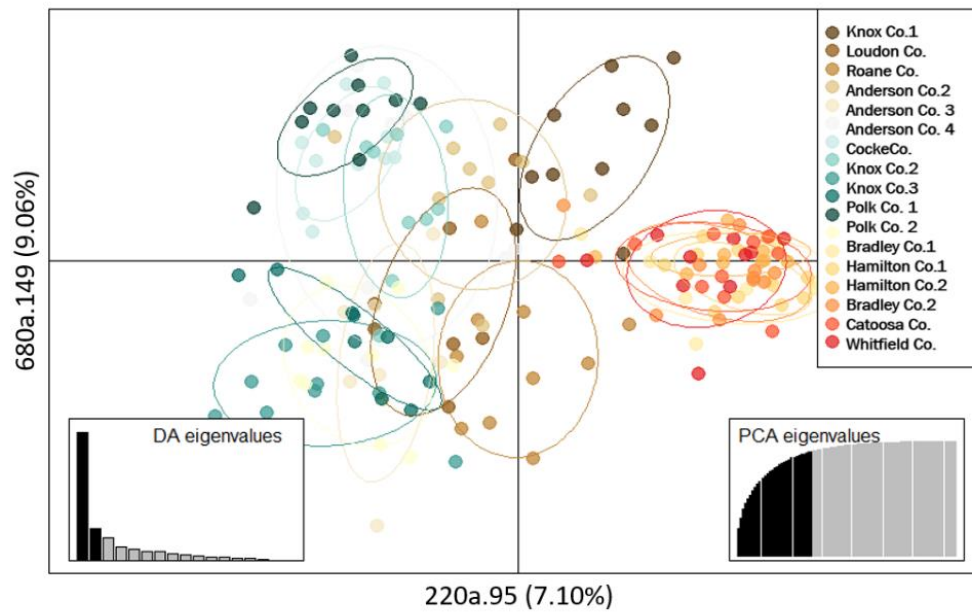


Figure 1.5 (B). The Minimum Spanning Networks (MSN) of *Cercis canadensis* genotypes were calculated using Nei's Genetic distance. Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.

(A)



(B)

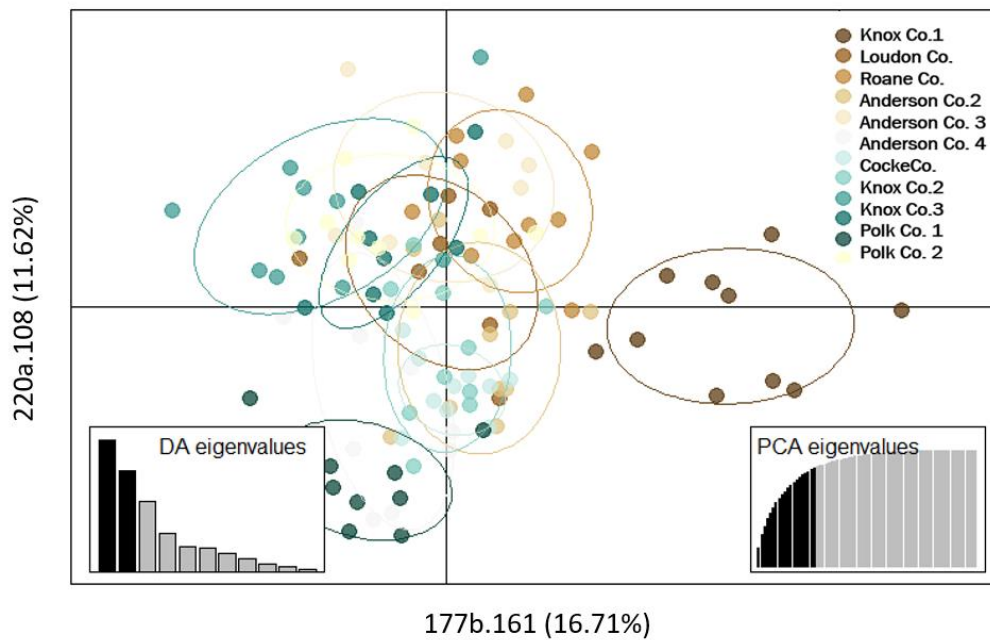


Figure 1.6. Discriminant Analysis of Principal Components (DAPC) plot of *Cercis canadensis* individuals with 18 collection sites (A) and 12 collection sites (B). The alleles with respective locus explaining the most variance is indicated at plot axes, with percentage contributions in parenthesis.

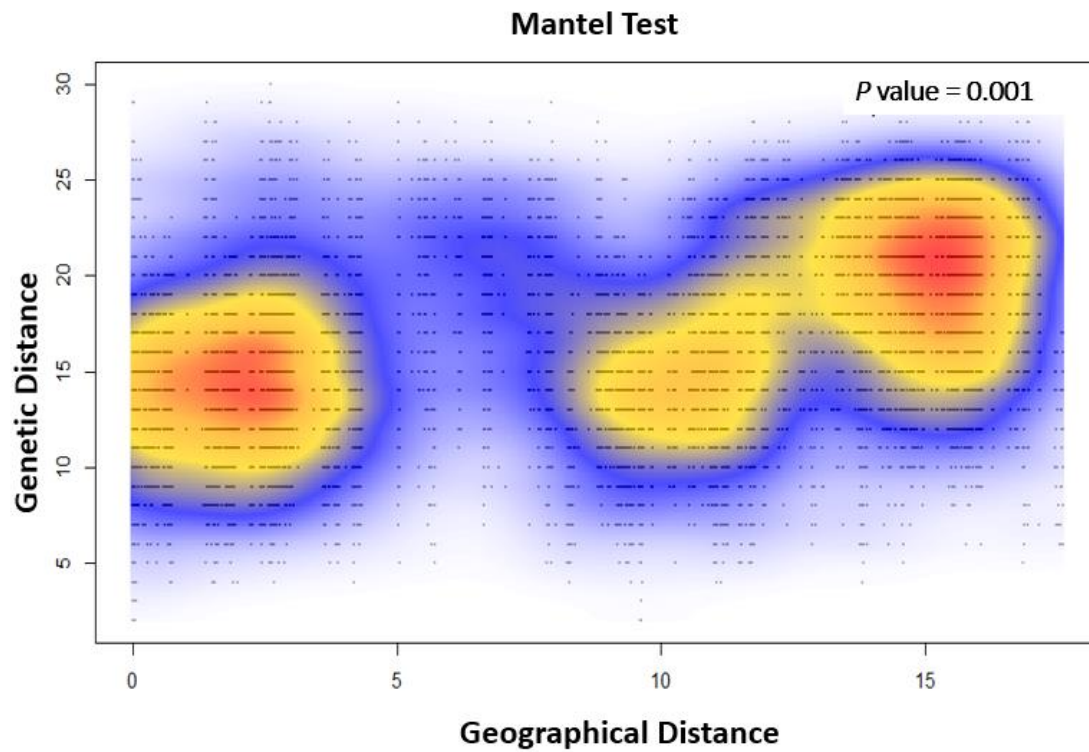


Figure 1.7. Mantel test of *Cercis canadensis* individuals across 18 collection sites in east Tennessee and around Georgia-Tennessee border. This test represents correlation between geographical and genetic distance.

CHAPTER 2

GENETIC DIVERSITY AND SPATIAL DISTRIBUTION OF *CERCIS*

***CANADENSIS* IN THE UNITED STATES**

Abstract

Cercis canadensis L. (Eastern Redbud) is a small to medium, understory tree native to 24 states in the eastern United States (U.S.) and Mexico. It is among the earliest deciduous trees that flower in the spring, has lush-green heart-shaped leaves and a wide range of fall foliage colors. Eastern Redbud has become a very popular landscape tree, contributing to more than \$27M annually in sales of native trees and cultivars. Because of the wide range of morphological variation, we hypothesized that high genetic diversity among the wild collections of *C. canadensis* would be evident across its native geographical distribution. We utilized 12 previously developed microsatellite loci in genotyping 691 wild individuals from 74 collection sites across 23 states to determine the population structure and genetic diversity across the Eastern U.S. Our results indicated a presence of population structure, high genetic diversity ($H_e = 0.67$), and high genetic differentiation ($F_{ST} = 0.19$) among the collection sites. Two major genetic clusters were inferred by STRUCTURE in our collection. The majority of variation was retained within collection sites (69.15%, $P < 0.001$). Only 16.97% ($P < 0.001$) of variation was retained between collection sites, and 13.87% of the variation ($P < 0.001$) among the two major population clusters identified by STRUCTURE. Our data suggested that high levels of genetic diversity and spatial structure of our *C. canadensis* collection could have important implications for habitat management efforts, and future breeding programs to improve desirable horticultural traits.

Introduction

Cercis canadensis L. (Fabaceae) or Eastern redbud is a deciduous tree native to regions of the mid-west to eastern United States (U.S.), as well as northeastern Mexico [1, 2]. Considered a relatively small ornamental tree, *C. canadensis* is characterized by its wide, colorful, umbrella-like crown [3]. It is considered a popular landscape tree due to its heart-shaped foliage, compact form, and early spring flowers [5]. It contributes to more than \$27M annually in sales of nursery stocks with more than three dozen commercially available cultivars [7, 8].

Early in the spring, either before or simultaneously with expansion of leaves, small, showy pea-like flowers emerge from old wood on branches. The petal color ranges from bright lavender to pink or white [132]. Bisexual flowers are present in early March to April in south-eastern states and late April to May in northern states, and blooming period lasts for about three weeks [14]. Increases in seasonal average minimum temperatures as a result of global warming have advanced the first flowering period of *C. canadensis* by about 9 days from years 1970 to 1999 [133]. Fruits and seeds of *C. canadensis* are consumed by several bird species such as quail, pheasants and goldfinch. Deer feed on the leaves and barks of the tree. Several small mammals such as grey squirrels and woodrats feed on the fruits and seed, and cache the seeds [14, 18, 124, 126]. *Osmia lignaria* Say (Hymenoptera: Megachilidae), a mason bee, depends on *C. canadensis* for early season pollen collections [16].

The taxonomy of genus *Cercis* was originally constructed on the basis of morphological and geographical data [9]. Because morphological characters are highly quantitative

and continuous in characters [13], it is unreliable to use only morphological information to distinguish among taxa and to construct a phylogenetic tree of the genus [6].

Therefore, morphological data coupled with molecular information have been used to more accurately define the taxonomy and to determine phylogenetic relationships among the *Cercis* taxa [6]. Although a wide range of morphological characters are present in the wild populations of *C. canadensis*, we have limited knowledge of the genetic diversity and spatial distribution of this species across its native range in the eastern U.S.

To address that knowledge gap, this study evaluated the genetic diversity among the wild populations of *C. canadensis* in its native range. Information gained here is expected to provide better understanding of the diversity patterns in the existing wild populations of *C. canadensis*, as well as help breeders improve the desired traits of selected and diverse cultivars. Using previously developed microsatellite loci [6], our research objectives were: 1) to evaluate the genetic diversity of the wild populations of *C. canadensis* within its native range of the U.S.; and 2) to infer patterns in the spatial distribution of *C. canadensis*. We hypothesized that *C. canadensis* wild populations will be genetically diverse and spatially clustered across its native range.

Materials and Methods

Sample Collection

Samples of *C. canadensis* were collected from 23 states in the mid-west and eastern U.S., which is the native range of this species (Fig. 2.1). In a collection site, at least 10 wild, open-pollinated trees within a mile radius of each other were selected. From each

tree, five to seven young and disease-free leaves were collected randomly and stored in a paper envelope with several pieces of paper in between to dry out the leaf tissue. For each tree, geographical coordinates were recorded. A total of 1193 individual tree samples were collected, representing trees in 117 collection sites in 23 states. To avoid the over representation of a geographical area, we randomly selected a subset of collection sites from areas with more than one collection site. This study incorporated a total of 790 individuals from 79 collection sites.

DNA Extraction

From each sampled tree, 60-100 mg of leaf tissue was used to isolate DNA. Samples were homogenized four times for 30 s each using a Beadmill 24 homogenizer (Fisher Scientific, Pittsburgh, Pennsylvania, U.S.). Samples were kept in liquid nitrogen for 5 min between each pulverization step. Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, California, U.S.) was used to isolate genomic DNA (gDNA) from the pulverized samples with the following minor modifications in the manufacturer's provided protocol. First, 2% w/v polyvinylpyrrolidone (PVP) was mixed into the lysis buffer (AP1). Then 8 µl of RNase was added into each sample tube, and the 65°C incubation step in water bath was increased to 45 min from 15 min. Every two minutes, each sample tube was inverted gently to mix the sample well and ensure that the cells were lysed. Lastly, the incubation step at -20°C was increased to one hour minimum from 20 min. Ethanol was used to wash the spin columns if there was any debris left from the sample tissue, and then the elution buffer was added. Elution buffer (AE) of 50 µl was pre-heated to 65°C before being added to the spin columns. The addition of elution buffer (AE) step was repeated twice instead of once. The concentrations of isolated gDNA were quantified

using ND1000 Ultraviolet-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, U.S.), and stored at -20°C.

Microsatellite Primers and Genotyping Conditions

Initially, gDNA was isolated from five *C. canadensis* individuals from the University of Tennessee Garden (Knoxville, TN, U.S.) and used to screen 68 microsatellite loci [6]. Twelve polymorphic microsatellite loci (Table 2.2) were selected for this particular population study based on successful amplification rate of the microsatellite loci.

Microsatellite loci regions were amplified using polymerase chain reaction (PCR) with a 10 µl reaction mixture containing the following: 1 µl undiluted gDNA, 1 µl of 10 µM of each forward and reverse primer, 0.5 µl of dimethyl sulfoxide, 4 µl of GoTaq G2 Hot Start Master Mix (Promega Corp, Madison, Wisconsin U.S.), and 2.5 µl water. To assure the validity of the produced data, both a negative control (reaction mixture with water instead of any DNA sample) and a positive control (a DNA sample from the initial screens that amplified with all microsatellite primers) were incorporated for every primer-pair tested. Amplification of DNA with twelve microsatellite loci across all samples was completed in 96 well plates using an Eppendorf Thermocycler (Eppendorf AG, Hamburg, Germany) plate format with the following thermal profile: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and an extension at 72°C for 30 seconds, with a final extension of 72°C for 4 minutes.

Amplified PCR products were visualized using QIAxcel Capillary Electrophoresis System (Qiagen) and analyzed with a 15/600 bp internal alignment marker and a 25 to

500 bp DNA size marker. All 790 *C. canadensis* gDNA samples were amplified and visualized against each of the twelve microsatellite loci using the procedure described above. Failed reactions were rerun once again before they were considered missing data. Samples with 40% missing data were discarded. Also, collection sites with more than four samples having 40% missing data were excluded from the dataset. As a result, five from the 79 collection sites were excluded and 49 individuals from the remaining 74 sites were discarded, resulting in 691 individuals from 74 collection to be included in the data analyses (Table 2.1).

Genetic Diversity

Using the Excel macro FLEXBIN [51], raw alleles were binned into statistically similar allelic classes. This binned genetic dataset was used for all the following statistical analyses. Clone-correction process was implemented to the dataset to identify possible clonal samples using POPPR version 2.8.2 [52, 134] in R version 3.5.3 (R Core Team, 2019). For each collection site, only multi-locus genotypes (MLG) were used to obtain unbiased estimates of allelic frequency from the dataset [53].

For 12 microsatellite loci and 74 collection sites, genetic diversity indices were calculated using package POPPR in R. POPPR was used to calculate the estimated total number of alleles per locus, allelic richness (A_r), population's long-term evolutionary potential to adapt and be persistent, observed heterozygosity (H_o), and expected heterozygosity (H_e). A_r was estimated using package HIERFSTAT version 0.04-22 [135] in R. Genetic fixation index (F_{ST}), inbreeding coefficient (F_{IS}), and allelic differentiation (F'_{ST}) [56, 136] were calculated using POPPR and HIERFSTAT packages

in R. Furthermore, the Shannon Weiner diversity index (H) was calculated for each collection sites using package POPPR in R.

Population Structure

Analysis of Molecular Variance (AMOVA) [70] was estimated using the R package POPPR with 10,000 permutations by sorting the individuals into hierarchical groups to assess the amount of molecular variance partitioned within and among the collection sites. The levels of population hierarchy included 74 collection sites as one hierarchical group and two groups on the basis of STRUCTURE analysis.

To investigate the presence of population structures in the native range of wild *C. canadensis* populations, the dataset was analyzed using the program STRUCTURE version 2.3.4 [58] with admixture model. A Monte Carlo Markov Chain (MCMC) approach, Bayesian clustering method was used with the following parameters: 500,000 burn-in period with 1,000,000 MCMC repetitions for 30 independent chains for K value from 1 to 18. The result output was visualized with STRUCTURE HARVESTER web version 06.94 [59]. The optimum K value, indicator of population clusters present in the dataset, was calculated utilizing Evanno method [60]. The estimation of ΔK criterion obtained from STRUCTURE HARVESTER were visualized using POPHELPER 2.2.6 [61].

The Evanno method implemented by STRUCTURE HARVESTER has a limitation in the estimation of a ΔK value lower than two [60]. Therefore, in order to verify the accuracy of the STRUCTURE result, we utilized InStruct version 1.0 [62], another Bayesian clustering method that unlike STRUCTURE is able to calculate ΔK value of 1 [62, 63].

This particular program takes into account the possibility of the collected dataset being a clonal population ($K = 1$) and overlooks the assumption of Hardy-Weinberg equilibrium within the studied populations. The following parameters were utilized in this analysis: for each K value (1 to 10), 20 independent MCMC chains with 100,000 burn-in followed by 500,000 MCMC repetitions. Additionally, admixture model was also incorporated to reveal the proportion of mixed ancestry present in an individual as well as to achieve improved clustering [64]. To obtain the optimum value of K , the Deviance Information Criterion (DIC) values resulted from InStruct program were calculated [63].

Additionally, we utilized several model-free clustering methods to investigate the population structure of *C. canadensis* samples. Both Nei's [65] and Bruvo's genetic distances (considers repeat motif length of loci in determining genetic distance [66, 67]) among the different collection sites were visualized by producing dendrograms with the Minimum Spanning Networks (MSN) function implemented in the package POPPR in R. Discriminant Analysis of Principal Components (DAPC) as implemented in the R package ADEGENET version 2.1.1 [69] was also employed to investigate the underlying genetic structure of this species in its wide geographical range. This is a two-step multivariate analysis that investigates the genetic variations within populations without considering any prior clustering probability among the sampled collection sites [68]. At first, a principal component analysis (PCA) was conducted and then the number of PCA vectors (to explain majority of variance with minimizing over-fit of the DAPC) was selected. Then, a selected number of PCAs were used to investigate differences between groups while minimizing within group variations as well as ordination of collection sites into different groups using discriminant analysis [68, 137].

Isolation by distance (IBD) was estimated using the Mantel test [71, 72] with 10,000 permutations in R package MASS version 7.3-51.1 [73]. IBD reveals if there is a correlation between genetic distance and geographical distance among the individuals in studied dataset. The Mantel test was implemented across 74 collection sites while considering the whole dataset as one population.

Demography

To investigate the occurrence of a recent bottleneck in *C. canadensis* populations, the program BOTTLENECK version 1.2.02 [74] was utilized. For this purpose, the dataset was screened with the following three mutation models: infinite allele model (I.A.M.), stepwise-mutation model (S.M.M.), and two-phase mutational model (T.P.M.).

Significance of this bottleneck test in these three models was estimated by the Wilcoxon sign-ranked test with 10,000 simulations [74, 75].

To investigate and interpret the population history of *C. canadensis*, the genotypic data was analyzed using the program DIYABC version 2.1 [138, 139] that implements approximate Bayesian Computation statistical methods. For this analysis, collected individuals were divided into three major groups on the basis of geographic origin of the collected individuals. This dataset grouping was also supported by the DAPC analysis output. The DIYABC program sorts populations with shared geographical histories into similar groupings to simplify and interpret the evaluated scenarios and therefore, has been implemented into many studies [139-142].

First, we tested the dataset for 35 possible scenarios and for each scenario analysis, program considered of different population histories such as divergence, admixture,

effective population size, and presence of unsampled original population (ghost population). For each scenario, over 600,000 pseudo-observed datasets (PODs) were simulated from the dataset under the assumed prior parameters ranges and their uniform distribution for all scenarios. Also, we presumed a generalized stepwise mutation model with setting a mutation rate of 5×10^{-4} (ranged from 10^{-6} to 10^{-3} mutations in each generation per locus). The DIYABC program calculates summary statistics for each dataset - number of alleles, average genetic diversity, average size variance of a population, pairwise F_{ST} , and $(d\mu)^2$ (distance between pairs of populations).

The summary statistics of the simulated dataset were compared with the estimated statistics of the observed dataset. In the program, the ‘pre-evaluate scenario-prior combinations by Principal Component Analysis (PCA)’ step was selected to check if in at least one scenario, simulated dataset produced from the prior dataset were close enough to the observed dataset. Then, logistic regression analysis was implemented on 1% of the simulated dataset that was closely related to the observed dataset. Posterior probability of the compared scenarios were estimated through regression analysis to select the scenario with the best posterior probability and the subsequent parameters [138]. Using “confidence in scenario choice” option in the program option, goodness of fit for each considered scenario was tested by evaluating type I and type II errors based on 1,000 pseudo-observed data sets [138]. Small value of type II error can render good confidence to a particular scenario although type I error is relatively high.

Results

Microsatellite Genetic Diversity and Hierarchical Fixation Indices

Twelve microsatellite loci were amplified in 790 *C. canadensis* samples in this study. After discarding the individuals that failed across 40% of the loci and had clonality present (two clonal samples were detected overall), 74 collection sites with 689 unique multilocus genotypes remained for further data analyses (Table 2.1). The number of alleles per locus ranged from 6 to 13 with a mean of about 10 alleles per locus (Table 2.2). Overall allelic richness (A_r) ranged from 1.23 for locus 995a to 1.79 for locus 780b, with a mean of 1.55 implying a presence of high allelic richness in wild *C. canadensis* individuals. Observed heterozygosity (H_o) across all loci was 0.32, ranging from 0.01 (995a) to 0.99 (127spa). The overall expected heterozygosity (H_e) across all 12 microsatellite loci was high ($H_e = 0.69$), ranging from 0.52 (locus 995a) to 0.86 (locus 780b).

Furthermore, overall Shannon-Weiner diversity index (H) for the 12 loci was 1.42 and ranged from 0.84 (locus 995a) to 2.19 (locus 780b) (Table 2.2). Additionally, high population fixation ($F_{ST} = 0.19$; ranging from 0.05 to 0.53; Table 2.2) and population differentiation ($F'_{ST} = 0.19$; ranging from 0.05 to 0.54) were identified among *C. canadensis* individuals. Also, we estimated inbreeding coefficient (F_{IS}) across all tested loci to be 0.43, indicating presence of homozygote excess (Table 2.2) among studied *C. canadensis* populations. The average estimated gene flow was 0.75 that suggests there has been a limited amount of gene flow present among the studied populations (Table 2.2).

A total of 19 private alleles were detected across 74 collection sites (Table 2.1). The Smith collection site from Texas had three private alleles and interestingly all three private alleles belonged to the locus 658a (Table 2.1). In total, four private alleles were identified from Texas, which was the highest number across all collection sites (Fig. 2.2). Also, 9 of the 12 microsatellite loci had private alleles with the highest number of private alleles from locus 168a ($P_a = 4$, Table 2.2). Nei's genetic diversity index (H_e) values in the studied 74 collection sites was 0.67, ranging from 0.32 (Anderson, TN) to 0.68 (Sharpy, NE) (Table 2.1).

Population Structure

Using Nei's genetic distance, we estimated pairwise distances between the 74 collection sites and the values ranged from 0.02 to 0.36. The Analysis of Molecular Variance (AMOVA) was completed by arranging the dataset into the following subsets: i) 74 collection sites as one hierarchical group; and ii) 74 collection sites into two major groups on the basis of the STRUCTURE results (Table 2.3). The first data arrangement showed that most of the genetic variation was present among individuals (74.24%, $P < 0.001$). A significant amount of variation was also partitioned in between collection site levels (25.76%, $P < 0.001$). When the dataset was divided into two clusters based on STRUCTURE results, 69.15% ($P < 0.001$) of the variation was partitioned among the individuals. There was also a significant amount of variation present between the two clusters (13.87%, $P < 0.001$) and between collection sites within clusters (16.97%, $P < 0.001$). The majority of the variation partitioned, belonged to individuals, as was expected because *C. canadensis* is an outcrossing species. Nevertheless, the significant amount of variation within collection sites and between clusters discerned the

presence of genetic structure in the studied dataset. Therefore, AMOVA results were congruent with the hierarchical fixation indices and indicated the presence of population structure.

STRUCTURE results revealed an optimum $\Delta K = 2$ (Fig 2.3), implying that across its wide native range, *C. canadensis* collection sites are divided into two major clusters (Fig. 2.3). Collection sites in the north region (Ohio to Nebraska) and mid-south to mid-north (from Texas to Nebraska) were part of the first cluster (Fig. 2.4). The rest of the collection sites from the northeast to all the way to mid-south (Mississippi) along the coastline, belonged to the second cluster. Also, 9.70% of individuals in the total dataset showed mixed ancestry from both clusters ($> 80\%$ proportion to one cluster as threshold for pure ancestry, Fig. 2.3). In contrast, analysis with InStruct (Deviance Information Criterion-based) indicated $\Delta K = 1$ which means the native *C. canadensis* populations were part of one single continuous cluster (Fig 2.5 (B)). This InStruct output was in contrast with the STRUCTURE results. Therefore, to identify the cluster number further analysis of the dataset was conducted with several alternative model free clustering methods.

First, a Neighbor-Joining dendrogram was constructed using pairwise F_{ST} values. The results revealed a presence of two major clades, which supported the STRUCTURE findings (Fig. 2.6). In addition, the collection site distribution in these two clades (NJ dendrogram) matched the distribution of collection sites in the two STRUCTURE-based clusters. Both analyses showed that the majority of collection sites (except two collection sites from Georgia) grouped in clusters based on their geographical origin.

The DAPC analysis plot was constructed without any prior group designation, further confirming the presence of genetic structure in the studied dataset. Collection sites of this study were found in two distinctly separate geographical groups (North-east and south-west group) in the DAPC plot, which corresponded to the geographical distance between them (Fig 2.7). The remaining collection sites from south to north, in the eastern U.S., gradually overlapped with each other. Therefore, the DAPC plot indicated the presence of two clusters separated by the first axis and a transition zone between these two clusters. This results further supported the STRUCTURE result of two possible geographical clusters.

The Minimum Spanning Network (MSN) with reticulations of the whole dataset congruently showed the presence of the genetic structure when generated with both Nei's genetic distance (Fig. 2.8 (A)) and Bruvo's genetic distance (Fig. 2.8 (B)). Specifically, with the Bruvo's genetic distance, a small isolated cluster was present in the middle, which may indicate the ancestral population of *C. canadensis*. Therefore, collectively almost all the independent population structure analyses (except the InStruct program) indicated the presence of geographical population structures in the studied *C. canadensis* populations in its wide geographical range.

The result of the isolation-by-distance analysis indicated that the geographical distance was weakly but linearly correlated ($r = 0.001$, $P < 0.001$) with genetic distance in the studied dataset (Fig. 2.9). As the dataset originated from a wide geographical range, a weak relationship between geographical and genetic distance was expected. When *C. canadensis* populations were studied at fine scale level, a stronger correlation ($r = 0.32$, $P < 0.001$) was revealed between these two parameters (Chapter 1).

Population Demography

The program BOTTLENECK used the S.M.M, I.A.M, and T.P.M. mutation models and was utilized to analyze our dataset that was partitioned into two major clusters according to the STRUCTURE results (Table 2.4). Sign test revealed no significant excess of heterozygosity and normal distribution of allelic frequencies in both clusters, further discarding any possibility of recent bottleneck occurrence in the native populations of *C. canadensis*.

DIYABC program with the Approximate Bayesian computation approach provided further support for presence of population structure in *C. canadensis* wild populations. The reconstruction of evolutionary scenarios implied that the original population (ancestral population) split into three current-state population structures. The top three most probable scenarios supported the presence of an unsampled ancestral population in the dataset. The most probable scenarios also indicated the occurrence of a bottleneck before the rise of the current populations (Fig. 2.10), following a typical population expansion expectation.

Based on DIYABC results, another probable scenario with marginally lower relative probability indicated absence of any bottleneck occurrence in the evolutionary history of these populations, with a similar triple split of the (unsampled) ancestral population. Estimated posterior parameters suggested that the split of that originating population occurred around 300 generations ago (ranging from 60 to 1,100 in simulated datasets) and the estimated mutation rate per generation was 10^{-4} . According to the bottleneck scenario, a population bottleneck occurred approximately (with the highest relative

probability) 3,000 generations ago (ranged from 130 to 9,000 generations) which is approximately 18,000 years ago. Also, the effective population size was estimated to be around 7,000 individuals, ranging from 5,000 to 8,000 individuals. Post-hoc analyses provided goodness-of-fit for this scenario, with the original dataset well embedded in the prior PODs population and nested in with the posterior PODs population.

Discussion

In this study, we determined the genetic diversity and spatial distribution of *C. canadensis*, the eastern U.S. native and an economically important landscape tree species across a wide geographical distribution. Our results revealed high levels of genetic diversity, moderate levels of allelic richness, high levels of population differentiation, population structure, and limited gene flow in the wild populations of *C. canadensis* across its native range. Also, we detected the presence of geographical clusters, longitudinally, in the south-eastern U.S. coastal plain and north-western U.S. region.

Many temperate as well as tropical plant and tree species sustain high genetic diversity and allelic richness across a wide geographical range, even in the presence of environmental stressors, pest and pathogen pressures, and human disturbance [143-149]. Eastern hemlock (*Tsuga canadensis* [L.] Carr.), for example, is a widely distributed North American tree species that has become infested by the invasive insect pest, Hemlock wooly adelgid, (*Adelges tsugae* Annand) (Hemiptera: Sternorrhyncha) across large portions of its native range since the introduction of the pest in 1951 in eastern U.S. [145, 150]. Despite high pest pressure, eastern hemlock continues to

present high genetic variation and allelic richness in its wide geographic range [145, 151]. *Fraxinus angustifolia* Vahl is a wind-pollinated tree species that is distributed across southern and eastern Europe that also maintains high genetic variation [152]. A genetic diversity study of five Asian *Cercis* spp. used microsatellite loci to reveal presence of four genetic population clusters that presented an average of 5.7 alleles per locus [7]. Similarly, we found that native *C. canadensis* maintained high genetic diversity and allelic richness, congruent with these studies and also consistent with observations among several hardwood tree species [40, 77-80, 99]. The ability of *C. canadensis* to maintain high genetic diversity can be influenced by several factors including wide and continuous geographic distribution, outcrossing reproduction system, and effective population size [42, 77, 97, 153].

Cercis canadensis populations, in addition to being highly genetically variable, also display a wide range of morphological variation across a diverse set of environmental conditions [6, 50, 106, 154, 155]. For example, leaf shape, size, surface pubescence, and other structural features are strongly related to environmental factors such as temperature and moisture content [155-158]. These characteristics are thought to originate through the process of local adaptation of *Cercis* spp. to different climatic pressures [6, 10, 159]. Also, when planted in the garden, the individuals from different climate regions revealed morphological characteristics related to the zones they belong to and thus these characteristics are genetically based [10, 160]. Our dataset further support this as we have found high levels of genetic diversity across the populations collected from wide native regions.

The basis of morphological variation observed in this species has resulted in taxonomic differentiation of *C. canadensis* species into three subspecies or varieties:

C. canadensis var. *canadensis*, *C. canadensis* var. *mexicana* (Rose) M. Hopkins (Mexican redbud), and *C. canadensis* var. *texensis* (S. Watson) M. Hopkins (Texas redbud) [6, 13, 161]. The first variety is found in mesophytic vegetation of the northeastern region of the U.S. and is naturally distributed from Connecticut to Nebraska in the North, and from Florida to Sierra Madre in the western U.S. [13]. *Cercis canadensis* var. *mexicana* and *C. canadensis* var. *texensis*, native in the southwestern part of Texas and northern to central Mexico [13], are found in semi-arid regions as well as more xeric conditions [12]. However, a highly continuous pattern of morphological variation in *C. canadensis* populations was documented across its wide native range, which questions the validity of these classifications [13, 162]. Also, several phylogenetic studies were unable to find high levels of support for the presence of these three varieties within the species level [155, 162].

A large number of studies documented that tree species with wide distributions and large population sizes usually have low genetic differentiation with weak population structure across their geographic range [38, 77, 97, 101, 144, 163]. For example, Bitternut hickory (*Carya cordiformis* [Wangenh.] K. Koch) and shagbark hickory (*Carya ovata* [Mill.] K. Koch) are two closely related species that present very weak genetic structure except one outlier population within *C. ovata* [144]. Similar results were found for populations of *Viburnum rufidulum* Raf., which is a temperate tree species and widely distributed across the eastern U.S. It was suggested that high levels of gene flow via pollen and seed dispersal, might help lower the genetic differences

among the populations of *V. rufidulum* [103]. In contrast, despite the wide distribution pattern, *C. canadensis* populations showed high genetic differentiation among populations with limited gene flow.

Like many other self-incompatible forest tree species [106], *C. canadensis* also depends on different pollen and seed dispersal modes for gene flow. Usually honeybees, megachilid bees, small sweat bees, beetles, and butterflies are the primary insect pollinators of *C. canadensis* [14, 18, 124-126]. Insect flight ranges among species vary from one to several miles [108, 109]. Therefore, long distance gene flow by pollen dispersal between trees can also be limited, which would reduce the overall gene flow within

C. canadensis populations. Our study indicates that gene flow has been limited across the breadth of the native range of *C. canadensis*. By contrast, within a localized genetic assessment of *C. canadensis* within a limited geographic distribution, there is evidence of high gene flow between local populations (Chapter 1). As geographical distance across populations of *C. canadensis* increases, gene flow among *C. canadensis* trees is reduced, which in turn increases the genetic differentiation among populations of *C. canadensis*.

Seed dispersal, along with pollen dispersal, plays an important role in maintaining gene flow between populations of a species [119]. Successful seed dispersal events depend on the behavior and foraging patterns of the animals that feed on the fruits of the species and disperse the seeds. Usually the rate of *C. canadensis* seed dispersal by animals (e. g., squirrels and small rodents) is limited by a potential disperser's forage range. Fruit consumption rate by animals is also restricted by reliance upon

C. canadensis fruits as emergency food in late fall or winter, and this behavior tends to lower the seed dispersal possibility [14, 18, 124-126].

Contrary to seed dispersal, heavy *C. canadensis* fruits typically fall in close proximity to the parental tree and progeny and can grow as non-reproductive seedlings during next several years [14, 120]. This condition results in creation of half-sibling neighborhoods at a localized spatial scale [103, 119, 121-123]. Small rodents and deer may repeatedly eat from same tree, thus carrying the closely related, half-sibling propagules to restricted distributions determined by retention time of scats [14, 18, 124-126]. These events likely limit the gene flow to short distances and create spatial genetic structures at a localized level, as was revealed in fine scale level assessments (Chapter 1) and supported by this assessment of large geographic genetic distribution in *C. canadensis* populations. Thus, behaviors of the seed carriers explain the reason of high genetic differentiation among studied dataset of *C. canadensis*.

Our results further indicated a positive inbreeding coefficient, which is an indicator of excess of homozygous alleles. Although *C. canadensis* is an outcrossing species, breeding among related individuals, as well as localized and limited seed dispersal may increase the likelihood of inbreeding at the local population level, which is evident from the increase in number of homozygous alleles [164]. The presence of spatial structure within *C. canadensis* populations further supports the occurrence of inbreeding between related individual trees [41, 165]. STRUCTURE and DAPC analyses of the *C. canadensis* dataset revealed the presence of two geographically distinct clusters that are longitudinally divided in northwest and southeast directions with a transition zone in the Kentucky-Tennessee-Mississippi area. This transition zone represents the rough

midpoint for *C. canadensis* distribution in the U.S. This distribution of geographical clusters potentially originates from one or several glacial refugia as most of the modern day eastern North American tree species originated from low density postglacial refugia.

The presence of only two genetic clusters is congruent with the simple postglacial lineage theory presented for eastern North American tree species [153, 166]. The most recent glacial event ended approximately 20,000 years ago and shifted boreal forest and temperate species closer to mid-latitudes within the eastern U.S. [167]. Several phylogenetic studies have indicated that the southeastern U.S. served as one prominent large scale glacial refugia for many temperate species [149, 166, 167]. Modern day temperate species including *Fagus grandifolia* (American beech), *Acer rubrum* (red maple), and *Cornus florida* (flowering dogwood), for example, likely originated from this southeastern refugium [149, 168, 169]. These studies also suggested the presence of several micro-refugia within this wide refugium that includes the Mississippi River Valley, Southern Appalachian, Ozark Plateau, the Atlantic Coast, Gulf Coast, Florida, Texas, and Cumberland Plateau in Tennessee [166, 167, 170]. *C. canadensis* also shares the same geographic distribution as these temperate tree species and modern-day wild populations of *C. canadensis* are ubiquitous across this region. Therefore, the southeastern refugium is also likely to be the major postglacial refugium for *C. canadensis*.

During the expansion of Laurentide ice sheet in last glacial period, a large area in the upper mid-west region including Minnesota, Wisconsin, and Iowa remained ice-free [149, 171]. Several phylogeographical studies suggested that many animals and floral species survived as northern glacial refugium in that area [143, 167, 172]. The

populations of *C. canadensis* from the northern region may have originated from northern refugium as their genetic diversity indices are higher in the collection sites in that region. This trend of higher genetic diversity is expected for the populations that are situated in a postglacial refugia zone [173, 174] or are considered the point of origin. Individuals from a postglacial northern refugia may have migrated from the northeastern U.S to the southward close to the Appalachian mountain region. As a consequence of long term population isolation within the refugial area, a post-glacial refugia can give rise to distinct spatial clusters due to genetic differentiation [119, 145]. Therefore, our STRUCTURE results yielding two geographical clusters, further support the presence of at least two post glacial refugia.

C. canadensis populations in the west of the *C. canadensis* native distribution range (in south from Texas to Nebraska in north, U.S.) may represent the remnant from the Ozark Plateau or Texas refugial populations. These populations are adapted to semi-arid to xeric environments and showed distinct adaptive characteristics [149, 166, 170]. Several studies suggest that North American *Cercis* species ancestors were of mesic origin (area with moderate or balanced moisture level) and may have dispersed into North America across the North Atlantic Land Bridge [13, 156, 175]. The ancestral *Cercis* population quickly adapted to a drier environment that was present during the Pleistocene and consequently adapted to a mesic environment allowing spread into the Northern hemisphere during the mid-Miocene period [13, 162]. These adaptational sequences could explain the wide variation in morphological characteristics observed across *C. canadensis* wild populations.

The DIYABC results also support the hypothesis that current *C. canadensis* populations originated from descendants of one or several refugia. According to our results, the ancestral population was split three ways from the ancestral population. Our results also indicated an occurrence of a bottleneck, based on DIYABC analysis. The last glacial period timeline (ca. 20,000 years ago) and the DIYABC suggested bottleneck occurrence timeline are matched closely. Therefore, *C. canadensis* populations might face a bottleneck event in the last glacial maximum period. In that case, three populations (after the split of the initial ancestral population) that endured bottleneck at the time of last glacial period might have later survived as postglacial refugia, from which present-day *C. canadensis* population clusters were derived. Also, as current populations of *C. canadensis* recovered from that ancient bottleneck and achieved an effective population size, the BOTTLENECK software analysis for the studied populations discarded recent occurrence of any bottleneck.

To better elucidate this current hypothesis of glacial refugia for *C. canadensis*, genetic screening using chloroplasts or mitochondrial DNA markers is needed. Haplotyping analysis could further discern the maternal lineages and evolutionary histories of this economically important and geographically significant species. A recent phylogeography and population demographic history study of *C. florida* using plastid DNA regions, has revealed the origin and time of the haplotype lineages of this deciduous temperate tree species. The study also suggested that the southeastern refugium is the most likely origin of postglacial populations that later rapidly expanded toward the north.

Appendix: Tables and Figures

Table 2.1. Estimation of diversity indices for seventy-four *Cercis canadensis* collection sites from the United States using 12 microsatellite loci.

States	Collection Sites	Latitude	Longitude	N	MLG	H	He	Pa
NY	Monroe	43.0339	-77.5503	10	10	2.30	0.53	0
NY	Tompkins	42.3988	-76.5549	9	9	2.20	0.61	0
PA	Centre	40.8038	-77.8058	10	10	2.30	0.58	0
PA	Huntingdon	40.5064	-77.9806	9	9	2.20	0.47	0
PA	Juniata	40.5658	-77.2557	6	6	1.79	0.52	0
WV	Cabell	38.3925	-82.4245	8	8	2.08	0.59	0
WV	Kanawha	38.4842	-81.4354	10	10	2.30	0.56	1
WV	Braxton	38.8735	-80.6293	8	8	2.08	0.48	0
WV	Monongalia	39.6034	-79.9922	7	7	1.95	0.50	0
VA	Prince William	38.8121	-77.5524	9	9	2.20	0.52	0
VA	King George	38.3293	-77.0908	7	7	1.95	0.57	1
VA	Brunswick	36.9225	-77.7582	10	10	2.30	0.55	0
VA	Washington	36.6563	-81.9050	10	10	2.30	0.55	0
VA	Radford	37.1345	-80.5218	10	10	2.30	0.47	0
MD	Carroll	39.3548	-76.8963	10	10	2.30	0.51	0
MD	Prince George	38.7473	-76.9924	6	6	1.79	0.56	0
NC	Guilford	36.0322	-79.7063	8	8	2.08	0.51	0
NC	Cabarrus	35.4578	-80.5947	9	9	2.20	0.47	1
NC	Mecklenburg	35.2415	-80.9845	10	10	2.30	0.47	0
NC	Wake	35.9789	-78.6368	10	10	2.30	0.58	0
OH	Hamilton 1	39.0857	-84.3589	9	9	2.20	0.67	0
OH	Hamilton 2	39.0390	-84.3478	10	10	2.30	0.63	1
MI	Washtenaw	42.3828	-83.9068	7	7	1.95	0.47	0
MI	Berrien 1	41.9052	-86.3701	9	9	2.20	0.63	0
MI	Berrien 2	41.7903	-86.7627	9	9	2.20	0.66	0
IN	Jefferson	38.7831	-85.3695	6	6	1.79	0.64	1
IN	Tippecanoe	40.5306	-86.9244	10	10	2.30	0.60	1
IN	Montgomery	40.0494	-86.9016	9	9	2.20	0.57	0
IN	Parke	39.8820	-87.2020	10	10	2.30	0.62	0
KY	Laurel	37.2145	-84.1943	7	7	1.95	0.51	0
KY	Madison	37.8949	-84.2743	10	10	2.30	0.46	2
KY	Bath	38.1049	-83.8265	10	10	2.30	0.59	0
KY	Carter	38.3641	-82.8029	8	8	2.08	0.48	0
TN	Cocke	35.8211	-83.1528	8	8	2.08	0.46	0
TN	Anderson	36.0054	-84.2077	10	10	2.30	0.32	0
TN	Knox	35.9462	-83.9178	9	9	2.20	0.38	0
TN	Hamilton 1	35.0758	-85.1288	10	10	2.30	0.50	0
TN	Hamilton 2	35.0235	-85.3791	10	10	2.30	0.54	0
GA	Whitfield	34.7287	-85.0314	10	10	2.30	0.53	0
TN	Shelby 1	35.0892	-89.8661	10	10	2.30	0.47	0

Table 2.1 continued

States	Collection Sites	Latitude	Longitude	N	MLG	H	He	Pa
TN	Shelby 2	35.2723	-89.6452	10	10	2.30	0.36	0
TN	Shelby 3	35.3255	-90.0515	10	10	2.30	0.55	0
TN	Cheatham	36.1222	-87.1416	10	10	2.30	0.52	0
TN	Wilson 1	36.0564	-86.4258	10	10	2.30	0.55	1
TN	Wilson 2	36.1683	-86.5654	10	9	2.20	0.44	0
GA	Clarke 1	33.8804	-83.3572	10	10	2.30	0.35	0
GA	Talbot	32.6655	-84.5126	10	10	2.30	0.34	0
GA	Clarke 2	33.9032	-83.3869	10	10	2.30	0.48	0
SC	Pickens	34.6415	-82.8241	10	10	2.30	0.57	0
SC	Lexington	34.1716	-81.3047	8	8	2.08	0.56	0
FL	Gadsden	30.6264	-84.8949	9	9	2.20	0.33	0
FL	Alachua	29.7761	-82.5081	10	10	2.30	0.56	1
FL	Dixie	29.6367	-82.7913	10	10	2.30	0.52	0
AL	Jefferson 1	33.7197	-86.7748	10	10	2.30	0.51	1
AL	Jefferson 2	33.5670	-86.6802	10	10	2.30	0.45	0
AL	Tallapoosa	32.7705	-85.7854	10	9	2.20	0.42	0
MS	Alcorn	34.9504	-88.8184	10	10	2.30	0.50	1
MS	Pontotoc	34.1447	-88.9972	10	10	2.30	0.53	0
MS	Oktibbeha	33.4258	-88.7004	10	10	2.30	0.53	1
AR	Pulaski	34.7601	-92.3173	10	10	2.30	0.48	0
AR	Washington 1	36.0878	-94.1679	10	10	2.30	0.52	0
AR	Washington 2	36.1316	-94.1336	10	10	2.30	0.50	0
MO	Boone 1	38.8401	-92.2923	10	10	2.30	0.59	0
MO	Genevieve	38.0137	-90.2203	10	10	2.30	0.54	0
MO	Boone 2	39.0788	-92.3066	10	10	2.30	0.52	1
MO	Dent	37.4552	-91.6869	10	10	2.30	0.52	0
IA	Fremont	40.6746	-95.6908	8	8	2.08	0.64	1
NE	Sarpy	41.1799	-95.9181	10	10	2.30	0.68	0
KS	Geary	38.8941	-96.8544	10	10	2.30	0.56	0
KS	Butler	38.8681	-96.8501	10	10	2.30	0.52	0
OK	Cherokee	35.7603	-94.9070	10	10	2.30	0.54	0
TX	Sabine	31.4106	-94.0225	10	10	2.30	0.44	1
TX	Smith	32.4520	-95.2534	10	10	2.30	0.55	3
TX	Collin	33.0794	-96.5488	9	9	2.20	0.51	0
Average/	Total			691	689	6.54	0.67	19

N = total number of samples per collection site; MLG = number of diploid individuals multi locus genotypes after clone correction; H = Shannon-Wiener index of MLG diversity; He = Nei's genotypic diversity corrected for sample size; Pa = number of private alleles in each collection site.

Table 2.2. Genetic diversity indices of 12 microsatellite loci across 74 collection sites of *Cercis canadensis*.

Locus	Repeat Motif	Allele No	Ar	Ho	He	H	Evenness	F _{ST}	F' _{ST}	F _{IS}	Nm	Pa
127spa	(TC)4	12	1.73	0.99	0.75	1.44	0.64	0.05	0.05	-0.39	0.82	1
168a	(CT)7	13	1.51	0.23	0.70	1.45	0.69	0.24	0.24	0.57	0.58	4
177b	(GA)6	9	1.64	0.23	0.78	1.70	0.80	0.15	0.15	0.66	0.84	0
199a	(GAGA)8	12	1.55	0.23	0.71	1.58	0.65	0.21	0.21	0.59	0.65	1
220a	(TATT)4	11	1.64	0.57	0.80	1.66	0.64	0.19	0.20	0.11	0.70	2
229a	(GAGAG)4	6	1.50	0.15	0.62	1.11	0.77	0.15	0.15	0.71	0.86	0
625a	(GA)4	11	1.57	0.24	0.68	1.32	0.59	0.13	0.13	0.59	0.81	2
658a	(CT)6	9	1.50	0.21	0.60	1.17	0.69	0.13	0.13	0.59	0.89	3
680a	(GT)8	8	1.37	0.01	0.59	1.15	0.66	0.33	0.33	0.98	0.37	0
762a	(TC)7	11	1.52	0.28	0.68	1.41	0.69	0.21	0.21	0.48	0.64	3
780b	(AG)12	16	1.79	0.70	0.86	2.19	0.75	0.07	0.07	0.12	1.70	1
995a	(AG)7	7	1.23	0.01	0.52	0.84	0.82	0.53	0.54	0.99	0.18	2
Overall/ Average		10	1.55	0.32	0.69	1.42	0.70	0.19	0.19	0.43	0.75	19

Ar = allelic richness corrected for sample size, Ho = observed heterozygosity, He = Nei's genotypic diversity, H = Shannon-Wiener index, F_{ST} = populations fixation index, F'_{ST} = population differentiation, F_{IS} = inbreeding coefficient, Nm = gene flow, Pa = number of private alleles

Table 2.3. Analysis of molecular variance of *Cercis canadensis* across 12 microsatellite loci for (i) 74 collection sites as one hierarchical cluster, ii) 74 collection sites into two groups according to two clusters of STRUCTURE.

i) 74 collection sites

Source of Variation	df	Sum Square	Variance Component	% of Variation	P value
Between collection sites	73	1441.17	1.72	25.76	0.001
Within collection sites	561	2785.17	4.96	74.24	0.001
Total	634	4226.35	6.69	100.00	

$F_{ST} = 0.26$

ii) 2 clusters (STRUCTURE)

Source of Variation	df	Sum Square	Variance Component	% of Variation	P value
Between two clusters	1	331.78	1.00	13.87	0.001
Between collection sites	72	1109.40	1.22	16.97	0.001
Within collection sites	561	2785.17	4.96	69.15	0.001
Total	634	4226.34	7.18	100.00	

$F_{IS} = 0.31$, $F_{SC} = 0.20$, $F_{CT} = 0.14$

* F_{ST} = variance among collection sites relative to the total variance

* F_{IS} = inbreeding coefficient of individuals relative to population

* F_{SC} = variance among collection sites within groups

* F_{CT} = variance among groups relative to the total variance

Table 2.4. Bottleneck determination by sign tests for *Cercis canadensis* samples utilizing 12 microsatellite loci. Seventy-four collection sites were divided into two clusters according to the STRUCTURE analysis.

Cluster	I.A.M	T.P.M	S.M.M	Mode-shift^b	P value
Cluster 1	9/3	4/8	0/12	Not shifted	NS
Cluster 2	10/2	6/6	0/12	Not shifted	NS

I.A.M. = infinite allele model; T.P.M. = two-phase mutation model; S.M.M. = stepwise mutation model,

a = Excess/deficit indicates the number of loci showing excess/deficit of gene diversity under mutation-drift in populations

b = A shift in the distribution of allelic frequency classes is expected in populations that experienced a recent bottleneck. NS - not significant

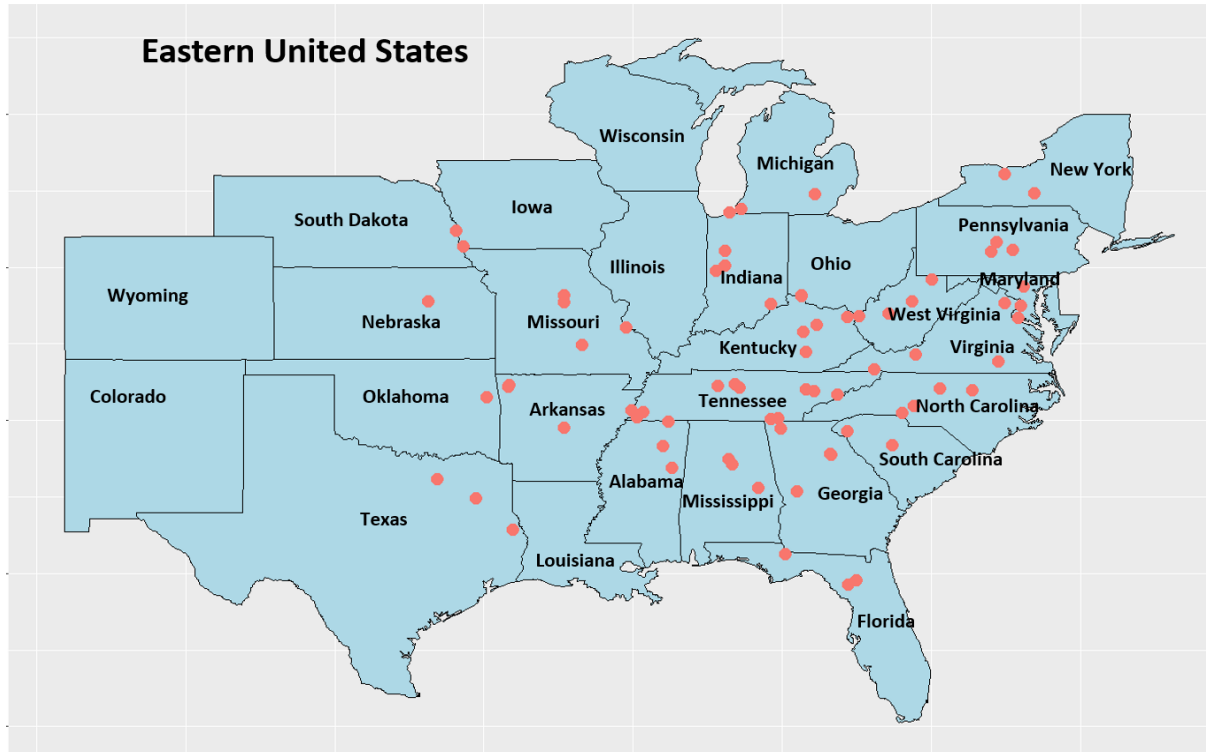


Figure 2.1. Map of 74 collection sites (red dots) of *Cercis canadensis* across its native distribution in the United States.

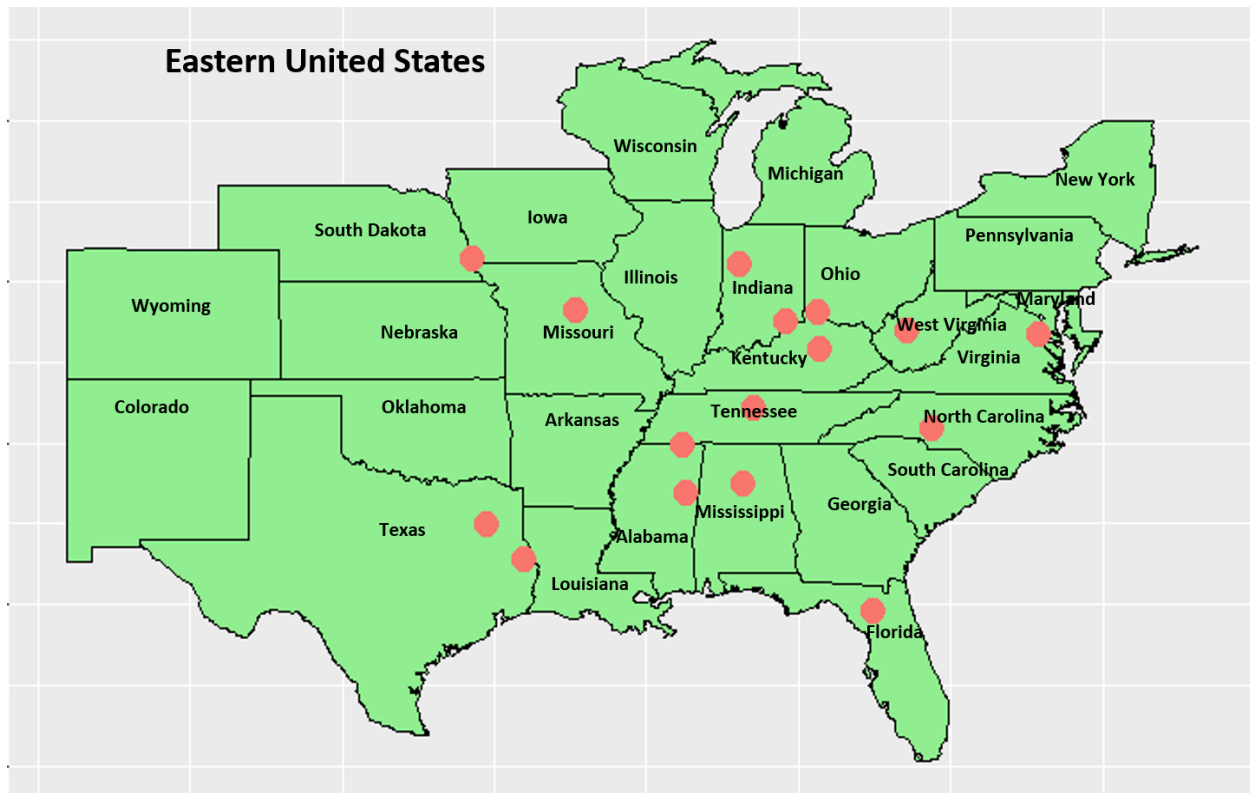


Figure 2.2. Map of private alleles found in *Cercis canadensis* populations across the eastern United States. Red dots represent sampling sites that contain private alleles.

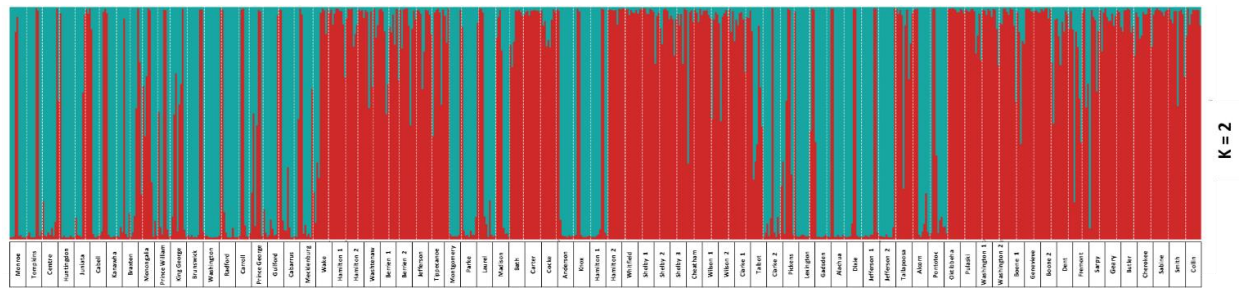




Figure 2.4. Map of 74 collection sites of *Cercis canadensis* across 23 states in the eastern United States. The color code represents designated cluster for each collection site according to the STRUCTURE analysis (red = cluster 1; blue = cluster 2).

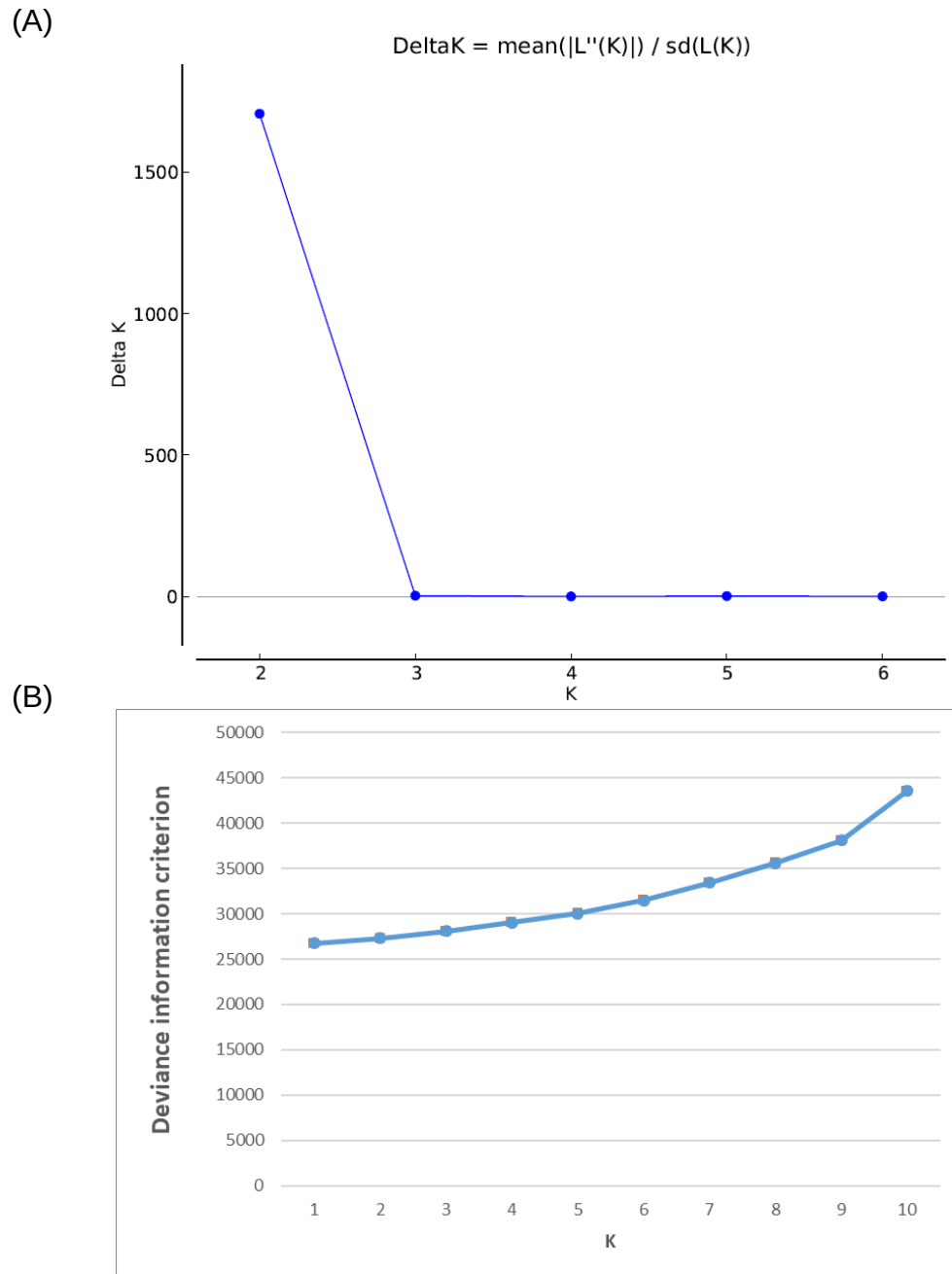
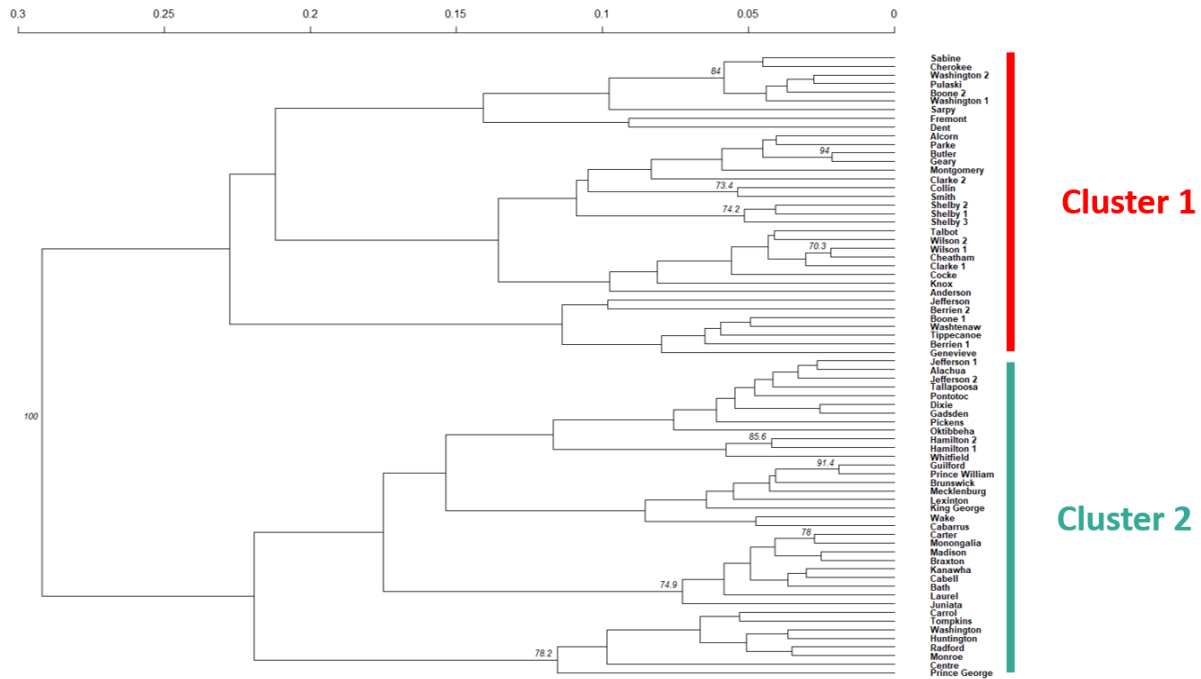


Figure 2.5. Bayesian clustering probabilities using programs: STRUCTURE and InStruct. STRUCTURE results were obtained using Delta K and graph results indicated two genetic clusters among 74 *Cercis canadensis* collection sites (A). InStruct were obtained using deviance information criterion calculations (B). Results indicated the highest probability of two different genetic clusters ($K = 1$) across 74 *Cercis canadensis* collection sites.



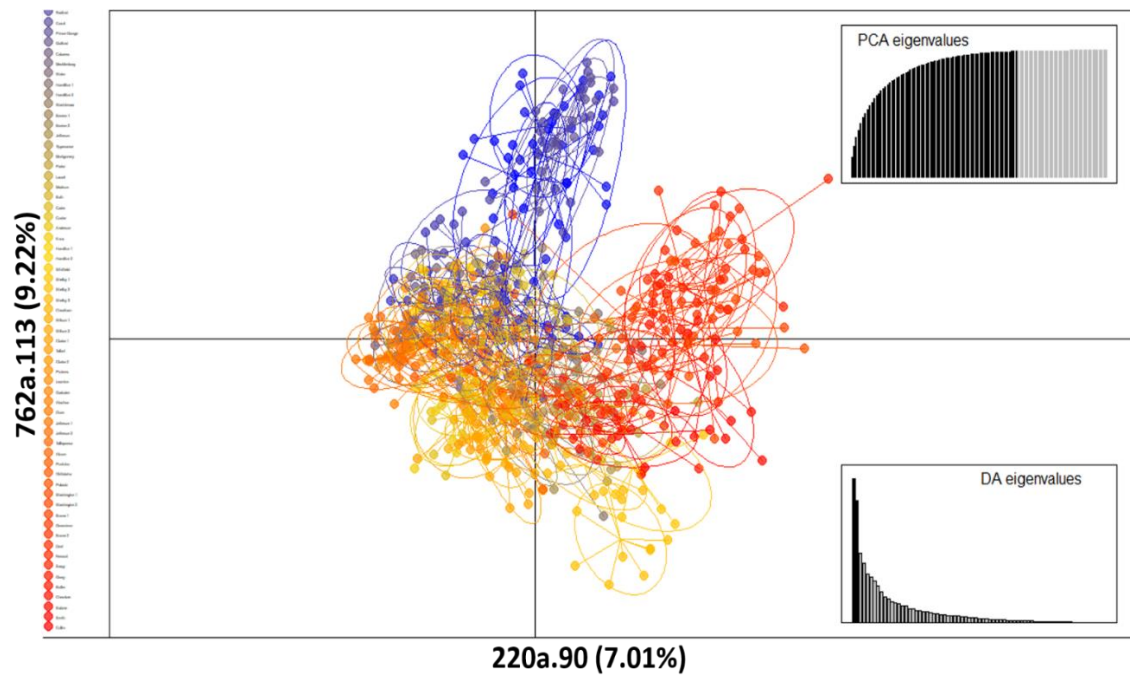


Figure 2.7. Discriminant Analysis of Principal Components biplot of *Cercis canadensis* individuals was constructed using 1,000 permutations. The alleles with loci explaining the most variance is indicated at plot axes, with percent contributions in parentheses. Discriminant Analysis (DA) eigenvalues are also presented.

(A)

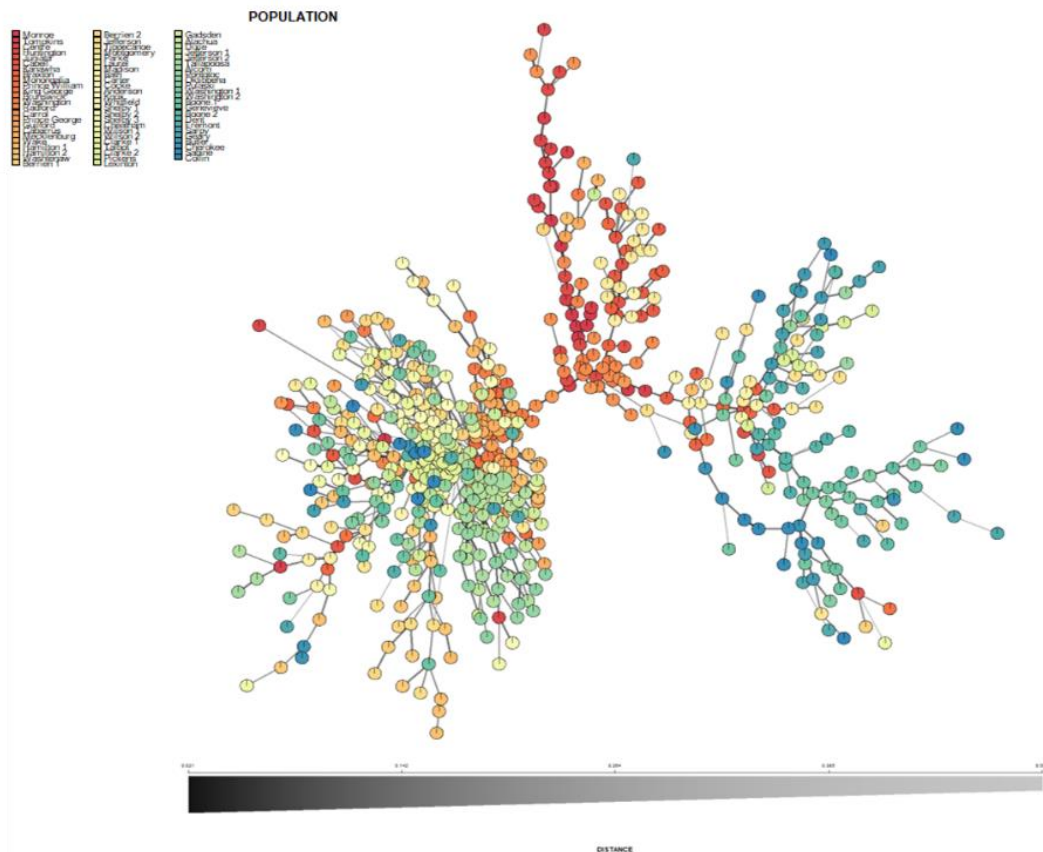


Figure 2.8. (A). The Minimum Spanning Networks (MSN) of *Cercis canadensis* genotypes were calculated using Bruvo's genetic distance (considering microsatellite motif length) (A). Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.

(B)

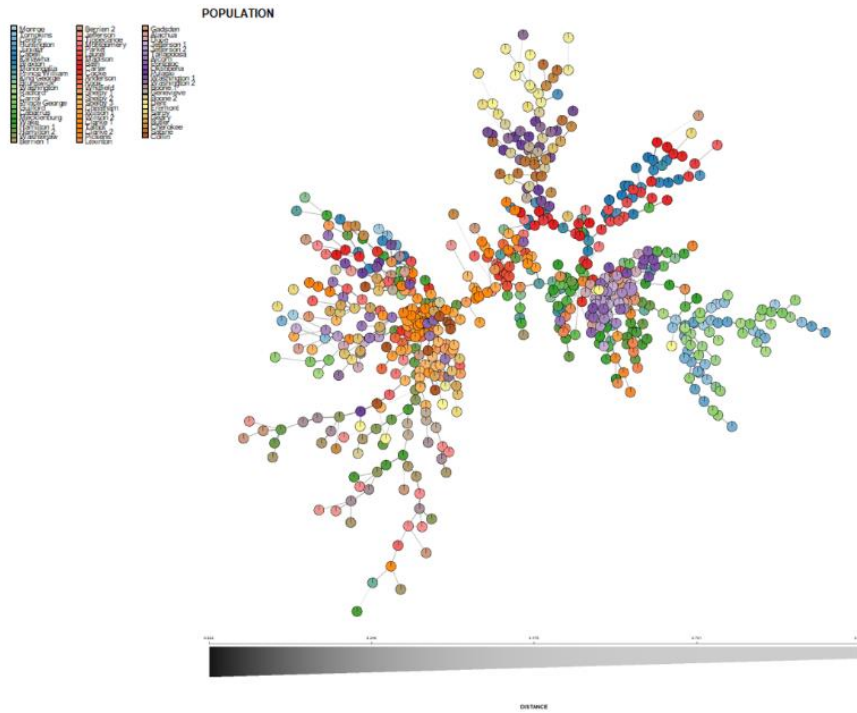


Figure 2.8. (B). The Minimum Spanning Networks (MSN) of *Cercis canadensis* genotypes were calculated Nei's genetic distance (B). Nodes are color coded for groups used in the analysis as represented in the legend. Widths are reversely proportional to the genetic distances between the respective nodes.

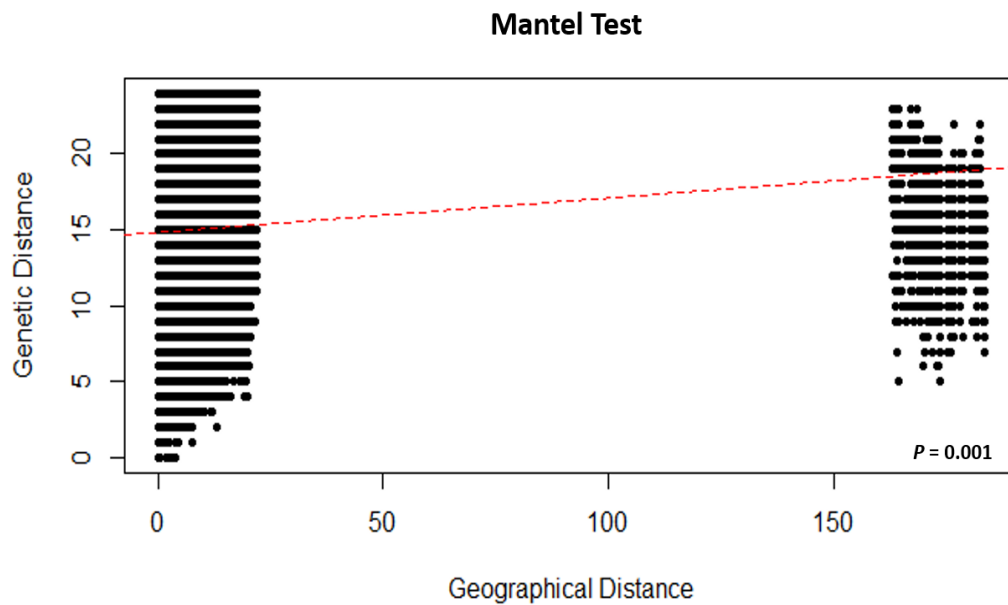


Figure 2.9. Mantel test of *Cercis canadensis* individuals from 74 collection sites across 23 states in the eastern United States. The test indicated the correlation between geographical and genetic distance.

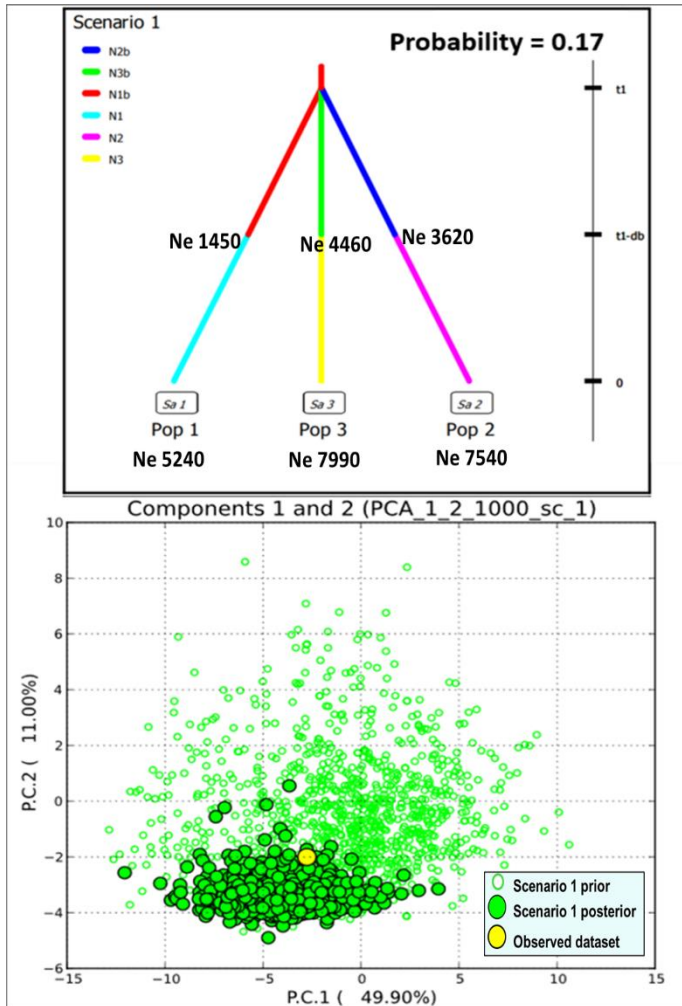


Figure 2.10. Most probable DIYABC evolutionary scenario for *Cercis canadensis*. Here, populations are labeled as northeast (N1), south (N2), and northwest (N3). Scenario 1 that suggested simultaneous triple split of the originating population (N1) at t_1 and occurrence of bottleneck at $t_1\text{-db}$, provided the best support. N_e = effective population size. Model checking analyses of the closest 1% of simulated prior and posterior datasets (bottom panel) using PCA further supported scenario.

CONCLUSION

This is the first study addressing the genetic diversity and population structure of *C. canadensis* wild populations both at fine and broad scale levels across its native distribution in the U.S. Previously, there was limited information about the genetic diversity and population structure of *C. canadensis*. Our fine-scale distribution results indicated high genetic diversity and high gene flow for sampled trees. Our findings are supported by other studies focused on temperate, self-incompatible tree species. This outcome suggests that reproductive isolation of *C. canadensis* that results from forest fragmentation may not be of great concern genetically and ecologically. Current isolated patches and remnant populations of trees are genetically stable, thus are able to maintain viable populations at a geographically fine scale. Our larger, broad scale population study indicated similar results, which further suggests that natural stands of *C. canadensis* are genetically diverse and are spatially clustered across a wide geographical range. The level of genetic diversity found here was supported by the presence of a number of morphological variations that have been documented across *C. canadensis* wild populations [6, 155, 161].

Although natural stands of *C. canadensis* seemingly maintain genetically fit populations, we cannot rule out the possible negative effect of forest fragmentation on *C. canadensis* population viability. As this is an outcrossing, self-incompatible, and animal-dispersed tree species, limitation or reduction in the numbers of seed dispersal agents (e.g. small rodents) in potentially fragmented populations can affect the fitness of *C. canadensis* populations. Therefore, to maintain healthy forests, it is important to ensure that healthy and diverse habitats are present for the animals responsible for seed and pollen dispersal of *C. canadensis*.

Moreover, this economically and ecologically significant deciduous shade tree species has a number of desirable ornamental characteristics including, but not limited to drought tolerance, disease resistance, and a wide range of tree architectures across its geographical distributions [6, 14]. The uniqueness of these adaptive traits across different geographic locations can be used as genetic reservoirs of desired genetic variations. Also, we can identify certain populations and/or individual trees for controlled breeding programs to improve certain traits or provide an opportunity to develop additional cultivars in the nursery stocks.

REFERENCES

1. Couvillon G: ***Cercis canadensis* L. Seed size influences germination rate, seedling dry matter, and seedling leaf area.** *HortScience* 2002, **37**(1):206-207.
2. Li S, Shi T, Kong F, Ma Q, Mao Y, Yin T: **Methods for breaking the dormancy of eastern redbud (*Cercis canadensis*) seeds.** *Seed Science and Technology* 2013, **41**(1):27-35.
3. Trigiano R, Beaty R, Graham E: **Somatic embryogenesis from immature embryos of redbud (*Cercis canadensis*).** *Plant Cell Reports* 1988, **7**(2):148-150.
4. Pooler M, Jacobs K, Kramer M: **Differential resistance to *Botryosphaeria ribis* among *Cercis* taxa.** *Plant Disease* 2002, **86**(8):880-882.
5. Cheong E, Pooler M: **Micropropagation of chinese redbud (*Cercis yunnanensis*) through axillary bud breaking and induction of adventitious shoots from leaf pieces.** *In Vitro Cellular & Developmental Biology-Plant* 2003, **39**(5):455-458.
6. Wadl P, Trigiano R, Werner D, Pooler M, Rinehart T: **Simple sequence repeat markers from *Cercis canadensis* show wide cross-species transfer and use in genetic studies.** *Journal of the American Society for Horticultural Science* 2012, **137**(3):189-201.
7. Thammina C, Kidwell-Slak D, Lura S, Pooler M: **SSR markers reveal the genetic diversity of asian *Cercis* taxa at the US National Arboretum.** *Hortscience* 2017, **52**(4):498-502.
8. USDA NASS: **Census of Horticultural Specialties.** 2014, **3.**

9. Coşkun Fh, Parks C: **A molecular phylogenetic study of red buds (*Cercis* L., Fabaceae) based on ITS nrDNA sequences.** *Pakistan Journal of Botany* 2009, **41**(4):1577-1586.
10. Raulston J: **Redbud.** *American Nurseryman* 1990, **171**(5):39-51.
11. Geneve R: **Eastern redbud (*Cercis canadensis* L.) and judas tree (*Cercis siliquastrum* L.).** In: *Trees III*. Springer; 1991: 142-151.
12. Griffin J, Ranney T, Pharr D: **Heat and drought influence photosynthesis, water relations, and soluble carbohydrates of two ecotypes of redbud (*Cercis canadensis*).** *Journal of the American Society for Horticultural Science* 2004, **129**(4):497-502.
13. Davis C, Fritsch P, Li J, Donoghue M: **Phylogeny and biogeography of *Cercis* (Fabaceae): Evidence from nuclear ribosomal ITS and chloroplast ndhF sequence data.** *Systematic Botany* 2002, **27**(2):289-302.
14. Dickson J: ***Cercis canadensis* L.** In: *Silvics of North America*. Edited by Burns R, Honkala B, vol. 2. Hardwood: United States Department of Agriculture, Forest Service, Washington, DC; 1990: 266-269.
15. Hyche LL: **Redbud Insects: A Guide to Recognition and Habits of Species Damaging Foliage of Ornamental Redbud Trees in Alabama;** 2002.
16. Kraemer M, Favi F: **Emergence phenology of *Osmia lignaria* subsp *lignaria* (Hymenoptera: Megachilidae), its parasitoid *Chrysura kyrae* (Hymenoptera: Chrysididae), and bloom of *Cercis canadensis*.** *Environmental Entomology* 2010, **39**(2):351-358.

17. Roy D: **Cercis L. (redbud)**. In: *Seeds of woody plants in the United States*. Edited by Schopmeyer C: Washington, DC: U.S. Department of Agriculture, Forest Service; 1974: 305-308.
18. Sullivan J: **Cercis canadensis**. In. Edited by U.S. Department of Agriculture FS, Rocky Mountain Research Station: Fire Sciences Laboratory; 1994.
19. Brown R, Brown M: **Woody plants of Maryland.**: University of Maryland; 1972.
20. Kwak M, Velterop O, Andel J: **Pollen and gene flow in fragmented habitats**. *Applied Vegetation Science* 1998, **1**(1):37-54.
21. Kolb A, Diekmann M: **Effects of life-history traits on responses of plant species to forest fragmentation**. *Conservation Biology* 2005, **19**(3):929-938.
22. Hamberg L, Lehvävirta S, Kotze J: **Forest edge structure as a shaping factor of understorey vegetation in urban forests in Finland**. *Forest Ecology and Management* 2009, **257**(2):712-722.
23. Cuartas-Hernández S, Núñez-Farfán J: **The genetic structure of the tropical understory herb *Dieffenbachia seguine* L. before and after forest fragmentation**. *Evolutionary Ecology Research* 2006, **8**(6):1061-1075.
24. Suárez-Montes P, Chávez-Pesqueira M, Núñez-Farfán J: **Life history and past demography maintain genetic structure, outcrossing rate, contemporary pollen gene flow of an understory herb in a highly fragmented rainforest**. *PeerJ* 2016, **4**:e2764.
25. Weiher E, van der Werf A, Thompson K, Roderick M, Garnier E, Eriksson O: **Challenging Theophrastus: A common core list of plant traits for functional ecology**. *Journal of Vegetation Science* 1999, **10**(5):609-620.

26. Haila Y: **Islands and Fragments**. In: *Maintaining biodiversity in forest ecosystems*. Cambridge University Press; 1999: 234-264.
27. Santos T, Tellería J, Virgós E: **Dispersal of Spanish juniper *Juniperus thurifera* by birds and mammals in a fragmented landscape**. *Ecography* 1999, **22**(2):193-204.
28. Sato Y, Kudoh H: **Fine-scale genetic differentiation of a temperate herb: relevance of local environments and demographic change**. *Annals of Botany* 2014, **6**.
29. Kearns C, Inouye D, Waser N: **Endangered mutualisms: the conservation of plant-pollinator interactions**. *Annual Review of Ecology Systematics* 1998, **29**(1):83-112.
30. Van Geert A, Van Rossum F, Triest L: **Genetic diversity in adult and seedling populations of *Primula vulgaris* in a fragmented agricultural landscape**. *Conservation Genetics* 2007, **9**(4):845.
31. Sherwin W, Moritz C: **Managing and monitoring genetic erosion**. In: *Genetics, Demography and Viability of Fragmented Populations*. Cambridge University Press; 2000: 9-32.
32. Kramer A, Ison J, Ashley M, Howe H: **The paradox of forest fragmentation genetics**. *Conservation Biology* 2008, **22**(4):878-885.
33. Nason J, Herre A, Hamrick J: **The breeding structure of a tropical keystone plant resource**. *Nature* 1998, **391**(6668):685-687.

34. Byrne M, Elliott C, Yates C, Coates D: **Maintenance of high pollen dispersal in *Eucalyptus wandoo*, a dominant tree of the fragmented agricultural region in Western Australia.** *Conservation Genetics* 2008, **9**(1):97-105.
35. Ross-Davis A, Ostry M, Woeste K: **Genetic diversity of butternut (*Juglans cinerea*) and implications for conservation.** *Canadian Journal of Forest Research* 2008, **38**(4):899-907.
36. Elliott K, Swank W: **Long-term changes in forest composition and diversity following early logging (1919–1923) and the decline of American chestnut (*Castanea dentata*).** *Plant Ecology* 2008, **197**(2):155-172.
37. Mantooth K, Hadziabdic D, Boggess S, Windham M, Miller S, Cai G, Spatafora J, Zhang N, Staton M, Ownley B: **Confirmation of independent introductions of an exotic plant pathogen of *Cornus* species, *Discula destructiva*, on the east and west coasts of North America.** *PloS One* 2017, **12**(7):e0180345.
38. Hadziabdic D, Fitzpatrick B, Wang X, Wadl P, Rinehart T, Ownley B, Windham M, Trigiano R: **Analysis of genetic diversity in flowering dogwood natural stands using microsatellites: the effects of dogwood anthracnose.** *Genetica* 2010, **138**(9-10):1047-1057.
39. Daughtrey M, Hibben C: **Dogwood anthracnose: a new disease threatens two native *Cornus* species.** *Annual Review of Phytopathology* 1994, **32**(1):61-73.
40. Victory E, Glaubitz J, Rhodes Jr O, Woeste K: **Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites.** *American Journal of Botany* 2006, **93**(1):118-126.

41. Marquardt P, Echt C, Epperson B, Pubanz D: **Genetic structure, diversity, and inbreeding of eastern white pine under different management conditions.** *Canadian Journal of Forest Research* 2007, **37**(12):2652-2662.
42. Dubreuil M, Riba M, González-Martínez S, Vendramin G, Sebastiani F, Mayol M: **Genetic effects of chronic habitat fragmentation revisited: Strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability.** *American Journal of Botany* 2010, **97**(2):303-310.
43. Hall B, Motzkin G, Foster D, Syfert M, Burk J: **Three hundred years of forest and land-use change in Massachusetts, USA.** *Journal of Biogeography* 2002, **29**(10-11):1319-1335.
44. Noss R, LaRoe E, Scott M: **Endangered ecosystems of the United States: a preliminary assessment of loss and degradation.** In. Edited by US Department of the Interior NBSW, DC, USA., vol. 28; 1995: 1-58.
45. Russell C, Morse L: **Extinct and possibly extinct plant species of the United States and Canada.** In., vol. 13; 1992.
46. Cook R, Dixon P: **A review of recovery plans for threatened and endangered plant species: a report for the world wildlife fund.** In. Edited by World Wildlife Fund W, DC; 1989.
47. Young A, Boyle T, Brown T: **The population genetic consequences of habitat fragmentation for plants.** *Trends in Ecology Evolution* 1996, **11**(10):413-418.
48. Fischer M, Matthies D: **RAPD variation in relation to population size and plant fitness in the rare *Gentianella germanica* (Gentianaceae).** *American Journal of Botany* 1998, **85**(6):811-819.

49. Severns P: **Inbreeding and small population size reduce seed set in a threatened and fragmented plant species, *Lupinus sulphureus* ssp. *kincaidii* (Fabaceae).** *Biological Conservation* 2003, **110**(2):221-229.
50. Dirr M: **Manual of woody landscape plants: their identification, ornamental characteristics, culture, propagation and uses.** Champaign, Illinois: Stipes Publishing Co; 1990.
51. Amos W, Hoffman J, Frodsham A, Zhang L, Best S, Hill A: **Automated binning of microsatellite alleles: problems and solutions.** *Molecular Ecology Resources* 2007, **7**(1):10-14.
52. Kamvar Z, Tabima J, Grünwald N: **Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction.** *PeerJ* 2014, **2**:e281.
53. Tsui C, Roe A, El-Kassaby Y, Rice A, Alamouti S, Sperling F, Cooke J, Bohlmann J, Hamelin R: **Population structure and migration pattern of a conifer pathogen, *Grosmannia clavigera*, as influenced by its symbiont, the mountain pine beetle.** *Molecular Ecology* 2012, **21**(1):71-86.
54. Jost L: **GST and its relatives do not measure differentiation.** *Molecular Ecology* 2008, **17**(18):4015-4026.
55. Jost L, Archer F, Flanagan S, Gaggiotti O, Hoban S, Latch E: **Differentiation measures for conservation genetics.** *Evolutionary Applications* 2018, **11**(7):1139-1148.

56. Bird C, Karl S, Smouse P, Toonen R: **Detecting and measuring genetic differentiation.** *Phylogeography and Population Genetics in Crustacea* 2011, **19**(3):1-55.
57. Wright S: **Evolution in Mendelian populations.** *Bulletin of Mathematical Biology* 1990, **52**(1):241-295.
58. Pritchard J, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**(2):945-959.
59. Earl D: **STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.** *Conservation Genetics Resources* 2012, **4**(2):359-361.
60. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** *Molecular Ecology* 2005, **14**(8):2611-2620.
61. Francis R: **Pophelper: an R package and web app to analyse and visualize population structure.** *Molecular Ecology Resources* 2017, **17**(1):27-32.
62. Gao H, Williamson S, Bustamante C: **A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data.** *Genetics* 2007, **176**(3):1635-1651.
63. Gao H, Bryc K, Bustamante C: **On identifying the optimal number of population clusters via the deviance information criterion.** *PloS One* 2011, **6**(6):e21014.

64. Lind J, Gailing O: **Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* EJ Hill populations at gene-based EST-SSR and nuclear SSR markers.** *Tree Genetics & Genomes* 2013, **9**(3):707-722.
65. Nei M: **Estimation of average heterozygosity and genetic distance from a small number of individuals.** *Genetics* 1978, **89**(3):583-590.
66. Grünwald N, Everhart S, Knaus B, Kamvar Z: **Best practices for population genetic analyses.** *Phytopathology* 2017, **107**(9):1000-1010.
67. Bruvo R, Michiels N, D'souza T, Schulenburg H: **A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level.** *Molecular Ecology* 2004, **13**(7):2101-2106.
68. Jombart T, Devillard S, Balloux F: **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations.** *BMC Genetics* 2010, **11**(1):94.
69. Jombart T, Ahmed I: **adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.** *Bioinformatics* 2011.
70. Excoffier L, Smouse P, Quattro J: **Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.** *Genetics* 1992, **131**(2):479-491.
71. Diniz-Filho J, Soares T, Lima J, Dobrovolski R, Landeiro V, Telles M, Rangel T, Bini L: **Mantel test in population genetics.** *Genetics and Molecular Biology* 2013, **36**(4):475-485.
72. Mantel N: **The detection of disease clustering and a generalized regression approach.** *Cancer Research* 1967, **27**:209-220.

73. Venables WR, Ripley B: **Modern Applied Statistics with S.** Springer, New York 2002, **Fourth Edition.**
74. Cornuet J, Luikart G: **Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data.** *Genetics* 1996, **144**(4):2001-2014.
75. Piry S, Luikart G, Cornuet J: **BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data.** *Journal of Heredity* 1999, **90**:502-503.
76. Leimu R, Mutikainen P, Koricheva J, Fischer M: **How general are positive relationships between plant population size, fitness and genetic variation?** *Journal of Ecology* 2006, **94**(5):942-952.
77. Hamrick J, Godt M: **Effects of life history traits on genetic diversity in plant species.** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 1996, **351**(1345):1291-1298.
78. Chang C, Bongarten B, Hamrick J: **Genetic structure of natural populations of black locust (*Robinia pseudoacacia* L.) at Coweeta, North Carolina.** *Journal of Plant Research* 1998, **111**(1):17-24.
79. Marquardt P, Epperson B: **Spatial and population genetic structure of microsatellites in white pine.** *Molecular Ecology* 2004, **13**(11):3305-3315.
80. Hadziabdic D, Wang X, Wadl P, Rinehart T, Ownley B, Trigiano R: **Genetic diversity of flowering dogwood in the Great Smoky Mountains National Park.** *Tree Genetics & Genomes* 2012, **8**(4):855-871.

81. Foster D, Motzkin G, Slater B: **Land-use history as long-term broad-scale disturbance: regional forest dynamics in central New England.** *Ecosystems* 1998, **1**(1):96-119.
82. LaBonte N, Tonos J, Hartel C, Woeste K: **Genetic diversity and differentiation of yellowwood [*Cladrastis kentukea* (Dum.Cours.) Rudd] growing in the wild and in planted populations outside the natural range.** *New Forests* 2017, **48**(2):263-274.
83. Orwig D, Foster D: **Forest response to the introduced hemlock woolly adelgid in southern New England, USA.** *Journal of the Torrey Botanical Society* 1998, **1**:60-73.
84. Ramsfield T, Bentz B, Faccoli M, Jactel H, Brockerhoff E: **Forest health in a changing world: effects of globalization and climate change on forest insect and pathogen impacts.** *Forestry* 2016, **89**(3):245-252.
85. Trumbore S, Brando P, Hartmann H: **Forest health and global change.** *Science* 2015, **349**(6250):814-818.
86. Boraks A, Broders K: **Population genetic diversity of the rare hardwood butternut (*Juglans cinerea*) in the northeastern USA.** *Tree Genetics & Genomes* 2016, **12**(3).
87. Millar C, Stephenson N: **Temperate forest health in an era of emerging megadisturbance.** *Science* 2015, **349**(6250):823-826.
88. Allen C, Breshears D, McDowell N: **On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene.** *Ecosphere* 2015, **6**(8):1-55.

89. Klos R, Wang G, Bauerle W, Rieck J: **Drought impact on forest growth and mortality in the southeast USA: an analysis using forest health and monitoring data.** *Ecological Applications* 2009, **19**(3):699-708.
90. Mutch R, Arno S, Brown J, Carlson C, Ottmar R, Peterson J: **Forest health in the Blue Mountains: a management strategy for fire-adapted ecosystems.** In. Edited by US Department of Agriculture FS, vol. 310; 1993.
91. Wear D, Greis J: **Southern forest resource assessment: summary of findings.** *Journal of Forestry* 2002, **100**(7):6-14.
92. Hayes M, Moody A, White P, Costanza J: **The influence of logging and topography on the distribution of spruce-fir forests near their Southern limits in Great Smoky Mountains National Park, USA.** *Plant Ecology* 2007, **189**(1):59-70.
93. Pyle C: **Pre-park disturbance in the spruce-fir forests of Great Smoky Mountains National Park.** In: *The Southern Appalachian Spruce-Fir Ecosystem: Its Biology, Its Threats.* Edited by US Department of Interior NPS, Research/Resource Management; 1984: 115–130.
94. McLaughlin S, Andersen C, Hanson P, Tjoelker M, Roy W: **Increased dark respiration and calcium deficiency of red spruce in relation to acidic deposition at high-elevation southern Appalachian Mountain sites.** *Canadian Journal of Forest Research* 1991, **21**(8):1234-1244.
95. White P, Van de Gevel S, Soulé P: **Succession and disturbance in an endangered red spruce– Fraser fir forest in the southern Appalachian**

- Mountains, North Carolina, USA. *Endangered Species Research* 2012, 18(1):17-25.**
96. Petit R, Hampe A: **Some evolutionary consequences of being a tree.** *Annual Review of Ecology, Evolution, and Systematics* 2006, **37**:187-214.
 97. Hamrick J, Godt M, Sherman-Broyles S: **Factors influencing levels of genetic diversity in woody plant species.** 1992, **6**:95-124.
 98. Wang Z, Kang M, Liu H, Gao J, Zhang Z, Li Y, Wu R, Pang X: **High-level genetic diversity and complex population structure of Siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers.** *PloS One* 2014, **9**(2):e87381.
 99. Nybom H: **Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants.** *Molecular Ecology* 2004, **13**(5):1143-1155.
 100. Brunet J, Zalapa J, Guries R: **Conservation of genetic diversity in slippery elm (*Ulmus rubra*) in Wisconsin despite the devastating impact of Dutch elm disease.** *Conservation Genetics* 2016, **17**(5):1001-1010.
 101. Chung M, Chung M, Oh G, Epperson B: **Spatial genetic structure in a *Neolitsea sericea* population (Lauraceae).** *Heredity* 2000, **85**(5):490.
 102. Hardesty B, Dick C, Kremer A, Hubbell S, Bermingham E: **Spatial genetic structure of *Simarouba amara* Aubl. (Simaroubaceae), a dioecious, animal-dispersed Neotropical tree, on Barro Colorado Island, Panama.** *Heredity* 2005, **95**(4):290.

103. Dean D, Wadl P, Hadziabdic D, Klingeman W, Ownley B, Rinehart T, Dattilo A, Scheffler B, Trigiano R: **Analysis of genetic diversity and population structure for the native tree *Viburnum rufidulum* occurring in Kentucky and Tennessee.** *Journal of the American Society for Horticultural Science* 2015, **140**(6):523-531.
104. Duan H, Hu R, Wu B, Chen D, Huang K, Dai J, Chen Q, Wei Z, Cao S, Sun Y: **Genetic characterization of red-colored heartwood genotypes of Chinese fir using simple sequence repeat (SSR) markers.** *Genetics and Molecular Research* 2015, **14**(4):18552-18561.
105. Bednorz L, Kosiński P: **Genetic variability and structure of the wild service tree (*Sorbus torminalis* L. (crantz)) in Poland.** *Silvae Genetica* 2006, **55**(1-6):197-202.
106. Roberts D, Werner D, Wadl P, Trigiano R: **Inheritance and allelism of morphological traits in eastern redbud (*Cercis canadensis* L.).** *Horticulture Research* 2015, **2**:15049.
107. Tucker S: **Floral ontogeny of *Cercis* (Leguminosae : Caesalpinioideae : Cercideae): Does it show convergence with papilionoids?** *International Journal of Plant Sciences* 2002, **163**(1):75-87.
108. Pasquet R, Peltier A, Hufford M, Oudin E, Saulnier J, Paul L, Knudsen J, Herren H, Gepts P: **Long-distance pollen flow assessment through evaluation of pollinator foraging range suggests transgene escape distances.** *Proceedings of the National Academy of Sciences* 2008, **105**(36):13456-13461.

109. Hagler J, Mueller S, Teuber L, Machtley S, Van Deynze A: **Foraging range of honey bees, *Apis mellifera*, in alfalfa seed production fields.** *Journal of Insect Science* 2011, **11**(1).
110. White G: **A study of population genetics of *Swietenia humilis* zucc. in fragmented forest.** *Doctoral dissertation.* University of Dundee; 1998.
111. Lowe A, Jourde B, Breyne P, Colpaert N, Navarro C, Wilson J, Cavers S: **Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (*Swietenia macrophylla*).** *Heredity* 2003, **90**(3):268-275.
112. Sork V, Smouse P: **Genetic analysis of landscape connectivity in tree populations.** *Landscape Ecology* 2006, **21**(6):821-836.
113. Dick C, Etchelecu G, Austerlitz F: **Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest.** *Molecular Ecology* 2003, **12**(3):753-764.
114. Nason J, Hamrick J: **Reproductive and genetic consequences of forest fragmentation: two case studies of neotropical canopy trees.** *Journal of Heredity* 1997, **88**(4):264-276.
115. White G, Boshier D, Powell W: **Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini.** *Proceedings of the National Academy of Sciences* 2002, **99**(4):2038-2042.
116. Bacles CF, Burczyk J, Lowe A, Ennos R: **Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior* L.** *Evolutionary Ecology Research* 2005, **59**(5):979-990.

117. Colabella F, Gallo L, Moreno A, Marchelli P: **Extensive pollen flow in a natural fragmented population of Patagonian cypress *Austrocedrus chilensis*.** *Tree Genetics & Genomes* 2014, **10**(6):1519-1529.
118. Wang R, Stephen C, Xiao-Yong C: **Fragmentation can increase spatial genetic structure without decreasing pollen-mediated gene flow in a wind-pollinated tree.** *Molecular Ecology* 2011, **20**(21):4421-4432.
119. Vekemans X, Hardy O: **New insights from fine-scale spatial genetic structure analyses in plant populations.** *Molecular Ecology* 2004, **13**(4):921-935.
120. Hayden W: **Redbud seedpods hold surprises.** *Bulletin of the Virginia Native Plant Society* 2013, **32**(5):1-6.
121. Schnabel A, Laushman R, Hamrick J: **Comparative genetic structure of two co-occurring tree species, *Maclura pomifera* (Moraceae) and *Gleditsia triacanthos* (Leguminosae).** *Heredity* 1991, **67**(3):357-364.
122. Nakanishi A, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S: **Effects of seed- and pollen-mediated gene dispersal on genetic structure among *Quercus salicina* saplings.** *Heredity* 2008, **102**:182.
123. Gonzales E, Hamrick J, Smouse P, Trapnell D, Peakall R: **The impact of landscape disturbance on spatial genetic structure in the Guanacaste tree, *Enterolobium cyclocarpum* (Fabaceae).** *Journal of Heredity* 2009, **101**(2):133-143.
124. Post D: **Change in nutrient content of foods stored by eastern woodrats (*Neotoma floridana*).** *Journal of Mammalogy* 1992, **73**(4):835-839.

125. Wright V, Fleming E, Post D: **Survival of *Rhyzopertha dominica* (Coleoptera, Bostrichidae) on fruits and seeds collected from woodrat nests in Kansas.** *Journal of the Kansas Entomological Society* 1990:344-347.
126. Wakeland B, Swihart R: **Ratings of white-tailed deer preferences for woody browse in Indiana.** *Proceedings of the Indiana Academy of Science* 2009, **118**(1):96-101.
127. Setoguchi M: **Food Habits of Red-Bellied Tree Squirrels on a Small Island in Japan.** *Journal of Mammalogy* 1990, **71**(4):570-578.
128. Chung M, Nason J, Epperson B, Chung M: **Temporal aspects of the fine-scale genetic structure in a population of *Cinnamomum insularimontanum* (Lauraceae).** *Heredity* 2003, **90**(1):98.
129. Short H, Epps E: **Nutrient quality and digestibility of seeds and fruits from southern forests.** *The Journal of Wildlife Management* 1976, **40**(2):283.
130. Halls L, Crawford H: **Deer-forest habitat relationships in north arkansas.** *The Journal of Wildlife Management* 1960, **24**(4):387-395.
131. Koprowski J: **The response of tree squirrels to fragmentation: a review and synthesis.** In: *Animal Conservation Forum: 2005*. Cambridge University Press: 369-376.
132. Yusnita S, Geneve R, Kester S: **Micropropagation of white flowering eastern redbud (*Cercis canadensis* var. *alba* L.).** *Journal of Environmental Horticulture* 1990, **8**(4):177-179.

133. Abu-Asab M, Peterson P, Shetler S, Orli S: **Earlier plant flowering in spring as a response to global warming in the Washington, DC area.** *Biodiversity Conservation* 2001, **10**(4):597-612.
134. Kamvar Z, Brooks J, Grünwald N: **Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality.** *Frontiers in Genetics* 2015, **6**.
135. Goudet J: **Hierfstat, a package for R to compute and test hierarchical F-statistics.** *Molecular Ecology Notes* 2005, **5**(1):184-186.
136. Meirmans p, Hedrick p: **Assessing population structure: F_{ST} and related measures.** *Molecular Ecology Resources* 2011, **11**(1):5-18.
137. Cullingham C, Cooke J, Coltman D: **Effects of introgression on the genetic population structure of two ecologically and economically important conifer species: lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*Pinus banksiana*).** *Genome* 2013, **56**(10):577-585.
138. Cornuet J, Ravigné V, Estoup A: **Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1. 0).** *BMC Bioinformatics* 2010, **11**(1):401.
139. Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J, Estoup A: **DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data.** *Bioinformatics* 2014, **30**(8):1187-1189.

140. Kim J, Ni G, Kim T, Chun J, Kern E, Park J: **Phylogeography of the highly invasive sugar beet nematode, *Heterodera schachtii* (Schmidt, 1871), based on microsatellites.** *Evolutionary Applications* 2019, **12**(2):324-336.
141. Pedreschi D, Kelly-Quinn M, Caffrey J, O'Grady M, Mariani S: **Genetic structure of pike (*Esox lucius*) reveals a complex and previously unrecognized colonization history of Ireland.** *Journal of Biogeography* 2014, **41**(3):548-560.
142. Zhang J, Yao J, Hu Z, Jueterbock A, Yotsukura N, Krupnova T, Nagasato C, Duan D: **Phylogeographic diversification and postglacial range dynamics shed light on the conservation of the kelp *Saccharina japonica*.** *Evolutionary Applications* 2019, **12**(4):791-803.
143. Graignic N, Tremblay F, Bergeron Y, evolution: **Influence of northern limit range on genetic diversity and structure in a widespread North American tree, sugar maple (*Acer saccharum* Marshall).** *Ecology* 2018, **8**(5):2766-2780.
144. Bemmels JB, Dick CW: **Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species.** *Journal of Biogeography* 2018, **45**(8):1739-1750.
145. Potter K, Jetton R, Dvorak W, Hipkins V, Rhea R, Whittier W: **Widespread inbreeding and unexpected geographic patterns of genetic variation in eastern hemlock (*Tsuga canadensis*), an imperiled North American conifer.** *Conservation Genetics* 2012, **13**(2):475-498.
146. Thomson A, Dick C, Dayanandan S: **A similar phylogeographical structure among sympatric North American birches (*Betula*) is better explained by**

- introgression than by shared biogeographical history.** *Journal of Biogeography* 2015, **42**(2):339-350.
147. Hoban S, Borkowski D, Brosi S, McCleary T, Thompson L, McLachlan J, Pereira M, Schlarbaum S, Romero-Severson J: **Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: a product of range shifts, not ecological marginality or recent population decline.** *Molecular Ecology* 2010, **19**(22):4876-4891.
 148. Hoban S, McCleary T, Schlarbaum S, Romero-Severson J: **Spatial genetic structure in 21 populations of butternut, a temperate forest tree (*Juglans cinerea* L.), is correlated to spatial arrangement, habitat, and land-use history.** *Forest Ecology and Management* 2014, **314**:50-58.
 149. McLachlan J, Clark J, Manos P: **Molecular indicators of tree migration capacity under rapid climate change.** *Ecology* 2005, **86**(8):2088-2098.
 150. Havill N, Vieira L, Salom S: **Biology and control of hemlock woolly adelgid.** In. Edited by US Department of Agriculture FS, Forest Health Technology Enterprise Team., vol. 21; 2014: 1-21.
 151. Spaulding H, Rieske L: **The aftermath of an invasion: Structure and composition of central appalachian hemlock forests following establishment of the hemlock woolly adelgid, *Aelges tsugae*.** *Biological Invasions* 2010, **12**(9):3135-3143.
 152. Temunović M, Franjić J, Satovic Z, Grgurev M, Frascaria-Lacoste N, Fernández-Manjarrés J: **Environmental heterogeneity explains the genetic structure of**

- continental and mediterranean populations of *Fraxinus angustifolia* vahl.**
PloS One 2012, **7**(8):e42764.
153. Petit R, Aguinalalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M: **Glacial refugia: hotspots but not melting pots of genetic diversity.** *Science* 2003, **300**(5625):1563-1565.
 154. Donselman H: **Variation in native populations of eastern redbud (*Cercis canadensis* L.) as influenced by geographic location [USA].** *Proceedings, of the Florida State Horticulture Society* 1976, **89**:370-373.
 155. Fritsch P, Schiller A, Larson K: **Taxonomic implications of morphological variation in *Cercis canadensis* (Fabaceae) from Mexico and adjacent parts of Texas.** *Systematic Botany* 2009, **34**(3):510-520.
 156. Fritsch P, Nowell C, Leatherman L, Gong W, Cruz B, Burge D, Delgado-Salinas A: **Leaf adaptations and species boundaries in North American *Cercis*: implications for the evolution of dry floras.** *American Journal of Botany* 2018, **105**(9):1577-1594.
 157. Peppe D, Royer D, Cariglino B, Oliver S, Newman S, Leight E, Enikolopov G, Fernandez-Burgos M, Herrera F, Adams J: **Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications.** *New Phytologist* 2011, **190**(3):724-739.
 158. Nevo E, Pavlek T, Beharav A, Mariya A, Gennady I, Ludmila I, Konstatin M: **Drought and light anatomical adaptive leaf strategies in three woody species caused by microclimatic selection at Evolution Canyon, Israel.** *Israel Journal of Plant Sciences* 2000, **48**(1):33-46.

159. Robertson K: **Cercis: The Redbuds**. *Arnoldia* 1976, **36**(2):37-49.
160. Donselman H, Flint H: **Genecology of eastern redbud (*Cercis canadensis*)**. *Ecology* 1982, **63**(4):962-971.
161. Hopkins M: ***Cercis* in North America**. In: *Rhodora*. Edited by Fernald M, Eatherby C, Griscom L, Marris S, vol. 44: New England Botanical Club, Inc.; 1942: 193-211.
162. Fritsch P, Cruz B: **Phylogeny of *Cercis* based on DNA sequences of nuclear ITS and four plastid regions: Implications for transatlantic historical biogeography**. *Molecular Phylogenetics Evolution* 2012, **62**(3):816-825.
163. David R, Hamann A: **Glacial refugia and modern genetic diversity of 22 western North American tree species**. *Proceedings of the Royal Society B: Biological Sciences* 2015, **282**(1804):20142903.
164. Aldrich P, Hamrick J: **Reproductive dominance of pasture trees in a fragmented tropical forest mosaic**. *Science* 1998, **281**(5373):103-105.
165. Weir B, Ott J: **Genetic data analysis II**. *Trends in Genetics* 1997, **13**(9):379.
166. Magni C, Ducousso A, Caron H, Petit R, Kremer A: **Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae**. *Molecular Ecology* 2005, **14**(2):513-524.
167. Peterson B, Graves W: **Chloroplast phylogeography of *Dirca palustris* L. indicates populations near the glacial boundary at the Last Glacial Maximum in eastern North America**. *Journal of Biogeography* 2016, **43**(2):314-327.

168. Shaw J, Small RL: **Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, *Rosaceae*)**. *American Journal of Botany* 2005, **92**(12):2011-2030.
169. Call A, Sun Y, Yu Y, Pearman P, Thomas D, Trigiano R, Carbone I, Xiang Q: **Genetic structure and post-glacial expansion of *Cornus florida* L.(Cornaceae): integrative evidence from phylogeography, population demographic history, and species distribution modeling**. *Journal of Systematics Evolution* 2016, **54**(2):136-151.
170. Soltis D, Morris A, McLachlan J, Manos P, Soltis P: **Comparative phylogeography of unglaciated eastern North America**. *Molecular Ecology* 2006, **15**(14):4261-4293.
171. Li P, Li M, Shi Y, Zhao Y, Wan Y, Fu C, Cameron K: **Phylogeography of North American herbaceous *Smilax* (Smilacaceae): combined AFLP and cpDNA data support a northern refugium in the driftless area**. *American Journal of Botany* 2013, **100**(4):801-814.
172. Rowe K, Heske E, Brown P, Paige K: **Surviving the ice: Northern refugia and postglacial colonization**. *Proceedings of the National Academy of Sciences* 2004, **101**(28):10355-10359.
173. Potter K, Jetton R, Bower A, Jacobs D, Man G, Hipkins V, Westwood M: **Banking on the future: progress, challenges and opportunities for the genetic conservation of forest trees**. *New Forests* 2017, **48**(2):153-180.

174. Taberlet P, Fumagalli L, Wust-Saucy A, Cosson J: **Comparative phylogeography and postglacial colonization routes in Europe.** *Molecular Ecology* 1998, **7**(4):453-464.
175. Jia H, R S: **Fossil leaves and fruits of *Cercis* L. (Leguminosae) from the Eocene of western North America.** *International Journal of Plant Sciences* 2014, **175**(5):601-612.

VITA

Meher Afroze Ony was born in Chittagong, Bangladesh on June 24, 1991. She graduated with a Bachelor of Science (BSc) in Zoology from University of Chittagong, Bangladesh in 2016. After completing her BSc, she joined the University of Tennessee in the Department of Entomology and Plant Pathology under the mentorship of Dr. Denita Hadziabdic Guerry to pursue her Masters in Plant Pathology.