PHYLOGEOGRAPHY AND POPULATION GENETICS OF WILD AND ANTHROPOGENIC POPULATIONS OF A HIGHLY CLONAL TREE SPECIES, *ASIMINA TRILOBA* (ANNONACEAE)

by

GRAHAM EDWARD WYATT

(Under the Direction of DORSET W. TRAPNELL)

ABSTRACT

To date, the relatively scarce large-scale phylogeographic studies of eastern North American flora have primarily focused on canopy trees with wind-dispersed seeds. Given the varied topography of eastern North America, possible locations of Pleistocene refugia, and differential rates of post-glacial migration from those refugia, overall phylogeographic patterns are less well-studied than other temperate regions around the world and appear to be largely species-dependent. It had also been hypothesized that human-mediated dispersal of useful plant species may have affected the rate of post-LGM migration and range expansion of some valued species during the Holocene epoch. It seems timely to conduct a regional phylogeographic study on patterns of genetic diversity of the highly clonal, mammal-dispersed, useful understory tree species, *Asimina triloba*. The goals of this work are to 1) estimates levels and patterns of genetic diversity and structure across the species' range, 2) infer locations of Pleistocene refugia, 3) compare patterns of genetic diversity of putative anthropogenic populations with that of wild populations, and 4) investigate levels of clonality within populations across the species range. For a tree species, genetic diversity and structure among *A. triloba* populations was moderately

high, though isolation-by-distance was not significant. The Appalachian Mountains appear to have acted as a barrier to dispersal during northward post-glacial migration and range expansion. Another east-west genetic discontinuity was detected on either side of the Tombigbee River in Alabama. These results are consistent with the inference by ecological niche modeling of two principal Pleistocene refugia, separated by peninsular Florida. Most measures indicated that genetic diversity was significantly lower in putative anthropogenic populations than wild populations, with the exception of heterozygosity, which was higher in anthropogenic populations. This work represents the first genetic evidence for human manipulation of a valued eastern North American tree species. Clonality was observed to vary greatly across the range, with more cloning on the edges of the species range and high heterozygosity of dominant clones. Spatial genetic structure within populations appeared to be strongly linked to clonal growth. This study represents the first regional phylogeographic study on a mammal-dispersed understory tree in eastern North America.

INDEX WORDS: Phylogeography; *Asimina triloba*; population genetics; clonality

PHYLOGEOGRAPHY AND POPULATION GENETICS OF WILD AND ANTHROPOGENIC POPULATIONS OF A HIGHLY CLONAL TREE SPECIES, *ASIMINA TRILOBA* (ANNONACEAE)

by

GRAHAM EDWARD WYATT

B.S., Clemson University, 2006

M.S., Universidade Estadual de Campinas, Brazil, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

PHYLOGEOGRAPHY AND POPULATION GENETICS OF WILD AND ANTHROPOGENIC POPULATIONS OF A HIGHLY CLONAL TREE SPECIES, *ASIMINA TRILOBA* (ANNONACEAE)

by

GRAHAM EDWARD WYATT

Major Professor:	Dorset W. Trapnell
Committee:	James L. Hamrick
	Shu-mei Chang
	James H. Leebens-Mack
	Chris J. Peterson

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
December 2019

DEDICATION

I dedicate this dissertation to my parents, Ann Stoneburner and Robert Wyatt. This work would not have been possible without their perennial love and support.

ACKNOWLEDGEMENTS

I would like to acknowledge the following people and institutions for their assistance throughout the process of forming this dissertation. I am grateful to Dorset Trapnell and Jim Hamrick for providing tremendous support and ideas in the development of projects, analysis of data, and writing of this dissertation. Also, I am grateful for the contribution of ideas and advice from committee members Shu-mei Chang, Jim Leebens-Mack, and Chris Peterson. I acknowledge the Department of Plant Biology at the University of Georgia for providing a home-base and support, especially Susan Watkins, Chelsea Hargroves, and Richard Hare. I am thankful for the analytical contributions from Rodney Dyer at Virginia Commonwealth University. I would also like to acknowledge Dat Hoang, Patrick Smallwood, and Piotrek Tuczapski for their assistance in the laboratory, along with Patrick Mclean, Kirsten Jackson, Jessica Brown, Dan Bui, and Lura Lockard. The extensive fieldwork undertaken in this study was made lighter on several occasions with assistance from Cody Barber and Earl Kang.

This dissertation would not be complete without acknowledging the substantial contribution of all my family and friends. My parents, to whom this work is dedicated, provided unfaltering love and support at every juncture. Jenny Kim brought a light into my life, and since the day we met it has guided me during the best of times and carried me at the worst. To them and the rest of my friends and family, I am eternally grateful.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Dispersal and colonization are among the most important processes for the evolutionary persistence of species. The importance of dispersal was recognized by Darwin (1859) having observed many geographically disjunct species during his travels. These processes can have a strong effect on species ranges through time, which shift in response to changing environmental conditions. This in turn has cascading effects on patterns of genetic diversity within a species, as range shifts are often not uniform due to the spatial heterogeneity of landscapes (Lyford et al. 2003). As such, patterns of genetic diversity within a species can provide inferences into past environmental changes, the effect of landscapes on dispersal and colonization, as well as predictions for species persistence in landscapes where they occur.

Studies of phylogeographic patterns following glacial retreat after the Last Glacial Maximum (LGM) in eastern North America are scarce, less well defined than in other regions of the world, and until recently have been primarily focused on animals (Soltis 2006; Park & Donoghue 2019). Phylogeographical structure within species of eastern North America have been suggestive of complex patterns operating over various spatial and temporal scales, rather than reflective of few geographic barriers (Soltis 2006; Park & Donoghue 2019).

The means by which a species disperses its seeds, along with the rate of reproduction, can determine the frequency of long-distance dispersal and colonization of new habitats, potentially leading to range expansion. This adds to the complexity of plant species responses to climate shifts following the LGM. Dispersal typically occurs only during short phases of a plant's life cycle. Plants have evolved numerous mechanisms by which seeds are dispersed. Endozoochory is a relatively successful means of long-distance seed dispersal, in which animals ingest seeds and later excrete them at potentially long distances from the maternal plant (Jordano et al. 2007). Phylogeographic studies in Europe and North America have largely focused on wind-dispersed seeds (Cain et al. 1998; Petit et al. 2003; Roberts & Hamann 2015; Park & Donoghue 2019). Thus, many inferences regarding rate of post glacial range expansion and resulting phylogeography of tree species are biased towards wind-dispersed taxa. These studies on species with gravity- or animal-dispersed seeds have found that they migrated exceptionally long distances over the last ~20,000 years, a phenomenon termed as "Reid's paradox" (Bennett 1985; Webb 1986; Johnson & Webb 1989; Betancourt et al. 1991; Cain et al. 1998; Clark et al. 1998).

Other life history traits certainly affect the ability of a plant species to persist within habitats, disperse to new habitats, and successfully colonize following dispersal. Among these are mating system, growth form, ability to asexually propagate, generation time, and rate of reproduction. Given the diversity of life history traits among the flora of eastern North America, phylogeographic responses to climatic shifts following the LGM are likely varied and highly species dependent and likely influenced differentially by an assortment of barriers (Soltis et al. 2006).

Despite the varied responses to climatic shifts and multitude of possible physical barriers to dispersal, the effect of eastern North American physiography has produced some recurrent phylogeographic patterns across a variety of plant taxa (Soltis et al. 2006). Rivers such as the Mississippi, Apalachicola, and Tombigbee along with the Appalachian Mountains appear to have served as barriers to dispersal and gene flow in some taxa as ranges expanded northward during glacial retreat (Soltis et al. 2006; Miller & Parker 2009).

The effect of historic human-mediated dispersal has been rarely studied for the flora of eastern North America, likely owing to the difficulty of obtaining precise knowledge of interactions between Native Americans and the plants they relied on. Historical accounts of plant use and dispersal by indigenous peoples often contain observer bias and limited botanical knowledge on the part of the observer, making it difficult to document dispersal by anthropogenic means during pre-contact eastern North America (MacDougal 2003). Associations between useful plant species and pre-contact indigenous settlements have been observed in several parts of the New World (Yarnell 1964, 1965; Parker et al. 2010; Larranaga et al. 2017; Levis et al. 2017). It appears likely that pre-Columbian civilizations influenced the dispersal of some desirable species (White 1906; Yarnell 1964; Brooks & Johannes 1990; Wykoff 1991; Asch 1995; Keener & Kuhns 1997; Abrams & Nowacki 2008; Duran et al. 2012), especially given the long distances (>1,000 km) over which goods were exchanged as far back as 4,000 YBP (Sanger et al. 2018). Molecular markers have provided new tools in understanding the dispersal of species following the last glacial maximum, and in addition aid in the disentangling of human-mediated versus natural dispersal.

Several authors have noted the importance of *Asimina triloba* to pre-Columbian humans as both a valued food source and perennial source of textile material and medicine (Peattie 1950; Krochmal & Krochmal 1973; Brooks & Johannes 1990; Peterson 1991; Wykoff 1991; Keener & Kuhns 1997). Given the probable importance of *A. triloba*, many of the same authors have suggested that pre-Columbian humans were likely responsible for the appearance of *A. triloba* at the northern limits of its current range (White 1906; Brooks & Johannes 1990; Peterson 1991; Wykoff 1991; Keener & Kuhns 1997). Murphy (2001), however, counters that idea, contending that present-day mammalian dispersers are more than sufficient to account for the spread of *A. triloba* into the northern limits of its range.

Preliminary investigations into the genetic variation of *A. triloba* have focused on cultivars (Pomper et al. 2003; Pomper et al. 2010), yet still found moderate to high levels of genetic diversity. This suggests that high levels of diversity should exist in wild populations. As a highly clonal tree species, little work has been done on the degree of clonality across the range of *A. triloba*. Pomper et al. (2009) investigated six patches in central Kentucky, finding that half of those consisted of a single multiclocus genotype (MLG). Hosaka et al. (2005) conducted a three-year observational study on the production of clonal stems of *A. triloba* in one Maryland population and found that root-suckering outpaced stem mortality and may be an important means of persistence in habitats. They also found that newly produced clones grow at the same rate as established stems.

Asimina triloba possess a number of life history traits that made it a particularly interesting candidate for phylogeographic investigation. Fruits appear to have evolved for

mammalian dispersal (Willson 1993), although the best co-evolved seed dispersers may have become extinct by 12,000 YBP, making the fruits of *A. triloba* and example of a temperate evolutionary anachronism (Janzen & Martin 1982). Within recent times, seeds of *A. triloba* have been observed to have been excreted by raccoons (*Procyon lotor*) (Yeager & Elder 1945; Willson & Schemske 1980; Willson 1993), red fox (*Vulpes vulpes*), opossums (*Didelphis virginiana*) (Willson 1993), and coyote (*Canis latrans*) (Cypher & Cypher 1999). Observations of *A. triloba* seeds in the scat of black bear (*Ursus americanus*) have not been seen, although they have been observed to eat the fruit from Virginia and North Carolina (Willson 1993). Captive whitetail deer have been observed to consume fruits of *A. triloba* (Murphy 2001), although no record of *A. triloba* seeds in whitetail deer scat or observations of consumption of the fruits in the wild exist. Turkey have been observed to consume the fruits (Murphy 2001) and seeds (G. Wyatt pers. obs.). Principal pollinators of *A. triloba* are weak-flying insects such as flies and beetles, and have not been thought to move pollen long distances (Willson & Schemske 1980; Goodrich et al. 2006; G. Wyatt pers. obs.).

When considering the following regarding *A. triloba*: 1) the current geographic range, 2) the extinction of eastern North American megafauna (hypothesized seed dispersers) approximately 12,000 YBP and the current reliance on small to medium-sized mammals as primary seed dispersers, and 3) the apparent likely and widely-reported spread of *A. triloba* through anthropogenic means, along with the scarcity of phylogeographical studies on highly clonal, animal-dispersed tree species in eastern North America, an investigation into the range-wide phylogeography, history of anthropogenic dispersal, and clonality of this interesting and understudied species appears well-warranted and overdue.

The goals of this study of *A. triloba* were to 1) estimate range-wide patterns of genetic diversity and structure as well as infer Pleistocene refugia through ecological niche modeling, 2) investigate differences in genetic diversity between populations of putative anthropogenic origin (i.e., those occurring on or adjacent to known Native American settlements) and wild populations, 3) determine the degree of clonality across the range and the effect of clonality on the spatial genetic structure within populations. To address these questions, the nuclear genetic diversity of *A. triloba* was characterized using nine microsatellite markers developed by Lu et al. (2011).

LITERATURE CITED

Abrams MD, Nowacki GJ (2008) Native Americans as active and passive promoters of mast and fruit trees in the eastern USA. *The Holocene* 18: 1123-1137.

Asch DL (1995) Aboriginal specialty-plant propagation: Illinois prehistory and an eastern North American post-contact perspective. *Unpublished Ph.D. dissertation*, *Department of Anthropology*, University of Michigan, Ann Arbor.

Bennett KD (1985) The spread of *Fagus grandifolia* across eastern North America during the last 18,000 years. *Journal of Biogeography* 12: 147 – 164.

Bentancourt JL, Schuster WS, Mitton JB, Anderson RS (1991) Fossil and genetic history of a pinyon pine (*Pinus edulis*) isolate. *Ecology* 72: 1685-1697.

Brooks RR, Johannes D (1990) *Historical, ethno-, and economic botany series*, Vol. 3. Dioscorides Press, Portland, OR.

Cain ML, Damman H, Muir A (1998) Seed dispersal and the Holocene migration of woodland herbs. *Ecological Monographs* 68: 325 – 347.

Clark JS, Fastie C, Hurtt G, Jackson ST, Johnson C, King GA, Lewis M, Lynch J, Pacala S, Prentice C, Schupp EW, Webb III T, Wyckoff P (1998) Reid's Paradox of rapid plant migration. *BioScience* 48: 13-24.

Cypher BL, Cypher EA (1999) Germination rates of tree seeds ingested by coyotes and raccoons. *American Midland Naturalist* 142: 71 – 76.

Darwin C, Duthie JF, Hopkins W (1859) On the origin of species by means of natural selection: Or, the preservation of favoured races in the struggle for life. London: John Murray.

Duran KL, Pardo A, Mitton JB (2012) From middens to molecules: phylogeography of the piñon pine, *Pinus edulis*. *Journal of Biogeography* 39: 1536–1544.

Goodrich KR, Zjhra M, Ley CA, Raguso RA (2006) When flowers smell fermented: The chemistry and ontogeny of yeasty floral scent in pawpaw (*Asimina triloba*: Annonaceae). *International Journal of Plant Sciences* 167: 33 – 46.

Hosaka N, Gómez S, Kachi N, Stuefer JF, Whigham DF (2005) The ecological significance of clonal growth in the understory tree, Pawpaw (*Asimina triloba*). *Northeastern Naturalist* 12: 11 – 22.

Janzen DH, Martin PS (1982) Neotropical anachronisms: The fruits the Gomphotheres ate. *Science* 215: 19 – 27.

Jordano P, García C, Godoy JA, García-Castaño JL (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences* 104: 3278 – 3282.

Johnson CT, Webb III T (1989) The role of blue jays (*Cyanocitta cristata* L.) in the postglacial dispersal of Fagaceous trees in eastern North America. *Journal of Biogeography* 16: 561 – 571.

Keener C, Kuhns E (1997) The impact of Iroquoian populations on the northern distribution of pawpaws in the Northeast. *North American Archaeologist* 18: 327–342.

Krochmal A, Krochmal C (1973) A guide to the medicinal plants of the U.S. New York: New York Times Books Co.

Larranaga N, Albertazzi FJ, Fontecha G, Palmieri M, Rainer H, van Zonneveld M, Hormaza JI (2017) A Mesoamerican origin of cherimoya (*Annona cherimola* Mill.): Implications for the conservation of plant genetic resources. *Molecular Ecology* 26: 4116 – 4130.

Levis C, Costa F, Bongers F, Peña-Claros M (2017) Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* 355: 925 – 931.

Lu L, Pomper KW, Lowe JD, Crabtree SB (2011) Genetic variation in pawpaw cultivars using microsatellite analysis. *Journal of the American Society of Horticultural Science* 136: 415–521.

Lyford ME, Jackson ST, Betancourt JL, Gray ST (2003) Influence of landscape structure and climate variability on a late Holocene plant migration. *Ecological Monographs* 73: 567 – 583.

MacDougall A (2003) Did Native Americans influence the northward migration of plants during the Holocene? *Journal of Biogeography* 30: 633–647.

Miller DM, Parker KC (2009) The impacts of the Appalachian Mountains on the post-glacial migration pathways and gene flow of sugar maple (*Acer saccharum*). *Physical Geography* 30: 89 – 104.

Murphy JL (2001) Pawpaws, persimmons, and 'possums: On the natural distribution of pawpaws in the northeast. *North American Archaeologist* 22: 93-115.

Park B, Donoghue MJ (2019) Phylogeography of a widespread eastern North American shrub, *Viburnum lantanoides*. *American Journal of Botany* 106: 389 – 401.

Parker KC, Trapnell DW, Hamrick JL, Hodgson WC, Parker AJ (2010) Inferring ancient Agave cultivation practices from contemporary genetic patterns. Molecular Ecology, 19, 1622-1637.

Peattie DC (1950) A natural history of trees of eastern and central North America. Houghton Mifflin Co., Boston.

Peterson RN (1991) Pawpaw (*Asimina*). *Acta Horticulturae* 290: 567 – 600.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563 – 1565.

Pomper KW, Crabtree SB, Brown SP, Snake CJ, Bonney TM (2003) Assessment of genetic diversity of pawpaw (*Asimina triloba*) cultivars with intersimple sequence repeat markers. *Journal of American Society of Horticultural Science* 128: 521 – 525.

Pomper KW, Lowe JD, Lu L (2009) Clonality of pawpaw (*Asimina triloba*) patches in Kentucky. *Journal of the Kentucky Academy of Science* 70: 3 – 11.

Pomper KW, Lowe JD, Lu L, Crabtree SB, Dutta S, Schneider K, Tidwell J (2010) Characterization and identification of pawpaw cultivars and advanced selections by simple sequence repeat markers. *Journal of the American Society for Horticulture Science* 135: 143 – 149.

Roberts DR, Hamann A (2015) Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B* 282: 20142903.

Sanger MC, Hill MA, Lattanzi GD, Padgett BD, Larsen CS, Culleton BJ, Kennett DJ, Dussubieux L, Napolitano MF, Lacombe S, Thomas DH (2018) Early metal use and crematory practices in the American Southeast. *Proceedings of the National Academy of Sciences* 115: 7672 – 7679.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15: 4261-4293.

Webb SL (1986) Potential role of passenger pigeons and other vertebrates in the rapid Holocene migrations of nut trees. *Quaternary Research* 26: 367 – 375.

White CA (1906) The northern limit of the Papaw tree. *Science* 23: 749-751.

Willson MF (1993) Mammals as seed-dispersal mutualists in North America. *Oikos* 67: 159 – 176.

Willson MF, Schemske DW (1980) Pollinator limitation, fruit production, and floral display in Pawpaw (*Asimina triloba*). *Bulletin of the Torrey Botanical Club* 107: 401 – 408.

Wykoff MW (1991) Black walnut on Iroquoian landscapes. *Northeast Indian Quarterly* 8: 4 – 17.

Yarnell RA (1964) Aboriginal relationships between culture and plant life in the Upper Great Lakes region. *Museum of Anthropology*, *University of Michigan*, *Anthropology* paper No. 23. Ann Arbor: University of Michigan.

Yarnell RA (1965) Implications of distinctive flora on Pueblo ruins. *American Anthropologist* 67: 662 – 674.

Yeager LE, Elder WH (1945) Pre- and post-hunting season foods of raccoons on an Illinois goose refuge. *Journal of Wildlife Management* 9: 48 – 53.

CHAPTER 2

PHYLOGEOGRAPHY OF *ASIMINA TRILOBA* (ANNONACEAE)[1]

[1]Graham E. Wyatt. To be submitted to Journal of Heredity.

ABSTRACT

Among the most important ecological processes for the evolutionary persistence of a species are dispersal and colonization, which enable the species to undergo range modifications in response to changing environmental conditions and to persist over evolutionary time. Understanding historical patterns of plant dispersal and range expansion during the warming climate that ensued after the Last Glacial Maximum (~20,000 years before present.) provides insight into species' capacity to modify their ranges. Some tree species with apparently limited seed dispersal migrated remarkably long distances after the late Pleistocene glacial retreat, a phenomenon known as Reid's Paradox (Clark 1998). Large-scale phylogeographic studies of Eastern North American plant taxa and their response to post-glacial climate change are relatively scarce. We employed nuclear microsatellite loci to explore this phenomenon in a North American tree species. *Asimina triloba* (L.) Dunal (pawpaw). Pawpaw occurs in 26 states extending from the eastern U.S. westward to Nebraska and Texas, and spanning a latitudinal gradient from Florida to the Great Lakes region and southern Ontario, well north of the southern limit of the Laurentide Ice Sheet. We sampled 62 populations covering an east-west distance of 1,635 km and a north-south distance of 1,102 km. DNA was extracted from leaf samples and characterized with nine microsatellite loci. Global G_{ST} = 0.341 among all populations, which was higher than that for most tree species. No biologically significant relationship between genetic and geographic distances was observed. The number of alleles per population did not decrease significantly with increasing latitude. Private alleles were observed in 19 populations. Principal components and STRUCTURE analyses reveal a phylogeographic discontinuity of populations east and west of the Appalachian Mountains and Tombigbee River (ALA), which may explain the absence of a significant isolation-by-distance. Ecological niche modeling indicated that there were likely two Pleistocene (~22,000 years before present) refugia;

one on the Gulf of Mexico and the other on the southeastern Atlantic Coast which is consistent with present-day genetic structuring of the *A. triloba* populations.

INTRODUCTION

Among the most fundamentally important ecological processes for the evolutionary persistence of a species are dispersal and colonization. These mechanisms allow the species to undergo range modifications in response to continually changing environmental conditions and to persist over evolutionary time. During ice ages, less cold-tolerant northern hemisphere species typically retreated to southern, lower latitude refugia, and during warming periods following glacial melting there was often a leading edge of northward migration of these taxa. Understanding historical patterns of plant dispersal and range expansion during the warming climate that ensued after the Last Glacial Maximum (LGM) ~20,000 years before present (YBP) allows inference regarding their capacity to modify their ranges in response to shifting modern climatic regimes. Interestingly, some tree species with apparently limited seed dispersal abilities migrated remarkable distances after glacial retreat in the late Pleistocene, a phenomenon known as 'Reid's Paradox' (Clark et al. 1998; MacDougall 2003).

To achieve range modifications, most plants rely on their seeds for dispersal and colonization of newly available habitats. While cryptic within the span of a human lifetime, plant species are in fact undergoing continuous movement and changes in their distributions. It is only through the lens of time that natural range shifts of plant species generally become evident. For most of the earth's history, environmental change has been a gradual process allowing incremental shifts in distribution, although there is evidence that there have been episodes of extremely rapid change in the past that

were not linked to continent-wide extinctions (Hof et al. 2011). The difference in recent centuries and decades is not only that climate change has accelerated but that land-use change, urbanization, habitat degradation, and the modification of ecosystems have been rapid. Gradual environmental change leads to evolution, but too rapid environmental change can lead to extinction (e.g., the asteroid impact that led to mass extinction at the Cretaceous-Paleogene boundary). Understanding how species disperse, colonize new habitats, and alter their ranges is essential for predicting their response to accelerating global climate change associated with the Anthropocene (Zalasiewicz et al. 2011). Predictions of future responses to a changing environment are possible based on understanding how genetic patterns within and among extant populations have been shaped by historical processes. There is a critical need for such insights and predictive power in the face of anthropogenic threats to biodiversity that are both direct (e.g., over harvesting) and indirect (e.g., landscape disturbance, habitat loss and climate modification).

Understanding historical patterns of plant dispersal and movement during the warming climate that ensued after the LGM 33,000 to 26,500 years before present (YBP) and onset of deglaciation 20,000 to 19,000 YBP (Clark et al. 2009) allows inference regarding the capacity of plants to alter their ranges in response to shifting climatic regimes. Some tree species have migrated impressive distances after glacial retreat in the late Pleistocene despite apparent limited dispersal ability. Examples include species in the genera *Quercus*, *Fagus*, and *Castanea* (Johnson & Webb 1989; Clark et al. 1998). This disparity between expected and observed dispersal is also found in some herbaceous taxa, e.g., *Asarum canadense* (Cain et al. 1998). In 1899 Clement Reid was the first of many to express perplexity at the surprisingly rapid rate of northward migration from glacial refugia by large-seeded trees, such as oaks, into formerly glaciated regions. A diffusion model of migration was developed by Skellam (1951) who found that the observed migration rate in oaks greatly exceeded that expected. Rates of late Pleistocene range expansion observed in trees, on the order of

100-1000 m per year, appeared contrary to their life history traits and expected natural migration rate. Fossil pollen sequence data has been used to estimate migration rates and document this phenomenon in a variety of tree species (e.g., Clark 1998). Not only are surprising rates of migration often found but the rate of reproduction and propagule vagility appear to lack predictive power regarding the speed of migration. Various researchers have modelled the dispersal kernel that would fit observed rates of range expansion (e.g., Clark 1998; Clark et al. 1998; Phillips et al. 2008; Roques et al. 2010) and examined the role that seed mass, temperature and germinability play in migration speed (Graae et al. 2009). Some animal vectors have been found to be effective long-distance seed dispersers. For example, small mammals and larger carnivorous mammals dispersed *Prunus mahaleb* seeds 495 m and 990 m respectively (Jordano et al. 2007), while blue jays dispersed acorns up to 1.9 km (Darley-Hill & Johnson 1981). Furthermore, while not often discussed in regard to seed dispersal, Pleistocene megafauna that are now extinct or extripated from their former range could have played a significant role in dispersing large seeds. Mammoths (*Mammuthus columbi*), tapirs (*Tapirus veroensis*), bison (*Bison bison*), horses (*Equus sp*.), and mastodons (*Mamut americanum*) were present in Southeastern North America, as recently as 12,700 YBP in the case of mastodons (Perrotti 2018). There is also evidence that horses and mastodons underwent annual migrations in Southeastern North America that spanned distances in excess of 150 km (Hoppe & Koch 2007). Humans may have also influenced dispersal rates after the LGM (Bentancourt et al. 1991; MacDougal 2003; Abrams & Nowacki 2008).

MATERIAL AND METHODS

Study Species

Asimina is one of two temperate genera of the large (~128 genera and 2,300 species [Kral 1997]), and otherwise tropical custard apple family (Annonaceae) in the basal order Magnoliales (Huang et

al. 2000). Some species of *Asimina* are valued for their edible fruit, pharmaceuticals, pesticides, and fiber (Peattie 1950; Mabberley 1978; Ratanyake et al. 1992; McLaughlin & Hui 1993). In the southeastern U.S. there are eight native species of *Asimina*, most of which are narrowly distributed across Florida and southern Georgia. A notable exception is *A. triloba*, which occurs in 26 eastern U.S. states extending west beyond the Mississippi River to Nebraska, Kansas, and Texas and spanning a latitudinal range from Florida to the Great Lakes region and southern Ontario (Weakley 2012). This range extends well north of the southern limits of the Laurentide Ice Sheet ~20,000 YBP, whose uneven southern margin extended as far south as approximately 39° latitude (Dyke & Prest 1987).

Asimina triloba is an understory tree that spreads clonally via root suckers (Keener & Kuhns 1997) and tends to grow at lower elevations (< 350 m; Freeman & Hulbert 1985) in rich, moist soils next to streams (Murphy 2001). It is less frequently found growing in upland terrain (Keener & Kuhns 1997). Flowers emit a yeasty odor (Wyatt pers. obs.) and are pollinated by weak-flying beetles and flies (Muscidae and Sacrophagidae) (Kral 1997; Willson & Schemske 1980; Goodrich et al. 2006) which carry pollen only short distances. It is thought to be a strongly self-incompatible species (Pomper et al. 2010; unpublished data, G. Wyatt). *Asimina triloba* has the largest (5-16 cm long, 3-7 cm wide) edible fruits (berries) of any native North American tree. Fruits have a flavor that has been described as a cross between banana, mango and pineapple. The desirability of these fruits has fueled interest in cultivar development for the horticultural trade, but a major drawback for its commercial value is the fact that the fruits do not store or transport well (Murphy 2001). Seeds are typically dispersed by raccoons, coyotes, and foxes (Yeager & Elder 1945; Willson 1993; Cypher & Cypher 1999; obs pers, G. Wyatt).

Sampling

Leaf samples were collected from a mean of 29.6 individuals (range = 10 to 62) from each of 62 populations of *A. triloba* (Table 2.1; Figure 2.1) located throughout its range in the U.S. The mean pairwise geographic distance separating populations is 660.3 km (range = 0.31 to 1834.9 km). Sampled populations were typically in closed-canopy oak-hickory or alluvial forest along creeks or rivers. Leaves were collected from stems separated by ≥10 m to try to avoid sampling multiple ramets belonging to the same genet. The size of the sampling area was dependent on stem-density and varied considerably among populations. Where possible, young leaves with little herbivore damage were collected from mature, flower-bearing stems (≥3 m in height) and snap-frozen in liquid nitrogen for transport to the University of Georgia for analysis.

Microsatellite Genotyping

Frozen leaf tissue (~ 0.05 g) was pulverized and genomic DNA was extracted using a modified CTAB protocol (Doyle & Doyle 1983). DNA quality and quantity were evaluated using an ND-1000 Nanodrop® spectrophotometer. Nine nuclear microsatellite loci were amplified with markers developed by Lu et al. (2011; Table 2.2). A 3-primer PCR protocol was used whereby a CAG-tag sequence (Hauswaldt and Glenn 2003) was added to the 5' end of one primer and a third fluorescently labeled (FAM, HEX, or NED) primer identical to the CAG-tag was included. PCR amplification was carried out in 12.5 μL reaction volumes containing 3.25 μL molecular grade ddH2O, 2.5 μL 5X One Taq® standard reaction buffer (New England Biolabs, Ipswich, MA), 0.35 μL 25 mM MgCl solution (Sigma-Aldrich), 1.25 μL 10X (1 mg/mL) bovine serum albumin (Thermo-Fisher), 1 μL 2.5 mM dNTP mix (New England Biolabs, Ipswich, MA), 1 μL primer mix (0.5 μM CAG-tagged primer and 5 μM unlabeled primer), 0.45 μL 10 μM universal fluorescently

labaled CAG primer, 0.2 µL One Taq® Hot Start DNA polymerase (New England Biolabs, Ipswich, MA), and 2.5 µL of diluted template DNA (20 ng/µL).

Genetic Analyses

A test for null alleles was performed in Genepop version 4.2 using Brookfield's method (Raymond & Roussett 1995; Brookfield 1996). Because *A. triloba* is a clonal species it was necessary to check for the presence of duplicate MLGs within populations, assess whether these represented clones or resulted from chance alone, and if clones remove the duplicate MLGs from subsequent genetic analyses. The likelihood that identical MLGs represent ramets of the same genet (i.e., clones) and are not the result of chance, can be ascertained based on the number of loci, allele frequencies, and the number of sampled stems in that population. The multilocus probability of identity (PI_m) was calculated as $PI_m = \Pi_s (PI_s)$, where

$$PI_s = \sum_i x_i^4 + \sum_i \sum_{j>i} (2x_i x_j)^2$$

and x_i and x_j are the frequencies of the *i*th and jth allele at a locus, respectively. The probability of identity (P_I) was adjusted for sample size (N) by $P_I = 1 - [\,(1 - PI_m)^N]$. Duplicate genotypes within populations that represented clones were removed such that only one individual per MLG was included in subsequent analyses.

Genetic diversity statistics were estimated using GenAlEx version 6.51b2 (Peakall & Smouse 2006; Peakall & Smouse 2012). Measures of genetic diversity were percent polymorphic loci (P), total number of alleles per population (A_T), effective number of alleles (A_E), private alleles

(P_A; i.e., alleles found in a single population), rare alleles (A_{RARE}; alleles ≤ 0.1 frequency in a population), observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}). Population values were estimated and averaged across all populations while species level values were calculated by pooling data from all populations.

Observed heterozygosity (H_O) and Hardy-Weinberg expected heterozygosity (H_E) were compared for each locus in each population, using Wright's inbreeding coefficient (F_{IS}; Wright 1922, 1951) and deviations from Hardy-Weinberg expectations were tested for significance using χ^2 (Li & Horvitz 1953). The Bonferroni correction for multiple comparisons was applied using FSTAT (Goudet 1995).

Genetic Structure Analyses

A Bayesian clustering approach (STRUCTURE version 2.3.4; Pritchard et al., 2000) was used to estimate levels of genetic admixture among populations and the number of genetically distinct clusters (K). Ten independent simulations at each K-value from 1 to 20 were run, using a burn-in of 500,000 repetitions and a run length of 1,000,000 Markov chain Monte Carlo (MCMC) iterations. The admixture model was chosen to infer alpha (α). A model based on correlated allele frequencies with no *a priori* assumptions regarding sampling locations was selected. The Evanno et al. (2005) method was used to determine the optimal ΔK using Structure Harvester (Earl and vonHoldt 2012). A principal components analysis (PCA) of genotypes was used to check the validity of these findings through the R package "gstudio" (R Core Team 2013; Dyer 2016). Genetic structure among populations was estimated using Nei's (1973) G_{ST} in GenAlEx version 6.51b2 (Peakall & Smouse 2006; Peakall & Smouse 2012). G_{ST} was calculated for all

possible pairs of populations, among populations within the two genetic clusters identified in STRUCTURE, and globally (i.e., among all 62 populations).

Geographic Patterns

Various population-level statistics (MLG, A_T, A_E, H_O, and H_E) were linearly regressed against both latitude and longitude, and tested for significance, to determine if there was a north-south or east-west patterning of genetic diversity. Multiple regression analyses involving latitude and longitude were performed in order to determine the relative contribution of each to population-level statistics that appeared to be significant.

Isolation-by-Distance

Four estimates of isolation-by-distance (IBD; Wright 1943) were performed using regression: i) (G_{ST} / 1 - G_{ST}) and log geographic distance (Rousset 1997), ii) a Mantel (1967) test, iii) G_{ST} and geographic distance, and iv) conditional genetic distance with geographic distance (Dyer et al. 2010), which takes into account the total genetic covariance among all populations. Each test was evaluated for significance. Additional tests for IBD were performed within the genetic clusters identified by the STRUCTURE analysis.

Ecological Niche Modelling

The realized present-day niche of *A. triloba* was modeled using maximum entropy in Maxent 3.4.0 (Phillips et al. 2006; Phillips & Dudík 2008). Location data for all 62 populations in addition to 20 populations not included in this study were used to compile the presence-only data required by Maxent. These data were supplemented with GPS coordinates from 18 post-1992

herbarium specimens housed in seven herbaria: DEK, ILLS, IND, KSP, MO,TROY, and USF. Although there were 100 presence records widely distributed across species range of *A. triloba*, because Maxent can only use one occurrence record per grid cell (~1 km^2) of climatic layers in BIOCLIM, the number of usable records was reduced to 96.

Nineteen climate variables at 30 arc second resolution were obtained from WorldClim (Fick & Hijmans 2017) and elevation data was accessed from the USGS (Archuleta et al. 2017). Twelve climate variables were highly correlated (Pearson correlation coefficient > |0.7|) with the seven most explanatory variables and were therefore excluded from further analyses. Thus, with elevation data, eight variables were used for model construction (Table 2.3). Twenty-five percent of the location data were reserved for training. We used a convergence threshold of 10^{-5} and a maximum of 500 iterations. A regularization parameter of 3.0 was selected as it produced the model with the best AICc score. Ten replicates were performed for each model and averaged.

To construct models for inference of optimal *A. triloba* habitat historically, the present-day model was projected onto rasters of six bioclimatic variables representing the Last Glacial Maximum (LGM; ~ 22,000 YBP) and the Mid Holocene (~6,000 YBP) as inferred by the Community Climate System Model (CCSM4; Gent et al. 2011), the Model for Interdisciplinary Research on Climate (MIROC-ESM; Watanabe et al. 2011), and the Max Planck Institute for Meteorology Earth Systems Model (MPI-ESM; Giorgetta et al. 2013). All but one (BIO02) of the bioclimatic variables used for the present-day projection were used for modelling of historical habitat. BIO02 was omitted in accordance with Varela et al. (2015), due to strongly divergent predictions between models for this variable. To reduce variability among the three

models, the inferred probability of suitable habitat at was averaged across the three models (CCSM4, MIROC-ESM, MPI-ESM) for the Holocene and LGM (Varela et al. 2015). Predicted habitat suitability for 2070 was also modelled using the same climactic variables as well as the Representative Concentration Pathway 4.5 (RCP; Vuuren et al. 2011). The RCP trajectory was selected because it is intermediate between extremes of predicted greenhouse gas concentration. All models were evaluated by the area under the ROC curve (AUC), with scores of ≥0.7 indicating a robust model (Fielding & Bell 1997).

RESULTS

Genetic Diversity

Frequencies of null alleles were found to be non-significant for all loci and all populations (p = 0.05). The mean probability of identity (P_I) of all sampled populations was 1.1 x 10^{-3} (range = 6.0 x 10^{-9} to 3.5 x 10^{-2}) (Table 2.4). In other words, for the average population there was only approximately one individual in one thousand that shared an identical MLG to another individual in that population due to chance rather than by descent. This result strongly suggests that duplicate MLGs within populations represent ramets of clonal individuals and as such were removed from further analyses. After duplicate genotypes were removed, the mean number of MLGs per population was 9.5 (range = 1 – 30), with only one population (LL3) having no duplicate MLGs.

The mean number of alleles per locus is 17.2 (range = 2 [Pp-G4] – 44 [Pp-B103]) (Table 2.2). Negative F_{IS} values were found in 32 populations, and the overall mean F_{IS} was -0.153 (range = -1.00 to 0.316) (Table 2.2) which indicates excess heterozygosity. The number of MLGs within

populations and F_{IS} values are significantly correlated (r = 0.68; $p = 2.1 \times 10^{-7}$; Figure 2.2), suggesting that populations that are highly clonal (i.e., low number of MLGs) tend to have higher heterozygosity values than less clonal populations.

The mean number of alleles per population was 32.5 (range = 10 [MAN] to 69 [BCL]). The mean effective number of alleles was 2.6 (range = 0.8 [MS2] to 4.1 [LEE]). A total of 28 private alleles were detected in 20 populations, with each containing one to four (CKW) unique alleles. The mean observed heterozygosity was 0.530 (range = 0.111 [MS2] to 0.778 [SAN]), while the mean expected heterozygosity was 0.496 (range = 0.056 [MS2] to 0.703 [LEE]).

Genetic Structure

Simulations from STRUCTURE yielded an optimal $K = 2$, with mean Ln P(D) ranging from −54197.6 to −38049.0. The Q-matrix of the simulation with the highest estimated Ln P(D) for $K = 2$ was converted into pie frequency charts for each population (Table 2.5; Figure 2.3). The two genetic clusters identified fell primarily on either side of the Appalachian Mountains. There is also an east-west phylogeographic discontinuity along the Tombigbee drainage in Alabama (Figure 2.3). A number of populations had assignments to the cluster located on the opposite side of the Appalachians as its neighboring populations. Examples of this are populations CAG (Texas), PIA (Illinois), KNX, PER, and SHO (Tennessee), GCR, CKW, and CCC (Kentucky), LOC (Maryland), and SUS (Pennsylvania). The two clusters identified were consistent with a principal components analysis of genotypes (Figure 2.4). Though there was a high degree of overlap of genotypes in PCA-space, there appeared to be distinct clustering of eastern and western genotypes.

Genetic structure (G_{ST}) among all 62 populations was 0.341. Mean pairwise G_{ST} = 0.243, ranging from 0.032 (CAD and HAM) to 0.665 (LOC and MAN). The mean pairwise G_{ST} for each population with the remaining 61 populations ranged from 0.147 (WSL) to 0.422 (MAN). Global GST for the eastern and western cluster was 0.332 and 0.333. Mean pairwise G_{ST} values within the eastern cluster and western clusters identified by STRUCTURE were 0.226 and 0.216 respectively; both less than the overall mean pairwise G_{ST} of 0.243. The mean pairwise G_{ST} value between populations in the eastern and western clusters was higher at 0.251.

Geographic Patterns

Most measures of genetic diversity were not significantly correlated with latitude: number of MLGs per population (r = -0.13, p = 0.31), A_T (p = 0.46), A_E (p = 0.24), and H_O (p = 0.24). However, two measures were nearly significant; both the percent of polymorphic loci (r = 0.24; p = 0.056) and expected heterozygosity (Figure 2.5; r = 0.23, p = 0.068) decreased with increasing latitude. A multiple regression model of expected heterozygosity as a function of latitude and longitude revealed that 94 % of the variance in the model was explained by latitude. In the case of percent of polymorphic loci (P %), a multiple regression model with latitude and longitude revealed that latitude explained 88 % of the variance in P % and longitude explained 11 %.

Although private alleles occurred throughout the range of *A. triloba*, the majority (79 %) were found in 68% of the study populations at latitudes less than 38° N, and the number of rare alleles (A_{RARE}) per population decreased significantly with increasing latitude (Figure 2.6, p = 0.03).

Low frequency alleles (i.e., found in <10 % of populations) were more prevalent in mid- and lower-latitude populations and showed a significant decrease with increasing latitude ($p = 0.02$). Some populations located near or within inferred Pleistocene refugial areas for *A. triloba* (e.g., SAN, CFR, and NAG) had H_O values (0.778, 0.689, and 0.667; Table 2.6) that exceeded expected heterozygosity values as well as the mean H_O (0.530) and H_E (0.496). Populations with the highest number of MLGs, A_R, A_T, A_E, P_A and A_{RARE} were typically centered around 35° N, rather than at lower latitudes nearer to likely refugial habitats.

Regarding the levels of genetic diversity in the eastern and western clusters identified by STRUCTURE, all diversity statistics except H_O indicated that eastern populations generally had greater genetic variation than western populations (Table 2.6), although the differences were not significant.

Significant isolation-by-distance was identified in the regression of G_{ST} and geographic distance ($r = -0.006$; $p = 3.4$ x 10-13) as well as the regression of conditional genetic distance against geographic distance ($r = 0.0001$; $p = 5.7$ x 10-4). Given the small r-values for these two tests, the relationship is not likely to be biologically meaningful. The regression of G_{ST} / (1 - G_{ST}) and log geographic distance, and the Mantel test were both non-significant, with $p = 0.4$ and 0.19 respectively. Interestingly, all four analyses revealed significant isolation-by-distance among populations in the eastern cluster (i.e., populations with >50 % posterior probability assignment to cluster 1; Table 2.6) ($r > 0.2$, $p < 0.01$). Within the western cluster, while the analyses indicated a trend towards isolation-by-distance, the regressions were not significant ($p > 0.1$).

Ecological Niche Modelling

The AUC scores for all models (LGM, Holocene, and present-day) were high (≥ 0.859). The most influential bioclimatic variables were minimum temperature of coldest month (BIO06) and elevation, contributing 54 % and 21 % to the model respectively. The realized present-day niche model for *A. triloba* (Figure 2.7) closely matches the species range presented by Little (1977), as well as the range of sampled populations (Figure 2.1). A largely contiguous area from New Jersey to Northern Arkansas was shown to have high habitat suitability, along with areas in the Southeast and Mid-Atlantic with a maximum probability of occurrence (*PrO*) of 0.999982. Grid cells containing high habitat suitability (> 0.75) occupied a total area of approximately 645,726 km^2.

The Pleistocene projection (Figure 2.8) indicates a narrow southerly band of suitable habitat, likely reflective of the cooler, drier conditions at that time, and consistent with species range shifts reported by Davis et al. (1980). There was 86 % less highly suitable habitat (*PrO* > 0.75) relative to the present-day model. The optimal habitat was in coastal areas along the Gulf of Mexico (maximum *PrO* = 1.0; area of high suitability = ~ 34,900 km^2) and Atlantic (maximum *PrO* = 1.0; area of high suitability = ~ 55,075 km^2). A significant portion of optimal habitat (and likely refugia) occurred in areas that are currently below sea level (Figure 2.8). Additional in-land sites may have also served as refugia during retreat and advance of the Laurentide Ice Sheet prior to the Last Glacial Maximum. Most notably, an area of inferred suitable habitat existed between the two modeled coastal refugia, in present-day Southwest Georgia (maximum *PrO* = 0.919; area of high suitability = ~1,700 km^2). Upon glacial retreat and a shift towards more

favorable climatic conditions, during northward range expansion of *A. triloba* there was likely some admixture of Gulf and Atlantic refugial lineages in this area of Southwest Georgia.

In stark contrast to the LGM projection, the niche projection on climatic layers for the Mid-Holocene (Figure 2.9) shows suitable habitat across a range very similar to that of the present-day model (maximum *PrO* = 0.99974), although areas above 40° N latitude had significantly lower suitability (mean *PrO* = 0.099) than the present-day model (mean *PrO* = 0.110) ($p < 0.01$). The Mid-Holocene niche projection overall showed 14 % less highly suitable habitat (i.e., *PrO* > 0.75) compared to the present-day model.

The projected suitable *A. triloba* habitat in the year 2070 under the RCP 4.5 model suggests an overall range expansion and a northeastward shift (Figure 2.10). Areas of high habitat suitability (*PrO* > 0.75) north of 40 ° N latitude increased 412 %, from 65,025 km^2 to 332,900 km^2. Furthermore regions currently occupied by *A. triloba* are projected to remain suitable, with several northern areas enlarging and increasing in suitability, particularly Western Michigan, New York State, and Southern Ontario. Far northern areas formerly inhospitable to *A. triloba*, such as the mouth of the St. Lawrence River, are predicted to become suitable for *A. triloba* by 2070 (Figure 2.10).

DISCUSSION

Genetic Diversity

At the species level, both percent polymorphic loci (*P* % = 100) and expected heterozygosity (H_E = 0.765) indicate that *Asimina triloba* possesses high genetic diversity, even when considering

that its life history traits are associated with high genetic variation (i.e., a wide-spread distribution, outcrossing, long-lived perennial with gravity-dispersed or animal dispersed seeds [Hamrick & Godt 1996]). Mean population levels of expected and observed heterozygosity were also higher-than-expected (mean H_E = 0.496 and mean H_O = 0.530; Table 2.6). This phenomenon has been observed in many long-lived woody species (Frascaria et al. 1993; Huang & Layne 1998) as well as clonal species (Birky 1996; Pappert et al. 2000; Balloux et al. 2003; Brzyski & Culley 2011) and may arise through a number of processes: selection favoring strongly heterozygous individuals (overdominance), large reproductive population size, or disassortative mating between individuals with different alleles (Stoeckel et al. 2006). While no single explanation may be ruled-out, a progeny array of 197 seeds from a single population (BTH) had a nearly identical H_O and F_{IS} value as that of 21 adult individuals. In the same population, individuals smaller than 4 cm dbh also had nearly identical H_O and F_{IS} values as compared to adults (unpublished data, G. Wyatt). Individuals in populations with many MLGs tended to be less heterozygous than individuals in populations characterized by fewer MLGs (Figure 2.2), which may indicate spatial dominance of highly heterozygous clones in populations with few MLGs or bi-parental inbreeding in populations with many MLGs. Populations with many MLGs tended to have positive F_{IS} values.

Genetic Structure

The results from the STRUCTURE v 2.3.4 analysis (Figure 2.3) along with a principal components analysis of genotypes appear to corroborate a general genetic distinction between *A. triloba* populations west of the Appalachian Mountains and Tombigbee drainage (hereafter referred to as “western”) and east of those geographic features (hereafter referred to as

"eastern"). The populations (e.g., CAD, LOC, NAG, PIA, and SUS) that stand in exception to this rule may reflect more recent establishment, which has not allowed time for admixture of surrounding populations to occur. All of these populations are on the periphery of the range of *A. triloba* (Figure 2.1). As with the populations MAN and LOC, the watershed assignment could explain some of the imperfect delimitation observed between eastern and western clusters. Further investigation would be necessary to determine if and to what degree they are indeed genetically distinct from nearest neighboring populations.

The global G_{ST} of 0.341 suggests moderate population differentiation in comparison with other tree species. This pattern typically results from low levels of gene flow among populations resulting from several non-mutually exclusive processes: 1) limited or short-distance pollen movement between populations, 2) a history of relatively small populations both in the Pleistocene refugia and in modern populations, 3) a history of founder effects and low interpopulation gene flow, and 4) competition between clones after populations have been established by relatively few MLGs. Given the weak-flying fly and beetle pollinators of *A. triloba* (Willson & Schemske 1980), pollen is presumably rarely moving long distances between populations.

The eastern and western clusters identified by STRUCTURE analysis had within pairwise G_{ST} values less than 0.243, which would be expected for genetically differentiated clusters and appears to agree with both the STRUCTURE and principal components analyses. Similarly, the mean pairwise G_{ST} values among the identified clusters was greater than 0.243, as expected.

Geographic Patterns

There are several explanations for the lack of significant or weak overall isolation-by-distance. It may be difficult to detect isolation-by-distance among populations 1) over a large geographic area, 2) with large mean pairwise Euclidean distances (in the case of this study, 690.13 km), and 3) showing moderate levels of differentiation, especially when gene flow is likely occurring on a much finer scale and is likely discontinuous across landscapes. Historical vicariance, long-distance seed dispersal, or some combination of these two factors may explain the lack of isolation-by-distance. It is perhaps worth noting that the highest observed pairwise G_{ST} (0.665) occurred between populations LOC and MAN (Figure 2.1) though they are only separated by 114 km. This confounding result can be explained by the abutment of the North Branch of the Potomac River watershed, containing population LOC, to a low-elevation section of the Eastern continental divide (elev. > 1,000 m). Population MAN is located in the Chesapeake watershed, much farther from the Eastern Continental Divide. STRUCTURE analysis also indicates that population LOC belongs to the western cluster of populations (Figure 2.3), which may also be explained by the proximity of the North Branch of the Potomac River to the Eastern Continental Divide. Considerable population differentiation owing to watershed assignment among geographically proximate populations of *Veratrum woodii* was observed by Zomlefer et al. (2018).

It is possible that the presence of strong isolation-by-distance in the inferred eastern cluster may have resulted from dispersal out of a single, mostly contiguous refugium along the Atlantic Coast (Figure 2.8). The absence of significant isolation-by-distance in the inferred western cluster is perplexing, though may have resulted from admixture between non-contiguous refugia along the

Gulf Coast during northward migration following glacial retreat, as seen in Petit et al. (2003) for several European trees and shrubs.

Greater genetic diversity, as measured by *P* %, was observed in populations at lower latitudes, which would be expected to result from a northward range expansion during glacial retreat (Hewitt 2000), especially in a species with large fruits, low fruit-set (Wyatt and Trapnell Chapter 3 in prep), and apparent limited seed dispersal. Similarly, the significant decrease in the number of rare alleles (frequency < 0.1) with increasing latitude (Figure 2.6, $p = 0.03$) is consistent with past northward range expansion characterized by successive founder events (Hewitt 1999; Hewitt 2000). Wyatt & Trapnell (in prep) showed that many measures of clonality (size, richness, diversity, evenness) are related to latitude, and that sampled populations of *A. triloba* tend to become more clonal with increasing latitude (i.e., larger, less diverse, less even). This may be reflective of local adaptation to harsher climates and less suitable habitat at the extreme edges of the species range.

The hypothesis that population genetic parameters A_T and A_E would show an inverse correlation with latitude was not supported, and appears to be contradictory to a number of studies in which find lower levels of genetic diversity in the northern part of species ranges after having expanded from glacial refugia (Hewitt 1996, Soltis et al. 1997, Conroy & Cook 2000). Our data regarding levels of genetic diversity in Pleistocene refugial areas are concordant with that of Petit et al. (2003), who found a pattern of higher genetic diversity in populations of 22 European tree species located at intermediate latitudes rather than in Pleistocene refugial areas. A similar pattern was reported by Duran et al. (2012): the highest genetic diversity of *Pinus edulis*

occurred at latitudes above southern Pleistocene refugia. In our study, areas with populations showing weak assignment to one particular cluster (Figure 2.3) were found primarily in Eastern Tennessee, Kentucky, and West Virginia. These areas, characterized by the Ridge-and-Valley physiography, would presumably present less of a topographic barrier to dispersal than the Blue Ridge Mountains to the Southeast, and may have allowed admixture between populations originating from separate refugia as *A. triloba* expanded northward. The higher genetic diversity at the intermediate latitudes of the range could have resulted from admixture between the two inferred refugia, as was hypothesized to explain higher genetic diversity of *Pinus edulis* in areas north of southern refugia (Duran et al. 2012) or in the European beech (Comps et al. 2001). Additionally, this pattern of higher diversity at intermediate latitudes make sense when considering that many of the inferred refugial areas for *A. triloba* during the Pleistocene are currently under sea-level. The lower genetic diversity observed in populations located near likely Pleistocene refugial areas may be a result of reduced gene flow and successive founder effects, a phenomenon observed at the trailing edge of a shifting species range (Nei 1975; Comps et al. 2001; Davis & Shaw 2001).

Northward expansion has been documented for many tree species through palynological studies and these data have been used to track the rate at which forests in temperate areas of the world have shifted following the Last Glacial Maximum (Delcourt & Delcourt 1987; Hewitt 1996; Jackson et al. 1997; Davis & Shaw 2001; Petit et al. 2002; Davis et al. 2005; Edwards et al. 2014). A decrease in the total number of alleles and effective number of alleles could be explained by the "sieve" effect of long-distance dispersal and subsequent adaptation of favorable genotypes to harsh conditions (Davis & Shaw 2001). Differential survival of genotypes that

become established through long-distance dispersal into a harsher climate may result in a lower number of alleles and multilocus genotypes than in the source population.

The ubiquity of private alleles across the range of *A. triloba* could be explained in several ways: 1) high mutation rates at the microsatellite loci employed, 2) a long historical presence of *A. triloba* at these sites, and/or 3) high levels of dispersal from refugial habitats northward to more recently established populations. It would seem unlikely that a long historical presence of *A. triloba* could explain the presence of private alleles for sampled populations on or near the greatest extent of the Laurentide ice sheet, such as populations SFP or BCP. Dispersal following glacial retreat of these rare alleles into these populations from more southern populations seems more likely, although high mutation rates at these loci cannot be discounted.

Ecological Niche Modelling

Our predicted present-day niche model for *A. triloba* was similar to a model created by Li et al. (2017). Southern areas of Eastern North America have long been hypothesized to have served as Pleistocene refugia for many temperate species that currently have much broader ranges (Hewitt 2000; Soltis et al. 2006). Based on the results from our ecological niche modeling, this appears to have been the case for *A. triloba* as well. Interestingly, the present-day phylogeographic discontinuity observed for this species along the Appalachian Mountains appears to be easily explained by inferred Pleistocene ecological niche models which show likely refugia along the Atlantic and Gulf coasts (Figure 2.8) below present-day sea level. It appears that *A. triloba* did not occupy a wide-spread geographic area during the LGM, as thought to have been the case for a number of tree species (Bennett 1985; Davis & Shaw 2001; Miller & Parker 2009; Lumibao et

al. 2017; Bemmels & Dick 2018). As temperatures warmed during glacial retreat, new habitat would have become available along the Atlantic Coast, which presumably would have been colonized by individuals from the Atlantic Coast refugium. At the same time, new habitat would have become available north of the Gulf Coast and into the Mississippi River valley, which would presumably become colonized by individuals from the Gulf Coast refugium. Due to elevation, the Appalachian Mountains would have continued through time to serve as a barrier to dispersal for this low elevation species between the east-west genetic clusters, as it did for other tree species (Soltis et al. 2006; Miller & Parker 2009). The pattern of separate Gulf and Atlantic coast refugia has been observed in other plant taxa, for example, *Fagus grandifolia* (Morris et al. 2010) and in the understory shrub *Viburnum lantanoides* (Park & Donoghue 2019). East-west structuring of populations has also been observed in *Pinus virginiana* (Parker et al. 1997) and *Carya ovata* (Bemmels & Dick 2018), though this species may have been more wide-spread than *A. triloba* during the LGM. Genetic discontinuity along the Tombigbee River drainage observed in this study (Figure 2.3), has previously been reported by other researchers (Bermingham & Avise 1986; Lawson 1987; Gill et al. 1993; Soltis et al. 2006; McKay 2009), although to date this pattern has only been observed in animal taxa.

Our future projection of habitat suitability for *A. triloba* for the year 2070 does not drastically differ from the present-day model, with the major difference being the strong increase in the expanse of northern suitable habitat. Given the low fruit-set observed in northern populations and seed dispersal currently mediated primarily by small mammals, it appears unlikely that *A. triloba* will rapidly colonize these new habitats without human-mediated dispersal. Also, there is no

guarantee that the species will be able to maintain its current level of genetic diversity given increased threats from habitat fragmentation and climate change.

CONCLUSIONS

This study presents the first evidence of patterns of range-wide genetic diversity and post-glacial range expansion for the highly clonal and geographically widespread understory tree species *Asimina triloba*. Populations sampled across the range of *A. triloba* show moderate genetic structuring and high heterozygosity. The results strongly suggest the existence of eastern and western genetic clusters divided along the Appalachian Mountains and continuing along the Tombigbee River in Alabama. These results are corroborated by ecological niche modeling, which predict refugial areas during the Last Glacial Maximum that occurred along the Gulf and Atlantic coasts in habitat currently submerged under sea-level. The area of suitable habitat for this species is expected to shift northeastward given current climate model predictions.

LITERATURE CITED

Archuleta CM, Constance EW, Arundel ST, Lowe AJ, Mantey KS, Phillips LA (2017) The national map seamless digital elevation model specifications. *U.S. Geological Survey Techniques and Methods*, book 11, chapter B9: 1 – 8.

Ballou F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635 – 1644.

Bemmels J, Dick CW (2018) Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two Eastern North American hickory species. *Journal of Biogeography* 45: 1739 – 1750.

Bennett KD (1985) The spread of *Fagus grandifolia* across Eastern North America during the Last 18,000 years. *Journal of Biogeography* 2: 147 – 164.

Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology* 7: 367 – 369.

Birky CW Jr. (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* 144: 427 – 437.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* 5: 453 – 455.

Brooks RR, Johannes D (1990) *Historical, ethno-, and economic botany series*, Vol. 3. Dioscorides Press, Portland, OR.

Brzyski J, Culley TM (2011) Genetic variation and clonal structure of the rare, riparian shrub *Spiraea virginiana* (Rosaceae). *Conservation Genetics* 12: 1323 – 1332.

Clark JS (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist* 152: 204–224.

Clark JS, Fastie C, Hurtt G, Jackson ST, Johnson C, King GA, Lewis M, Lynch J, Pacala S, Prentice C, Schupp EW, Webb III T, Wyckoff P (1998) Reid's Paradox of rapid plant migration. *BioScience* 48: 13-24.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The Last glacial Maximum. *Science* 325: 710-714.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European Beech. *Genetics* 157: 389 – 397.

Conroy CJ, Cook JA (2000) Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Molecular Ecology* 9: 165 – 175.

Cypher BL, Cypher EA (1999) Germination rates of tree seeds ingested by coyotes and raccoons. *American Midland Naturalist* 142: 71 – 76.

Darley-Hill S, Johnson C (1981) Acorn dispersal by the blue jay (*Cyanocitta cristata*). *Oecologia* 50: 231-232.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* 292: 673-679.

Davis MB, Shaw RG, Etterson JR (2005) Evolutionary responses to changing climate. *Ecology* 86: 1704-1714.

Davis MB, Spear RW, Shane LCK (1980) Holocene climate of New England. *Quaternary Research* 14: 240 – 250.

Delcourt PA, Delcourt HR (1987) Late Quaternary dynamics of temperate forests: applications of paleoecology to issues of global environmental change. *Quaternary Science Reviews* 6: 129 – 146.

Duran KL, Pardo A, Mitton JB (2012) From middens to molecules: phylogeography of the piñon pine, *Pinus edulis*. *Journal of Biogeography* 39: 1536 – 1544.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie physique et Quaternaire* 41: 237–263.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* 19: 3746 – 3759.

Dyer RJ (2016) gstudio: Tools related to the spatialanalysis of genetic marker data. R package version 1.5.0.

Earl DA, vonHoldt BM (2012) Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* 2: 359 – 361.

Edwards ME, Armbruster SW, Elias SE (2014) Constraints on post-glacial boreal tree expansion out of far-northern refugia. *Global Ecology and Biogeography* 23: 1198 – 1208.

Evanno GS, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 15: 2611 – 2620.

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302 – 4315.

Frascaria N, Santi F, Gouyon PH (1993) Genetic differentiation within and among populations of chestnut (*Castanea sativa* Mill) and wild cherry (*Prunus avium* L.). *Heredity* 70: 634 – 641.

Freeman CC, Hulbert LC (1985) An annotated list of the vascular flora of Konza Prairie Research Natural Area, Kansas. *Transactions of the Kansas Academy of Science* 88: 84-115.

Gill EB, Mostrom AM, Mack AL (1993) Speciation in North American chickadees. I. Patterns of mtDNA genetic divergence. *Evolution* 47: 195-212.

Gonzales E, Hamrick JL, Chang SM (2008) Identification of glacial refugia in Southeastern North America by phylogeographical analyses of a forest understory plant, *Trillium cuneatum*. *Journal of Biogeography* 35: 844-852.

Goodrich KR, Zjhra M, Ley CA, Raguso RA (2006) When flowers smell fermented: The chemistry and ontogeny of yeasty floral scent in pawpaw (*Asimina triloba*: Annonaceae). *International Journal of Plant Sciences* 167: 33 – 46.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86: 485 – 486.

Graae BJ, Verheyen K, Kolb A,Van Der Veken S, Heinken T, Chabrerie O, Diekmann M, Valtinat K, Zindel R, Karlsson E, Strom L, Decocq G, Hermy M, Baskin CC (2009) Germination requirements and seed mass of slow- and fast-colonizing temperate forest herbs along a latitudinal gradient. *Ecoscience* 16: 248-257.

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species, p. 43 – 63. In: Brown AHD, MT Clegg, AL Kahler, BS Weir (eds.). Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, MA.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philisophical Transactions of the Royal Society of London: Biological Sciences* 351: 1291 – 1298.

Hauswaldt SJ, Glenn TC (2003) Microsatellite DNA loci from the Diamondback terrapin (*Malaclemys terrapin*). *Molecular Ecology Notes* 3: 174 - 176

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247 – 276.

Hewitt GM (1999) Post-glacial recolonization of European biota. *Biological Journal of the Linnean Society* 68: 87 – 112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405: 907 – 913.

Hewitt GM (2001) Speciation, hybrid zones, and phylogeography—or seeing genes in space and time. *Molecular Ecology* 10: 537 – 549.

Hof C, Levinksy I, Araujo MB, Rahbek C (2011) Rethinking species' ability to cope with rapid climate change. *Global Change Biology* 17: 2987-2990.

Hoppe K, Koch PL (2007) Reconstructing the migration patterns of late Pleistocene mammals from northern Florida, USA. *Quaternary Research* 68: 347 – 352.

Huang H, DR Layne, Riemenschneider DE (1998) Genetic diversity and geographic differentiations in pawpaw [*Asimina triloba* (L.) Dunal] populations from nine states as revealed by allozyme analysis. *Journal of the American Society Horticultural Science* 123: 635 – 641.

Huang H, Layne DR, Kubisiak TL (2000) RAPD inheritance and diversity in Pawpaw (*Asimina triloba*). *Journal of the American Society of Horticultural Science* 125: 454 – 459.

Jackson ST, Overpeck JT, Webb-III T, Keattch SE, Anerson KH (1997) Mapped plant-macrofossil and pollen records of late Quaternary vegetation change in Eastern North America. *Quaternary Science Reviews* 16: 1-70.

Jordano P, García C, Godoy JA, García-Castaño JL (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences* 104: 3278 – 3282.

Kearse M, Moir R, Wilson A et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data: *Bioinformatics* 28(12), 1647-1649.

Keener C, Kuhns E (1997) The impact of Iroquoian populations on the northern distribution of pawpaws in the Northeast. *North American Archaeologist* 18: 327–342.

Kral R (1997) Annonaceae Jussieu - Custard-apple Family, pp. 11-20. In: Editorial Committee (eds.). Flora of North America North of Mexico. Oxford University Press, NY.

Lawson AD, Reeder TW (2002) Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Systematic Biology* 51: 44 – 68.

Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics* 5: 107 – 117.

Li PS, Thomas DC, Saunders MK (2017) Historical biogeography and ecological niche modelling of the *Asimina – Disepalum* clade (Annonaceae): role of ecological differentiation in Neotropical-Asian disjunctions and diversification in Asia. *Evolutionary Biology* 17: 188 – 206.

Little EL Jr. (1977) Atlas of the United States trees, Volume 4: minor Eastern hardwoods. Pub. 1342. Washington, DC: US Department of Agriculture. 17 p 230 maps.

Lu L, Pomper KW, Lowe JD, Crabtree SB (2011) Genetic variation in pawpaw cultivars using microsatellite analysis. *Journal of the American Society of Horticultural Science* 136: 415–521.

Lumibao CY, Hoban SM, McLachlan J (2017) Ice ages leave genetic diversity 'hotspots' in Europe but not in Eastern North America. *Ecology Letters* 20: 1459 – 1468.

McKay BD (2009) Evolutionary history suggests rapid differentiation in the yellow-throated warbler *Dendroica dominica*. *Journal of Avian Biology* 40: 181 – 190.

Mabberley DJ (1978) Annonaceae, the Sweetsop and the Soursop, pp. 30 - 31. In Heywood, V. H. (ed.). Flowering Plants of the World. Mayflower Books, New York.

MacDougall A (2003) Did Native Americans influence the northward migration of plants during the Holocene? *Journal of Biogeography* 30: 633–647.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209 – 220.

McLaughlin JL, Hui YH (1993) Chemotherapeutically active acetogenins. U.S. Patent No. 5,229,419.

Morris AB, Graham CH, Soltis DE, Soltis PS (2010) Reassessment of phylogeographical structure in an Eastern North American tree using Monmonier's algorithm and ecological niche modelling. *Journal of Biogeography* 37: 1657 – 1667.

Murphy JL (2001) Pawpaws, persimmons, and 'possums: On the natural distribution of pawpaws in the northeast. *North American Archaeologist* 22: 93-115.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70: 3321 – 3323.

Nei M (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29: 1- 10.

Pappert RA, Hamrick JL, Donovan LA (2000) Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany* 87: 1240 – 1245.

Parker KC, Hamrick JL, Parker AJ, Stacy EA (1997) Allozyme diversity in *Pinus virginiana* (Pinaceae): intraspecific and interspecific comparisons. *American Journal of Botany* 84: 1372 – 1382.

Miller DM, Parker KC (2009) The impacts of the Appalachian Mountains on the post-glacial migration pathways and gene flow of sugar maple (*Acer saccharum*). *Physical Geography* 30: 89 – 104.

Patterson WA III, Sassaman KE (1988) Indian fires in the prehistory of New England. *In*: Nicholas GP (Ed) *Holocene human ecology in northeastern North America*, pp. 107-135. Plenum Publishing, NY.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology Notes* 6: 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537 – 2539.

Peattie DC (1950) A Natural History of Trees of Eastern and Central North America. Houghton Mifflin Co., Boston.

Perrotti AG (2018) Pollen and *Sporomiella* evidence for terminal Pleistocene vegetation change and megafaunal extinction at Page-Ladson, Florida. *Quaternary International* 466: 256-268.

Petit RJ, Brewer S, Bordacs S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PB, Jensen JS, Konig AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu JL, Kremera A (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 156: 49–74.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563 – 1565.

Phillips SJ, Dudík M, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modeling* 190: 231 – 259.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161 – 175.

Phillips BL, Brown GP, Travis JMJ, Shine R (2008) Reid's Paradox revisited: the evolution of dispersal kernels during range expansion. *The American Naturalist* 172: S34-S48.

Pomper KW, Lowe JD, Lu L (2009) Clonality of pawpaw (*Asimina triloba*) patches in Kentucky. *Journal of the Kentucky Academy of Science* 70: 3 – 11.

Pomper KW, Lowe JD, Lu L, Crabtree SB, Dutta S, Schneider K, Tidwell J (2010) Characterization and identification of pawpaw cultivars and advanced selections by simple sequence repeat markers. *Journal of the American Society for Horticulture Science* 135: 143 – 149.

Pritchard JK, Stephens JK, Donnelly M (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945 – 959.

R Core Team (2013) R: A language and environment for statistical computing.

Ratanyake S, Rupprecht JK, Potter WM, McLaughlin JL (1992) Evaluation of various parts of the paw paw tree, *Asimina triloba* (Annonaceae), as commercial sources of the pesticidal annonaceous acetogenins. *Journal of Economic Entomology*, 85: 2353 – 2356.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86:248-249

Reid C (1899) *The origin of the British flora*. Dulau, London, UK.

Roques L, Hamel F, Fayard J, Fady B, Klein EK (2010) Recolonisation by diffusion can generate increasing rates of spread. *Theoretical Population Biology* 77: 205-212.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 145: 1219 – 1228.

Ruggiero MV, Reusch TBH, Procaccini G (2005) Local genetic structure in a clonal dioecious angiosperm. *Molecular Ecology* 14: 957 – 967.

Skellam JG (1951) Random dispersal in theoretical populations. *Biometrika* 38: 196-218.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated Eastern North America. *Molecular Ecology* 15: 4261-4293.

Stoeckel S, Grange J, Fernández-Manjarres JF, Bilger I, Frascaria-Lacoste N, Mariette S (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species – *Prunus avium* L. *Molecular Ecology* 15: 2109 – 2118.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* 166: 581-591.

Varela S, Lima-Ribeiro MS, Terribile LC (2015) A short guide to the climatic variables of the Last Glacial Maximum for Biogeographers. *PLOS One* 10: e0129037.

Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. *Climatic Change* 109: 5 – 31.

Weakley AS. Flora of the Southern and Mid-Atlantic States. The University of North Carolina Herbarium, Chapel Hill. Working draft of 30 November 2012.

White CA (1906) The northern limit of the Papaw tree. *Science* 23: 749-751.

Wright S (1922) Coefficients of inbreeding and relationship. *The American Naturalist* 56: 330 – 338.

Wright S (1943) Isolation by distance. *Genetics* 28: 114 – 138.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* 15: 323 – 354.

Willson MF, DW Schemske (1980) Pollinator limitation, fruit production, and floral display in Pawpaw (*Asimina triloba*). *Bulletin of the Torrey Botanical Club* 107: 401 – 408.

Willson MF (1993) Mammals as seed-dispersal mutualists in North America. *Oikos* 67: 159 – 176.

Yeager LE, Elder WH (1945) Pre- and post-hunting season foods of raccoons on an Illinois goose refuge. *Journal of Wildlife Management* 9: 48 – 53.

Zalasiewicz J, Williams M, Haywood A, Ellis M (2011) The Anthropocene: a new epoch of geologic time? Philosophical Transactions of the Royal Society 369: 835 – 841.

Zomlefer WB, Comer JR, Lucardi RD, Hamrick JL, Allison JR (2018) Distribution and genetic diversity of the rare plant *Veratrum woodii* (Liliales: Melanthiaceae) in Georgia: A preliminary study with AFLP fingerprint data. *Systematic Botany* 43: 858 – 869.

Table 2.1. Location and elevation data for 62 sampled *Asimina triloba* populations in the United States.

Population	County, State	Latitude	Longitude	Elevation (m)
BCB	Coffee, TN	35.485	-85.961	337
BCL	Lee, AR	34.706	-90.698	69
BCP	Brown, IN	39.175	-86.222	291
BLU	Carter, MO	36.961	-90.987	137
BSF	Christian, MO	36.858	-93.222	286
BTH	Oglethorpe, GA	33.706	-83.019	156
CAD	Harrison, TX	32.693	-94.176	71
CAT	Catawba, NC	35.821	-81.177	269
CCC	Carter, KY	38.379	-83.123	256
CFR	Harnett, NC	35.469	-78.898	38
CHR	Clark, IN	38.398	-85.635	146
CKW	Madison, KY	37.632	-84.196	309
COL	Richland, SC	33.947	-80.629	39
DCR	Harrison, WV	39.146	-80.391	335
DRP	Johnson, IL	37.430	-88.942	126
ER2	Orange, NC	36.083	-79.021	134
EUT	Greene, AL	32.785	-87.837	34
GCR	Anderson, KY	37.975	-84.825	178
HAM	West Feliciana, LA	30.786	-91.255	21
HEM	Sabine, TX	31.482	-93.936	95
HOL	Bourbon, KS	37.777	-94.827	252
HPC	Platte, MO	39.228	-94.778	280

KAN	Kanawha, WV	38.224	-81.625	249
KNX	Knox, TN	35.941	-83.931	266
LAW	Douglas, KS	38.991	-95.315	264
LCC	McCormick, SC	34.054	-82.343	129
LEE	Lee, AR	34.760	-90.733	61
LHP	Davidson, TN	36.087	-86.551	173
LIL	Fairfield, SC	34.253	-81.233	70
LL1	Stewart, TN	36.581	-87.947	150
LL2	Trigg, KY	36.792	-88.096	135
LL3	Trigg, KY	36.900	-88.040	131
LOC	Morgan, MD	39.692	-78.199	177
MAN	Prince William, VA	38.805	-77.525	82
MLK	Boone, MO	38.936	-92.351	195
MS1	Shelby, TN	35.340	-90.051	103
MS2	Shelby, TN	35.341	-90.055	85
MWR	Stoddard, MO	36.970	-90.151	105
NAG	St. Augustine, TX	31.480	-93.995	89
OFR	Jackson, KY	37.542	-84.174	259
PCB	Pike, AL	31.907	-86.174	125
PER	Davidson, TN	36.065	-86.880	226
PIA	Jersey, IL	38.939	-90.285	137
PMQ	Jersey, IL	38.984	-90.490	162
PWP	Prince William, VA	38.566	-77.346	43
RCA	Laurel, KY	36.962	-84.351	248
RL	Davidson, TN	36.063	-86.799	233
ROC	Boone, MO	38.875	-92.327	207
SAL	Pickens, SC	35.073	-82.587	349
SAN	Adams, MS	31.460	-91.211	55

SFP	Guernsey, OH	40.128	-81.510	251
SHE	Paulding, GA	34.022	-84.911	281
SHO	Cannon, TN	35.872	-85.991	408
STO	York, SC	34.881	-81.104	167
SUS	Lancaster, PA	39.808	-76.286	140
TAL	Cleburne, AL	33.859	-85.524	240
TER	Saluda, SC	34.147	-81.824	142
TYG	Taylor, WV	39.308	-80.027	360
WBF	Winston, AL	34.284	-87.396	183
WDP	Marshall, MS	34.668	-89.467	116
WES	Platte, MO	39.384	-94.880	280
WSL	St. Charles, MO	38.666	-90.732	151

Table 2.2. Nuclear microsatellite loci used to characterize sampled populations of *A. triloba*. A_T = total number of alleles, A_E = effective number of alleles, H_O = observed heterozygosity, H_E = expected heterozygosity, F_{ST} = Wright's F_{ST}, and F_{IS} = inbreeding coefficient.

Locus	**Motif**	**Fragment size (bp)**	A_T	A_E	H_O	H_E	F_{ST}	F_{IS}
Pp-G4	AAT	187 - 190	2	1.50	0.34	0.28	0.28	-0.21
Pp-G119	AAT	158 - 191	13	2.68	0.70	0.56	0.57	-0.26
Pp-G124	AAT	198 - 225	10	2.50	0.61	0.54	0.55	-0.14
Pp-C108	ATG	183 - 204	10	2.26	0.64	0.49	0.31	-0.31
Pp-G103	AAT	278 - 338	23	2.02	0.33	0.35	0.43	0.04
Pp-G121	AAT	183 - 240	14	2.72	0.69	0.56	0.33	-0.28
Pp-B103	GA	262 - 360	44	3.68	0.68	0.62	0.35	-0.11
Pp-B108	GA	114 - 140	14	2.52	0.59	0.49	0.44	-0.21
Pp-B128	GA	137 - 193	25	2.33	0.36	0.43	0.55	0.16
Mean	**-**	**-**	**17.2**	**2.47**	**0.55**	**0.48**	**0.42**	**-0.15**

Table 2.3. Magnitude of variation in seven bioclimatic variables and elevation for 62 wild populations of *A. triloba* in North America used for species niche modeling in Maximum Entropy. BIOCLIM temperature data are stored in °C * 10 but have been converted to actual temperatures here.

Code	Environmental Variable	Range	Unit
BIO02	Mean Diurnal Range (Mean of monthly (max temp - min temp))	5.9 - 18.4	° Celsius
BIO06	Min Temperature of Coldest Month	-2.5 - 19.4	° Celsius
BIO08	Mean Temperature of Wettest Quarter	-6.1 - 30.1	° Celsius
BIO13	Precipitation of Wettest Month	39 - 282	mm
BIO16	Precipitation of Wettest Quarter	95 - 677	mm
BIO17	Precipitation of Driest Quarter	21 - 523	mm
BIO19	Precipitation of Coldest Quarter	24 - 541	mm
GTOPO30	Global digital elevation model	0 - 2035	m

Table 2.4. Probability of identity (*PI*) values for each population of *A. triloba*. N = number of samples and MLG = number of multilocus genotypes.

Population	**N**	**MLG**	P_I
BCB	25	8	1.3×10^{-4}
BCL	48	27	6.0×10^{-9}
BCP	48	2	2.5×10^{-4}
BLU	25	9	9.5×10^{-5}
BSF	23	21	1.2×10^{-5}
BTH	37	10	3.4×10^{-7}
CAD	39	18	2.1×10^{-4}
CAT	16	1	8.4×10^{-5}
CCC	26	13	1.4×10^{-4}
CFR	20	5	1.2×10^{-5}
CHR	30	18	6.2×10^{-6}
CKW	24	13	1.7×10^{-4}
COL	48	12	7.5×10^{-5}
DCR	15	1	7.9×10^{-5}
DRP	23	13	5.1×10^{-6}
ER2	10	5	5.3×10^{-5}
EUT	33	30	5.0×10^{-7}
GCR	48	6	2.5×10^{-4}
HAM	47	2	2.5×10^{-4}
HEM	40	5	2.1×10^{-4}
HOL	10	8	5.3×10^{-5}

HPC	25	14	1.1×10^{-3}
KAN	43	10	2.3×10^{-4}
KNX	25	3	1.3×10^{-4}
LAW	30	12	9.6×10^{-3}
LCC	24	18	2.2×10^{-5}
LEE	24	12	1.3×10^{-4}
LHP	20	4	1.1×10^{-4}
LIL	24	21	2.2×10^{-5}
LL1	25	1	1.3×10^{-4}
LL2	25	7	1.3×10^{-4}
LL3	24	24	1.3×10^{-4}
LOC	38	3	2.0×10^{-4}
MAN	48	1	2.5×10^{-4}
MLK	24	2	3.5×10^{-2}
MS1	24	4	1.3×10^{-4}
MS2	23	1	1.2×10^{-4}
MWR	25	14	1.3×10^{-4}
NAG	30	3	1.6×10^{-4}
OFR	33	6	1.7×10^{-4}
PCB	48	12	7.2×10^{-5}
PER	48	5	2.5×10^{-4}
PIA	14	1	7.4×10^{-5}
PMQ	60	4	3.2×10^{-4}
PWP	48	27	2.5×10^{-4}
RCA	25	16	1.3×10^{-4}
RL	27	12	1.4×10^{-4}
ROC	25	1	1.3×10^{-4}
SAL	30	14	4.1×10^{-5}

SAN	21	1	1.1×10^{-4}
SFP	25	7	1.3×10^{-4}
SHE	24	5	4.0×10^{-3}
SHO	24	16	1.3×10^{-4}
STO	10	5	5.3×10^{-5}
SUS	30	8	1.6×10^{-4}
TAL	24	3	8.9×10^{-3}
TER	34	22	3.2×10^{-6}
TYG	48	2	2.5×10^{-4}
WBF	43	22	5.1×10^{-6}
WDP	27	3	6.1×10^{-3}
WES	22	9	1.2×10^{-4}
WSL	11	9	4.3×10^{-7}

Table 2.5. STRUCTURE Q-matrix for *A. triloba* from the run with highest Ln P(*D*) for $\Delta K = 2$ (-50942.4). Values in columns represent the posterior probabilities of assignment to one of two ΔK clusters.

Population	**Cluster 1**	**Cluster 2**
COL	0.991	0.009
PCB	0.991	0.009
TAL	0.989	0.011
CAT	0.984	0.016
PIA	0.981	0.019
SHE	0.980	0.020
MAN	0.975	0.025
ER2	0.968	0.032
LCC	0.938	0.062
TER	0.934	0.066
CFR	0.922	0.078
HEM	0.920	0.080
KNX	0.887	0.113
LIL	0.885	0.115
SAL	0.855	0.145
WBF	0.820	0.180
PER	0.701	0.299

STO	0.686	0.314
BTH	0.677	0.323
GCR	0.657	0.343
SHO	0.643	0.357
EUT	0.578	0.422
CKW	0.573	0.427
CCC	0.561	0.439
PWP	0.558	0.442
NAG	0.524	0.476
LL2	0.428	0.572
RL	0.360	0.640
RCA	0.357	0.643
KAN	0.261	0.739
WSL	0.233	0.767
LL3	0.216	0.784
LEE	0.165	0.835
BCL	0.143	0.857
SUS	0.104	0.896
CHR	0.103	0.897
WES	0.093	0.907
HOL	0.091	0.909
LAW	0.084	0.916
SFP	0.083	0.917

DRP	0.079	0.921
HPC	0.075	0.925
BSF	0.074	0.926
WDP	0.070	0.930
OFR	0.062	0.938
HAM	0.056	0.944
CAD	0.052	0.948
BLU	0.047	0.953
MWR	0.031	0.969
LHP	0.031	0.969
BCP	0.029	0.971
TYG	0.027	0.973
MS1	0.026	0.974
BCB	0.021	0.979
LL1	0.016	0.984
MLK	0.015	0.985
DCR	0.014	0.986
MS2	0.012	0.988
PMQ	0.012	0.988
ROC	0.010	0.990
LOC	0.010	0.990
SAN	0.008	0.992

Table 2.6. Summary statistics of *A. triloba* nuclear genetic diversity arranged by membership in genetic clusters identified by STRUCTURE. N = sample size, MLG = number of multilocus genotypes, P (%) = percent of polymorphic loci, A_T = total number of alleles, A_E = effective number of alleles per locus, P_A = number of private alleles, A_{RARE} = number of rare alleles (< 0.1 frequency within a population), H_O = observed heterozygosity, H_E = expected heterozygosity, and F_{IS} = inbreeding coefficient. All F_{IS} values are non-significant (p = 0.05).

Population	***N***	**MLG**	***P* (%)**	A_T	A_E	P_A	A_{RARE}	H_O	H_E	F_{IS}
Eastern Cluster										
BTH	37	10	100	54	3.4	2	5	0.620	0.665	0.066
CAT	16	1	67	14	1.7	-	-	0.667	0.333	-1.000
CCC	26	13	89	43	3.2	-	4	0.477	0.614	0.222
CFR	20	5	89	30	2.5	1	3	0.689	0.550	-0.238
CKW	24	13	100	40	2.5	4	5	0.523	0.526	-0.024
COL	48	12	78	29	1.9	-	1	0.394	0.397	-0.015
ER2	10	5	67	17	1.7	-	-	0.500	0.326	-0.527
EUT	33	30	100	57	3.9	-	9	0.463	0.650	0.231
GCR	48	6	78	26	2.4	1	2	0.356	0.471	0.316
HEM	40	5	89	26	2.8	-	-	0.578	0.544	-0.016
KNX	25	3	100	31	2.8	2	3	0.537	0.600	0.133
LCC	24	18	89	38	2.8	-	4	0.476	0.555	0.147
LIL	24	21	89	44	2.7	-	4	0.452	0.562	0.194
MAN	48	1	33	10	1.1	-	-	0.333	0.167	-1.000
NAG	30	3	78	20	3.1	-	-	0.667	0.586	-0.158
PCB	48	12	89	29	1.9	-	5	0.397	0.391	0.013

PER	48	5	100	37	3.4	-	-	0.583	0.642	0.044
PIA	14	1	44	15	1.4	1	1	0.444	0.222	-1.000
PWP	48	27	100	56	3.6	-	4	0.540	0.670	0.169
SAL	30	14	100	43	3	-	1	0.563	0.612	0.030
SHE	24	5	89	26	2.3	-	1	0.511	0.479	-0.045
SHO	24	16	89	49	3.4	-	5	0.596	0.566	-0.062
STO	10	5	89	25	2.3	-	2	0.454	0.498	0.129
TAL	24	3	89	25	2.2	-	1	0.611	0.480	-0.266
TER	34	22	89	41	2.9	-	5	0.529	0.579	0.081
WBF	43	21	89	53	3.5	-	10	0.578	0.623	0.082
Eastern Mean	**30.8**	**10.7**	**85.1**	**33.8**	**2.6**	**0.4**	**2.9**	**0.521**	**0.5112**	**-0.096**
Eastern Pooled	**800**	**277**	**100**	**129**	**6.1**	**27**	**99**	**0.529**	**0.767**	**0.297**
Western Cluster										
BCB	25	8	78	21	1.9	-	-	0.597	0.404	-0.367
BCL	48	27	100	69	2.7	1	8	0.503	0.573	0.125
BCP	48	2	67	17	1.7	1	1	0.444	0.292	-0.547
BLU	25	9	78	28	2.5	1	1	0.460	0.506	0.062
BSF	23	21	89	43	3.1	-	2	0.546	0.543	-0.010
CAD	39	18	100	31	3.3	1	1	0.633	0.657	0.085
CHR	30	18	89	45	3.2	-	1	0.580	0.598	0.030
DCR	15	1	44	13	1.4	-	1	0.444	0.222	-1.000
DRP	23	13	100	41	3.3	2	7	0.562	0.599	0.070
HAM	47	2	100	28	2.2	-	-	0.444	0.528	0.170
HOL	10	8	89	34	2.9	-	1	0.545	0.495	-0.098
HPC	25	14	89	31	2.1	-	-	0.467	0.469	-0.020
KAN	43	10	89	43	3.4	-	-	0.559	0.599	0.090
LAW	30	12	89	29	1.9	-	1	0.492	0.392	-0.185
LEE	24	12	100	56	4.1	-	6	0.643	0.703	0.096

LHP	20	4	89	22	1.9	-	2	0.611	0.406	-0.442
LL1	25	1	56	14	1.6	-	-	0.556	0.278	-1.000
LL2	25	7	78	38	3.4	-	1	0.613	0.580	-0.055
LL3	24	24	78	51	3.7	1	7	0.495	0.577	0.141
LOC	38	3	56	16	1.6	-	-	0.278	0.264	-0.134
MLK	24	2	67	18	1.8	1	1	0.611	0.375	-0.622
MS1	24	4	78	18	3.5	-	-	0.611	0.674	0.072
MS2	23	1	67	15	0.8	1	1	0.111	0.056	-1.000
MWR	25	14	78	38	2.3	1	1	0.524	0.442	-0.092
OFR	33	6	89	36	2.9	-	2	0.550	0.512	-0.082
PMQ	60	4	89	26	2.2	-	1	0.583	0.441	-0.313
RCA	25	16	100	49	3.8	1	4	0.490	0.645	0.273
RL	27	12	100	44	3.6	2	5	0.681	0.639	-0.064
ROC	25	1	56	14	1.6	-	-	0.556	0.278	-1.000
SAN	21	1	78	16	1.8	1	2	0.778	0.389	-1.000
SFP	25	7	100	27	2.1	1	1	0.456	0.482	0.014
SUS	30	8	100	37	3	-	1	0.602	0.603	0.088
TYG	48	2	89	20	2	-	-	0.722	0.444	-0.600
WDP	27	3	89	26	2.5	1	2	0.444	0.531	0.192
WES	22	9	100	45	3.1	-	3	0.537	0.590	0.078
WSL	11	9	100	40	3.5	-	5	0.600	0.647	0.078
Western Mean	**28.8**	**8.7**	**84.4**	**31.6**	**2.6**	**0.4**	**1.9**	**0.537**	**0.484**	**-0.194**
Western Pooled	**1037**	**313**	**100**	**126**	**5.9**	**24**	**94**	**0.574**	**0.732**	**0.226**
Species Mean	**29.6**	**9.5**	**84.7**	**32.5**	**2.6**	**0.4**	**2.3**	**0.530**	**0.496**	**-0.153**
Species Pooled	**1837**	**590**	**100**	**153**	**6.5**	**-**	**144**	**0.553**	**0.765**	**0.273**

Figure 2.1. Map of 62 *A. triloba* study populations. The reported range of *A. triloba* from Little (1977) is outlined in blue and filled with grey.

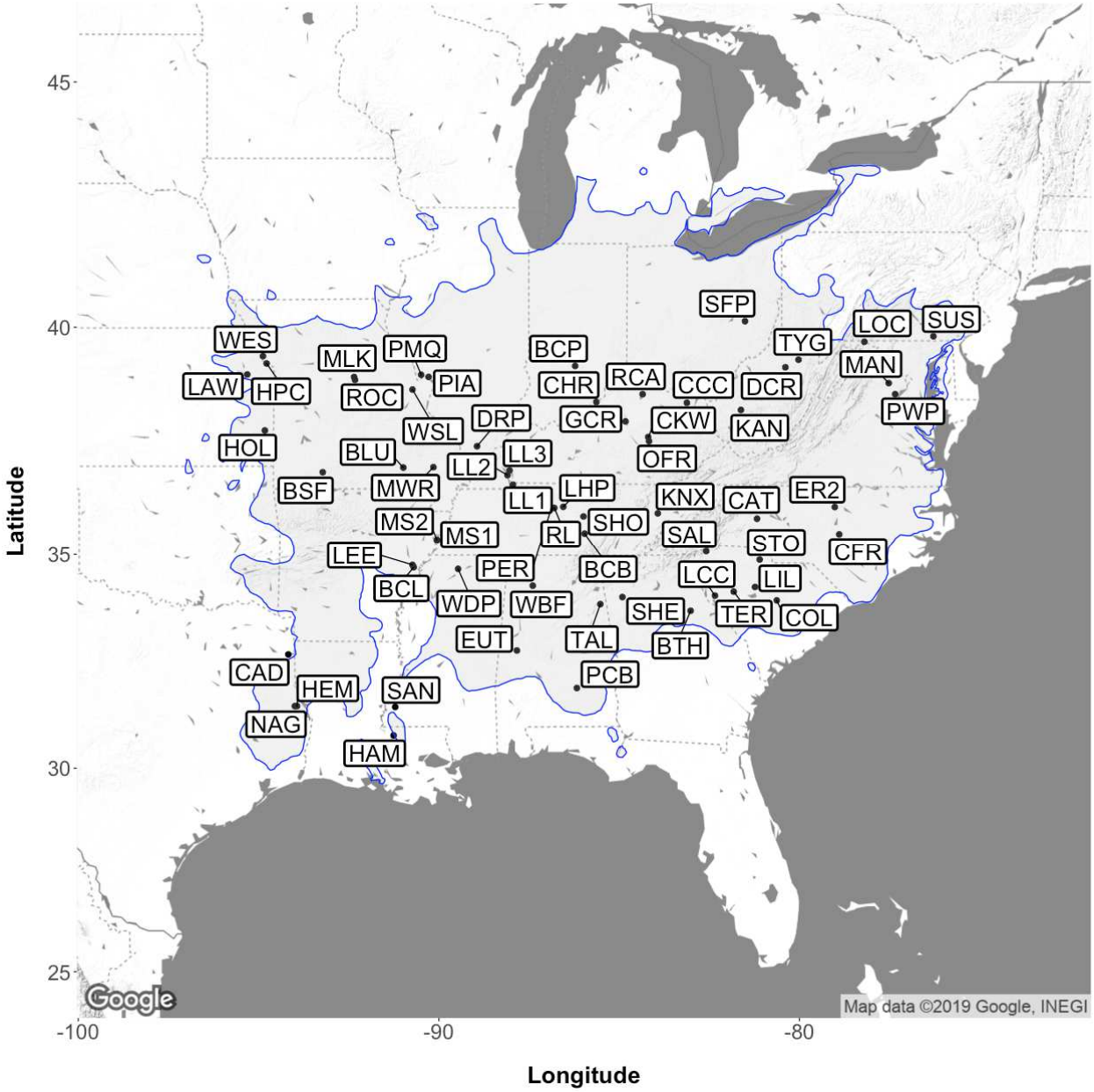


Figure 2.2. Relationship between inbreeding coefficient (F_{IS}) and number of multilocus genotypes in *A. triloba* (MLG) (r = 0.67, $p = 2.0 \times 10^{-9}$).

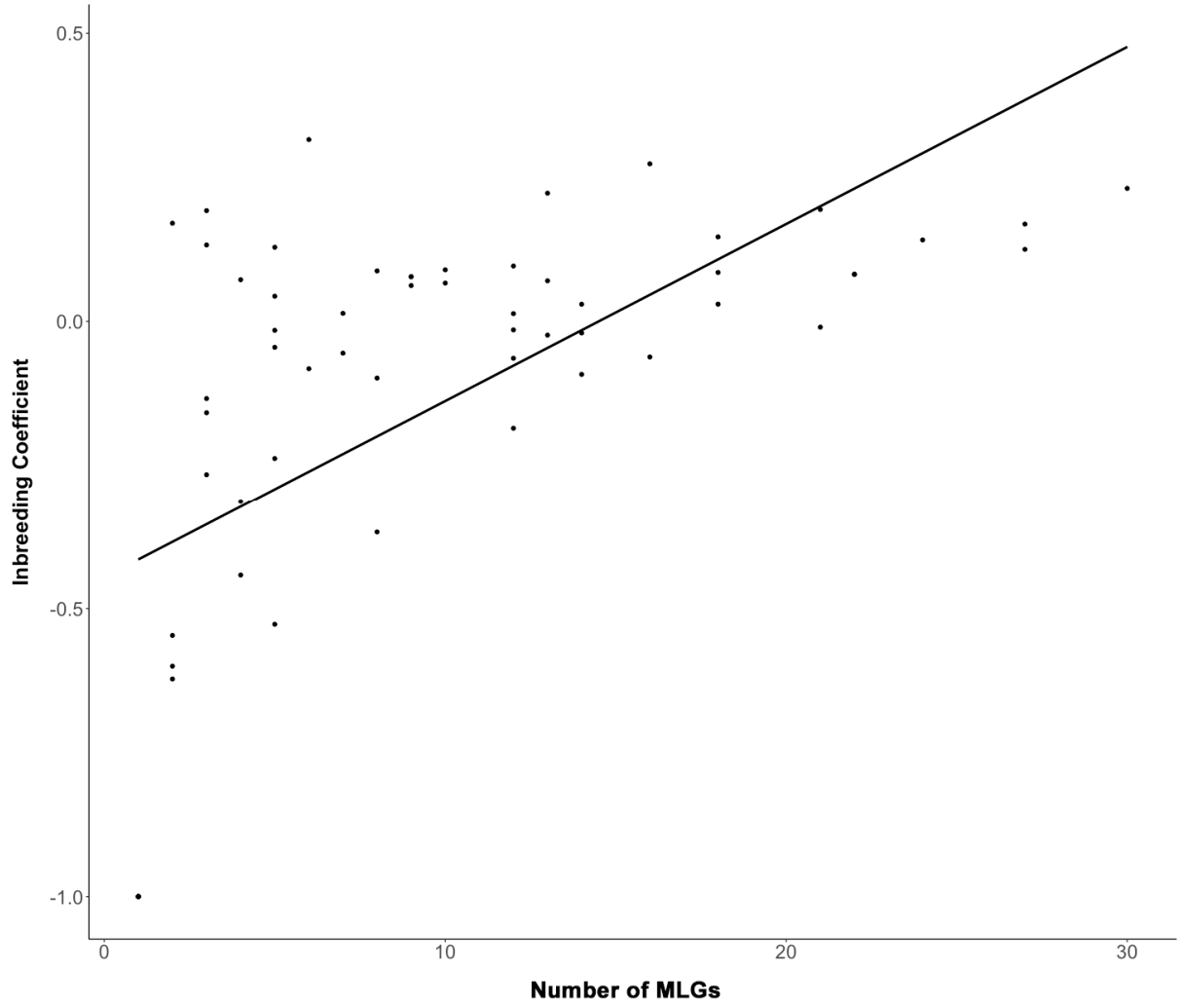


Figure 2.3. Pie charts for 62 populations of *A triloba* indicating the proportion of individuals in each population assigned to the eastern cluster (red) and western cluster (blue) identified by STRUCTURE. The Tombigbee River in Alabama is indicated in blue.

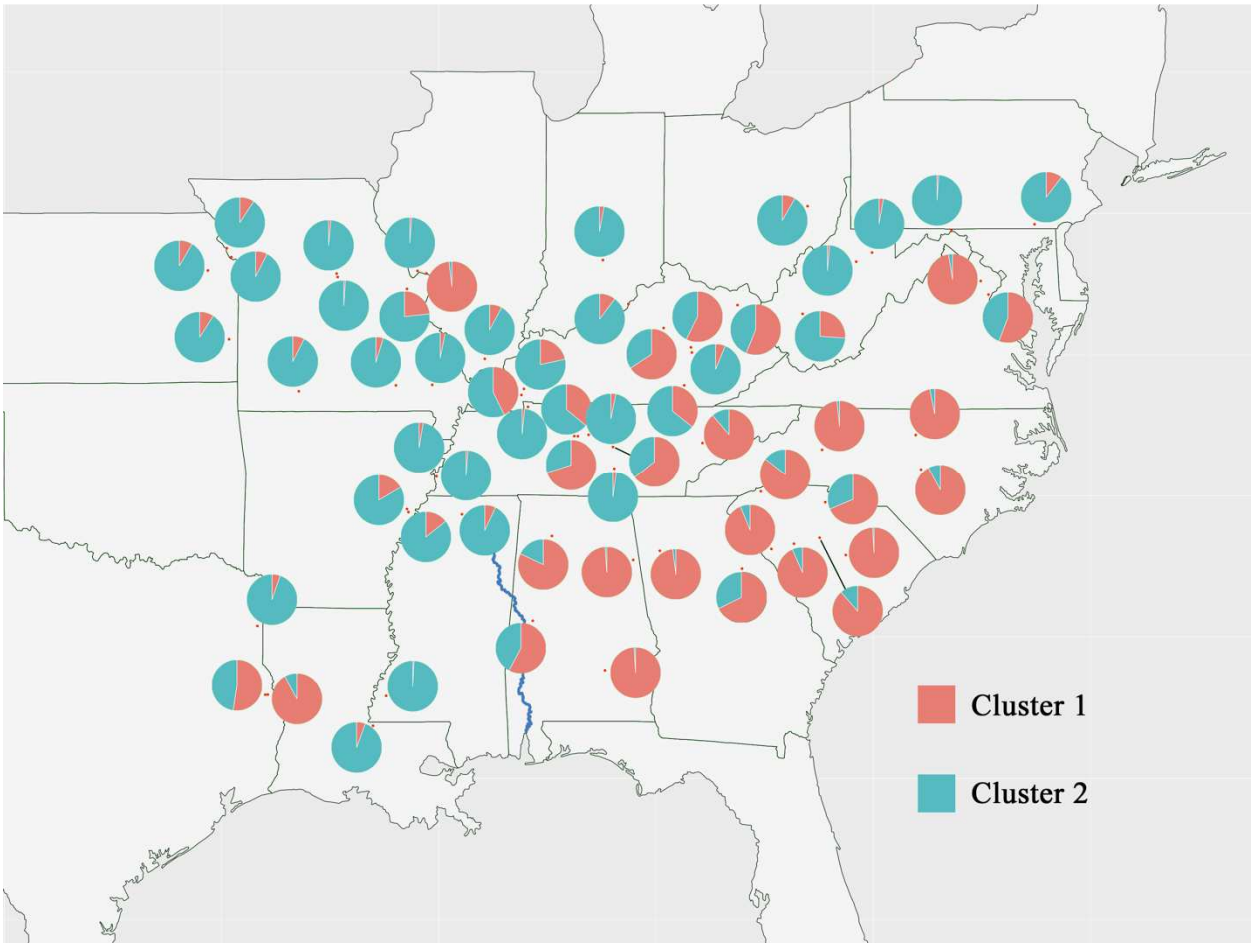


Figure 2.4. Results of a principal components analysis of genotypes from 62 *A. triloba* populations.

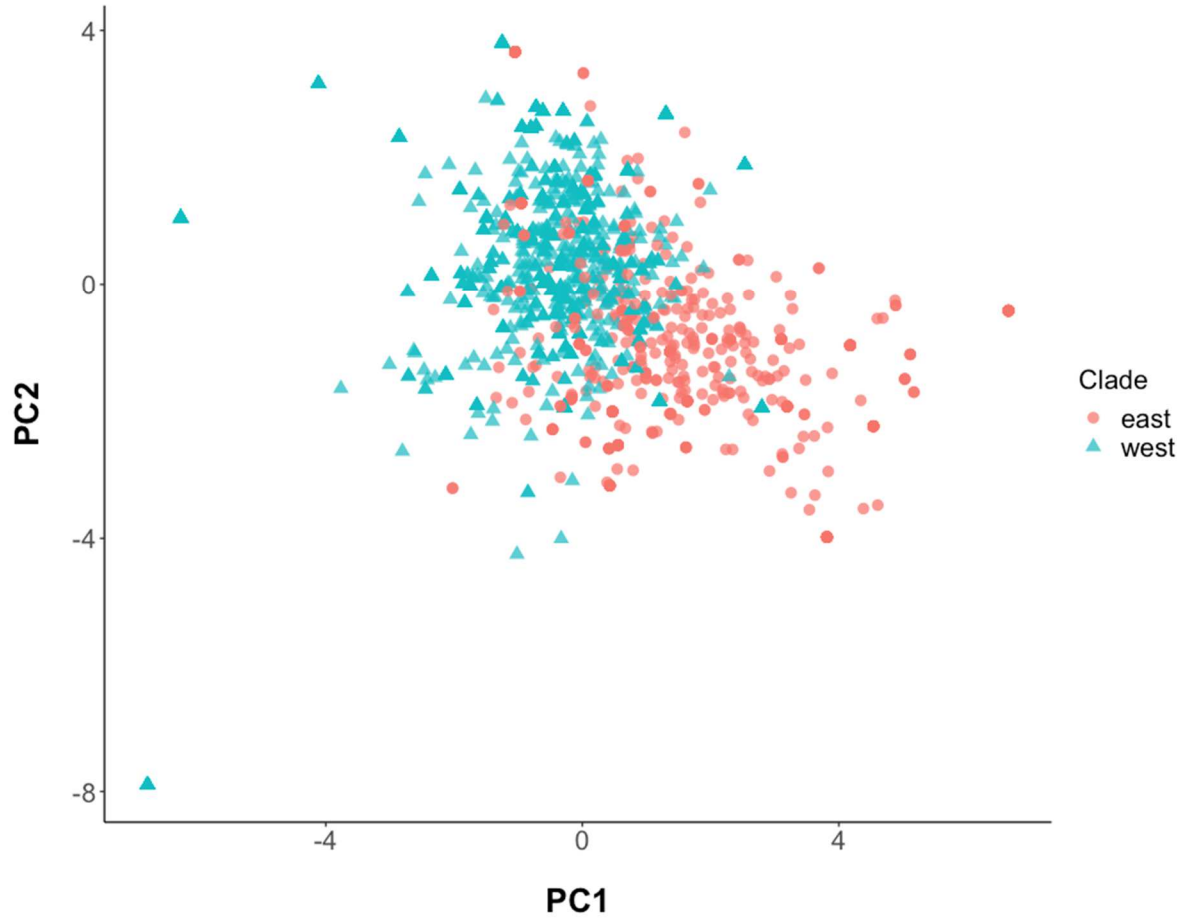


Figure 2.5. Relationship between expected heterozygosity (H_E) and latitude of *A. triloba* populations ($p = 0.068$).

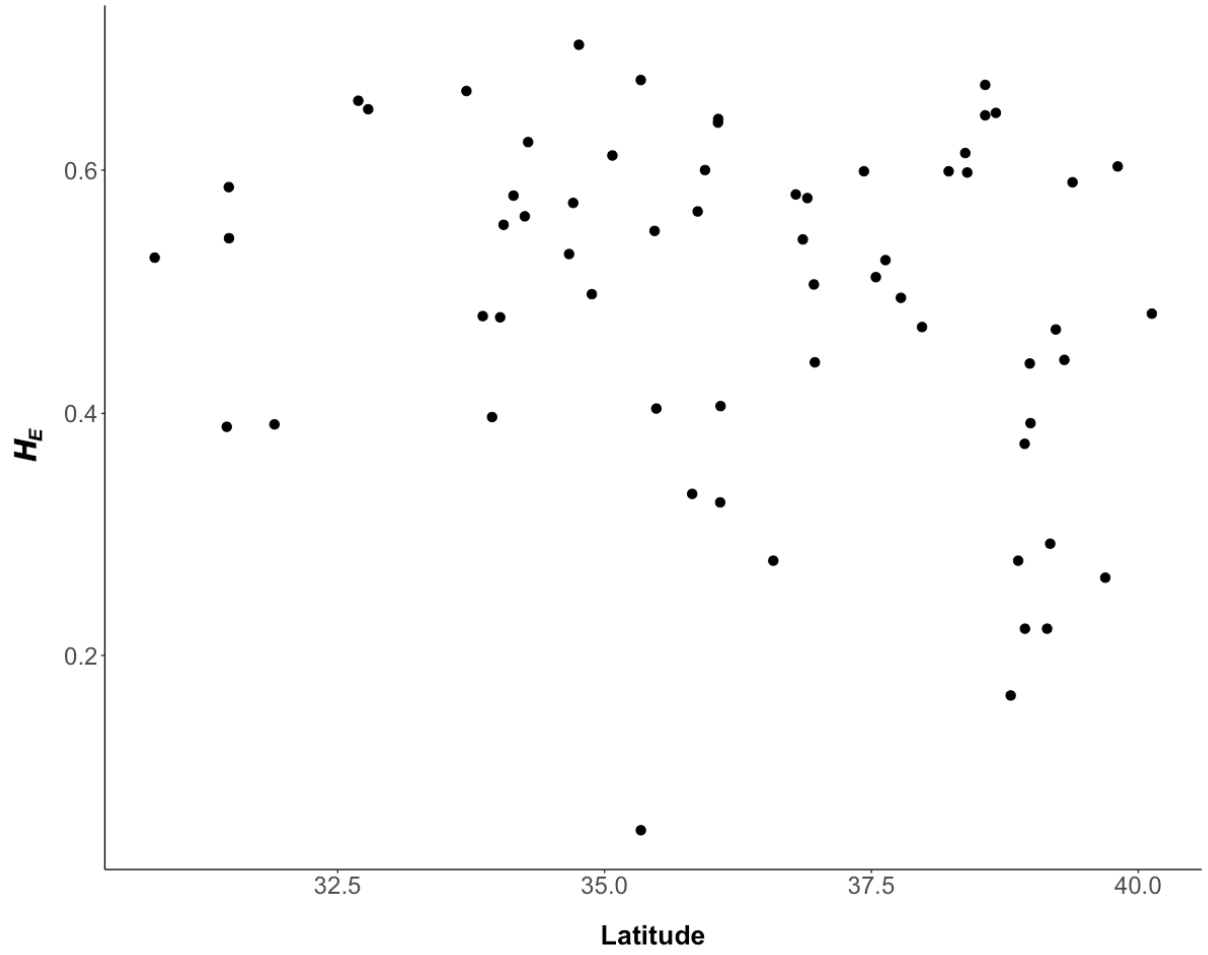


Figure 2.6. Relationship between number of rare alleles (alleles with a frequency < 0.1 within populations) and latitude of *A. triloba* populations (r = 0.28, $p = 0.03$).

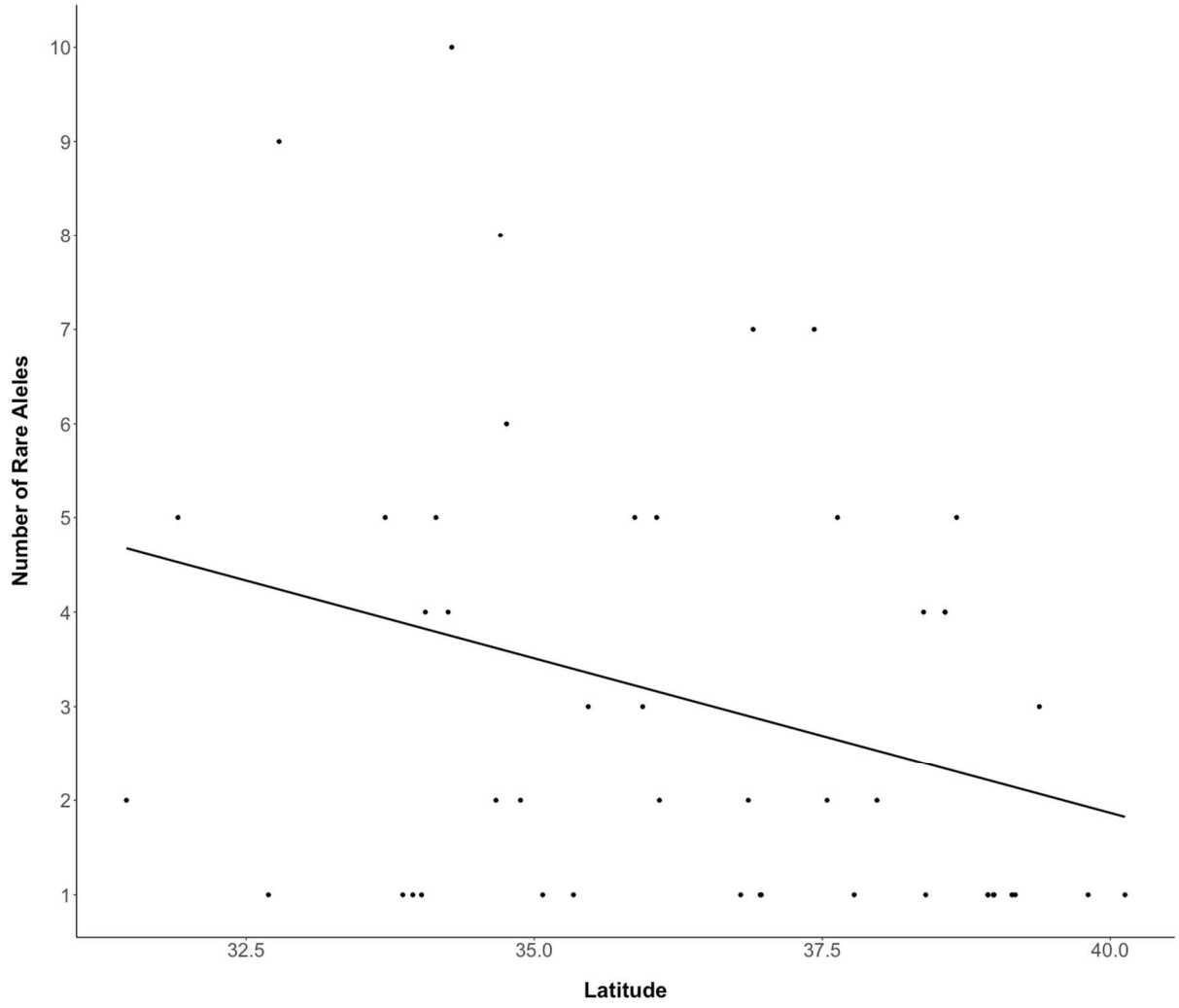


Figure 2.7. Realized present-day niche of *A. triloba* modeled through maximum entropy using seven climatic variables (Worldclim) and elevation (USGS). Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*.

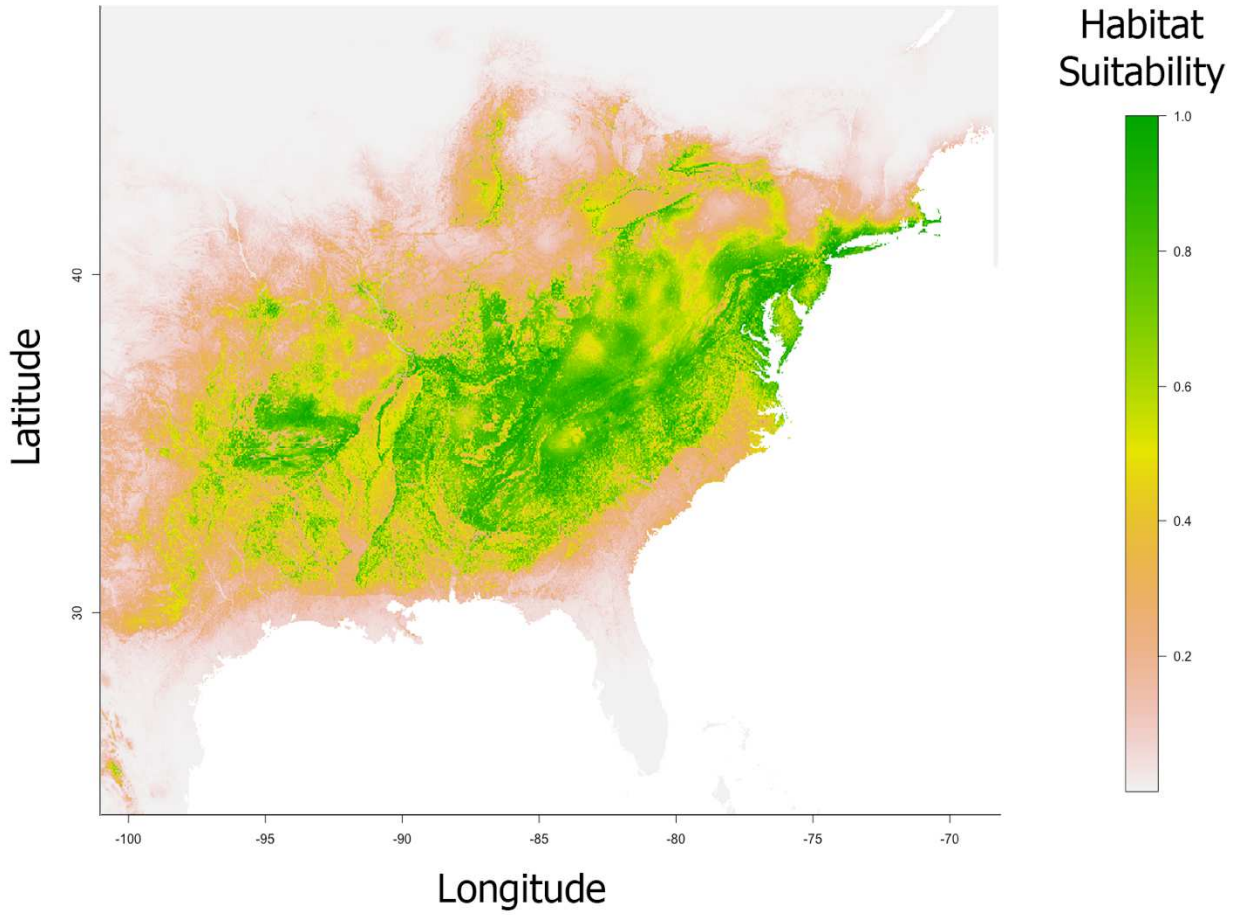


Figure 2.8. Projection of habitat suitability for *Asimina triloba* during the Last Glacial Maximum in the Pleistocene Epoch (~22,000 YBP), as inferred from the average of three climatic data sets (CCSM4, MIROC, and MPI-ESM) from six bioclimatic variables as well as elevation. Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*.

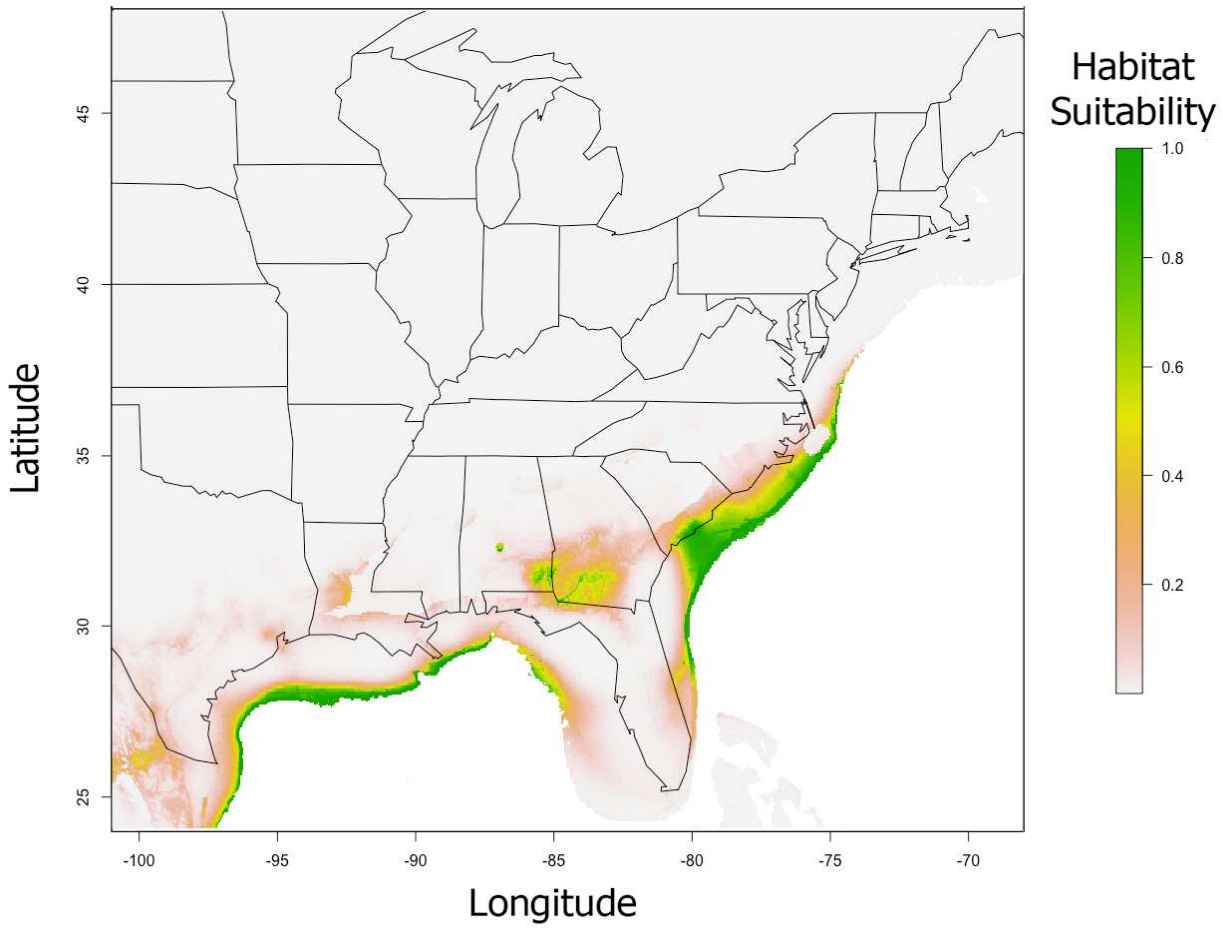


Figure 2.9. Predictions of *A. triloba* habitat suitability during the Mid Holocene Epoch (~6,000 YBP), as inferred from the average of three climatic data sets (CCSM4, MIROC, and MPI-ESM) from six bioclimatic variables as well as elevation. Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*.

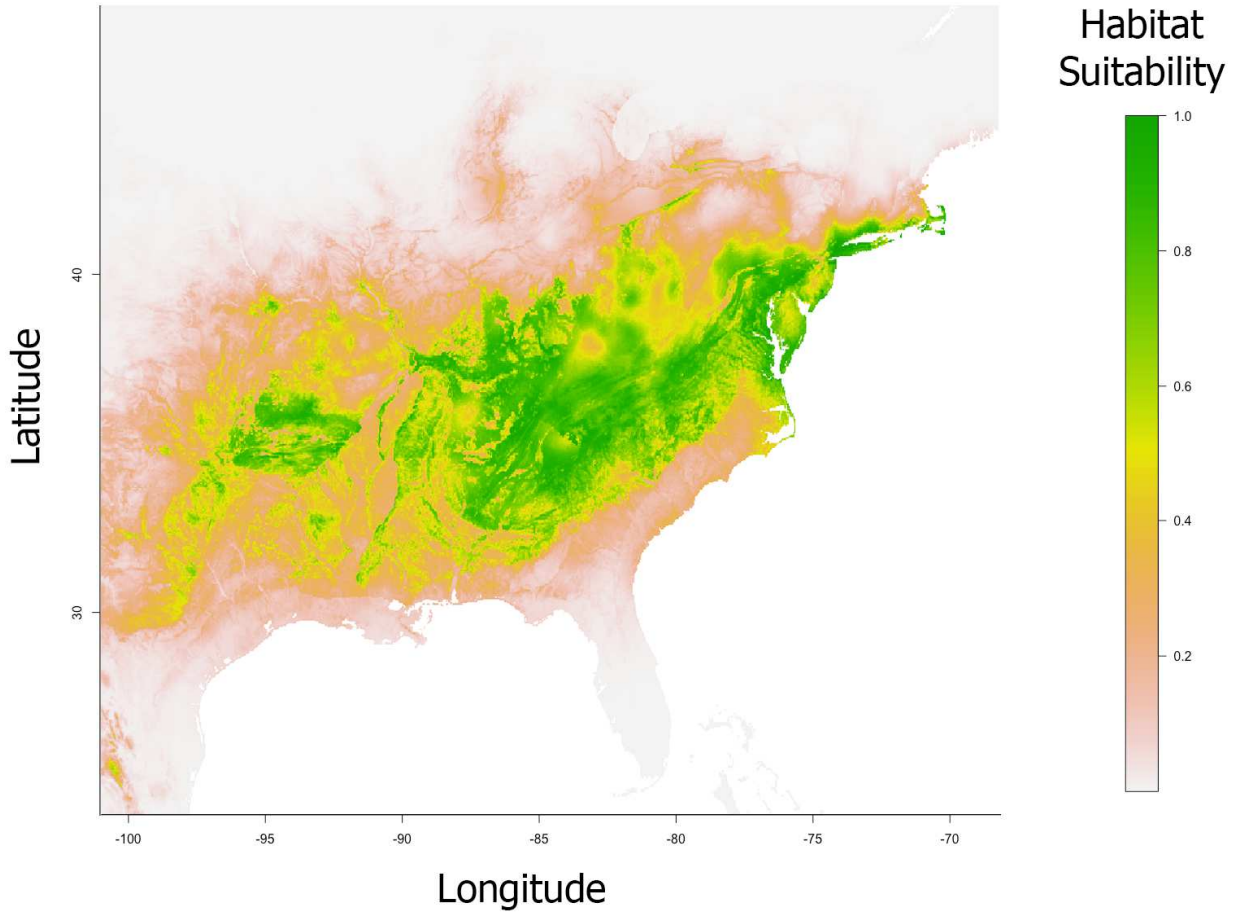
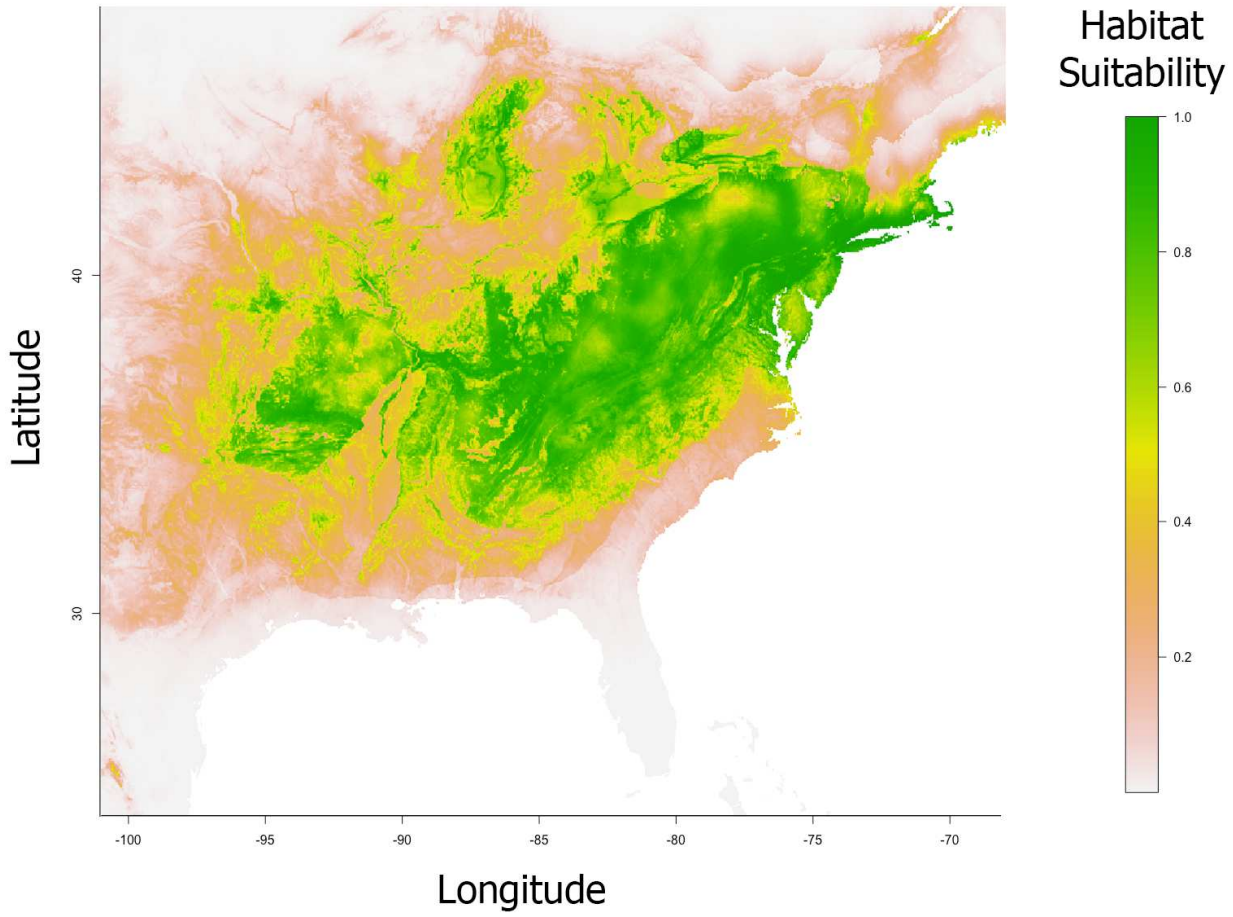


Figure 2.10. Projection of habitat suitability for *Asimina triloba* for the year 2070, as inferred from the average of three climatic data sets (CCSM4, MIROC, and MPI-ESM) and two RCP pathways (2.6 and 8.5) from seven bioclimatic variables as well as elevation. Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*.



CHAPTER 3

GENETIC DIFFERENCES BETWEEN ANTHROPOGENIC AND WILD POPULATIONS OF *ASIMINA TRILOBA* (ANNONACEAE), A NORTH AMERICAN TREE SPECIES[1]

[1] Graham E. Wyatt. To be submitted to Molecular Ecology

ABSTRACT

Humans have long been implicated in the dispersal of North American flora during the Holocene. Human-mediated dispersal is a viable explanation for the apparent rapid long-distance dispersal of species lacking mechanisms to reliably achieve long-distance dispersal (i.e., gravity or small mammal-dispersed seeds), yet historic accounts provide little to no evidence of dispersal of plants by humans. Associations between pre-Columbian Native Americans and the useful flora surrounding their settlements have been observed throughout the New World and suggest interactions between utilized plants and humans. Frugivorous Pleistocene megafauna that went extinct ~12,000 YBP may have also played a key role in the dispersal of fruit-producing trees lacking extant analog dispersers. However, this hypothesis is difficult to test and cannot account for more recent (~12,000 YBP to present) dispersal events. *Asimina triloba*, which possesses the largest fruits of any native tree species, is thought to be an example of both an evolutionary anachronism with regard to seed dispersal (i.e., extinct co-evolved seed dispersers) and a tree species valued by humans. Seeds of this understory tree have been found in ancient Native American middens in the Ohio River Valley, and its fibrous bark was utilized as well. The results of this study found significantly lower levels of genetic diversity in putative anthropogenic populations (i.e., populations located < 1 km from pre-Columbian settlements) than wild populations for most genetic diversity parameters. However, observed heterozygosity was significantly higher in anthropogenic populations. This result appears to be due to the multilocus genotypes of these populations being heterozygous at a higher proportion of its loci. High levels of genetic structure among these northernmost anthropogenic populations is indicative of independent introductions from separate source populations followed by low gene flow among

these northern anthropogenic populations. This study provides the first direct genetic evidence of human manipulation of a valued eastern North American tree species, *A. triloba*.

INTRODUCTION

During ice age glacier advances, Northern Hemisphere plant species retreated to southern refugia and during subsequent warming periods there was extensive northward migration (Hewitt 2000; Davis & Shaw 2001; Lyford et al. 2003). Understanding historical patterns of plant dispersal and mechanisms responsible for movement during the warming climate that ensued after the last glacial maximum (LGM) (Clark et al. 2009) allow inferences regarding their capacity for range modification in response to future climatic shifts. Interestingly some tree species with limited apparent seed dispersal migrated considerable distances after glacial retreat in the late Pleistocene (~19,000 YBP), a phenomenon known as 'Reid's Paradox' (Clark et al. 1998; MacDougall 2003). In 1899 Clement Reid was the first of many to note the surprisingly rapid rate of northward migration from glacial refugia by large-seeded trees, such as oaks, into formerly glaciated regions of Great Britain.

While some animal vectors can be effective long-distance seed dispersers, it has been hypothesized that humans may have also played a role in dispersing valued tree species beyond the southern limits of their glacial refugia (MacDougal 2003; Delcourt & Delcourt 2004; Abrams & Nowacki 2008). There is considerable evidence that pre-Columbian peoples had a marked impact on the North American landscape prior to the arrival of the first Europeans (Pyne 1983, Denevan 1992; Stewart 2002; Abrams & Nowacki 2008). However, little attention has been paid

to the possible role of native peoples in the dispersal of useful plant species (but see Petit 2002; Cottrell et al. 2002; MacDougall 2003; Delcourt & Delcourt 2004; Abrams & Nowacki 2008) and some population genetic studies have intentionally avoided species valued or impacted by humans (Petit et al. 2003). One of the few studies involving a North American tree species used genetic markers and packrat middens (pollen and macrofossils) to investigate an isolated (by > 200 km) northern stand of *Pinus edulis* (Colorado pinyon pine). Bentancourt et al. (1991) concluded that this isolated population originated by long-distance dispersal and postulated that Native Americans, who valued the pine nuts as a food source, served as the dispersal vectors. Additional evidence of human-mediated dispersal in North America comes from tobacco (*Nicotania* sp.) and bottle gourd (*Lagenaria siceraria*) (Asch 1995), which were thought to be absent from prehistoric Illinois and were later introduced by external contacts and subsequently cultivated as useful specialty-plant species. Wykoff (1991) has also asserted that black walnut (*Juglans nigra*), probably some hickories (*Carya* spp.), oaks (*Quercus*), and medicinal species were likely introduced into New York State prior to European contact.

It is generally agreed that indigenous peoples inhabiting North America manipulated their environment to manage resources since their arrival ~15,000 (Morse & Morse 1983; Delcourt & Delcourt 2004) to 17,000 years before present (YBP) (Adovasio et al. 1978; Richland et al. 2007). Yet debate remains about the uniformity of their influence over space and time (Muñoz et al. 2014). At least 15,000 YBP an ice-free corridor connected Alaska and the rest of the continental United States, permitting human migration into this region. Earlier arrival of humans in North America ~ 16,000 YBP also could have been possible along the coasts of the North Pacific (Erlandson et al. 2015). While Denevan (2011) estimated 2 million indigenous people

lived in North America in 1492, there is much controversy over the extent of anthropogenic impact on the present distribution of plant species.

Some investigators maintain that impacts by indigenous peoples were ubiquitous, with few ecosystems escaping their impact (Denevan 1992; Krech 2000; Kay 2002; Abrams & Nowacki 2008), but that after the introduction of Old World diseases by Spanish explorers and the decimation of indigenous populations (Lovell 1992), North American ecosystems recovered somewhat and may have obscured indigenous imprints on the landscape (Denevan 1992). Others have argued that landscape alteration near densely settled areas showed greater impacts, while more remote sparsely inhabited areas exhibited little anthropogenic alteration (Vale 1998; 2002; Parker 2002; Muñoz et al. 2014). Muñoz et al. (2014) contended that agricultural and silvicultural impacts were localized to the immediate vicinity of settlements (< 1 km to tens of kilometers) and were greatest along riparian corridors, near settlements, and along trade routes. By approximately 2,000 YBP, indigenous societies throughout the river valleys of Eastern North America were interacting widely, erecting monuments, and producing sophisticated material culture (Wright 2017). Furthermore, plants used by indigenous people may show signs of manipulation in proximity to settlements but may exist in an un-manipulated state farther away. Variable land-use patterns were exhibited by indigenous peoples, including: 1) semi-nomadic hunter-gatherers with temporary encampments, 2) seasonal movements and periodic occupation of resource-specific sites (e.g., hunting grounds, fish and shellfish, nut trees) to 3) stable villages associated with agricultural fields or aquatic food sources (Abrams & Nowacki 2008). Highlands and Piedmont Lowlands were characterized by more temporary seasonal encampments (Pagaoulatos 1992). More stable villages were located along major rivers and tributaries of Ridge

and Valley areas and the Coastal Plain, while foraging destinations were scattered across the landscape. Hunter-gatherer groups were more typical in the north, and more agricultural-based subsistence was practiced from Massachusetts southward especially along major river lowlands (Patterson & Sassaman 1988). By 2,000 YBP, large agricultural fields associated with settlements were established in the floodplains of the larger Midwest and southern rivers (Fritz 1990; Scarry & Scarry 2005). Some cultures relied heavily on nutritious nuts (i.e., balanocultures) of mast-producing species (chestnut, hickory, and oaks) for sustenance and are thought to have planted orchards and manipulated the distribution of some tree species (Abrams & Nowacki 2008). Acorns served as a particularly important dietary staple for many indigenous peoples (Bainbridge 1985; Logan 2005), and widespread acorn use almost certainly predates the widespread use of corn (Bainbridge 1985). Inadvertent or intentional anthropogenic seed dispersal, particularly in these more heavily impacted portions of the landscape, may have contributed to post-glacial maximum range expansion (e.g., White 1906; Wykoff 1991; Keener & Kuhns 1997; Abrams & Nowacki 2008) and may help explain Reid's Paradox.

The overall goal of this study is to investigate the role of indigenous pre-Columbian people on the genetic characteristics of *Asimina triloba*. *Asimina triloba*, commonly called Pawpaw or Indian Banana, has the largest (5-16 cm long, 3-7 cm wide) edible fruits of any native North American tree and have a flavor described as a cross between banana, mango and pineapple. Its fruits were prized as a food source by Native Americans (e.g., Brooks & Johannes 1990; Peterson 1991; Keener & Kuhns 1997), and the fibrous bark was used as textile material to make rope and cloth (Peattie 1950). *Asimina triloba* may have been utilized medicinally as well (Krochmal & Krochmal 1973; Peterson 1991). Seeds of *A. triloba* have been found in middens

from the Ozarks and throughout the Ohio River Valley (Gilmore 1931; Jones 1936; Yarnell & Watson1966) and may have been dispersed by these and other cultures (Brooks & Johannes 1990; Wykoff 1991; Peterson 1991; Keener & Kuhns 1997). The first written record of *A. triloba* comes from Hernando de Soto's expedition across Southeastern North America, which in 1541 reported widespread planting of the tree by indigenous tribes of the southeast and mentions its flavor and fragrance (Pickering 1879). Some authors have asserted that *A. triloba* was grown by pre-Columbian humans (Peattie 1950; Peterson 1991; Hormaza 2014; Cai et al. 2019), although no known evidence for the cultivation of *A. triloba* orchards by Native Americans has been discovered. There is however molecular evidence for pre-Columbian anthropogenic dispersal of *Annona cherimola*, between Central and South America (Larranaga et al. 2017). *Annona cherimola* shares a number of similarities with that of *A. triloba* in terms of being a woody perennial species with large, edible and nutritious fruits.

The specific goals of this study are to 1) estimate the genetic diversity of *A. triloba* populations of putative anthropogenic origin (i.e., those occurring on or adjacent to known Native American settlements), 2) compare levels and patterns of genetic diversity between putative anthropogenic and wild populations (i.e., those not occurring near known Native American settlements).

MATERIAL AND METHODS

Study Species

Asimina triloba is the most geographically wide-spread species within its genus, occurring from eastern Texas to New Jersey and as far northwest as southeastern Iowa, with the northern extent of the range being southern Ontario, Canada (Little 1977). *Asimina triloba* occurs in mesic, alluvial

forests along rivers and creeks and typically at less than 350 m elevation (Murphy 2001; Pomper et al. 2010; Freeman & Hulbert 1985), resulting in a patchy distribution across landscapes. *Asimina triloba* is an understory tree, with stems rarely exceeding a dbh greater than 20 cm (Wyatt pers. obs.). Information on the longevity of individual stems is scarce. A mature stem with a dbh of 11.4 cm from a population in Georgia was observed to have 32 annual growth rings (G. Wyatt unpubl. data). Owing to its clonal growth through root-suckers (Keener & Kuhns 1997), populations of *A. triloba* are often dense and patchily distributed within their habitat, with larger stems often at the center of a patch. Also it is likely that individual genets persist for long periods of time (Hosaka et al. 2005; Wyatt & Trapnell Chapter 3 in prep). Mature stems produce flowers in the spring, typically after the last frost, and flowering on a given stem continues for roughly a month. Pollinators are weak flying insects such as flies, beetles, or small lepidoptera (Willson & Schemske 1980; Goodrich et al. 2006; Wyatt pers. obs.). *Asimina triloba* produces the largest edible fruits of any native tree in North America that can weigh up to 1 kg (Darrow 1975). Long-distance dispersal along rivers may be facilitated by the fact that the fruits float (Bowden & Miller 1951; Peterson (1991); Wyatt obs. pers.). Seeds are also dispersed by racoons and coyotes (Cypher & Cypher 1999; Wyatt pers. obs.). Deer do not appear to disperse the seeds in the wild (Wyatt pers. obs.); only in captivity have they been shown to consume *A. triloba* fruits (Murphy 2001). It has been suggested that the highly nutritious fruits of *A. triloba* may be an evolutionary anachronism, in that frugivorous seed dispersers that co-evolved with *A. triloba* (e.g., mastodons, ground sloths, horses, etc.) are no longer extant (Janzen & Martin 1982; Poor 1984).

Sampling

Leaf samples were collected from 82 populations distributed throughout the species' range. Populations were separated from one another by a mean of 681.3 km (range = 0.2 to 1845.8 km).

Twenty populations are classified as of anthropogenic origin based on their location < 1 km from documented pre-Columbian habitations (i.e., villages or mound sites) (Muñoz et al. 2014). Sixty-two populations are designated as 'wild' as there were no known indigenous sites or trade routes located within a 1 km radius of these populations. It is important to note that some populations designated as 'wild' may in fact be anthropogenic, due to incomplete documentation regarding pre-Columbian population sites and the dating of possible occupation of sites that in some cases may date from 15,000 – 17,000 YBP (Adovasio et al. 1978; Morse & Morse 1983; Delcourt & Delcourt 2004; Williams et al. 2018).

Leaf tissue samples were obtained from a mean of 28.1 individuals (range = 5 to 50) from each of the 20 anthropogenic *A. triloba* populations (Table 3.1; Figure 3.1) and a mean of 29.6 individuals (range = 10 to 62) from the 62 wild populations (Figure 3.1). Leaves were collected from mature (≥ 3 m in height) stems separated by ≥ 10 m when possible to avoid sampling multiple ramets belonging to the same genet. Samples were snap frozen in liquid nitrogen for transport to the University of Georgia for analysis.

Microsatellite Genotyping

Frozen leaf tissue (~ 0.05 g) was pulverized and genomic DNA was extracted using a modified CTAB protocol (Doyle & Doyle 1983). DNA quality and quantity were evaluated using an ND-1000 Nanodrop® spectrophotometer. Nine nuclear microsatellite loci were amplified with markers developed by Lu et al. (2011). PCR amplification was carried out in 12.5 µL reaction volumes containing 3.25 µL molecular grade ddH2O, 2.5 µL 5X One Taq® standard reaction buffer (New England Biolabs, Ipswich, MA), 0.35 µL 25 mM MgCl solution (Sigma-Aldrich), 1.25 µL 10X (1 mg/mL) bovine serum albumin (Thermo-Fisher), 1 µL 2.5 mM dNTP mix (New

England Biolabs, Ipswich, MA), 1 µL primer mix (0.5 µM CAG-labeled primer and 5 µM unlabeled primer), 0.45 µL 10 µM universal dye-labeled primer, 0.2 µL One Taq® Hot Start DNA polymerase (New England Biolabs, Ipswich, MA), and 2.5 µL of diluted template DNA (20 ng/µL). A 3-primer PCR protocol was used whereby a CAG-tag sequence (Hauswaldt and Glenn 2003) was added to the 5' end of one primer and a third fluorescently labeled (FAM, HEX, or NED) primer identical to the CAG-tag was included.

Genetic Analyses

Duplicate genotypes within populations that represented clones were removed such that only one individual per multilocus genotype (MLG) was included in subsequent genetic analyses. To be relatively certain that two individuals sharing an identical MLG were clones and not representing the identical MLG by chance, it was necessary to calculate the probability of identity (P_I). This parameter is dependent on the number of loci, allele frequencies within a population, and the number of stems sampled in that population. P_I was calculated for each population and gave the likelihood of two individuals within a population having an identical MLG by chance. The multilocus probability of identity (PI_m) was calculated as $PI_m = \Pi_s (PI_s)$, where

$$PI_s = \sum_i x_i^4 + \sum_i \sum_{j>i} (2x_i x_j)^2$$

and x_i and x_j are the frequencies of the ith and jth allele, respectively. The probability of identity (P_I) was adjusted for sample size (N) by $P_I = 1 - [\ (1 - PI_m)^N]$.

Null alleles are a problem associated with microsatellite analyses and can potentially bias estimates of heterozygosity (Jarne & Lagoda 1996). A test for null alleles was performed in Genepop version 4.2 using Brookfield's method (Raymond & Roussett 1995; Brookfield 1996). Genetic diversity statistics were estimated by GenAlEx version 6.51b2 (Peakall & Smouse 2006; Peakall & Smouse 2012). Measures of genetic diversity within populations were percent of polymorphic loci (*P %*), total number of alleles (A_T), effective number of alleles (A_E), number of private alleles (P_A; i.e., an allele found in a single population), rare alleles (A_{RARE}; alleles in ≤ 10 % of populations), observed heterozygosity (H_O), expected heterozygosity (H_E), and the inbreeding coefficient (F_{IS}). Population values were estimated and averaged across all populations while species level values were calculated by pooling data from both wild and anthropogenic populations. Observed heterozygosity (H_O) and Hardy-Weinberg expected heterozygosity (H_E) were compared for each locus in each population, using Wright's inbreeding coefficient (F_{IS}; Wright 1922, 1951) and deviations from Hardy-Weinberg expectations were tested for significance using $\chi 2$ (Li & Horvitz 1953). The Bonferroni correction for multiple comparisons was applied using FSTAT (Goudet 1995).

Rarefaction of the total number of alleles (A_T) and effective number of alleles (A_E) was performed through the R-package "gstudio" version 1.5.0 using the default 999 re-sampling iterations (R Core Team 2013; Dyer 2016). This was done because the total number of individuals from sampled wild populations (N = 62) was greater than that of the anthropogenic populations (N = 20). Differences between these genetic parameters for wild and anthropogenic populations following rarefaction were analyzed for statistically significant differences.

The number of total sampled populations (82), samples per population (mean = 28.1), broad geographic sampling area, and the total number of alleles observed (163) allowed the identification of rare alleles, defined as alleles present in fewer than 10 % of the populations. Differences between numbers of rare alleles between wild and anthropogenic populations were estimated, as well as differences in pairwise geographic distances between populations in networks of rare alleles (i.e., populations sharing a rare allele). The pairwise distances between populations sharing a rare allele were expected to be smaller than pairwise distances between randomly chosen populations. The number of rare alleles was linearly regressed against latitude and tested for significance, to determine if there was a south-to-north pattern.

Genetic structure among populations was estimated using Nei's (1973) G_{ST} and pairwise G_{ST} values were calculated for all possible pairs of populations in GenAlEx.version 6.5.1b2 (Peakall & Smouse 2006; Peakall & Smouse 2012). Additionally, the mean number of shared alleles within a population (i.e., the average number of alleles that it shares with all other populations) was determined for each population. This number gives an estimate of genetic similarity to other populations and was shown to be analogous to pairwise G_{ST} values. These values were compared between wild and anthropogenic populations. STRUCTURE version 2.3.4 (Pritchard et al., 2000), a Bayesian clustering approach, was used to estimate levels of genetic admixture among populations and the number of genetically distinct clusters (K) of all 82 sampled populations. Ten independent simulations at each K-value from 1 to 20 were run, using a burn-in of 500,000 repetitions and a run length of 1,000,000 Markov chain Monte Carlo (MCMC) iterations. The admixture model was chosen to infer alpha (α). A model based on correlated allele frequencies with no prior assumptions regarding sampling locations was selected. The Evanno et al. (2005)

method was used to determine the optimal ΔK using Structure Harvester (Earl and vonHoldt 2012).

Estimates of isolation by distance (IBD; Wright 1943) were carried out by using regression: i) (G_{ST} / 1 - G_{ST}) and log geographic distance (Rousset 1997) for anthropogenic and wild populations. The results were evaluated for significance.

Alleles that were present in anthropogenic populations were investigated for patterns that suggest origin or ancestry with wild populations. For example, alleles in low frequency among wild populations that increase in frequency with latitude and/or are present in a "stepping-stone" pattern of anthropogenic populations with increasing latitude would be indicative of human-mediated dispersal.

RESULTS

Genetic Diversity

Null allele frequencies were found to be non-significant for all loci and all populations (p = 0.05). This is consistent with the excess of heterozygosity and negative F_{IS} values observed in both anthropogenic and wild populations. Most population genetic statistics (MLG, P %, A_T, A_E, P_A, A_{RARE}) indicated that anthropogenic populations had less genetic diversity than wild populations (Table 3.2). In fact MLG (p = 0.01), A_T (p = 0.01), A_E (p = 0.09), A_{RARE} (p = 0.03), and F_{IS} (p = 0.02) were all significantly greater in wild populations than anthropogenic populations. P % and H_E were not significantly different between wild and anthropogenic populations (p = 0.17; p = 0.14, respectively) but the wild populations had higher levels of

genetic diversity for both parameters. Observed heterozygosity (H_O) was the only statistic that was higher in anthropogenic populations (mean = 0.586) than wild populations (mean = 0.530) at a significance level of 0.1 (p = 0.08; Table 3.2). Anthropogenic populations had significantly fewer total alleles (A_T) (mean = 25, pooled = 95) than wild populations (mean = 32.5, pooled = 153) (p = 0.01; Figure 3.2). The rarefaction of genetic data revealed that the differences in the total number of alleles (A_T) and effective number of alleles (A_E) between anthropogenic and wild populations are not likely due to the lower number of anthropogenic populations ($p < 0.01$). The results of rarefaction on these statistics are shown in Figures 3.3 and Figure 3.4. The mean F_{IS} value for wild populations (-0.153) was significantly higher than that of anthropogenic populations (-0.426) (p = 0.02; Table 3.2). This indicates that the anthropogenic population had a greater excess of heterozygosity relative to Hardy-Weinberg expectations that of the wild populations.

A total of 27 private alleles (P_A) were found in 20 wild and eight private alleles were found in five anthropogenic populations. These large differences may be due to differences in the number of sampled populations between anthropogenic (N = 20) and wild populations (N = 62). The large number of sampled populations (82) and highly polymorphic markers, which yielded 163 alleles in total across all 9 loci, permitted analyses into networks of rare alleles. A total of 54 rare alleles (i.e., those occurring in 2-8 populations of the 82 total) were found in 53 wild populations and 14 anthropogenic populations. Rare alleles occurring exclusively in networks of wild populations (N = 34 [63 %]) were more numerous than those occurring in mixed networks of wild and anthropogenic populations (N = 20 [37 %]). No rare alleles were exclusively within anthropogenic populations. The number of rare alleles decreased with increasing latitude (Figure

3.5; $p = 0.0001$). The number of rare alleles was significantly greater in wild populations than anthropogenic populations ($p = 0.03$). The mean pairwise distances between populations within networks of all rare alleles was 623.6 km. This differed significantly ($p = 0.002$) from a random distribution of equal size generated in R (R Core Team 2013), having the same mean and standard deviation as all pairwise distances for all 82 populations. The mean pairwise distance for networks of wild populations sharing a rare allele was 589.5 km, significantly less ($p < 0.01$) than the overall mean population pairwise distance (681.3 km). For mixed networks of wild and anthropogenic populations sharing a rare allele, the mean pairwise distance was 714.9 km, significantly greater than the distance for wild only networks of rare alleles (Figure 3.6).

Global G_{ST} among the 62 wild populations was 0.341, while the global G_{ST} among the 20 anthropogenic populations was 0.381. Mean pairwise GST for wild populations was 0.243 and was 0.293 for anthropogenic populations. The mean number of shared alleles for wild populations was significantly higher than that of anthropogenic populations ($p = 0.03$). Northern populations of *A. triloba* associated with pre-Columbian population centers had high levels of genetic structure (G_{ST}) between each other (Figure 3.7). The mean G_{ST} between the four northernmost anthropogenic populations was 0.408. Additionally, the four populations with the lowest pairwise G_{ST} for each of the four northern anthropogenic populations are all located long distances (> 450 km) away (Figure 3.7). Simulations from STRUCTURE yielded an optimal $K = 2$, with mean Ln P(D) ranging from −63469.2 to −42744.8. The Q-matrix of the simulation with the highest estimated Ln P(D) for $K = 2$ was converted into pie frequency charts for each population (Figure 3.8). The two genetic clusters identified were consistent with results using only the 62 wild populations in Wyatt & Trapnell Chapter 1 (in prep). As with that analysis,

these two clusters fell primarily on either side of the Appalachian Mountains, along with an east-west phylogeographic discontinuity along the Tombigbee drainage in Alabama (Figure 3.8).

Estimates of isolation by distance using the Rousset (1997) method revealed a weak but significant relationship between G_{ST} and geographic distance in wild populations (r = -0.006; p = 3.4.x 10^{-13}). No relationship between G_{ST} and geographic distance was observed for anthropogenic populations (r = 0.06; p = 0.37).

Patterns of specific alleles in anthropogenic populations that may indicate origin or ancestry with wild populations were observed. Among these are allele 157 at locus Pp-B128 (Figure 3.9), which appears to increase in frequency and becomes only associated with anthropogenic populations above approximately 39° N latitude. A similar pattern is observed in allele 177 at locus Pp-B128 (Figure 3.10). Interestingly, while the anthropogenic populations LM1 and LM2 have close to the same latitude as the anthropogenic population SU1 (Figure 3.1; Figure 3.10), they are 682.5 km away from SU1. Those three populations are the only to share allele 177 at locus Pp-B128 above 39° N latitude. Allele 207 at locus Pp-124 is less common, found only in four populations, but appears in the anthropogenic population FCP in Florida and at a much higher frequency 1242.6 km northward in anthropogenic population LCN in Michigan (Figure 3.11). Allele 240 at locus Pp-121 is a more common allele, found in 27 populations, although it shows a similar pattern to the aforementioned alleles. It is present in a southern anthropogenic population and many wild southern populations, with five out of six of its more northern host populations being anthropogenic (Figure 3.12). Allele 354 at locus B103 appears likely to have

originated in the western part of the range, although it is found in very distant anthropogenic populations (LM1, PET, and FCP) (Figure 3.13).

DISCUSSION

Comparisons of Genetic Diversity in Anthropogenic and Wild Populations

Lower levels of genetic diversity for populations occurring near sites historically occupied by humans has been observed in other plant genera, such as *Agave* (Parker et al. 2007), *Spondias* (Miller & Schaal 2006), *Phaseolus* (Bitocchi et al. 2012), *Theobroma* (Risterucci et al. 2002), and *Annona* (Larranaga et al. 2017) (see also Levis et al. 2017). The significantly lower total number of alleles in anthropogenic populations (mean = 25) as compared to wild populations (mean = 32.5) suggests that human impacts have decreased the genetic diversity of anthropogenic populations of *A. triloba* (Figure 3.2). These results suggest that seeds and/or propagules were removed from wild populations to pre-Columbian population centers and/or were potentially transported long distances. Seeds or propagules obtained from wild populations were likely selected for favorable traits (i.e., fruit production, fruit size, and/or fiber quality) (Peattie 1950; Peterson 1991). While the large difference in the number of sampled anthropogenic and wild populations (20 vs. 62) may explain the difference in the number of private alleles observed between the two, differences in other genetic diversity measures, such as MLG and total number of alleles, are likely not a reflection of the number of sampled populations. The number of sampled stems for anthropogenic and wild populations are nearly identical (28.1 vs. 29.6), and the method of sampling did not differ between the two population

types. Rarefaction of the mean number of alleles also showed that anthropogenic populations had less genetic diversity than wild populations (Figure 3.3).

F_{IS} has been observed to be low in populations of organisms that reproduce asexually (Birky 1996; Ballou et al. 2003; Bryzski & Culley 2011; Meloni et al. 2013). This observation is consistent with findings from Wyatt & Trapnell (in prep), in that the anthropogenic populations sampled in that study are more clonal than wild populations. The significant excess heterozygosity in anthropogenic populations ($p = 0.021$) may have resulted from the dominance of well-adapted genotypes over time, eventually leading to the wide-spread establishment of one or a few individuals within these populations. Another explanation for the excess heterozygosity observed in anthropogenic populations, especially those from more northern latitudes, is that seeds or propagules were dispersed from multiple source populations. This trend of decreasing genetic diversity with increasing heterozygosity has been observed to occur in wild populations of *Fagus* in Europe (Comps et al. 2001), though human-interactions could have accelerated this trend in the case of *A. triloba*.

G_{ST} values of anthropogenic populations were slightly higher than those of wild populations, indicating higher population differentiation among those populations. This observation is consistent with the result that wild populations have significantly greater mean number of shared alleles than anthropogenic populations. The four northernmost populations (CHA, CON, LCN, and WEN) are anthropogenic, have high G_{ST} values amongst each other (mean = 0.41), and the overall pairwise G_{ST} among all anthropogenic populations is much lower (mean = 0.29). Each of these four northern populations have low G_{ST} values with other populations (both wild and

anthropogenic) farther south (Figure 3.7). This may be indicative of long-distance dispersal events followed by periods of low or non-existent gene flow between these four populations. Populations also become very patchy and infrequent north of 40° N (Keener & Kuhns 1997; Wyatt pers. obs.). Seeds of *A. triloba* have not been found in trash middens from Archaic sites from areas around Lake Erie and northward, which may be an indication that the introduction of *A. triloba* to these areas may have occurred more recently than 3,000 YBP, or perhaps that the fruits of *A. triloba* were not the primary reason that these populations were established.

Regardless of the dispersal vector, be it natural or anthropogenic, the high G_{ST} values among sampled populations in the north is indicative of multiple separate introductions from genetically distinct source populations followed by low gene flow between populations, rather than a single introduction that subsequently spread. Only one ethnohistorical account exists of *A. triloba* being consumed for food by indigenous tribes of the northeast (Keener & Kuhns 1997). Populations at this latitude tend to be isolated as compared to lower latitudes, and at present appear to have not expanded from locations associated with historically-known village sites (Keener & Kuhns 1997; Wyatt pers obs). Almost no fruit-set was observed in the four northernmost and anthropogenic populations (one fruit on 127 sampled stems), likely due to the low numbers of MLGs present in those populations and/or climatic conditions (Table 3.2). This raises the question of why Native Americans would disperse *A. triloba* to the edge of its present range if it does not yield fruit. A possible answer is that the tree provided some resource other than fruit, perhaps fibrous bark, which can be easily processed into rope or other useful materials (Peattie 1950; Wyatt pers. obs.) or medicinal use (Krochmal & Krochmal 1973). An alternate and compatible explanation may be that these four populations once consisted of more MLGs, but the lack of sexual reproduction

coupled with the loss of poorly adapted MLGs led to the low numbers of MLGs currently found in those populations. Associations between certain useful plant species and known pre-Columbian settlements have been observed in other locations in the Americas (Yarnell 1965; Parker et al. 2010; Larranaga et al. 2017; Levis et al. 2017).

The results from the STRUCTURE analysis on all 82 populations are concordant with Wyatt & Trapnell (in prep), in that they show an east-west discontinuity, thought to have arisen from post-glacial range expansion from separate refugia. Strong differentiation between anthropogenic and wild populations were not observed overall, however some anthropogenic populations (FCP, OSF, SEL, and SU1) are located notably outside the geographic boundaries of their inferred genetic cluster, which could have resulted from long-distance human-mediated dispersal.

Populations along the Little Miami River in southwest Ohio (LM1, LM2, and LM3) are among the best examples of anthropogenic populations. Not only are the populations located less than a kilometer from numerous earthwork mounds from the Fort Ancient culture (~2,000 YBP), but seeds of *A. triloba* have been found in the trash middens of those sites (Keener & Kuhns 1997). Two of the Little Miami River populations (LM2 and LM3) were characterized by low numbers of MLGs (one and three), low numbers of total alleles (A_T), and low F_{IS} values (-1.00 and -0.383). LM1 was located farther from the earthworks (~900 m) and had a much higher number of MLG (14), 43 total alleles, and an F_{IS} value of 0.125, traits that are more like those of wild populations.

The findings in this study could provide evidence to inform archaeological investigations of sites previously unexcavated or unexamined. Several populations designated as wild have genetic diversity parameters that are more similar to populations in anthropogenic sites, possessing low genetic diversity and appearing to share few alleles with surrounding wild populations. An example of this is population PIA along the Mississippi River north of St. Louis, which consisted of only one MLG and appeared to have a high genetic similarity to populations east of the Appalachians (Wyatt & Trapnell, in prep).

The overall decrease in the number of rare alleles (i.e., alleles found in ≤ 10 % of populations) with increasing latitude (Figure 3.5) are consistent with findings in Wyatt & Trapnell (in prep), which saw decreases in genetic diversity with increasing latitude in wild *A. triloba* populations. This decrease in rare alleles present in southern refugia likely reflects the loss of alleles due to founder effects associated with northward migration and relatively low levels of gene flow following colonization (Hewitt 1999).

Allelic Evidence

An allele that is common or present only in certain anthropogenic populations and/or separated from populations sharing that particular allele provide possible evidence for long-distance dispersal by humans. Dispersal of an allele over 1,242 km (as in the case of allele 207 at locus 124; Figure 3.11) would be difficult to account for via animal dispersers. Given the high mutation rate of microsatellites relative to other genomic regions, mutation cannot be discounted as a possible explanation for these patterns. However, there are additional examples of alleles showing similar pattern expected via long-distance dispersal (Figures 3.9, 3.10, 3.12, and 3.13).

Recently, knowledge regarding movement and trade of indigenous people in eastern North America has increased. Humans in North America were exchanging goods over distances as great as 1,000 km as early as the Archaic Period (~ 8000 – 3000 YBP) (Sanger et al. 2018). Taking into account the possible long-distance exchange among indigenous peoples of Eastern North America and the long period of time during which these exchanges may have occurred, human-mediated dispersal is a viable explanation for these allelic patterns.

The high degree of genetic structure between the northernmost anthropogenic populations would suggest low rates of gene flow between those populations and may be a result of the especially patchy distribution of *A. triloba* at that latitude. Fruit-set is low in these populations (G. Wyatt, obs. pers.) and this could contribute to low rates of gene flow between these populations. Additionally, it could reflect the genetically distinct source populations that colonized those areas (Figure 3.7).

CONCLUSIONS

This study represents the first investigation into human-mediated dispersal of a useful tree species in Eastern North America using genetic markers. Perhaps the most salient finding was that genetic diversity statistics (e.g., MLG, A_T, A_E, P_A, A_{RARE} and F_{IS}) were significantly higher in wild populations than putative anthropogenic populations, while H_o was significantly lower. Lower genetic diversity in anthropogenic populations of *A. triloba* could have occurred through early domestication of this wild tree species and/or through repetitive founder events. Along with these possibilities, the high levels of genetic structure observed in northern anthropogenic populations suggests independent dispersal and colonization events from southern source

populations and low levels of subsequent gene flow, which is consistent with the expectation of human-mediated dispersal into widespread northern habitats. The presence and distribution of certain alleles in regards to frequency in putative anthropogenic populations may also provide evidence to human-mediated dispersal.

LITERATURE CITED

Abrams MD, Nowacki GJ (2008) Native Americans as active and passive promoters of mast and fruit trees in the eastern USA. *The Holocene* 18: 1123-1137.

Adovasio JM, Gunn JD, Donahue J, Stuckenrath R (1978) Meadowcroft Rockshelter, 1977: An overview. *American Antiquity* 43: 632-651.

Asch DL (1995) Aboriginal specialty-plant propagation: Illinois prehistory and an eastern North American post-contact perspective. *Unpublished Ph.D. dissertation*, *Department of Anthropology*, University of Michigan, Ann Arbor.

Bainbridge DA (1985) The rise of agriculture: a new perspective. *Ambio* 14: 148-151.

Ballou F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635 – 1644.

Bedell J, Katz G, Shellenhamer J, Kraus L, Groesbeck S (2011) The people of the mountain: archeological overview, assessment, identification, and evaluation study of Catocin Mountain Park, Maryland Vol 1. Final Report. *Prepared by:* Louis Berger Group Inc., Washington, DC. *Prepared for*: National Park Service, Washington, DC.

Bennett BA (1986) Geology of the Mallory Site, Fredonia, New York. Master's Thesis, Department of Anthropology and Sociology, SUNY College, Fredonia, NY.

Bentancourt JL, Schuster WS, Mitton JB, Anderson RS (1991) Fossil and genetic history of a pinyon pine (*Pinus edulis*) isolate. *Ecology* 72: 1685-1697.

Birky CW Jr. (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* 144: 427 – 437.

Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Spagnoletti Z, Gioia T, Logozzo G, Attene G, Nanni L, Papa R (2012) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist* 197: 300 – 313.

Bowden WM, Miller B (1951) Distribution of the Pawpaw, *Asimina triloba* (L.) Dunal, in southern Ontario. *The Canadian Field Naturalist* 65: 27 – 31.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* 5: 453 – 455.

Brooks RR, Johannes D (1990) *Historical, ethno-, and economic botany series*, Vol. 3. Dioscorides Press, Portland, OR.

Brose DA (1976) Conneaut Fort, a prehistoric Whittlesey Focus Village in Ashtabula County, OH. *Pennsylvania Archaeologist* 46: 29 – 77.

Butler BM (1981) Sellars: a small mound center in the Hinterlands. *Tennessee Anthropologist* 6: 37 – 60.

Cai Z, Gold M, Brannan R (2019) An exploratory analysis of US consumer preferences for North American pawpaw. *Agroforest Systems* 93: 1673 – 1685.

Carr C (2008) *Chapter 2: Environmental setting, natural symbols, and subsistence.* In: Case DT, Carr C (eds): The Scioto Hopewell and their neighbors: Bioarchaeological documentation and cultural understanding. Springer, New York.

Clark JS, Fastie C, Hurtt G, Jackson ST, Johnson C, King GA, Lewis M, Lynch J, Pacala S, Prentice C, Schupp EW, Webb III T, Wyckoff P (1998) Reid's Paradox of rapid plant migration. *BioScience* 48: 13-24.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The last glacial maximum. *Science* 325: 710-714.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European Beech. *Genetics* 157: 389 – 397.

Connolly R (2004) Time, space, and function at Fort Ancient. *In*: Connolly R, Lepper BT (Eds) *The Fort Ancient earthworks: prehistoric lifeways of the Hopewell culture in Southwestern Ohio*. Ohio Historical Society, Columbus, OH.

Cottrell JE, Munro RC, Tabbener HE, Gillies ACM, Forrest GI, Deans JD, Lowe AJ (2002) Distribution of chloroplast DNA variation in British oaks (*Quercus robur* and *Q. petraea*): the influence of postglacial colonization and human management. *Forest Ecology and Management* 156: 181 – 195.

Cypher BL, Cypher EA (1999) Germination rates of tree seeds ingested by coyotes and raccoons. *American Midland Naturalist* 142: 71 – 76.

Darrow GM (1975) Minor temperate fruits, pp 276 - 277. In: Janick J, Moore JN (eds) *Advances in fruit breeding*. Purdue University Press, West Lafayette, IN.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. Science 292: 673-679.

Delcourt PA, Delcourt HR (2004) Prehistoric Native Americans and ecological change. Cambridge University Press, Cambridge, U.K.

Denevan WM (1992) The pristine myth: the landscape of the Americas in 1492. *Annals of the Association of American Geographers* 82: 369-385.

Denevan WM (2011) The "pristine myth" revisited. *The Geographical Review* 101: 576-591.

Dyer RJ (2016) gstudio: Tools related to the spatial analysis of genetic marker data. R package version 1.5.0.

Earl DA, vonHoldt BM (2012) Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* 2: 359 – 361.

Erlandson JM, Braje TJ, Gill KM, Graham MH (2015) Ecology of the kelp highway: Did marine resources facilitate human dispersal from Northeast Asia to the Americas? *Journal of Island and Coastal Archaeology* 10: 392 – 411.

Fritz GJ (1990) Multiple pathways to farming in precontact Eastern North America. *Journal of World Prehistory* 4: 387 – 435.

Freeman CC, Hulbert LC (1985) An annotated list of the vascular flora of Konza Prairie Research Natural Area, Kansas. *Transactions of the Kansas Academy of Science* 88: 84-115.

Gardner WM (1966) The Waddell's Mill Pond Site. *Florida Anthropologist* 19: 43 – 64.

Goodrich KR, Zjhra M, Ley CA, Raguso RA (2006) When flowers smell fermented: The chemistry and ontogeny of yeasty floral scent in pawpaw (*Asimina triloba*: Annonaceae). *International Journal of Plant Sciences* 167: 33 – 46.

Hewitt GM (1999) Post-glacial recolonization of European biota. *Biological Journal of the Linnean Society* 68: 87 – 112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405: 907 – 913.

Higgins III TF, Downing CM, Linebaugh DW, Opperman AF, Turner III RE (1995) Archaeological investigations of site 44CF7, Falling Creek ironworks, and vicinity Chesterfield County, Virginia. William and Mary Center for Archaeological Research, Department of Anthropology, College of William and Mary, Williamsburg, Virginia.

Hinsdale WB (1931) Archaeological atlas of Michigan. University of Michigan press, Ann Arbor, MI.

Hormaza JI (2004) The pawpaw, a forgotten North American fruit tree. *Arnoldia* 72: 13 – 23.

Janzen DH, Martin PS (1982) Neotropical anachronisms: The fruits the Gomphotheres ate. *Science* 215: 19 – 27.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424 – 429.

Kandle PL, Downing CM, Linebaugh DW, Deitrick V (1996) Archaeological investigations at site 44CF102 Osbornes/Gatesville Town Chesterfield County, Virginia. William and Mary Center for Archaeological Research, Department of Anthropology, College of William and Mary, Williamsburg, Virginia.

Kay CE (2002) Are ecosystems structured from the top-down or bottom-up? p. 215. *In*: Kay CE, Simmons RT (eds) *Wilderness and political ecology: Aboriginal influences and the original state of nature*. University of Utah Press, Salt Lake City, Utah.

Keener C, Kuhns E (1997) The impact of Iroquoian populations on the northern distribution of pawpaws in the Northeast. *North American Archaeologist* 18: 327–342.

Krech S (2000) *The ecological Indian: Myth and history*. W. W. Norton, NY.

Krochmal A, Krochmal C (1973) A guide to the medicinal plants of the U.S. New York: New York Times Books Co.

Larranaga N, Albertazzi FJ, Fontecha G, Palmieri M, Rainer H, van Zonneveld M, Hormaza JI (2017) A Mesoamerican origin of cherimoya (*Annona cherimola* Mill.): Implications for the conservation of plant genetic resources. *Molecular Ecology* 26: 4116 – 4130.

Levis C, Costa F, Bongers F, Peña-Claros M, et al. (2017) Persistent effects of pre-Columbian plant domestication on Amazonian forest composition. *Science* 355: 925 – 931.

Little, EL Jr. (1977) Atlas of the United States trees, Volume 4: minor eastern hardwoods. Pub. 1342. Washington, DC: US Department of Agriculture. 17 p 230 maps.

Logan WB (2005) *Oak: The frame of civilization*. W.W. Norton & Company, NY.

Lovell WG (1992) Heavy shadows and black night: Disease and depopulation in colonial Spanish America. *Annals of the Association of American Geographers* 82: 426-443.

Lovis WA (1974) The Moccasin Bluff Site and the woodland cultures of Southwestern Michigan. *American Anthropologist* 76: 463 – 464.

Lu L, Pomper KW, Lowe JD, Crabtree SB (2011) Genetic variation in pawpaw cultivars using microsatellite analysis. *Journal of the American Society of Horticultural Science* 136: 415–521.

MacNamara JM (1982) Summary of the 1981 excavations at the Conowingo site (18CE14). Maryland Geological Survey, Division of Archaeology File Report 172. Baltimore, MD.

MacDougall A (2003) Did Native Americans influence the northward migration of plants during the Holocene? *Journal of Biogeography* 30: 633–647.

Meloni M, Reid A, Caujapé-Castells J, Marrero Á, Fernández-Palacios JM, Mesa-Coelo RA, Conti E (2013) Effects of clonality on the genetic variability of rare, insular species: the case of *Ruta microcarpa* from the Canary Islands. *Ecology and Evolution* 6: 1569 – 1579.

Miller AJ, Schaal BA (2006) Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae) *Molecular Ecology* 15: 1467 – 1480.

Morse DE, Morse PA (1983) *Archaeology of the Central Mississippi Valley*. Academic Press, New York, NY.

Muñoz SE, Mladenoff DJ, Schroeder S, Williams JW (2014) Defining the spatial patterns of historical land used associated with the indigenous societies of eastern North America. *Journal of Biogeography*. DOI: 10.1111/jbi.12386

Murphy JL (2001) Pawpaws, persimmons, and 'possums: On the natural distribution of pawpaws in the northeast. *North American Archaeologist* 22: 93-115.

Parker KC (2002) Fire in the pre-European lowlands of the American Southwest. *In*: Vale TR (Ed) *Fire, native peoples, and the natural landscape*, pp. 101-142. Island Press, Washington, D.C.

Parker KC, Hamrick JL, Hodgson WC, Trapnell DW, Parker AJ, Kuzoff RK (2007) Genetic consequences of pre-Colombian cultivation for *Agave murpheyi* and *A. delamateri* (Agavaceae). *American Journal of Botany* 94: 1479-1490.

Parker KC, Trapnell DW, Hamrick JL, Hodgson WC, Parker AJ (2010) Inferring ancient Agave cultivation practices from contemporary genetic patterns. *Molecular Ecology*, 19, 1622-1637.

Patterson WA III, Sassaman KE (1988) Indian fires in the prehistory of New England. *In*: Nicholas GP (Ed) *Holocene human ecology in northeastern North America*, pp. 107-135. Plenum Publishing, NY.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537 – 2539.

Peattie DC (1950) A Natural History of Trees of Eastern and Central North America. Houghton Mifflin Co., Boston.

Perrault SL, Saucier RT, Heinrich P, Lee D, Ryan J, Hahn III THG (2008) Phase I cultural resource survey of the Lake Killarney ecosystem restoration project, West Feliciana Paris, Louisiana: Final report prepared by Coastal Environments, Inc. for U.S. Army Corps of Engineers, New Orleans District.

Peterson RN (1991) Pawpaw (*Asimina*). *Acta Horticulturae* 290: 567 – 600.

Petit RJ, Brewer S, Bordacs S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PB, Jensen JS, Konig AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu JL, Kremera A (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 156: 49–74.

Petit RJ, Aguinagalde I, de Beaulieu J, Bittaku C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* 300: 1563 – 1565.

Phelps DS (1983) The archaeology of the North Carolina coast and coastal plain: problems and hypotheses. pp. 1-51. *In*: Mathis MA, Crow JJ (Eds) *The Prehistory of North Carolina*: *An Archaeological Symposium*. North Carolina Division of Archives and History, Raleigh.

Pickering C (1879) *Chronological History of Plants*. Vol 1. Boston: Little, Brown and Company.

Pomper KW, Lowe JD, Lu L, Crabtree SB, Dutta S, Schneider K, Tidwell J (2010) Characterization and identification of pawpaw cultivars and advanced selections by simple sequence repeat markers. *Journal of the American Society for Horticulture Science* 135: 143 – 149.

Poor JM (1984) Plants that merit attention, vol. 1: Trees. Portland, OR: Timber Press.

Pritchard JK, Stephens JK, Donnelly M (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945 – 959.

Prufer OH (1964) New data on Hopewell hilltop enclosures in Southern Ohio. *In*: Dancey W, Pacheco PJ (Eds) *Ohio Hopewell Community Organization*, pp. 311 – 326. Kent State University Press, Kent, OH.

Pyne SJ (1983) Indian fires. *Natural History* 2: 6 – 11.

R Core Team (2013) R: A language and environment for statistical computing.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86:248-249.

Richland DM, Rundel PW, Jackson ST, Teskey RO, Aronson J, Bytnerowicz A, Wingfield MJ, Proches S (2007) Human impacts in pine forests: Past, present, and future. *Annual Review of Ecology, Evolution, and Systematics*. 38: 275 – 297.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 145: 1219 – 1228.

Sanger MC, Hill MA, Lattanzi GD, Padgett BD, Larsen CS, Culleton BJ, Kennett DJ, Dussubieux L, Napolitano MF, Lacombe S, Thomas DH (2018) Early metal use and crematory practices in the American Southeast. *Proceedings of the National Academy of Sciences* 115: 7672 – 7679.

Scarry MC, Scarry JF (2005) Native American 'garden agriculture' in southeastern North America. *World Archaeology* 37: 259 – 274.

Stewart OC (2002) The effects of burning of grasslands and forests by aborigines the world over. *In*: Lewis HT, Anderson MK (eds) *Forgotten fires: Native Americans and the transient wilderness*, pp. 67-338. University of Oklahoma Press, Norman, OK.

Vale TR (1998) The myth of the humanized landscape: An example from Yosemite National Park. *Natural Areas Journal* 18: 231-236.

Vale TR (2002) The pre-European landscape of the United States: Pristine or humanized? In: Vale TR (Ed) Fire, native peoples, and the natural landscape, pp. 1-40. Island Press, Washington, D.C.

Wallace PA (1952) Historic Indian Paths of Pennsylvania. *The Pennsylvania Magazine of History and Biography* 76: 411 – 439.

White CA (1906) The northern limit of the Papaw tree. *Science* 23: 749-751.

Williams TJ, Collins MB, Rodrigues K, Rink JW, Velchoff N, Keen-Zebert A, Gilmer A, Frederick CD, Ayala SJ, Prewitt ER (2018) Evidence of an early projectile point technology in North America at the Gault Site, Texas, USA. *Science Advances* 4: 1 – 7.

Willson MF, Schemske DW (1980) Pollinator limitation, fruit production, and floral display in Pawpaw (*Asimina triloba*). *Bulletin of the Torrey Botanical Club* 107: 401 – 408.

Wright S (1922) Coefficients of inbreeding and relationship. *The American Naturalist* 56: 330 – 338.

Wright S (1943) Isolation by distance. *Genetics* 28: 114 – 138.

Wright AP (2017) Local and "global" perspectives on the Middle Woodland Southeast. *Journal of Archaeological Research* 25: 37 – 83.

Wykoff MW (1991) Black walnut on Iroquoian landscapes. *Northeast Indian Quarterly* 8: 4 – 17.

Yarnell RA (1965) Implications of distinctive flora on Pueblo ruins. *American Anthropologist* 67: 662 – 674.

Yarnell RA, Watson PJ (1966) Archaeological and paleoethnobotanical investigations in Salts Cave, Mammoth Cave National Park, Kentucky. *American Antiquity* 31: 842 – 849.

Yerka SJ (2010) Geophysical study at Old Stone Fort State Archaeological Park, Manchester, Tennessee. Master's Thesis, Department of Anthropology, University of Tennessee, Knoxville, TN.

Table 3.1. Sampled anthropogenic populations and association with pre-Columbian settlements.

Population	Samples	Location	Latitude	Longitude	Culture or Period	Reference
FCP	35	Florida Caverns State Park, FL	30.8101	-85.227	Fort Walton	Gardner 1966
PET	24	Pettigrew State Park, NC	35.7930	-76.4126	Late Woodland	Phelps 1983
TUN	25	Tunica Hills WMA, LA	30.9303	-91.5136	Tunica	Perrault 2008
FHM	50	Fort Hill Memorial, OH	39.1230	-83.4006	Hopewell	Prufer 1964
LM1	17	Little Miami River, OH	39.4103	-84.099	Fort Ancient	Connolly 2004
LM2	24	Little Miami River, OH	39.4102	-84.1014	Fort Ancient	Connolly 2004
LM3	14	Little Miami River, OH	39.4088	-84.1001	Fort Ancient	Connolly 2004
MC1	16	Mammoth Cave, KY	37.204	-86.1394	Woodland	Yarnell & Watson 1966
MC2	18	Mammoth Cave, KY	37.1779	-86.1083	Woodland	Yarnell & Watson 1966
MC3	12	Mammoth Cave, KY	37.1483	-86.0929	Woodland	Yarnell & Watson 1966
OSF	36	Old Stone Fort State Park, TN	35.4852	-86.1034	Middle Woodland	Yerka 2010
SEL	5	Sellars Farm Archeological Site, TN	36.1694	-86.2363	Mississippian	Butler 1981
CFP	50	Cunningham Falls State Park, MD	39.6325	-77.4556	Late Woodland	Bedell et al. 2011
POC	50	Pocahontas State Park, VA	37.3812	-77.5834	Powhatan	Higgins et al. 1995; Kandle et al. 1996
SCI	25	Scioto Trail State Park, OH	39.2160	-82.9736	Middle Woodland	Case & Carr 2008
SU1	32	Susquehanna State Park, MD	39.6104	-76.1467	Late Woodland	MacNamara 1982
CON	30	Conneaut, OH	41.9334	-80.6107	Woodland	Wallace 1952; Brose 1976
LCN	48	Love Creek Nature Center, MI	41.9512	-86.3062	Late Woodland	Hinsdale 1944; Lovis 1974
CHA	40	Chautauqua Creek, NY	42.3388	-79.6040	Late Woodland	Wallace 1952; Bennett 1986
WEN	10	Wendt County Beach Park, NY	42.6792	-79.0516	Late Woodland	Bennett 1986

Table 3.2. Summary statistics of *A. triloba* nuclear genetic diversity in 20 putative anthropogenic populations. N = sample size, MLG = number of multilocus genotypes, P (%) = percent of polymorphic loci, A_T = total number of alleles, A_E = effective number of alleles, P_A = number of private alleles, A_{RARE} = number of rare alleles (i.e., alleles in ≤ 10% of populations), H_O = observed heterozygosity, H_E = expected heterozygosity, and F_{IS} = inbreeding coefficient.

Population	*N*	MLG	*P* (%)	A_T	A_E	P_A	A_{RARE}	H_O	H_E	F_{IS}
Anthropogenic										
CFP	50	1	56	13	1.4	-	1	0.556	0.278	-1.000
CHA	40	1	33	11	1.2	-	-	0.333	0.167	-1.000
CON	30	2	89	24	2.4	-	1	0.778	0.528	-0.467
FCP	35	7	56	26	2.5	1	6	0.381	0.396	0.088
FHM	50	6	100	29	2.4	3	3	0.491	0.545	0.058
LCN	48	4	89	18	1.8	-	3	0.778	0.448	-0.690
LM1	17	14	100	43	3.3	-	1	0.583	0.632	0.125
LM2	24	1	67	15	1.7	-	-	0.667	0.333	-1.000
LM3	14	3	78	20	2.1	-	-	0.593	0.444	-0.383
MC1	16	7	100	33	2.7	-	3	0.501	0.532	0.098
MC2	18	10	100	36	2.5	-	-	0.658	0.541	-0.194
MC3	12	1	44	13	1.4	-	1	0.444	0.222	-1.000
OSF	36	3	100	30	2.9	2	4	0.630	0.537	-0.179
PET	24	22	89	33	2.3	-	2	0.479	0.495	0.051
POC	50	6	67	21	2.0	-	-	0.537	0.403	-0.400
SCI	25	1	67	15	1.7	1	1	0.667	0.333	-1.000
SEL	5	2	78	22	2.3	-	2	0.722	0.472	-0.562
SU1	32	11	89	39	3.1	-	3	0.606	0.559	-0.111

TUN	25	7	100	45	3.9	1	8	0.651	0.668	0.044
WEN	10	1	67	15	1.7	-	-	0.667	0.333	-1.000
Mean	**28.1**	**5.5**	**78**	**25**	**2.3**	**0.40**	**2.0**	**0.586**	**0.443**	**-0.426**
Pooled	**561**	**105**	**100**	**95**	**5.7**	**8**	**-**	**0.59**	**0.706**	**0.17**
Wild Mean	**29.6**	**9.6**	**84.7**	**32.5**	**2.4**	**0.44**	**3.3**	**0.530**	**0.496**	**-0.153**
Wild Pooled	**1837**	**590**	**100**	**153**	**6.5**	**27**	**-**	**0.553**	**0.765**	**0.273**

Figure 3.1. Map of 82 *Asimina triloba* wild and putative anthropogenic populations. Red dots represent populations of putative anthropogenic origin while blue dots represent wild populations.

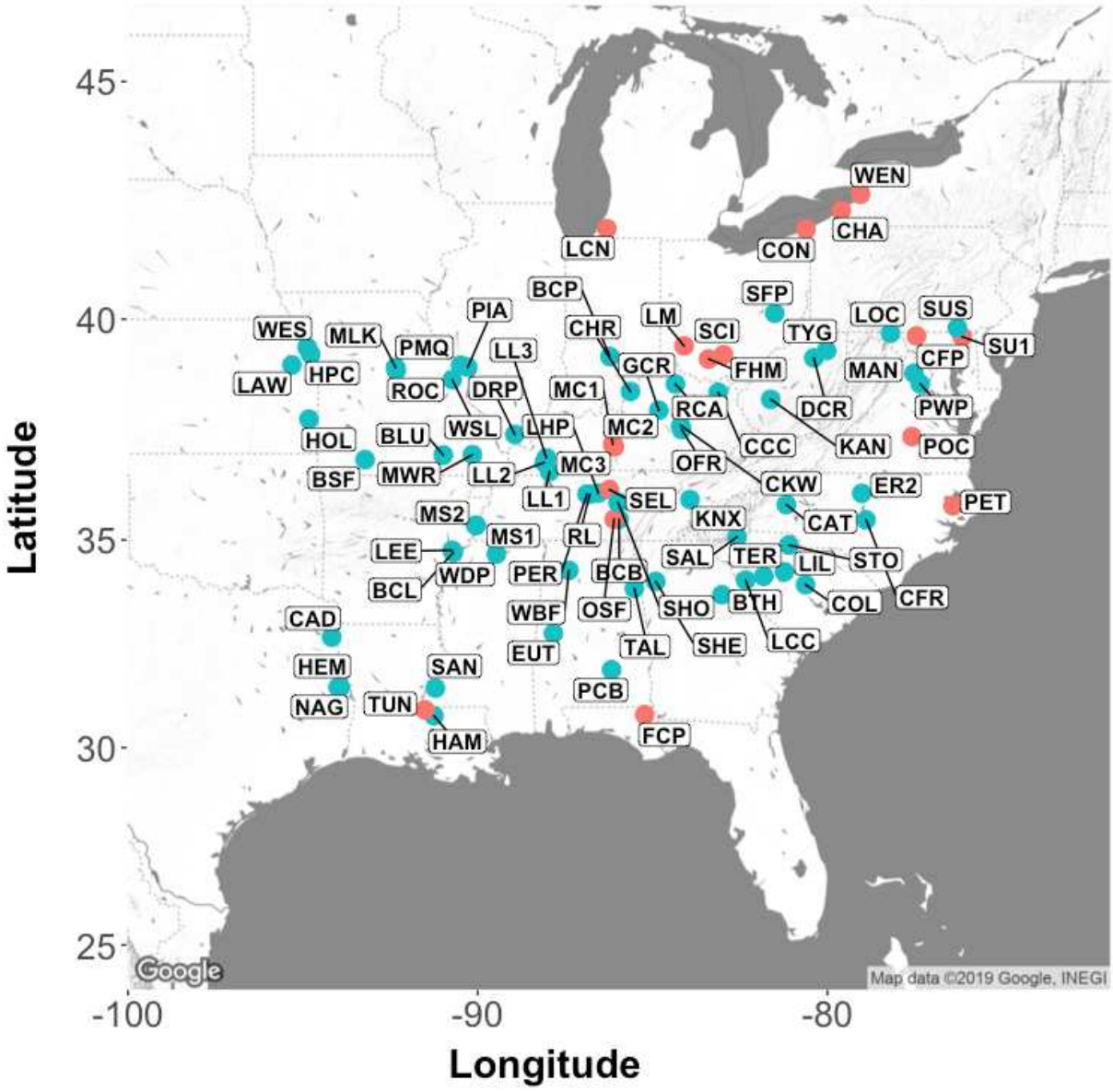


Figure 3.2. Comparison of total number of alleles for anthropogenic and wild populations (p = 0.01)

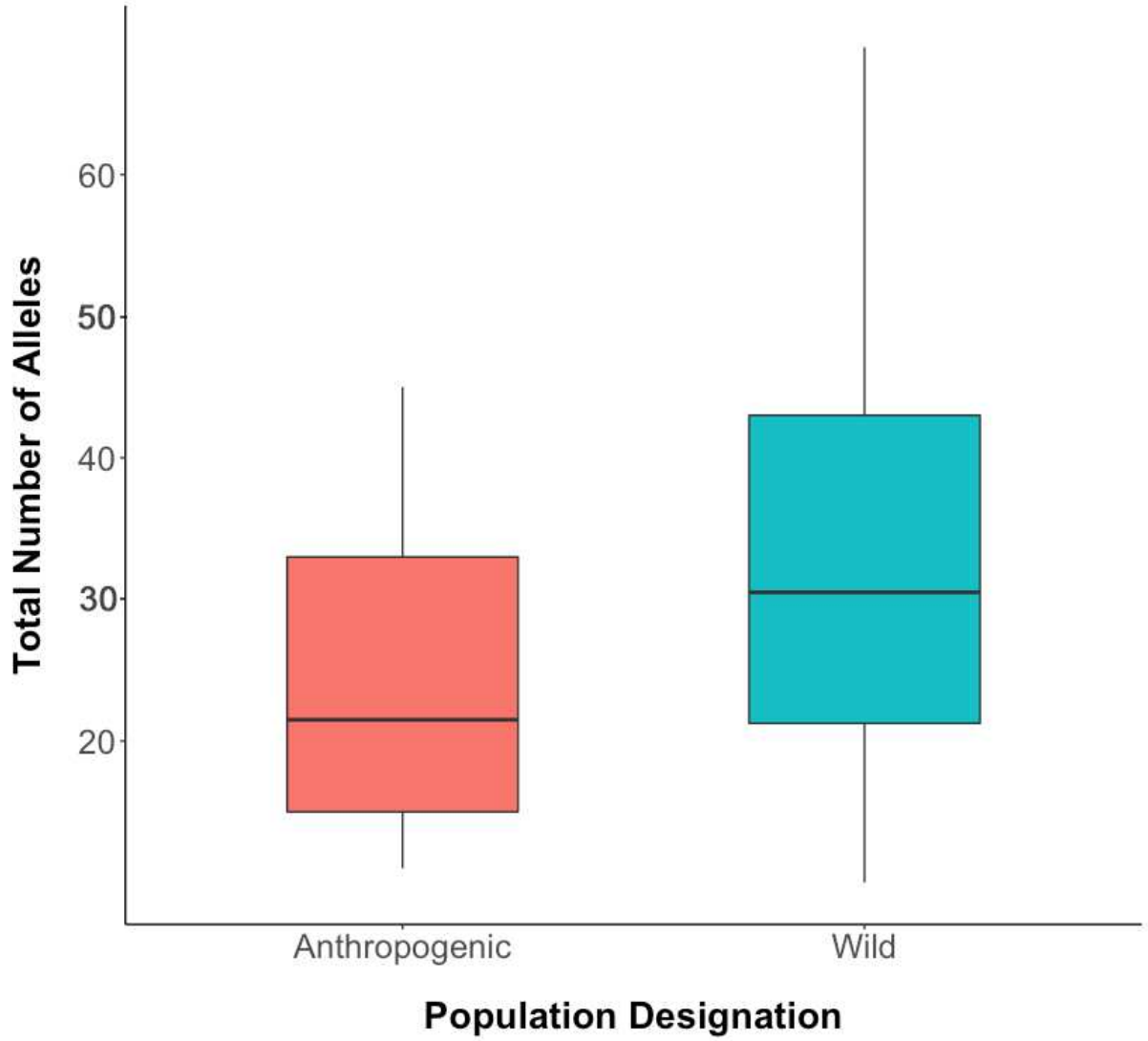


Figure 3.3. Histogram showing the distribution of the mean total number of alleles present through rarefaction (999 random sub-sampling of the data) for putative anthropogenic and wild populations of *A. triloba* ($p < 0.01$).

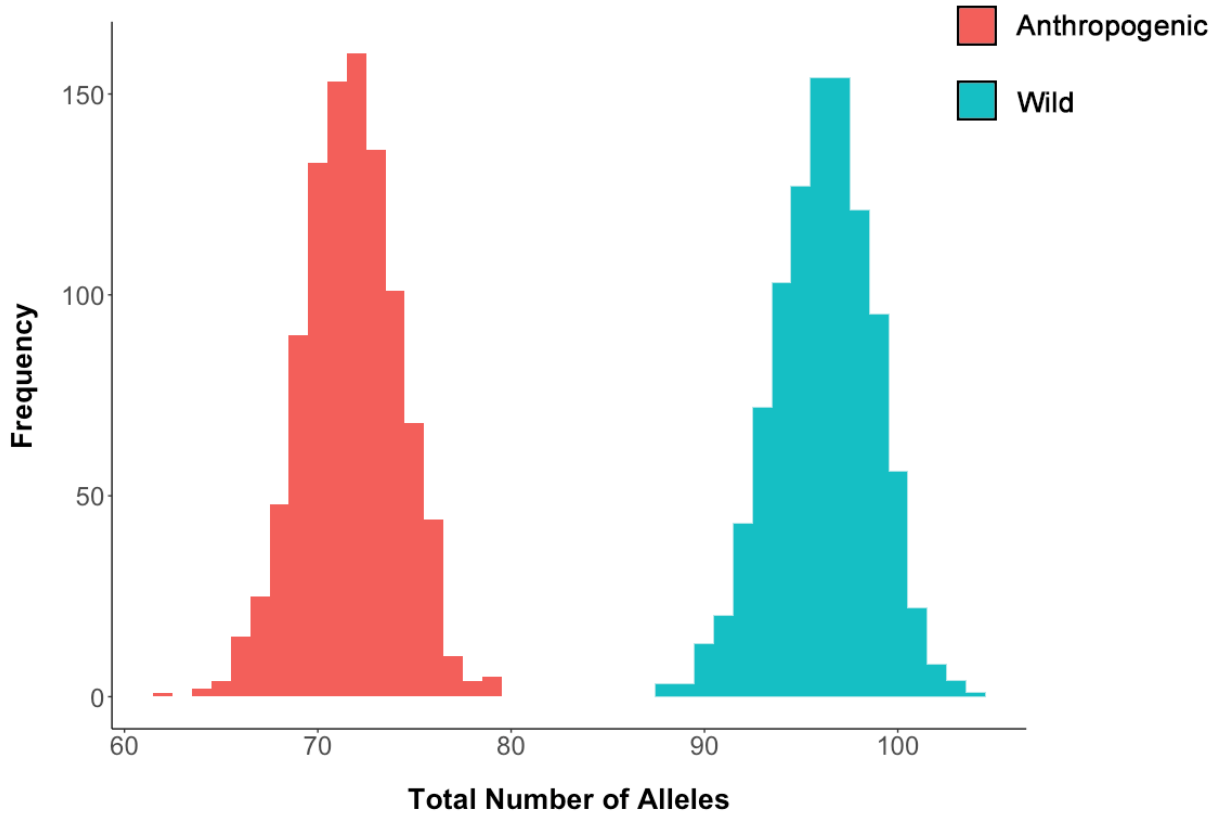


Figure 3.4. Histogram showing the mean effective number of alleles per locus after rarefaction (999 sub-sampling of the data) in both putative anthropogenic and wild populations ($p < 0.01$).

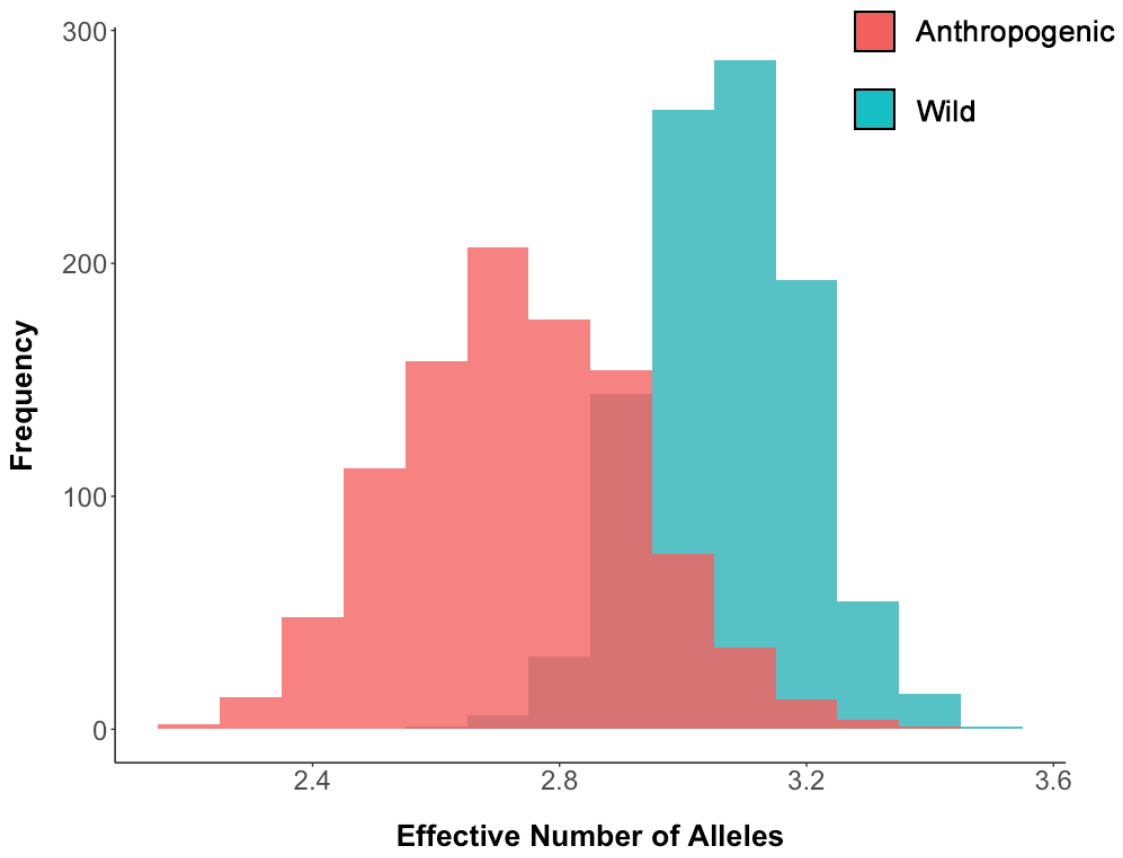


Figure 3.5. Number of rare alleles (present in less than 10% of populations) and latitude ($p < 0.05$).

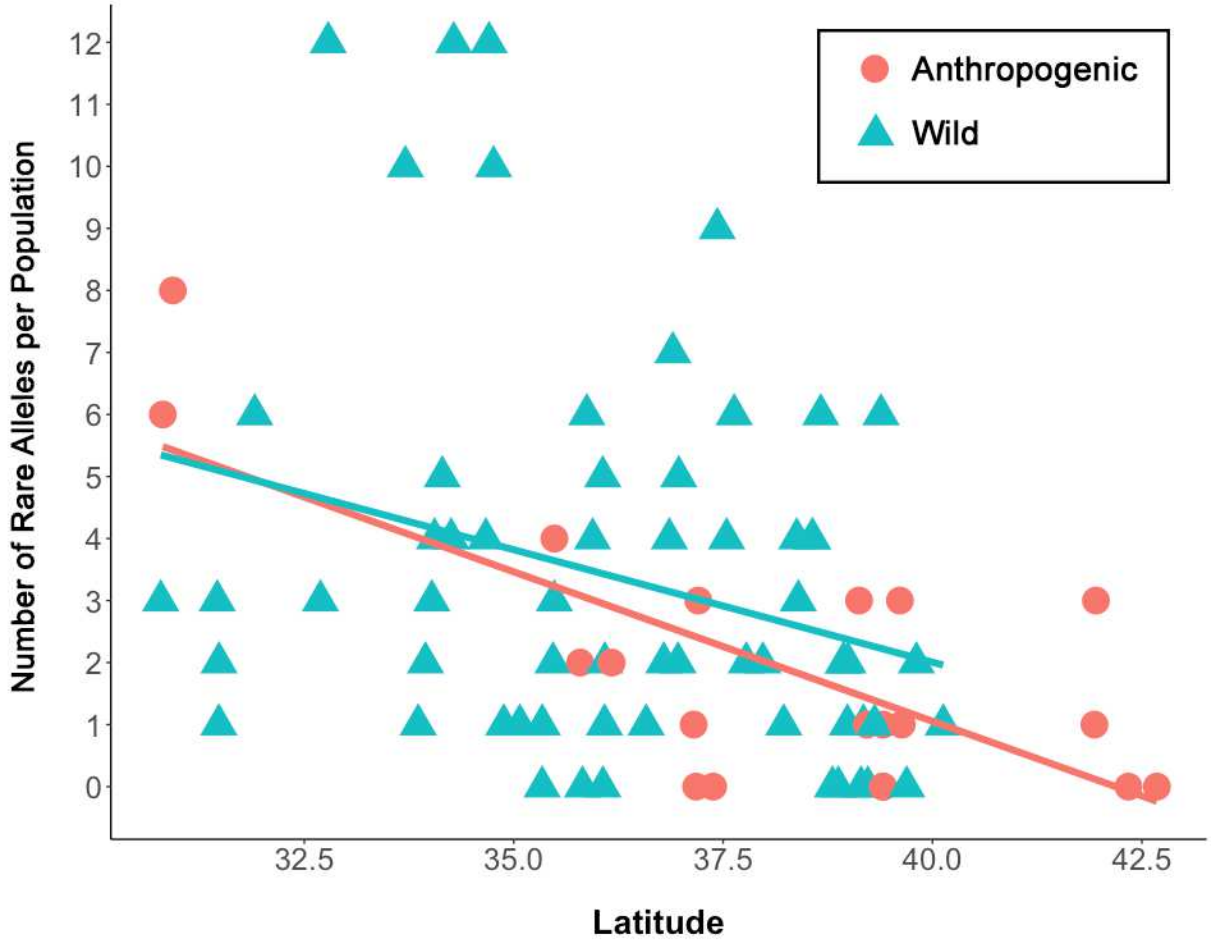


Figure 3.6. Distribution of pairwise geographic distances for various networks of rare alleles

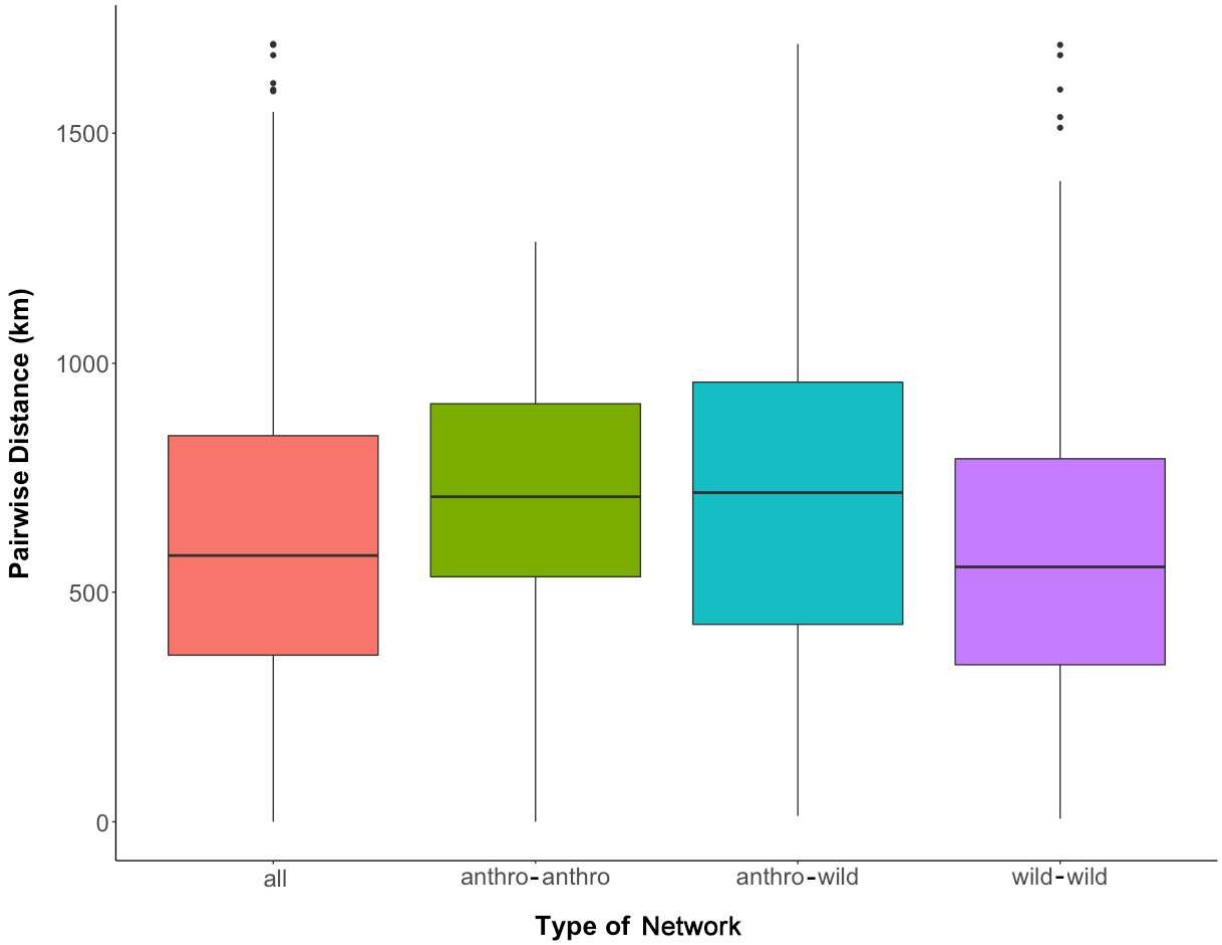


Figure 3.7. Map showing pairwise G_{ST} values among the four northernmost *A. triloba* populations, all of which are of putative anthropogenic origin (LCN, CON, CHA, and WEN), as well as between these four populations and the populations with which they share the lowest G_{ST} values. Red diamonds designate putative anthropogenic populations and turquoise designate wild populations. Black lines connect the anthropogenic populations while blue lines connect these populations with the populations that are most genetically similar. Straight-line distances (km) between populations are displayed in brackets under the G_{ST} values.

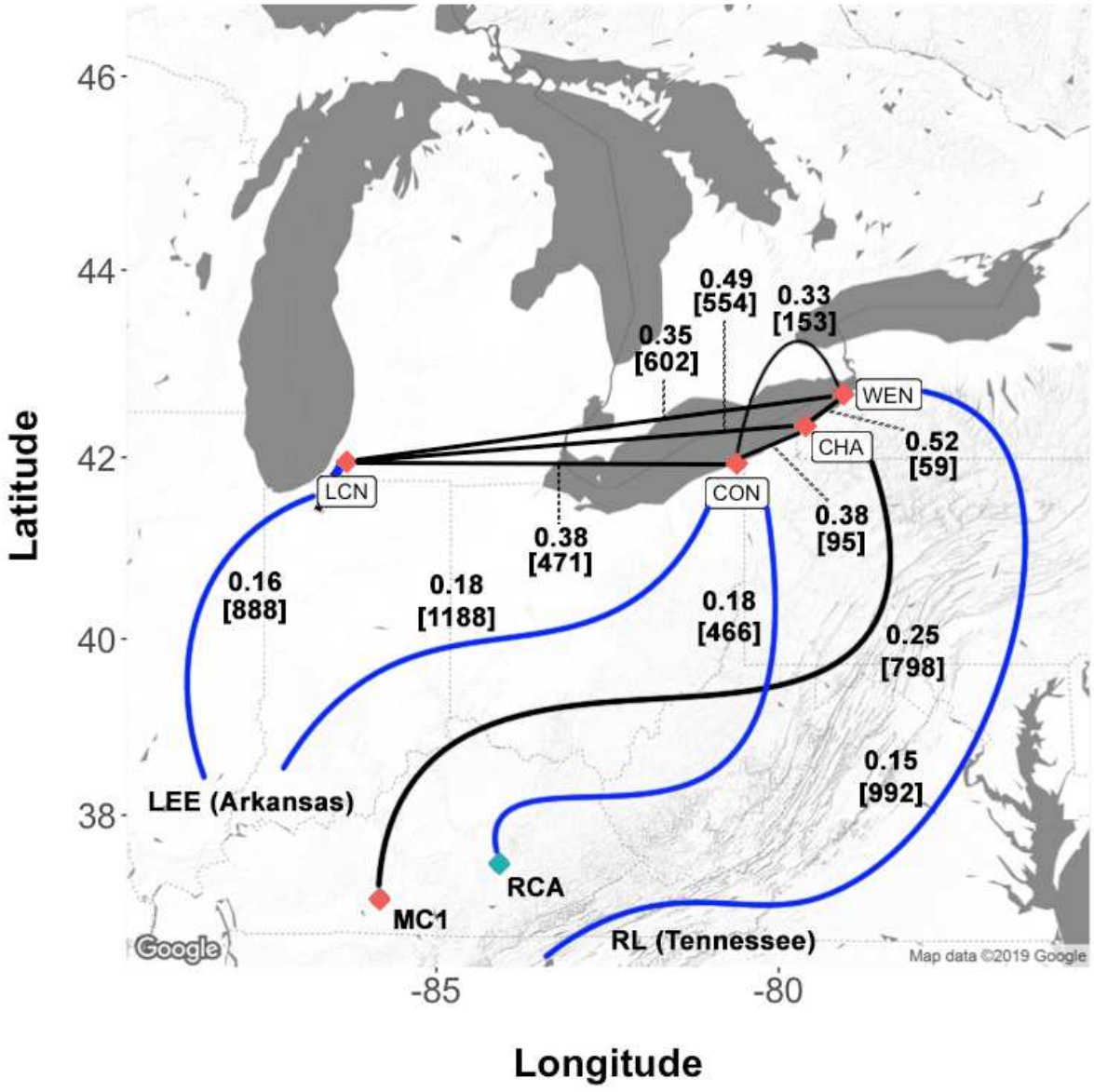


Figure 3.8. Pie frequency charts for all sampled populations showing percentage of populations assigned to one of the two *k*-clusters from STRUCTURE v 2.3.4. The Tombigbee River in Alabama is indicated in navy blue. Anthropogenic populations are designated with an "A." The reported range of *A. triloba* from Little (1977) is outlined in blue.

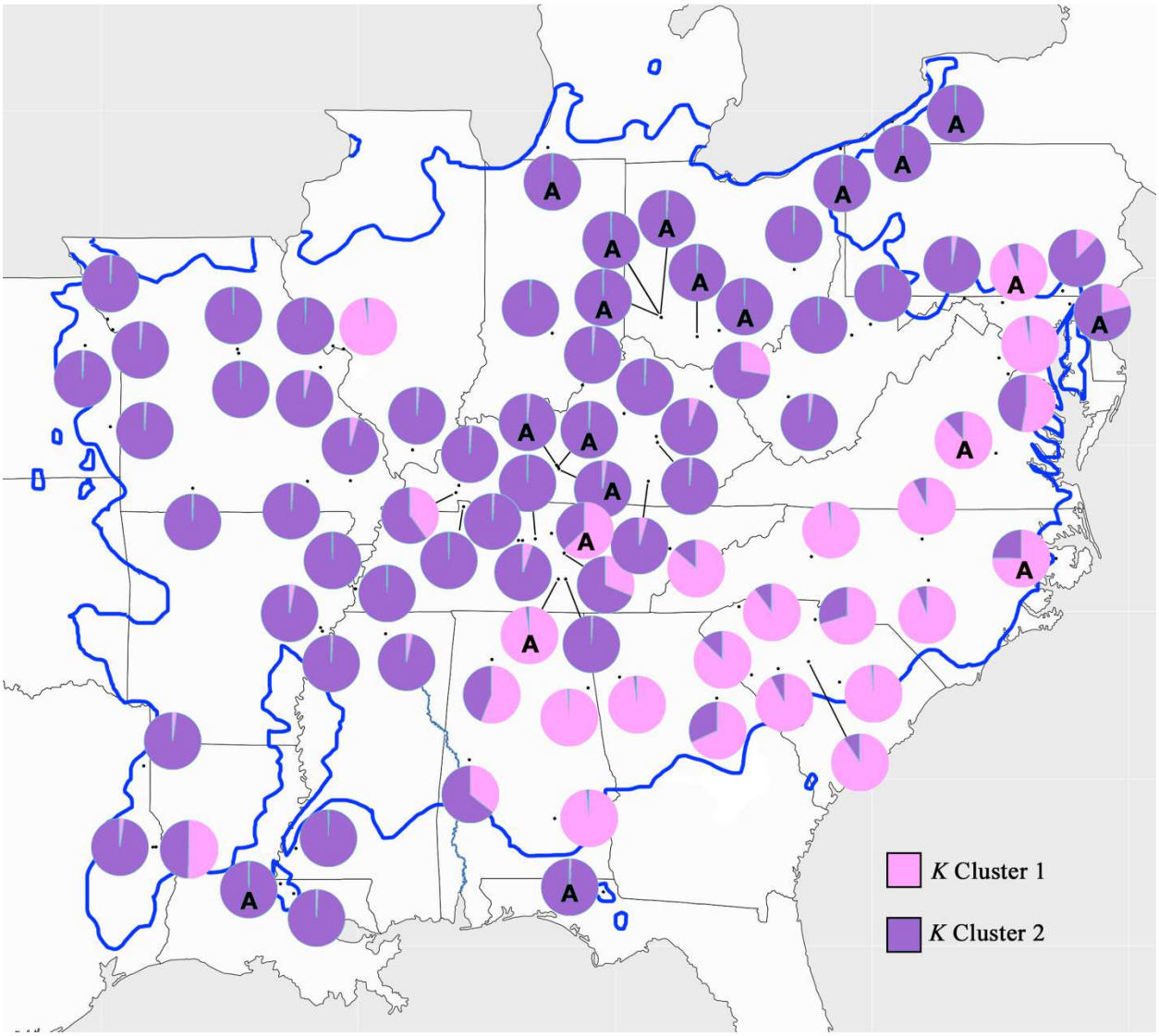


Figure 3.9. Map of *A. triloba* populations where allele 157 at locus Pp-B128 occurred. Colors indicate population designation.

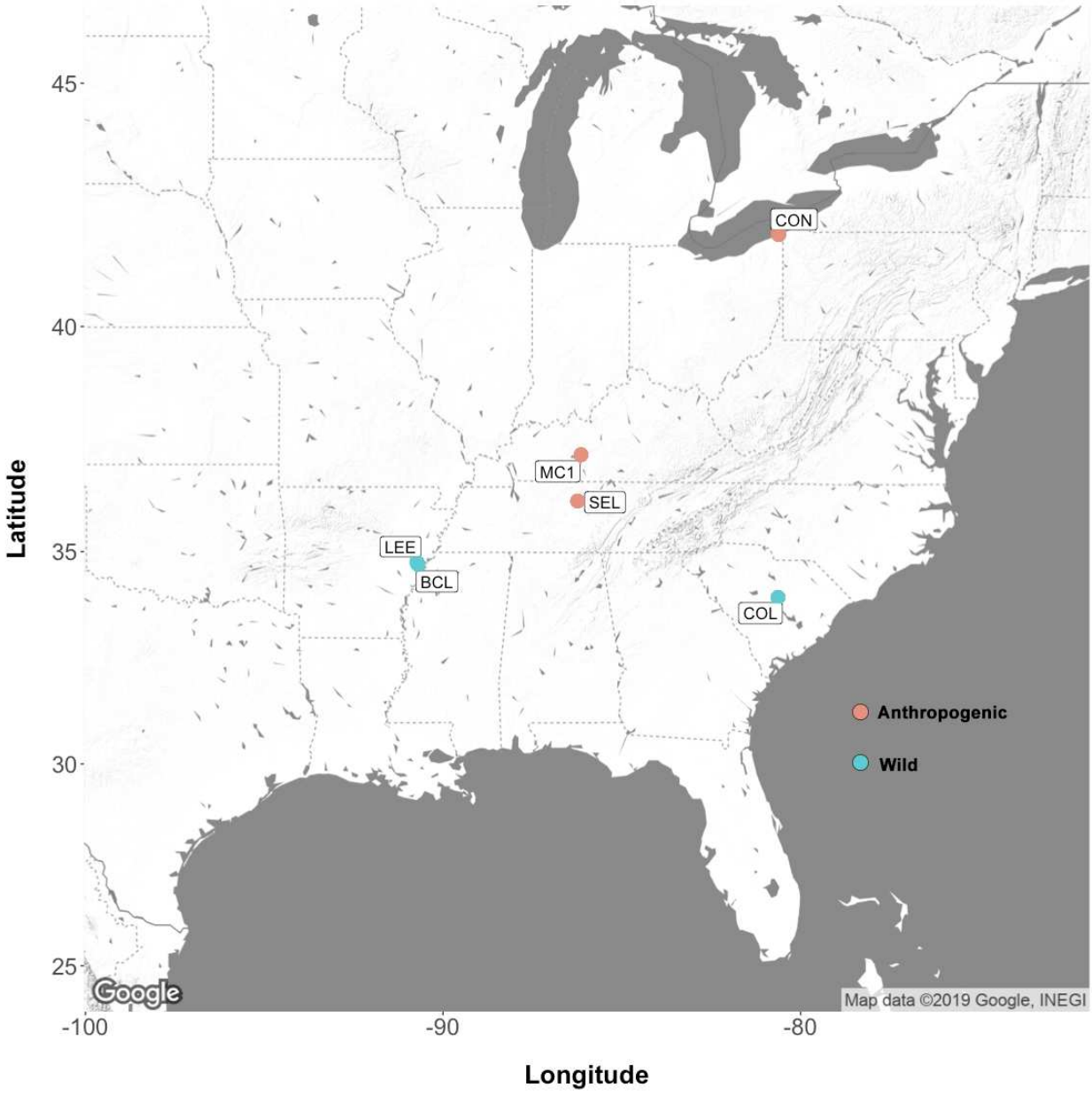


Figure 3.10. Map of *A. triloba* populations where allele 177 at locus Pp-B128 occurred. Colors indicate population designation.

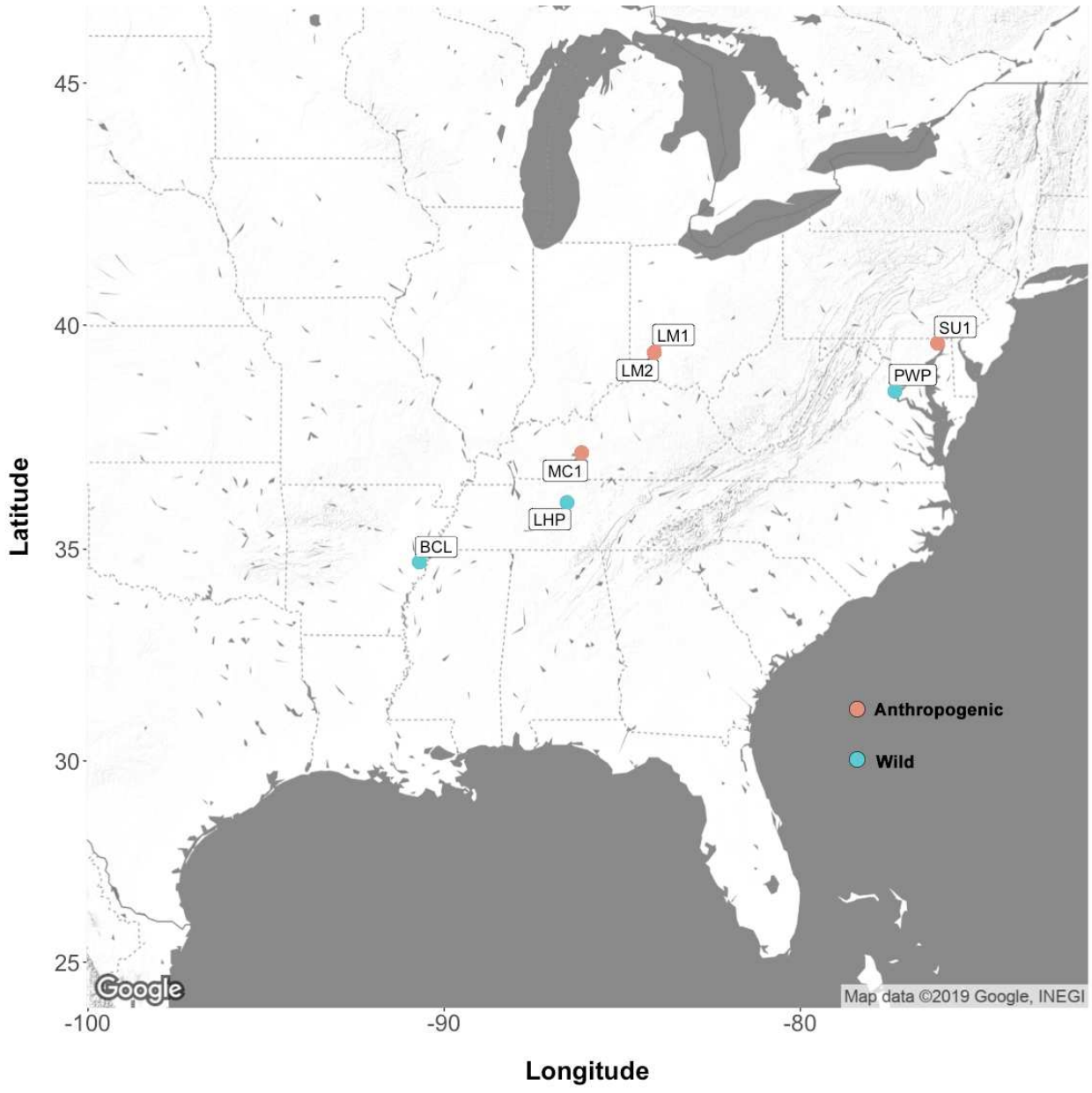


Figure 3.11. Map of *A. triloba* populations where allele 207 at locus Pp-124 occurred. Colors indicate population designation.

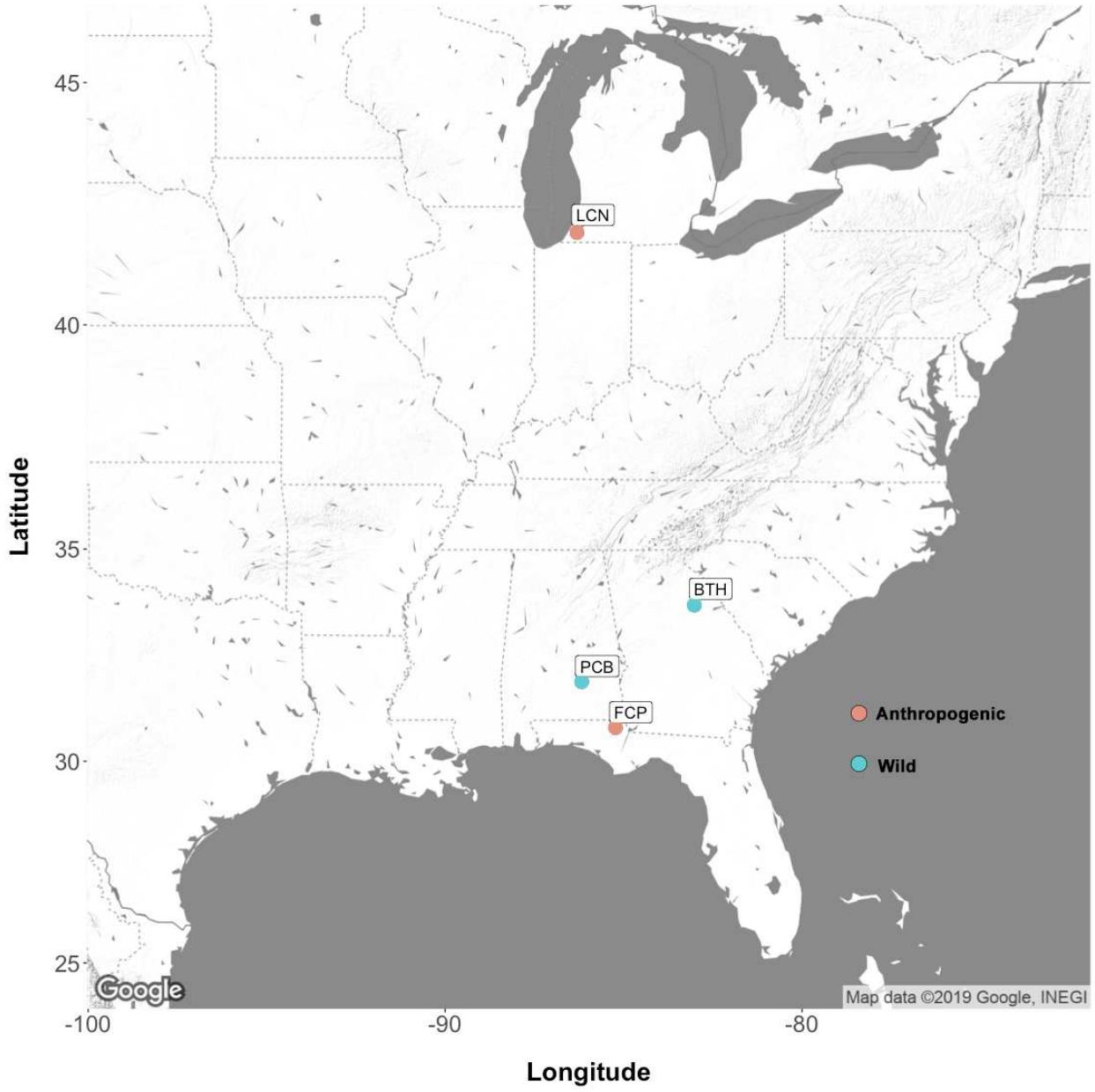


Figure 3.12. Map of *A. triloba* populations where allele 240 at locus Pp-121 occurred. Colors indicate population designation.

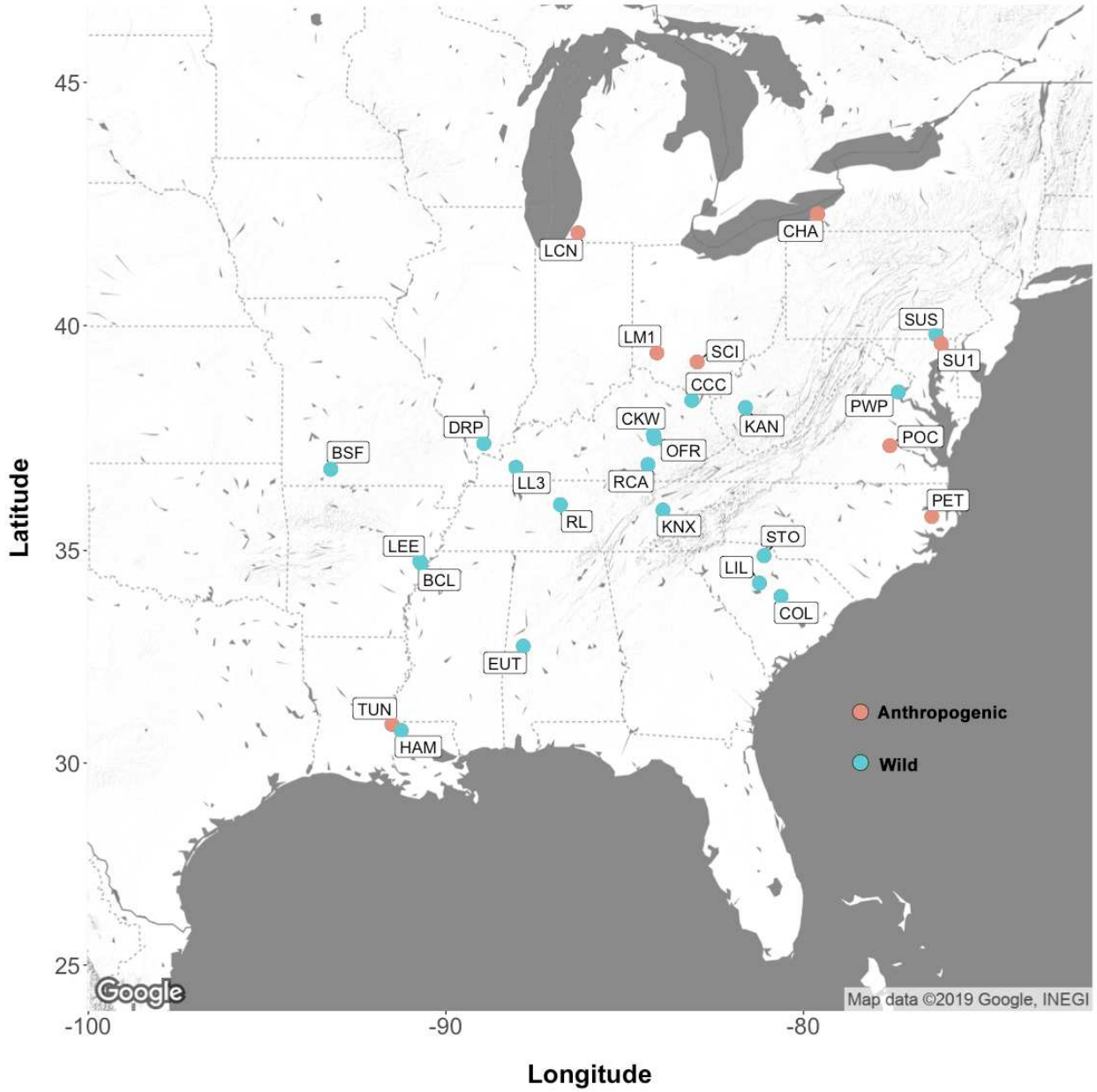
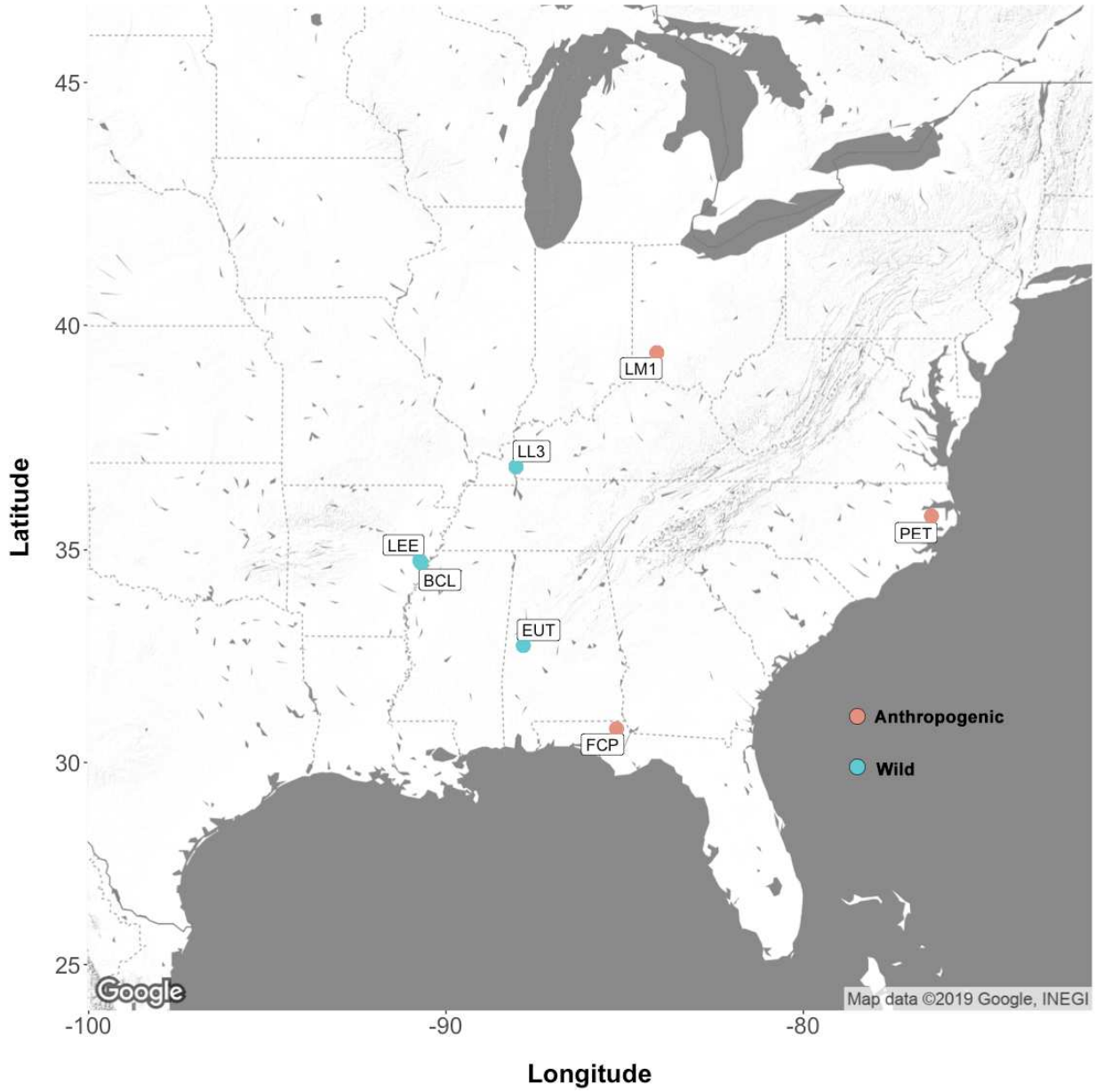


Figure 3.13. Map of *A. triloba* populations where allele 354 at locus Pp-B103 occurred. Colors indicate population designation.



CHAPTER 4

CLONALITY AND SPATIAL GENETIC STRUCTURE OF *ASIMINA TRILOBA* (ANNONACEAE) POPULATIONS ACROSS A BROAD GEOGRAPHIC RANGE[1]

[1] Graham E. Wyatt. To be submitted to American Journal of Botany

ABSTRACT

Clonality and the effects of clonal growth on spatial genetic structure were investigated in populations of *Asimina triloba*, a wide-ranging tree native to Eastern North America. Measures of clonality were observed across the native range of *A. triloba* for wild populations as well as those that may have been manipulated by pre-Columbian humans. Spatial genetic structure analyses were performed in 40 populations to understand how clonal growth effects the relatedness of individuals through space.

A broad trend in increased clonality at higher latitudes was observed, indicative of an increased dependence on clonal growth at the northern edge of the species' range. A similar, though less significant trend was observed with decreasing (westward) longitude. Several measures of clonality were significantly higher in putative anthropogenic populations than wild populations. Additionally, higher clonality was associated with strongly heterozygous individuals and may reflect adaptation to harsher climates found at the edges of the species' range. Given the size and extent of genets, especially at the edges of the species' range, it appears that clonal growth is an important means of growth and persistence of individuals. Spatial genetic structure of *A. triloba* is entirely linked to clonal growth, and is not significantly influenced by sexual reproduction and seed dispersal. Fruit-set was significantly greater in populations with a higher number of multilocus genotypes, as expected for a strongly self-incompatible species.

INTRODUCTION

Asexual propagation is common among angiosperms and serves to accomplish reproduction and assure persistence without incurring the potentially high costs associated with sexual reproduction, i.e., production of flowers and fruits. In populations that are locally adapted, this

form of reproduction also insures maintenance of co-adapted multilocus gene complexes. While ecological functions of clonal growth are reasonably well-understood for herbaceous species, this is less true for woody species (Hosaka et al. 2005). Woody species can undergo asexual propagation via collar sprouts, root sprouts, opportunistic sprouts from layered branches, or specialized underground stems (i.e., rhizomes). An additional advantage is that each ramet in a clone may obtain resources from the more established individual to which it is connected. A potential drawback of asexual propagation is that ramets may face direct competition for resources (e.g., light, water, soil nutrients) with the parent plant as well as other ramets of the parent plant. This may be a problem for sexually produced progeny as well, although less likely or less severe in tree species due to seed dispersal mechanisms. However, it is conceivable that ramets arising from root sprouts, or root suckering face less competition (Del Tredici 2001) as roots have the potential to tap resources beyond the canopy of the parental plant. Root connections between ramets may remain functional for decades or even hundreds of years (e.g., *Populus tremuloides*: Cook 1983).

The North American understory tree *Asimina triloba* (Annonaceae) is by far the most geographically wide-spread member of this temperate genus (Weakley 2012), occurring in 26 US states and extending into southern Canada. The fruits are edible and the largest of any native North American fruit tree (Pomper et al. 2010). *Asimina triloba* typically occurs in mesic hardwood and alluvial forests along rivers and creeks, at less than 350 m elevation (Pomper et al. 2010; Freeman & Hulbert 1985), resulting in a non-continuous distribution across landscapes. This species propagates via root suckering, a trait observed in approximately 25% of temperate angiosperm tree species (Pomper et al. 2009; Burns & Honkala 1990; Del Tredici 2001).

Consequently, stands of *A. triloba* are typically dense and patchy. Hosaka et al. (2005) found high survivorship of newly formed ramets, which likely contributes to population persistence. In three Maryland populations, ramet recruitment rates far exceed mortality rates (Hosaka et al. 2008).

Asimina triloba has been observed to form entirely clonal patches consisting of only one multilocus genotype (Rogstad et al. 1991; Pomper et al. 2009) however the degree of clonality of *A. triloba* populations across its broad geographic range has not been investigated before now. The extent of clonal versus sexual reproduction, and the size of clones have direct effects on the spatial genetic structure (SGS) of populations. In addition to clonality, the magnitude and extent of SGS can also be influenced by patterns of pollen movement and especially seed dispersal.

Our objective is to assess the extent of clonal reproduction in 82 populations and SGS in 40 populations of *A. triloba* from across its broad geographic range. The specific goals are to 1) determine the size and extent of genets within populations, and how patterns of clonality vary across the geographic range of the species 2) determine if putative anthropogenic populations (i.e., human manipulated) have the same patterns of clonality (number of unique multilocus genotypes, size of genets, diversity, etc.) as that of wild populations, and 3) examine SGS sans clones within populations to gain insights into localized seed movement and recruitment.

MATERIAL AND METHODS

Field Collections

Where possible, one leaf sample was collected per stem with a diameter at breast height (DBH) ≥ 1 cm and that was separated by at least 10 meters from the next stem. Five to 60 samples (mean = 29.1) were collected from each of 82 populations in 23 states of the USA, across the native range of *A. triloba* (Table 4.1; Figure 4.1). Populations within 1 km of documented archaeological sites, that indicated population centers for pre-Columbian Americans, were deemed of putative anthropogenic origin and were designated as such (N = 20). Populations without any known proximate archeological evidence were designated as 'wild' (N = 62; Table 4.1). In 40 of the 82 populations (anthropogenic = 10, wild = 30), sampled trees were mapped using a Brunton compass and Euclidean distances between sampled stems were determined by laser (Bosch DLR130). Leaf samples were snap-frozen in liquid nitrogen for transport to the University of Georgia and stored at -80° C until analysis. Additionally, the total number of fruits per sampled stem was recorded for all 82 populations. Fruit monitoring was conducted from June to August when fruits were developing.

Microsatellite Genotyping

Genomic DNA was extracted from approximately 0.05 g of frozen leaf tissue using a modified CTAB protocol (Doyle and Doyle, 1987). DNA quality and quantity were evaluated using an ND-1000 Nanodrop® spectrophotometer. Nine nuclear microsatellite loci (Pp-G4, Pp-G119, Pp-G124, Pp-C108, Pp-G103, Pp-G121, Pp-B103, Pp-B108, Pp-B128) were amplified with markers developed by Lu et al. (2011). PCR amplification followed a modified protocol from Lu et al. (2011) and was carried out in 12.5 µL reaction volumes containing: 3.25 µL molecular grade

ddH2O, 2.5 µL 5X One Taq® standard reaction buffer (New England Biolabs, Ipswich, MA), 0.35 µL 25 mM MgCl solution (Sigma-Aldrich), 1.25 µL 10X (1 mg/mL) bovine serum albumin (Thermo-Fisher), 1 µL 2.5 mM dNTP mix (New England Biolabs, Ipswich, MA), 1 µL primer mix (0.5 µM CAG-labeled primer and 5 µM unlabeled primer), 0.45 10 µM universal dye-labeled primer, 0.2 µL One Taq® Hot Start DNA polymerase (New England Biolabs, Ipswich, MA), and 2.5 µL of diluted template DNA (20 ng/µL). A 3-primer PCR protocol was used whereby a CAG-tag sequence (Hauswaldt & Glenn 2003) was added to the 5' end of one primer and a third fluorescently labeled (FAM, HEX, or NED) primer identical to the CAG-tag was included.

Clonality

Once all populations had been genotyped, the number of unique multilocus genotypes (MLG) in each population was determined. The probability of identity (P_I) was calculated for each population and corrected based on sample size. Clonal richness was calculated by $R = (G - 1)/(N - 1)$ (Dorken & Eckert, 2001; Kartzinel et al. 20015). Clonal diversity (D) was calculated using Simpson's diversity index (Simpson 1949), $D = 1 - \Sigma [N_j (N_j - 1) / N_r (N_r - 1)]$, where N_j is the number of samples of the jth MLG and N_r is the total number of samples from a population. In populations with a single clone, both R and D are zero, while in populations composed entirely of unique MLG both R and D will equal 1. Fager's evenness (E) was calculated as $E = (D - D_{min}) / (D_{max} - D_{min})$, where $D_{min} = (G - 1) (2 N_r - G) / N_r (N_r - 1)$, and $D_{max} = N_r (G - 1) / G (N_r - 1)$ (Fager 1972; Kartzinel et al. 2015). Spatial data for each mapped population was used to determine Euclidean distance between stems and area (convex hull, m2) of genets using the "sp" package in R version 3.5.1 (Pebesma & Bivand 2005; R Core Team 2013). This allowed

estimates of the area occupied by each MLG and the extent to which different genets overlap with one another. The maximum observed area of the convex hull occupied by stems belonging to the same MLG ($Area_{max}$) was measured in each population with a MLG consisting of three or more samples. The mean area of genets within populations was also calculated. The maximum observed distance in meters between stems belonging to the same MLG ($Distance_{max}$) was recorded for all populations with an MLG consisting of two or more stems. These estimates are likely an underestimate of the actual spatial extent of genets, because not all stems were sampled. Inbreeding coefficient (F_{IS}) was calculated in GenAlEx v 6.51b2 for all 82 populations (Peakall & Smouse 2006; Peakall & Smouse 2012). A Chi-square test was performed to determine the effect of the number of MLGs within a given population on fruit production.

Measures of clonality (MLG, R, D, E, $Area_{max}$, and $Distance_{max}$) were regressed against both latitude and longitude and assessed for statistical significance. Additionally, these measures were compared between putative anthropogenic and wild populations to determine if statistically significant differences exist among the two groups. The effect of sampling leaves from stems every 10 meters resulted in the collection of higher numbers of unique MLG individuals than would occur had sampling been done at smaller distance intervals, while still providing adequate spatial resolution to estimate the approximate extent of clonal growth for many individuals.

Spatial Genetic Structure

Spatial autocorrelation analyses were performed in GenAlEx v 6.51b2 in order to calculate the autocorrelation coefficient (r), an estimate of the genetic similarity of pairs of individuals relative to the distance separating them (Smouse & Peakall 1999; Peakall & Smouse 2006; Peakall &

Smouse 2012). The distance class at which the observed *r* value is no longer significant ($p < 0.05$) approximates the extent of detectable SGS (Peakall et al. 2003). Spatial autocorrelation analyses were performed on 36 of the 40 mapped populations using pairwise genetic and spatial distances among individuals. Spatial autocorrelation could not be performed on four of the mapped populations that consisted entirely of one MLG. Each of the smaller distance classes contained a minimum of 36 pairwise comparisons as recommended by Vekemans & Hardy (2004). For the 19 populations with more than 10 MLGs, all but one ramet per MLG were randomly removed from the analysis to remove the influence of clonality on the SGS. By performing the analyses with and without clones included, we could assess the mean relatedness and structure of individuals at specific distance classes and identify the role of clonality in determining genetic relatedness within populations. Significance of spatial autocorrelation values was tested with a two-tailed 95% confidence interval around the null hypothesis of no SGS (i.e., $r = 0$) by performing 999 random permutations of genotypes among geographic locations.

RESULTS

Clonality

Because of the high level of genetic variation yielded by the nine microsatellite loci (Wyatt and Trapnell, in prep) genetically unique individuals could be easily identified (Table 4.1). The mean within population P_I is 1.2×10^{-3} (range = 6.0×10^{-9} to 3.5×10^{-2}) (Table 4.1). Forty-three samples (1.7 % of all samples) had missing data from 6 or more loci consequently these samples were excluded from the analyses. Samples missing data for five or fewer loci were assigned to existing genets or designated as unique MLGs based on their alleles at the loci for which there were data.

For all 82 populations, the number of MLGs per population ranged from 1 to 30 (mean = 8.5; Table 4.1). The number of MLGS per population differed significantly between putative anthropogenic and wild populations (p = 0.007). Wild populations (N = 62) had 1 to 30 MLGs (mean = 9.5) while putative anthropogenic populations (N = 20) had 1 to 22 MLGs (mean = 5.3; Figure 4.9). There was no significant difference in mean sample size between the wild and anthropogenic populations (p = 0.71). Three anthropogenic populations (LM2, SCI, and MC3) and one wild population (ROC) each contained only one MLG (Table 4.2). Only in LL3 (a wild population) were all individuals characterized by a unique MLG (N = 24). For all 82 populations, the number of MLGs per population decreases with latitude (p = 0.06) (Figures 4.1, 4.3). Interestingly, in the four populations located above 40° N, all of which are designated as anthropogenic, the mean number of MLGs was only two.

Mean clonal richness (R) within populations was 0.37 and ranged from 0.00 (i.e., populations with only one MLG) to 1.00 (i.e., populations where every sample had a unique MLG). Anthropogenic populations had a lower mean R (0.23) than wild populations (0.42), a difference that was not significance of 0.1 (p = 0.1; Table 4.2). Clonal richness within all populations does decreased significant (p = 0.02) (Table 4.2; Figure 4.6). Mean clonal diversity, as measured by Simpson's diversity index (D), was 0.70 and ranged from 0.00 to 1.00. Mean D was not significantly lower for anthropogenic populations (0.53) than for wild populations (0.76; p = 0.14). A statistically significant trend of decreasing clonal diversity with latitude was observed across the range of sampled *A. triloba* populations (Figure 4.7; p = 0.003). Fager's Eveness (E) also varied from 0.00 to 1.00, with an overall mean of 0.59. Mean E for anthropogenic

populations was significantly lower (0.35) than that for wild populations (0.67; $p = 0.02$). As with other measures of clonality, *E* declined significantly with increasing latitude (Figure 4.8; $p = 0.003$).

As observed in Wyatt & Trapnell (in prep), the inbreeding coefficient (F_{IS}) was strongly correlated with the number of MLGs ($p = 2.1 \times 10^{-7}$). This indicates that populations comprised of a single MLG or with low numbers of MLGs (i.e., those dominated by a single or a few clones) tend to be highly heterozygous. Not surprisingly, Faeger's Evenness (*E*) was highly correlated with F_{IS} ($p = 6.3 \times 10^{-10}$; Figure 4.9). F_{IS} was strongly negatively correlated with latitude (Figure 4.4; $p < 0.001$).

Log of mean area of genets was increased with increasing latitude, but the relationship was non-significant ($p = 0.07$; Figure 4.10). The area of largest genet ($Area_{max}$, Table 4.2) within populations varied considerably from 13.9 m^2 (DRP) to 2,747.7 m^2 (MLK), with a mean area of 416.9 m^2. The mean $Area_{max}$ for anthropogenic populations (324.3 m^2) was less than that of wild populations (450.2 m^2), but the difference was not significant ($p = 0.46$). The maximum observed distance between ramets within populations ($Distance_{max}$) varied considerably from 4.8 m (EUT) to 163.4 m (ROC), with a mean of 34.4 m. $Area_{max}$, mean genet area per population, and $Distance_{max}$ both decreased with increasing latitude ($p < 0.1$). $Distance_{max}$ was the only measure of clonality to share a significant relationship with longitude: $Distance_{max}$ significantly decreased with increasing (eastward) longitude (Figure 4.5, $p = 0.052$).

Fruit-set varied widely among populations and was typically low. In 50 of the 82 populations (61%), no developing fruits were observed, while ≤ 10 developing fruits were observed in 79 % of populations. The mean number of developing fruits across all populations was 6.7. Population LIL had the highest fruit-set observed (149 developing fruits). The three populations with the highest fruit-set (LCC, TER, and LIL) were all located in South Carolina. The number of developing fruits within a population was strongly positively correlated with the number of MLGs within the population ($p = 0.0001$; Figure 4.11). A Chi-square test confirmed that the number of MLGs within a population had a highly significant effect on fruit-set ($\chi 2 = 20.89$, $p = 0.00003$) (Table 4.4).

Spatial Genetic Structure

The four populations possessing only one MLG were excluded from spatial autocorrelation analysis. Of the remaining 36 mapped populations, more than half (19) had ≥ 10 MLGs, which provided a sufficient number of pairwise comparisons at each distance interval. Spatial autocorrelation analyses were performed on the 36 populations with clones included and with clones removed. For the latter analyses, only one randomly selected individual per MLG was included for SGS analyses.

With the clones included, significant ($p < 0.05$) spatial autocorrelation (r) was observed at the smallest distance classes (2.5 - 45 m) for 28 of the 36 populations (80 %) (Table 4.2). Corrected mean relatedness (r) was 0.39 and ranged from -0.04 to 1.04. There was an absence of significant relatedness in seven populations.

In contrast, very few populations yielded significant r values when clones were removed, ranging from 0.07 to 0.18 (Table 4.3). Of the 19 populations that had enough MLGs to permit ≥ 36 pairwise comparisons per distance class, only three populations (16 %) had significant r values in the smallest distance class (10 – 37 m). Because of the high degree of clonality and the relatively small number of MLGs generally found, distance classes tended to be considerably larger than analyses where clones were included (i.e., mean minimum distance class = 30 m).

DISCUSSION

Clonality

The probability of identity (P_I) for all populations was small enough to allow for confident identification of unique MLGs and all ramets belonging to a specific MLG, even when sample size is taken into account. The decreasing numbers of MLGs within populations with increasing latitude (Figure 4.3; $p = 0.06$) would be consistent with northward range expansion from southern refugia after the last glacial maximum. This finding is also consistent with data presented by Wyatt & Trapnell (in prep), in which a similar pattern was observed in various genetic diversity parameters.

A highly significant difference (Figure 4.2; $p < 0.01$) exists in the number of MLGs within putative anthropogenic and wild populations. This may reflect historical human-mediated founder effects in anthropogenic *A. triloba* populations. Additionally, individuals in anthropogenic populations (mean $F_{IS} = -0.426$) tended to be significantly more heterozygous than individuals in wild populations (mean $F_{IS} = -0.152$; $p = 0.02$), even while anthropogenic

populations contain less genetic diversity overall as compared to wild populations (Wyatt & Trapnell, in prep).

Latitudinal trends in the measures of clonality (i.e., *R*, *D*, *E*, log of mean area of genets, and $Area_{max}$) observed in populations of *A. triloba* may reflect the greater importance of vegetative growth over sexual reproduction populations at higher latitudes. Populations at these higher latitudes are subject to longer, harsher winters and late frosts, which may negatively affect pollinators and consequently fruit-set (Wyatt pers. obs.). The inbreeding coefficient (F_{IS}) decreased significantly with increasing latitude ($p = 0.001$; Figure 4.4), which indicates that the individuals at the more northern parts of the range are more heterozygous and may be better suited to survive in harsher climates.

Hosaka et al. (2005) found that the number of newly produced ramets were higher in the interior of patches as compared to the periphery and surmised that the formation of ramets is not efficient for population expansion and habitat exploration. In this way *A. triloba* differs from most herbaceous and pioneer tree species. However, our results indicate that horizontal expansion through root suckering is the primary means by which a genet becomes established over a large spatial area. In population MLK, one genet covered an area of at least 2,747.7 m^2. Similarly, in population ROC, one genet was found that extended a linear distance of 163.4 m^2. There was no fruit-set was not observed in these populations, which reinforces the idea that clonal reproduction is of particular importance for these populations nearing the northwest extreme of the species' range. Given enough time and a periodic disturbance regime, it is possible that a population of *A.*

triloba consisting of several patches over a large area may in fact be one single genet. This could explain how populations consisting of a single MLG observed in this study arose.

The spatial extent of the largest genets observed ($Area_{max}$) appears to increase with latitude, though this trend was not significant at ($p = 0.085$). Increased genet size was also correlated with longitude: the maximum distance between stems belonging to the same MLG ($Distance_{max}$) significantly increased with westward longitude (Figure 4.5; $p = 0.052$). In other words, genets in the western part of the species' range tend to be larger. Again, this may be related to an increased reliance on clonal growth rather than sexual reproduction for populations near the edges of the range where populations are constrained by low temperatures and/or annual precipitation. With observed heterozygosity being significantly higher than expected throughout the species range (Wyatt & Trapnell in prep), it could be that highly adapted heterozygous genotypes become established more often in populations at the extreme edges of the range of *A. triloba* and persist principally through clonal growth, rather than through sexual reproduction and seed dispersal. Excess of heterozygosity has been shown to be common in long-lived woody species as well as clonal species (Birky 1996; Balloux 2003; Meloni et al. 2013). For such species, it is reasonable to expect that heterozygosity may be further increased in populations facing high abiotic stress, such that only the most vigorous individuals survive to maturity.

Clonal growth in *A. triloba*, a self-incompatible species, may serve to benefit colonization of new habitats following rare, though biologically significant long-distance dispersal events, such that it would allow an individual to grow and persist within a new habitat without reliance on sexual reproduction. Additionally, clonal growth tends to be common among riparian species (Brzyski & Culley 2011), which may reflect adaptation to flooding.

Populations that are primarily clonal were observed to have low fruit-set (Table 4.4; Figure 4.11). Why isn't there more asexual reproduction (i.e., increased clonality) in regions of high habitat suitability (Wyatt & Trapnell, in prep), such as the southeastern part of the range? This question remains unanswered, though a possible explanation is that individuals in areas of high habitat suitability are investing more resources into sexual reproduction relative to individuals in the northern and western portions of the species' range.

Spatial Genetic Structure

The results of the spatial genetic structure analyses further highlight the importance of vegetative growth over sexual reproduction in populations of *A. triloba* across the species' range. However, it cannot be ruled out that significant genetic structure exists at smaller spatial scales. In this study, stems were typically sampled at approximately 10 m intervals when possible (i.e., beyond the extent of the canopy of most stems). Sampling at a finer spatial scale may have allowed detection of more progeny that resulted from sexual reproduction underneath maternal trees, which may have yielded higher *r* values at smaller distance intervals. An alternative explanation for the apparent absence of significant SGS is that seeds are over-dispersed by mammals (e.g., racoons, coyotes), resulting in seedlings being established far from maternal trees and sibs. This appears unlikely, or at least very uncommon, since fruits typically fall and rot or are partially consumed under maternal trees (pers. obs.). It is possible that the primary seed dispersers that co-evolved with *A. triloba* are now extinct, a hypothesis first elaborated by Janzen & Martin (1982). The absence of megafauna seed dispersers could result in dramatic changes to patterns of SGS.

For a number of populations with low numbers of MLGs (BCP, CON, and LHP), *r* values were non-significant when clones were included in the analyses (Table 4.2). This is a bit surprising considering these populations possess only two to four MLGs. This suggests that there was considerable spatial overlap of genets that masked the degree of relatedness at small distance intervals.

CONCLUSIONS

Our research illustrates the importance of clonal growth for *A. triloba*. Clonality is evidently an important trait to the growth and persistence of this species, particularly at the edges of the species' range where presumably abiotic conditions are less optimal. Clonal growth appears to have profound effects on SGS of this species across its range. Studies into clonal tree species covering a wide geographic range in Eastern North America are scarce. The results of this study show that asexual propagation in trees can vary considerably across a wide geographic range and may be influenced by a number of biotic and abiotic factors, including history of manipulation by pre-Columbian cultures. Future studies on clonal growth of *A. triloba* might benefit from increased sample sizes and finer scale sampling of stems within populations, in order to better define the extent of clonal growth of individuals. Additionally, it would be interesting to quantify the relative investment of resources into asexual propagation versus sexual reproduction for trees across a latitudinal gradient.

LITERATURE CITED

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635 – 1644.

Birky CW Jr. (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* 144: 427 – 437.

Brzyski JR, Culley TM (2011) Genetic variation and clonal structure of the rare, riparian shrub *Spriaea virginiana* (Rosaceae). *Conservation Genetics* 12: 1323 – 1332.

Burns RM, Honkala BH (1990) Silvics of North America, Vol. 2, Hardwoods. Washington DC: USDA Forest Service Agriculture Handbook 654.

Cook RE (1983) Clonal plant populations: A knowledge of clonal structure can affect the interpretation of data in a broad range of ecological and evolutionary studies. *American Scientist* 71: 244 – 253.

Del Tredici P (2001) Sprouting in temperate trees: A morphological and ecological review. *The Botanical Review* 67: 121 – 140.

Dorken ME, Eckert CG (2001) Severly reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology* 89: 339 – 350.

Fager, EW (1972) Diversity: a sampling study. *The American Naturalist* 106: 293 – 310.

Freeman CC, Hulbert LC (1985) An annotated list of vascular flora of Konza Prairie Research Natural Area, Kansas. *Transactions of the Kansas Academy of Science* 88: 84 – 115.

Hauswaldt SJ, Glenn TC (2003) Microsatellite DNA loci from the Diamondback terrapin (*Malaclemys terrapin*). *Molecular Ecology Notes* 3: 174 – 176.

Hosaka N, Gómez S, Kachi N, Stuefer JF, Whigham DF (2005) The ecological significance of clonal growth in the understory tree, Pawpaw (*Asimina triloba*). *Northeastern Naturalist* 12: 11 – 22.

Hosaka N, Kachi N, Kudoh H, Stuefer JF, Whigham DF (2008) Patch Structure and ramet demography of the clonal tree, *Asimina triloba*, under gap and closed-canopy. *Plant Ecology* 197: 219 – 228.

Janzen DH, Martin PS (1982) Neotropical anachronisms: The fruits the Gomphotheres ate. *Science* 215: 19 – 27.

Meloni M, Reid A, Caujapé-Castells J, Marrero Á, Fernández-Palacios JM, Mesa-Coelo RA, Conti E (2013) Effects of clonality on the genetic variability of rare, insular species: the case of *Ruta microcarpa* from the Canary Islands. *Ecology and Evolution* 6: 1569 – 1579.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution* 57: 1182 – 1195.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology Notes* 6: 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537 – 2539.

Pebesma, EJ, RS Bivand (2005) Classes and methods for spatial data in R. R News 5 (2), https://cran.r-project.org/doc/Rnews/.

Pomper KW, Lowe JD, Lu L, Crabtree SB, Dutta S, Schneider K, Tidwell J (2010) Characterization and identification of pawpaw cultivars and advanced selections by simple sequence repeat markers. *Journal of the American Society for Horticulture Science* 135: 143 – 149.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rogstad SH, Wolff K, Schaal BA (1991) Geographical variation in *Asimina triloba* Dunal (Annonaceae) revealed by the M13 "DNA fingerprinting" probe. *American Journal of Botany* 76: 1391 – 1396.

Simpson EH (1949) Measurement of Diversity. *Nature* 163: 688 – 688.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual mutliallele and multilocus genetic structure. *Heredity* 82: 561 – 573.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13: 921 – 935.

Weakley AS. Flora of the Southern and Mid-Atlantic States. The University of North Carolina Herbarium, Chapel Hill. Working draft of 30 November 2012.

Table 4.1. Location of 82 study populations of *Asimina triloba*, their putative origin (anthropogenic versus wild), number of unique multilocus genotypes (MLG), probability of identity (P_I), and inbreeding coefficient (F_{IS}). None of the F_{IS} values were significant ($p > 0.05$).

Population	County, State	Latitude	Longitude	MLG	P_I	F_{IS}
Anthropogenic						
FCP	Jackson, FL	30.810	-85.227	3	1.7×10^{-3}	0.088
PET	Washington, NC	35.793	-76.413	21	3.0×10^{-4}	0.051
TUN	West Feliciana, LA	30.930	-91.514	7	3.8×10^{-6}	0.044
FHM	Highland, OH	39.123	-83.401	6	1.0×10^{-2}	0.058
LM1	Warren, OH	39.410	-84.099	14	9.0×10^{-5}	0.125
LM2	Warren, OH	39.410	-84.101	1	1.3×10^{-4}	-1.000
LM3	Warren, OH	39.409	-84.100	3	7.4×10^{-5}	-0.383
MC1	Edmonson, KY	37.204	-86.139	7	8.4×10^{-5}	0.098
MC2	Edmonson, KY	37.178	-86.108	10	9.5×10^{-5}	-0.194
MC3	Edmonson, KY	37.148	-86.093	1	6.3×10^{-5}	-1.000
OSF	Coffee, TN	35.485	-86.103	3	1.8×10^{-4}	-0.179
SCI	Ross, OH	39.216	-82.974	1	1.3×10^{-4}	-1.000
SEL	Wilson, TN	36.169	-86.236	2	2.6×10^{-5}	-0.562
CFP	Frederick, MD	39.633	-77.456	1	2.5×10^{-4}	-1.000
POC	Chesterfield, VA	37.381	-77.583	6	2.6×10^{-4}	-0.400
SU1	Harford, MD	39.610	-76.147	11	1.8×10^{-4}	-0.111
CON	Ashtabula, OH	41.933	-80.611	2	1.1×10^{-4}	-0.467
WEN	Erie, NY	42.679	-79.052	1	4.6×10^{-5}	-1.000
LCN	Berrien, MI	41.951	-86.306	4	1.3×10^{-2}	-0.690
CHA	Chautauqua, NY	42.339	-79.604	1	2.1×10^{-4}	-1.000
Mean				**5.3**	**1.4 x 10-4**	**-0.426**

Wild						
BTH	Oglethorpe, GA	33.706	-83.019	10	3.4×10^{-7}	0.066
CAT	Catawba, NC	35.821	-81.177	1	8.4×10^{-5}	-1.000
CFR	Harnett, NC	35.469	-78.898	5	1.2×10^{-5}	-0.238
COL	Richland, SC	33.947	-80.629	12	7.5×10^{-5}	-0.015
ER2	Orange, NC	36.083	-79.021	5	5.3×10^{-5}	-0.527
EUT	Greene, AL	32.785	-87.837	30	5.0×10^{-7}	0.231
LCC	McCormick, SC	34.054	-82.343	18	2.2×10^{-5}	0.147
LIL	Fairfield, SC	34.253	-81.233	21	2.2×10^{-5}	0.194
PCB	Pike, AL	31.907	-86.174	12	7.2×10^{-5}	0.013
SAL	Pickens, SC	35.073	-82.587	14	4.1×10^{-5}	0.030
SHE	Paulding, GA	34.022	-84.911	5	4.0×10^{-3}	-0.045
STO	York, SC	34.881	-81.104	5	5.3×10^{-5}	0.129
TAL	Cleburne, AL	33.859	-85.524	3	8.9×10^{-3}	-0.266
TER	Saluda, SC	34.147	-81.824	22	3.2×10^{-6}	0.081
WBF	Winston, AL	34.284	-87.396	22	5.1×10^{-6}	0.082
BCL	Lee, AR	34.706	-90.698	27	6.0×10^{-9}	0.125
CAD	Harrison, TX	32.693	-94.176	18	2.1×10^{-4}	0.085
HAM	West Feliciana, LA	30.785	-91.255	2	2.5×10^{-4}	0.170
HEM	Sabine, TX	31.482	-93.936	5	2.1×10^{-4}	-0.016
LEE	Lee, AR	34.760	-90.733	12	1.3×10^{-4}	0.096
MS1	Shelby, TN	35.340	-90.051	4	1.3×10^{-4}	0.072
MS2	Shelby, TN	35.341	-90.055	1	1.2×10^{-4}	-1.000
NAG	St. Augustine, TX	31.480	-93.995	3	1.6×10^{-4}	-0.158
SAN	Adams, MS	31.460	-91.211	1	1.1×10^{-4}	-1.000
WDP	Marshall, MS	34.668	-89.467	3	6.1×10^{-3}	0.192
BLU	Carter, MO	36.961	-90.987	9	9.5×10^{-5}	0.062
BSF	Christian, MO	36.858	-93.222	21	1.2×10^{-5}	-0.01

HOL	Bourbon, KS	37.777	-94.827	8	5.3×10^{-5}	-0.098
LAW	Douglas, KS	38.991	-95.315	12	9.6×10^{-3}	-0.185
MLK	Boone, MO	38.936	-92.351	2	3.5×10^{-2}	-0.622
MWR	Stoddard, MO	36.970	-90.151	14	1.3×10^{-4}	-0.092
PIA	Jersey, IL	38.939	-90.285	1	7.4×10^{-5}	-1.000
PMQ	Jersey, IL	38.984	-90.49	4	3.2×10^{-4}	-0.313
WES	Platte, MO	39.384	-94.88	9	1.2×10^{-4}	0.078
WSL	St. Charles, MO	38.666	-90.732	9	4.3×10^{-7}	0.078
BCB	Coffee, TN	35.485	-85.961	8	1.3×10^{-4}	-0.367
BCP	Brown, IN	39.175	-86.222	2	2.5×10^{-4}	-0.547
CCC	Carter, KY	38.379	-83.123	13	1.4×10^{-4}	0.222
CHR	Clark, IN	38.398	-85.635	18	6.2×10^{-6}	0.030
CKW	Madison, KY	37.632	-84.196	13	1.7×10^{-4}	-0.024
DCR	Harrison, WV	39.146	-80.391	1	7.9×10^{-5}	-1.000
DRP	Johnson, IL	37.430	-88.942	13	5.1×10^{-6}	0.070
GCR	Anderson, KY	37.975	-84.825	6	2.5×10^{-4}	0.316
HPC	Platte, MO	39.228	-94.778	14	1.1×10^{-3}	-0.020
KAN	Kanawha, WV	38.224	-81.625	10	2.3×10^{-4}	0.090
KNX	Knox, TN	35.941	-83.931	3	1.3×10^{-4}	0.133
LHP	Davidson, TN	36.087	-86.551	4	1.1×10^{-4}	-0.442
LL1	Stewart, TN	36.581	-87.947	1	1.3×10^{-4}	-1.000
LL2	Trigg, KY	36.792	-88.096	7	1.3×10^{-4}	-0.055
LL3	Trigg, KY	36.900	-88.040	24	1.3×10^{-4}	0.141
OFR	Jackson, KY	37.542	-84.174	6	1.7×10^{-4}	-0.082
PER	Davidson, TN	36.065	-86.880	5	2.5×10^{-4}	0.044
RCA	Laurel, KY	38.566	-84.351	16	1.3×10^{-4}	0.273
RL	Davidson, TN	36.063	-86.799	12	1.4×10^{-4}	-0.064
ROC	Boone, MO	38.875	-92.327	1	1.3×10^{-4}	-1.000

SFP	Guernsey, OH	40.128	-81.510	7	1.3×10^{-4}	0.014
SHO	Cannon, TN	35.872	-85.991	16	1.3×10^{-4}	-0.062
TYG	Taylor, WV	39.308	-80.027	2	2.5×10^{-4}	-0.600
LOC	Morgan, MD	39.692	-78.199	3	2.0×10^{-4}	-0.134
MAN	Prince William, VA	38.805	-77.525	1	2.5×10^{-4}	-1.000
PWP	Prince William, VA	38.566	-77.346	27	2.5×10^{-4}	0.169
SUS	Lancaster, PA	39.808	-76.286	8	1.6×10^{-4}	0.088
Mean				**9.5**	**1.1 x 10-3**	**-0.153**
Species Mean				**8.5**	**9.2×10^{-4}**	**-0.219**

Table 4.2. Locations and clonal diversity estimates for the 40 mapped populations of *A. triloba* of putative anthropogenic (10) and wild (30) origin. Measures include N, number of samples; MLG, number of unique multilocus genotypes; P_I, probability of identity; *R*, clonal richness; *D*, Simpson's diversity index; *E*, Fager's evenness; $Area_{max}$, area of largest genet (m^2); and $Distance_{max}$, maximum observed distance (m) between ramets.

Population	Latitude	Longitude	Sample area	N	MLG	P_I	*R*	*D*	*E*	$Area_{max}$	$Distance_{max}$
Anthropogenic											
LCN	41.951	-86.306	838.7	48	4	2.5×10^{-4}	0.064	0.197	0.116	838.7	44
CON	41.933	-80.611	111.4	30	2	1.6×10^{-4}	0.034	0.186	0.265	111	13.9
LM2	39.410	-84.101	152.9	11	1	1.3×10^{-4}	0	0	0	152.9	24.3
SCI	39.216	-82.974	902.4	24	1	1.3×10^{-4}	0	0	0	902.4	47
FHM	39.123	-83.401	851.9	48	6	1.0×10^{-2}	0.106	0.337	0.211	556.6	34.8
MC1	37.204	-86.139	529.8	16	7	8.4×10^{-5}	0.4	0.85	0.778	47.8	16.2
MC2	37.178	-86.108	2170.2	18	10	9.5×10^{-5}	0.529	0.889	0.66	91.7	29.4
MC3	37.148	-86.093	184.3	12	1	6.3×10^{-5}	0	0	0	184.3	23.6
PET	35.793	-76.413	5774.0	24	22	3.0×10^{-4}	0.913	0.993	0.559	-	17.3
TUN	30.930	-91.514	4133.0	25	7	1.3×10^{-4}	0.25	0.87	0.951	33.8	26.1
MEAN			**1565**	**25.6**	**6.1**	**1.4×10^{-4}**	**0.230**	**0.432**	**0.354**	**324.3**	**27.7**
Wild											
SUS	39.808	-76.286	401.3	30	8	1.6×10^{-4}	0.241	0.805	0.794	109.8	21.3
HPC	39.228	-94.778	8450.7	25	14	1.1×10^{-3}	0.542	0.893	0.603	239	70.2
BCP	39.175	-86.222	680.8	48	2	2.5×10^{-4}	0.021	0.383	0.727	517.2	33.7
LAW	38.991	-95.315	7060.9	30	12	9.6×10^{-3}	0.379	0.798	0.56	772.2	91.8
MLK	38.936	-92.351	2924.4	24	2	3.5×10^{-2}	0.043	0.159	0.173	2747.7	68.2
ROC	38.875	-92.327	1875.5	25	1	1.3×10^{-4}	0	0	0	1875.5	163.4
CHR	38.398	-85.635	2103.5	30	18	6.2×10^{-6}	0.586	0.952	0.84	34.2	7.9
CKW	37.632	-84.196	1746.5	24	13	1.7×10^{-4}	0.522	0.917	0.772	81.7	20.7
OFR	37.542	-84.174	717.0	33	6	1.7×10^{-4}	0.156	0.758	0.824	215	20.8
DRP	37.430	-88.942	2603.8	23	13	5.1×10^{-6}	0.545	0.938	0.852	13.9	8.1
BLU	36.961	-90.987	3635.7	25	9	9.5×10^{-5}	0.333	0.807	0.686	605	42.8
LL3	36.900	-88.040	9170.1	24	24	1.3×10^{-4}	1	1	1	-	-

BSF	36.858	-93.222	5627.6	23	21	1.2×10^{-5}	0.909	0.993	0.645	-	21.5
LL2	36.792	-88.096	890.3	25	7	1.3×10^{-4}	0.25	0.84	0.886	149.9	25.1
LHP	36.087	-86.551	1240.7	20	4	1.1×10^{-4}	0.158	0.647	0.718	850.3	53.9
SHO	35.872	-85.991	1925.3	24	16	1.3×10^{-4}	0.652	0.942	0.666	81.3	19.4
CFR	35.469	-78.898	1471.0	20	2	1.2×10^{-5}	0.053	0.1	0	1304.5	68.4
SAL	35.073	-82.587	5831.0	30	14	4.1×10^{-5}	0.448	0.894	0.756	420.2	51.5
BCL	34.706	-90.698	1422.1	15	12	6.0×10^{-9}	0.786	0.971	0.716	-	13.3
WDP	34.668	-89.467	212.1	27	3	6.1×10^{-3}	0.077	0.453	0.563	54.3	12.3
WBF	34.284	-87.396	7957.7	43	21	5.1×10^{-6}	0.476	0.971	0.984	20	13.7
LIL	34.253	-81.233	6689.1	24	19	2.2×10^{-5}	0.783	0.986	0.94	30.7	17.7
TER	34.147	-81.824	5945.4	34	22	3.2×10^{-6}	0.636	0.979	0.964	48.3	68
LCC	34.054	-82.343	7357.5	24	23	2.2×10^{-5}	0.957	0.996	0.179	-	15.5
SHE	34.022	-84.911	1246.2	24	5	4.0×10^{-3}	0.175	0.58	0.61	38.5	14.8
COL	33.947	-80.629	387.6	48	17	7.5×10^{-5}	0.234	0.801	0.743	15.4	15.4
TAL	33.859	-85.524	602.4	24	3	8.9×10^{-3}	0.087	0.489	0.612	134.4	14.6
BTH	33.706	-83.019	6837.5	60	21	3.4×10^{-7}	0.339	0.868	0.731	776.9	65.3
EUT	32.785	-87.837	4764.6	33	30	5.0×10^{-7}	0.906	0.996	0.894	-	4.8
PCB	31.907	-86.174	4083.3	48	12	7.2×10^{-5}	0.234	0.776	0.696	119	20.8
MEAN			**3528.7**	**28.9**	**12.5**	**1.9×10^{-3}**	**0.403**	**0.760**	**0.670**	**450.2**	**36.7**

Table 4.3. Spatial genetic structure statistics for 36 populations of *Asimina triloba* of putative anthropogenic (7) and wild (29) origin. N = number of samples and MLG = number of multilocus genotypes. Levels of relatedness (*r*) values at the three smallest distance classes (in meters) with clonal stems included and excluded. For 16 populations with < 12 MLGs, spatial genetic structure could not be estimated when clones were removed. N = number of samples and MLG = number of multilocus genotypes.

			Clones Included			Clones Removed		
Population	N	MLG	1st r (distance)	2nd r (distance)	3rd r (distance)	1st r (distance)	2nd r (distance)	3rd r (distance)
Anthropogenic								
LCN	48	4	0.43 (2.5)*	0.62 (5)	0.23 (7.5)*	-	-	-
CON	30	2	0.12 (2.2)	0.04 (4.4)	-0.06 (6.6)	-	-	-
FHM	48	6	0.20 (3.4)	0.04 (6.8)	-0.01 (10.2)	-	-	-
MC1	16	7	0.68 (10)*	0.69 (20)*	-0.46 (30)	-	-	-
MC2	18	10	0.37 (18)*	0.14 (36)*	-0.30 (54)	-	-	-
PET	24	21	0.05 (40)	0.03 (80)	0.00 (120)	0.00 (50)	0.01 (100)	0.07 (150)
TUN	25	7	0.54 (10.5)*	0.13 (21)	-0.05 (31.5)	-	-	-
Wild								
SUS	30	8	0.49 (3)*	0.20 (6)*	0.06 (9)			
HPC	25	14	0.35 (16.5)*	0.10 (33)	0.08 (49.5)	0.06 (45)	-0.02 (90)	-0.04 (135)
BCP	48	2	0.26 (2.8)	0.21 (5.6)*	0.23 (8.4)*	-	-	-
LAW	30	12	0.54 (14)*	0.39 (28)*	0.37 (42)*	0.05 (90)	-0.02 (180)	-0.09 (270)
MLK	24	2	0.53 (14)*	-0.07 (28)	-0.04 (42)	-	-	-
CHR	30	18	0.34 (5)*	0.13 (10)*	-0.03 (15)	0.09 (15)*	0.01 (30)*	-0.06 (45)
CKW	24	13	0.33 (12)*	0.08 (24)	-0.10 (36)	0.00 (28)	-0.05 (56)	0.20 (84)
OFR	33	6	0.60 (5)*	0.42 (10)*	0.13 (15)*	-	-	-
DRP	23	13	0.51 (14)*	-0.06 (28)	-0.15 (42)	0.07 (37)*	-0.04 (74)	-0.16 (111)
BLU	25	9	0.61 (13)*	0.13 (26)*	-0.09 (39)	-	-	-
LL3	24	24	0.03 (30)	0.02 (60)	-0.04 (90)	0.03 (30)	0.02 (60)	-0.04 (90)

BSF	23	21	0.27 (10)*	0.04 (20)	0.03 (30)	0.02 (30)	0.00 (60)	0.00 (90)
LL2	25	7	0.71 (8)*	0.18 (16)*	0.15 (24)*	-	-	-
LHP	20	4	-0.04 (12)	-0.02 (24)	0.06 (36)	-	-	-
SHO	24	16	0.21 (11)*	-0.02 (22)	-0.10 (33)	0.0 (23)	0.0 (46)	0.01 (69)
CFR	20	5	0.03 (22)	0.09 (44)	-0.04 (66)	-	-	-
SAL	30	14	0.45 (15)*	0.16 (30)*	0.12 (45)*	0.03 (55)	-0.02 (110)	0.02 (165)
BCL	15	12	0.03 (25)	-0.05 (50)	0.00 (75)	0.01 (47)	-0.02 (94)	0.01 (141)
WDP	27	3	1.04 (2.8)*	0.96 (5.6)*	0.79 (8.4)*	-	-	-
WBF	43	22	0.62 (7)*	0.24 (14)*	0.17 (21)*	0.03 (30)	0.01 (60)	-0.05 (90)
LIL	24	21	0.11 (20)*	-0.02 (40)	-0.04 (60)	0.03 (22)	-0.04 (44)	-0.01 (66)
TER	34	22	0.33 (13)*	0.01 (26)*	0.09 (39)	0.13 (20)	0.07 (41)	-0.08 (60)
LCC	24	18	0.04 (22)	0.06 (44)	-0.05 (66)	0.03 (22)	0.05 (44)	-0.05 (66)
SHE	24	5	0.71 (3.5)*	0.52 (7)*	0.50 (10.5)*	-	-	-
COL	48	17	0.22 (2.5)*	0.11 (5)*	0.03 (7.5)	0.00 (7)	0.05 (14)	-0.05 (21)
TAL	24	3	1.04 (5)*	0.87 (10)*	0.52 (15)*	-	-	-
BTH	30	14	0.35 (13) *	0.21 (26)*	-0.02 (39)*	-0.01 (65)	0.01 (130)	-0.03 (195)
EUT	33	30	0.19 (10)*	-0.01 (20)	-0.03 (30)	0.18 (10)*	0.00 (20)	-0.03 (30)
PCB	48	12	0.48 (3.5)*	0.19 (7)*	0.12 (10.5)*	0.03 (35)	0.04 (70)	-0.07 (105)

* Indicates significance at $p < 0.05$

Table 4.4. Comparison of fruit production in *A. triloba* populations possessing ≤ 2 MLGs versus populations with ≥ 3 MLGs. A heterogeneity chi-square test shows that the difference in fruit production is highly significant ($\chi^2 = 20.89$, $p = 0.00003$).

	Fruit Observed	No Fruit Observed	**Row Totals**
1 – 2 MLG	1	19	**20**
3 – 9 MLG	11	20	**31**
10+ MLG	19	8	**27**
Column Totals	**31**	**47**	**78**

Figure 4.1. Map of 82 sampled *A. triloba* populations, of both wild (blue circles) and putative anthropogenic (pink circles) origin. Size of the circles reflects the number of multilocus genotypes (MLGs) found in populations.

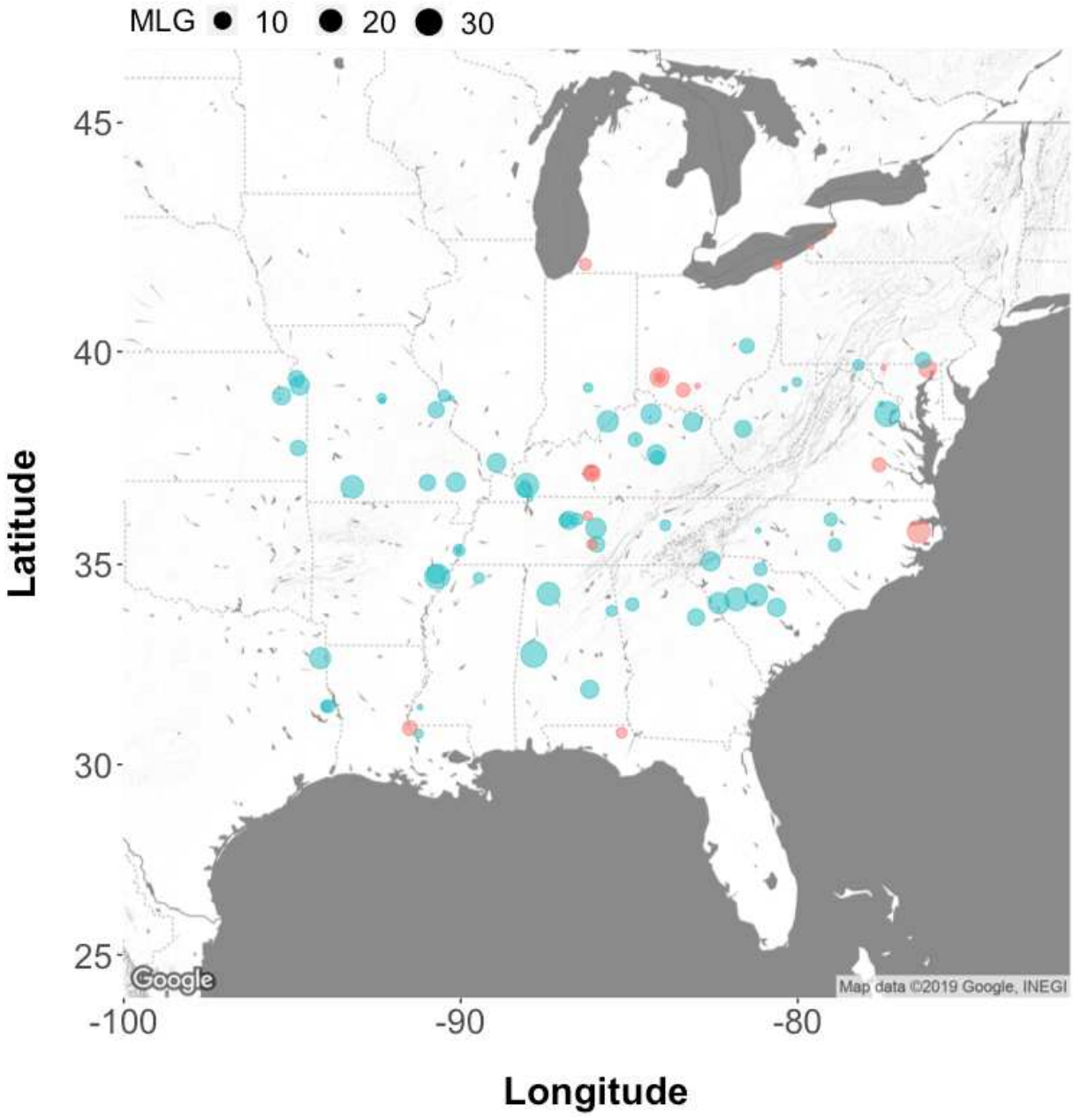


Figure 4.2. Comparison of the number of unique multilocus genotypes (MLG) found in anthropogenic and wild populations ($p = 0.007$).

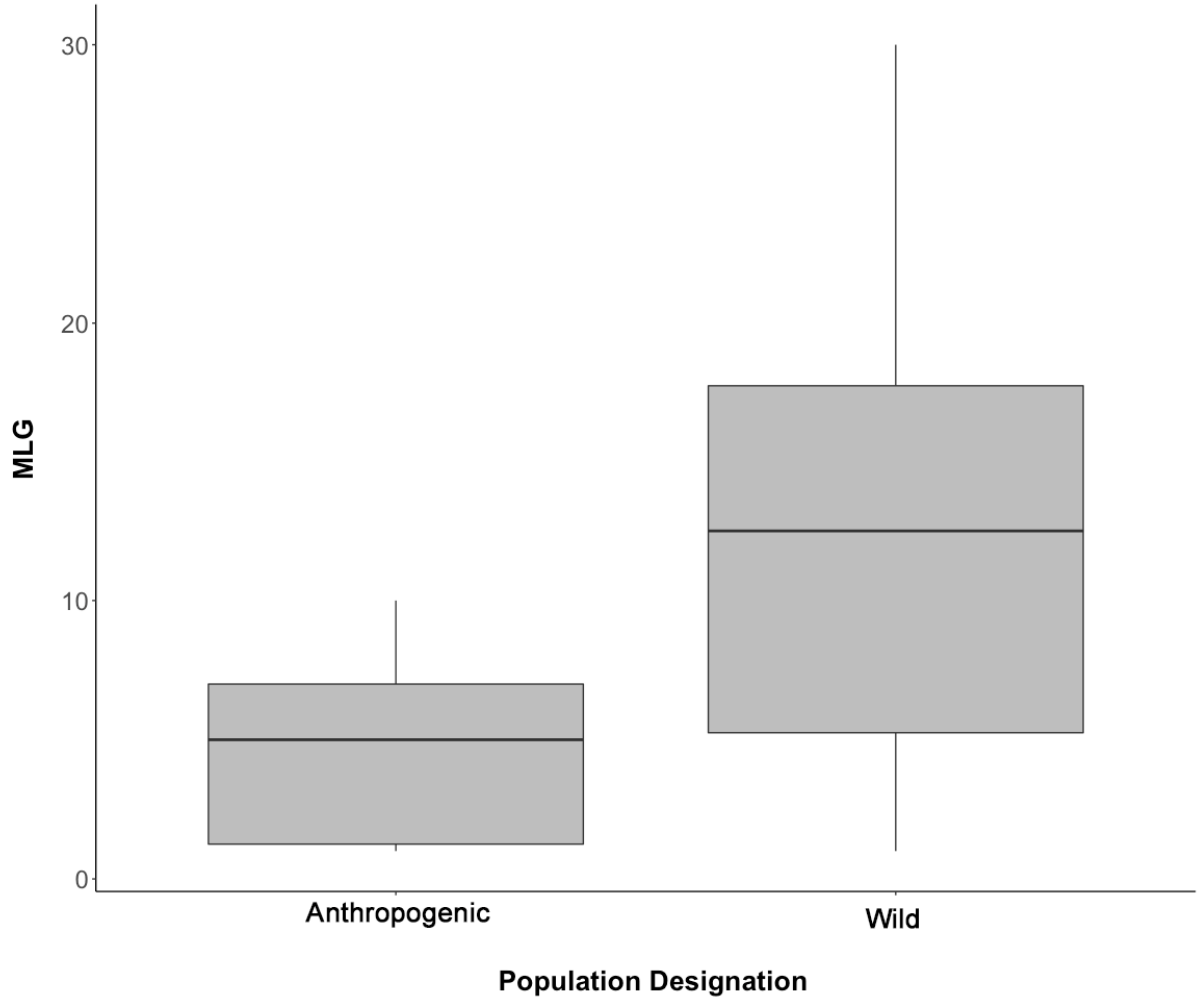


Figure 4.3. Relationship between number of multilocus genotypes (MLG) and latitude for the 82 sampled populations ($p = 0.06$).

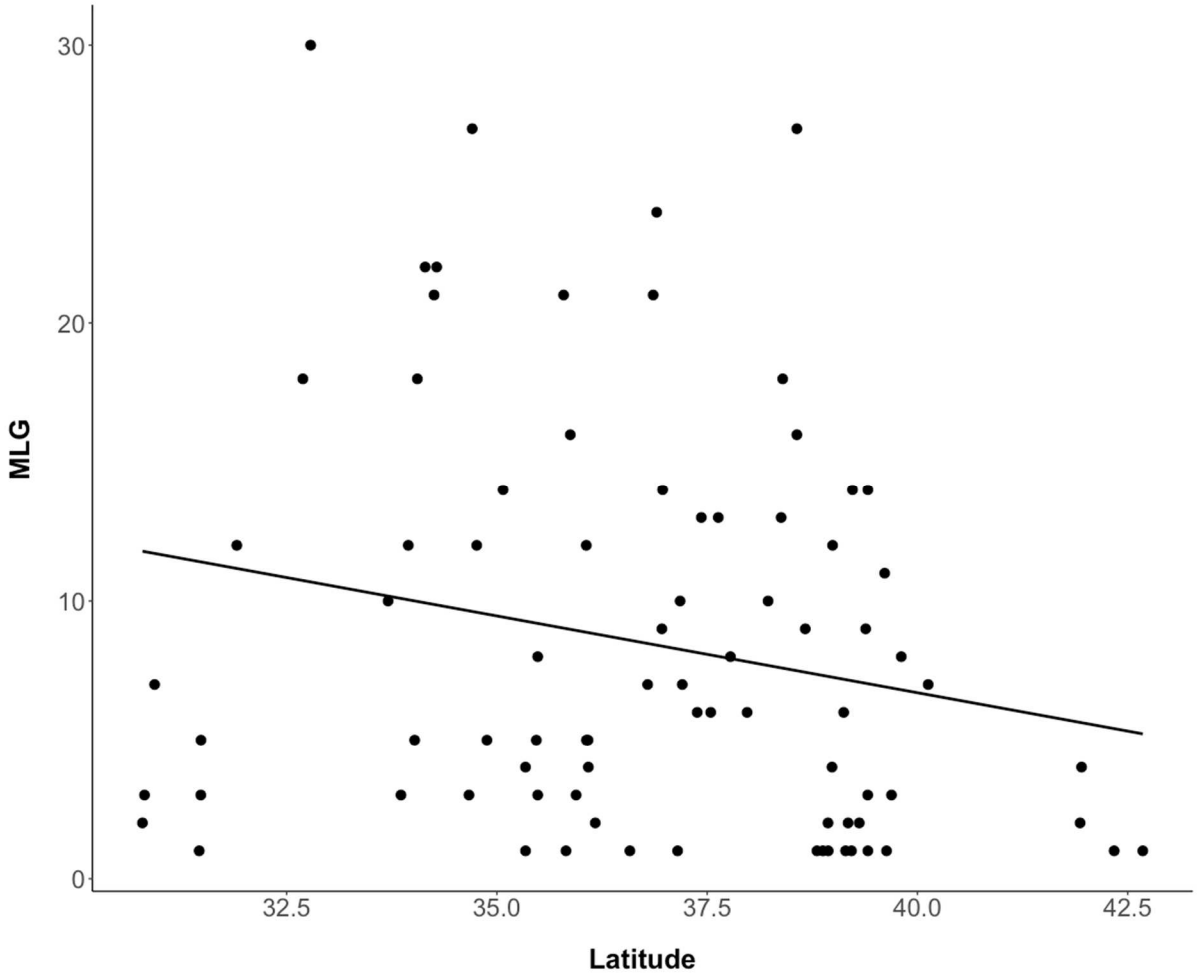


Figure 4.4. Relationship between inbreeding coefficient (F_{IS}) and latitude for the 82 sampled *A. triloba* populations ($p = 0.001$).

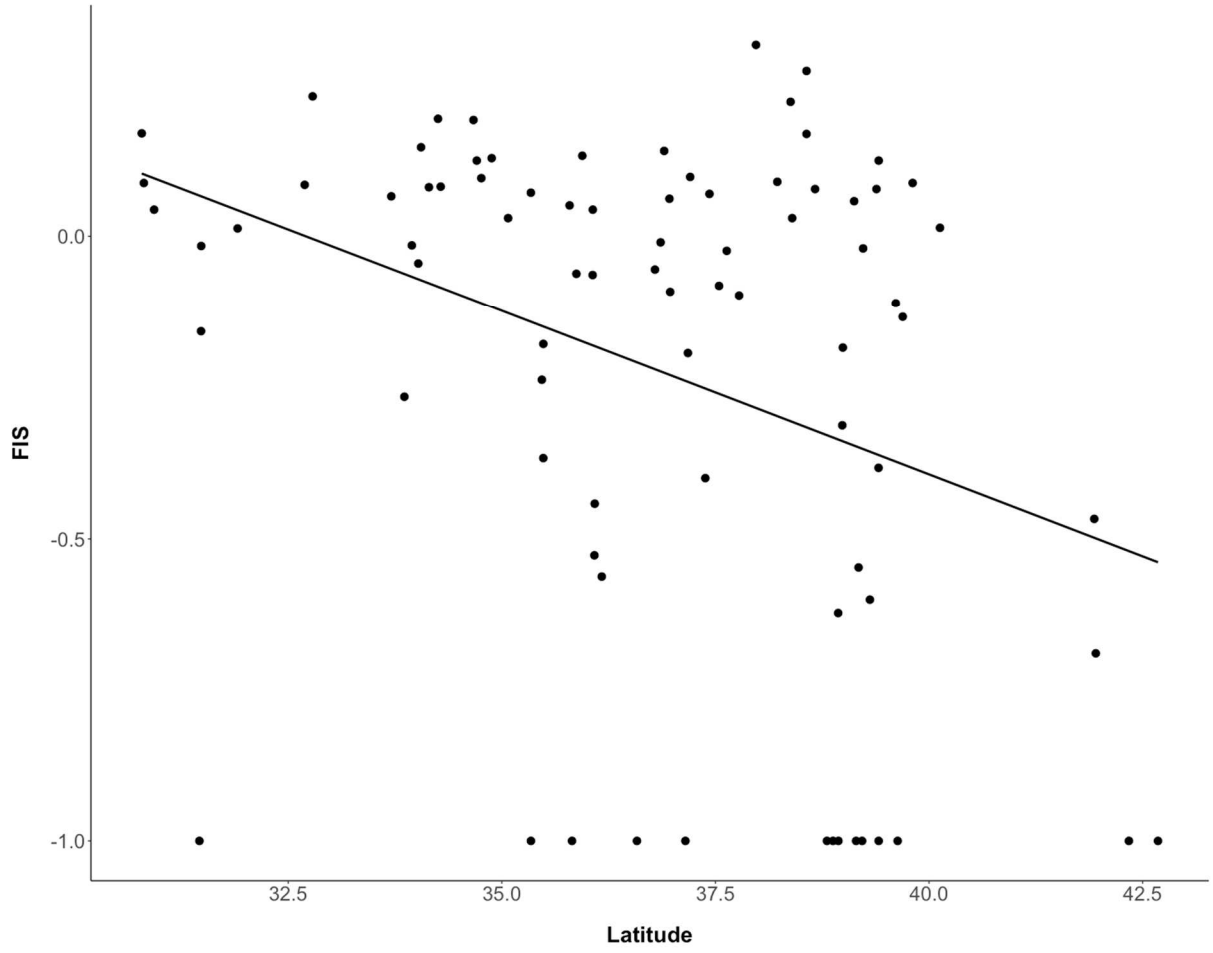


Figure 4.5. Relationship between the maximum observed distance in meters between stems belonging to the same MLG ($Distance_{max}$) and longitude ($p = 0.05$).

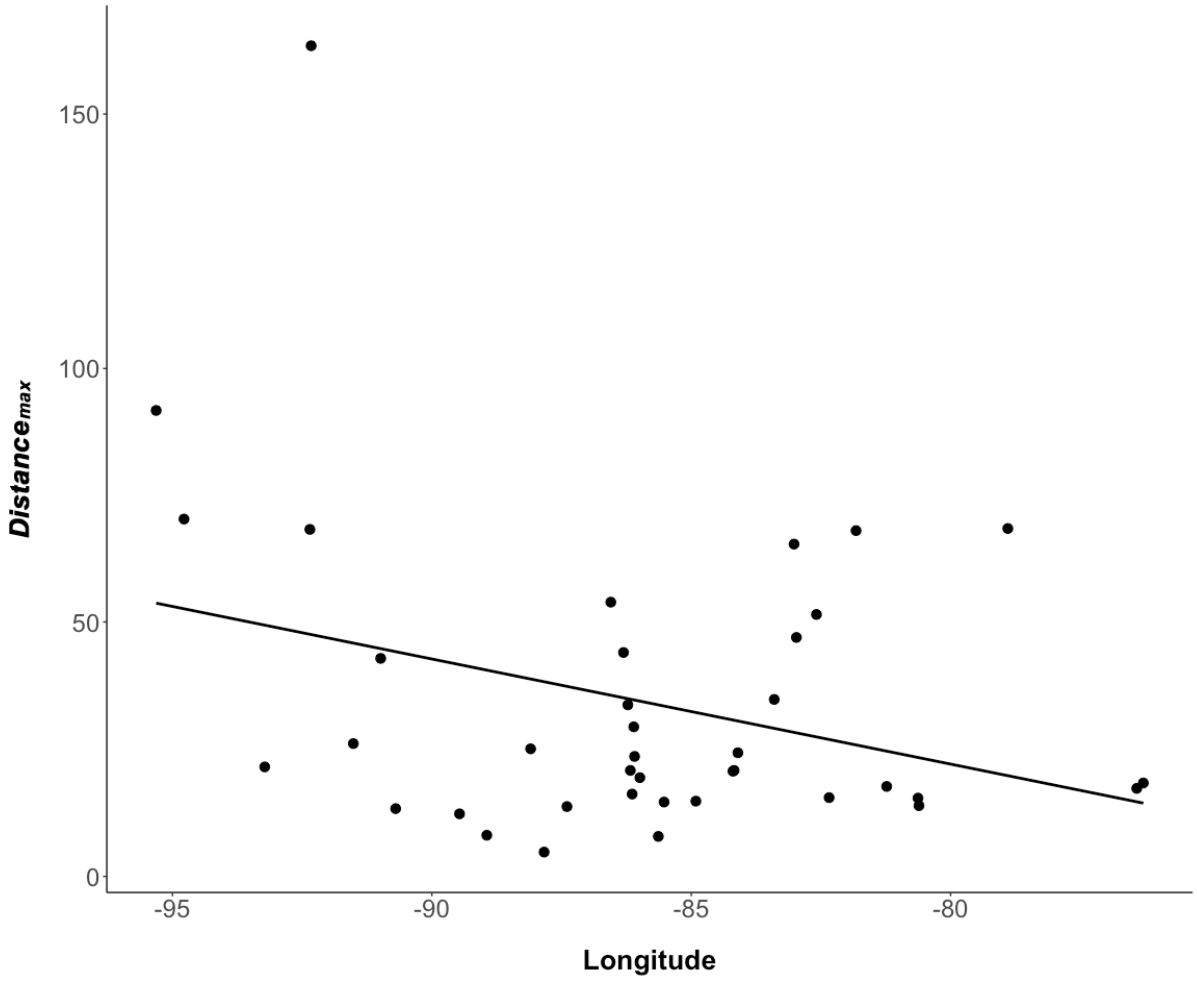


Figure 4.6. Relationship between clonal richness (*R*) and latitude in the 40 mapped populations ($p = 0.02$).

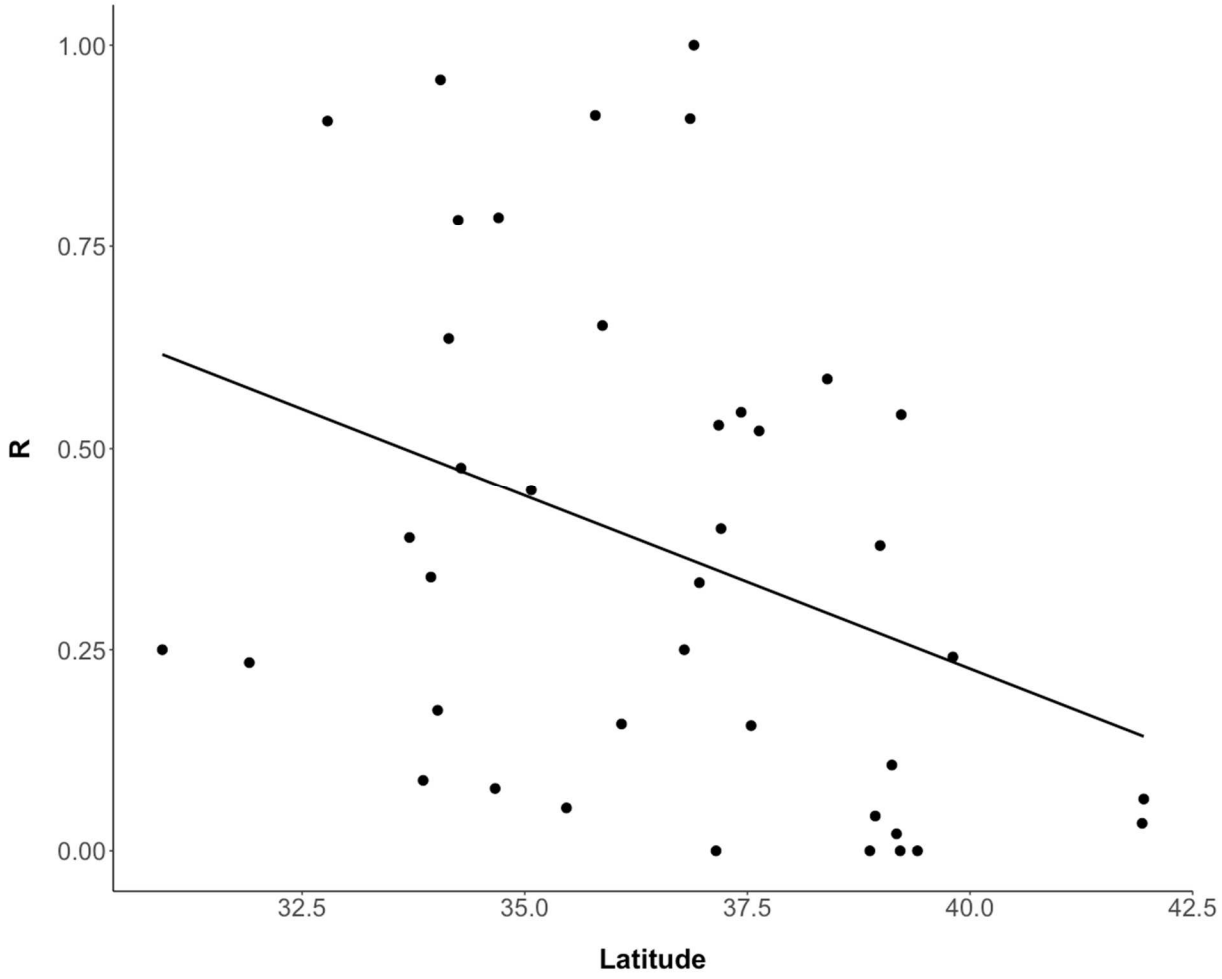


Figure 4.7. Relationship between clonal diversity (*D*), Simpson's diversity index, and latitude in the 40 mapped populations ($p = 0.003$).

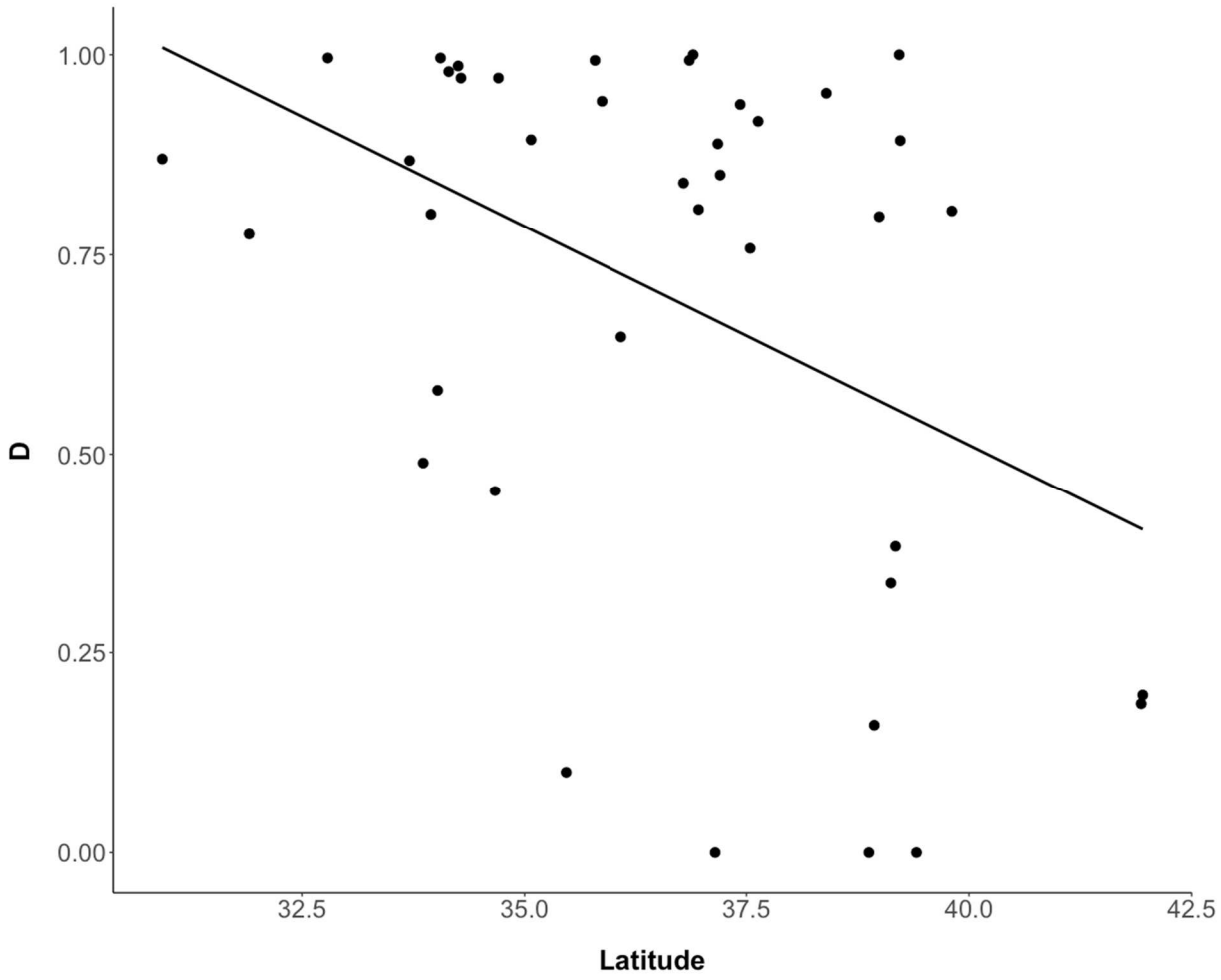


Figure 4.8. Relationship between Fager's evenness (*E*) and latitude in the 40 mapped populations ($p = 0.003$).

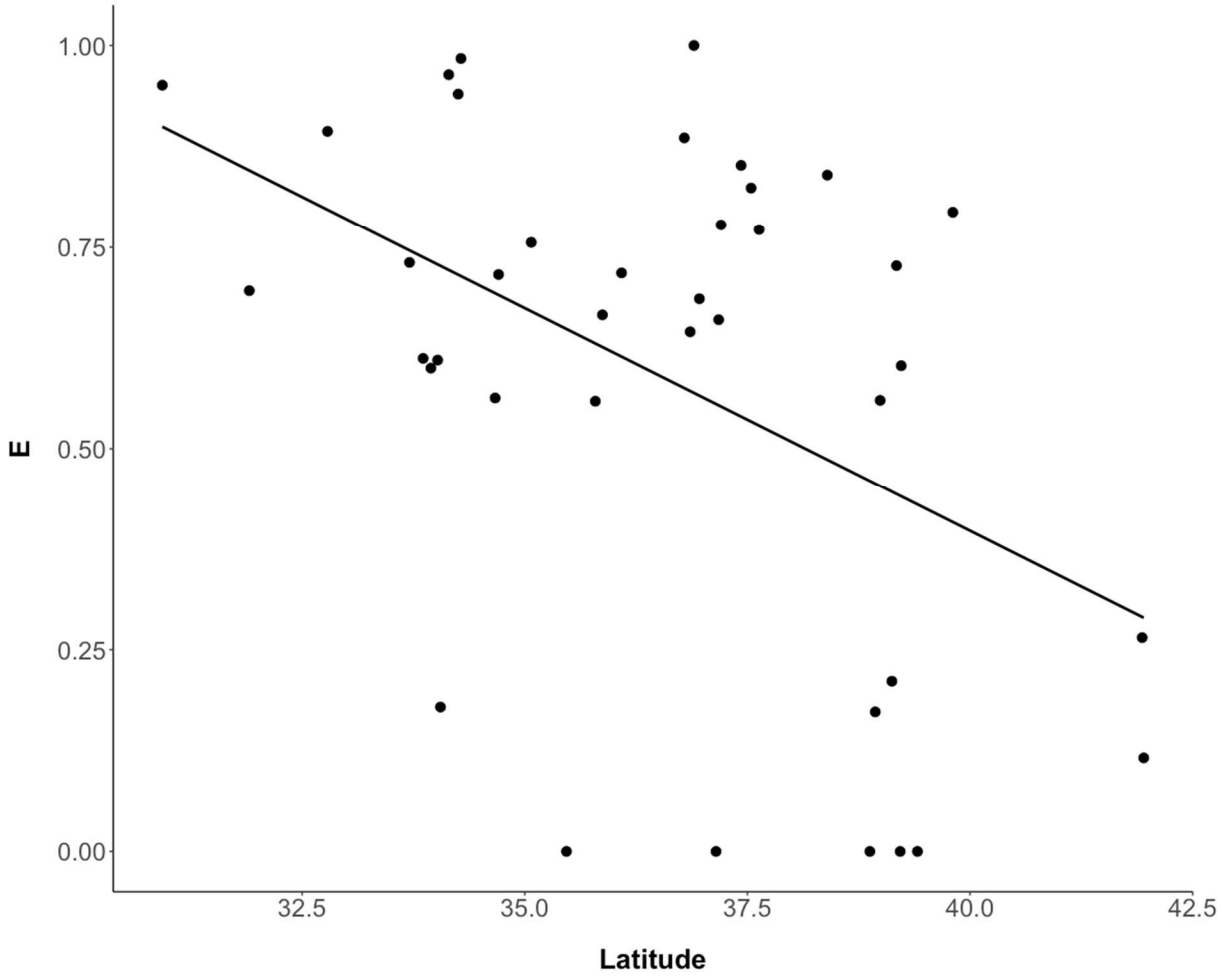


Figure 4.9. Relationship between Faeger's Evenness (E) and inbreeding coefficient (F_{IS}) in the 40 mapped populations ($p = 6.3 \times 10^{-10}$).

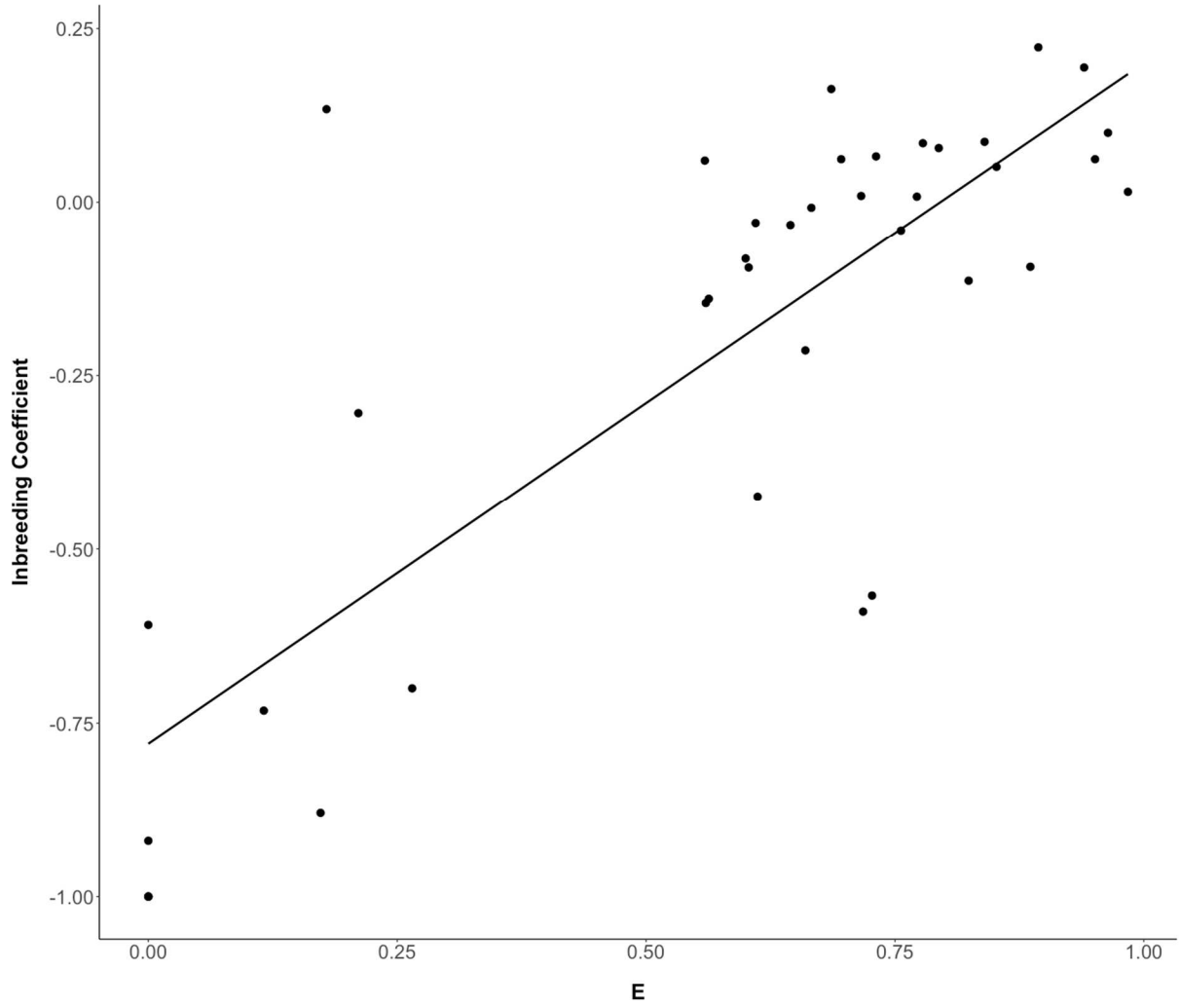


Figure 4.10. Relationship between log of mean area of genets (m^2) and latitude for the 40 mapped populations ($p = 0.07$).

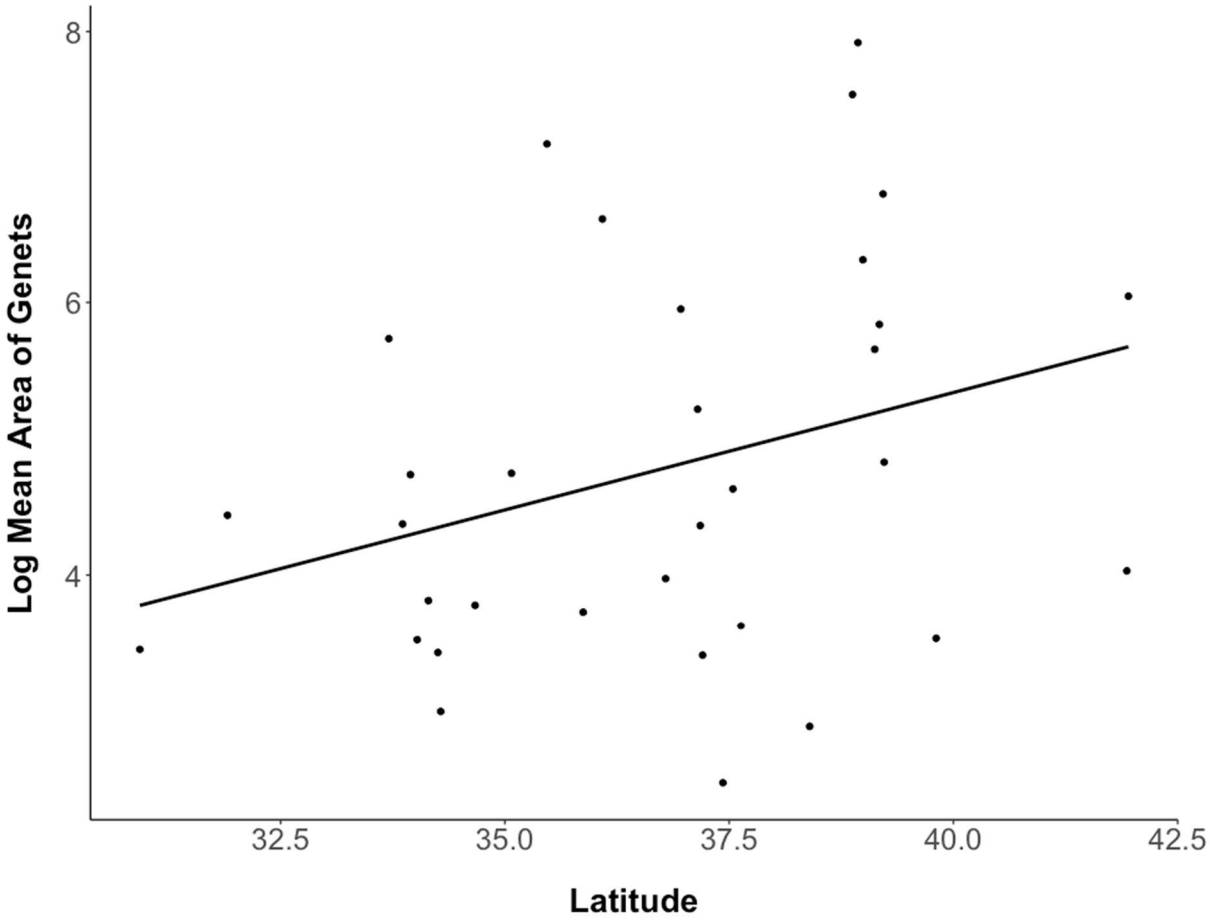
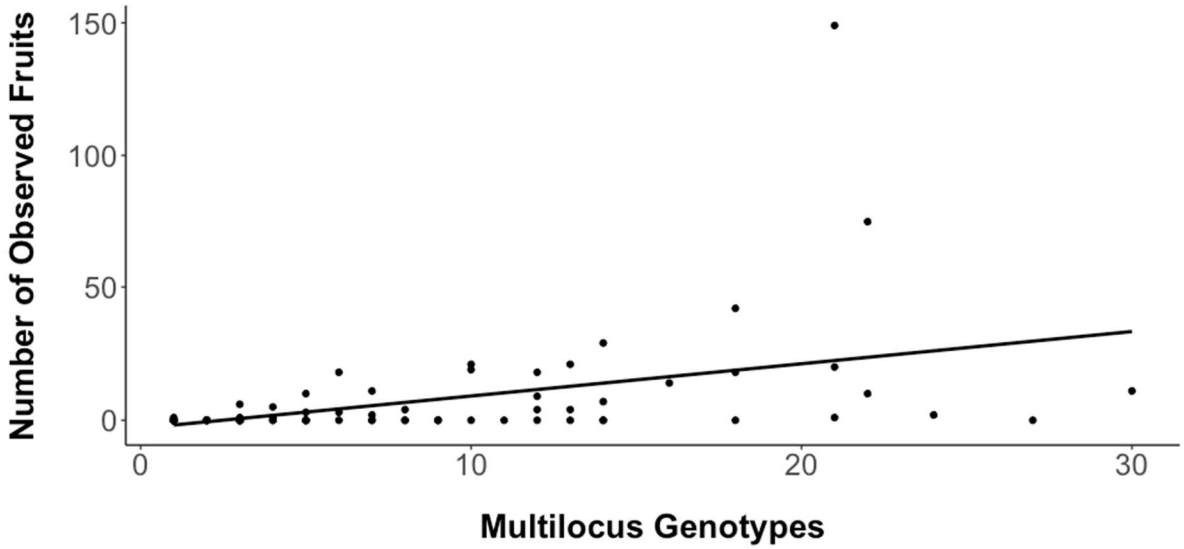


Figure 4.11. Total number of fruits observed on all sampled stems from 82 populations versus number of multilocus genotypes observed in a given population ($p = 0.0001$).



CHAPTER 5

CONCLUSIONS

This is the first investigation of range-wide patterns of genetic diversity and structure in *Asimina triloba*, a highly clonal understory tree species native to eastern North America. Genetic data provide support for the hypothesis that pre-Columbian humans played a role in the dispersal this valued tree species. High levels of genetic variation were found in the species and the degree of clonality within populations varied considerably across the species' range.

There was moderate genetic differentiation (G_{ST} = 0.341) was among 62 wild populations, which is higher than reported for other tree species (citation). High observed heterozygosity (H_O) was observed throughout the range (mean H_O = 0.530). No significant relationships were found between geographic and genetic distances. Genetic diversity did not decrease with increasing latitude, as would be expected for northward migrating plant species (Hewitt 2000). Two distinct genetic clusters, of populations separated primarily by the Appalachian Mountains and the Tombigbee River in Alabama were identified using a Bayesian clustering analysis (STRUCTURE; Pritchard et al. 2000) and principal components analysis of genotypes. These two clusters likely retain the genetic footprints of southern glacial refugia during the LGM (~22,000 YBP), the putative locations of which were modelled using ecological niche modeling in MAXENT. Modelling suggests that highly suitable habitat for *A. triloba* existed along the Gulf of Mexico and the southeastern Atlantic Coast in areas currently below

sea level. This finding is consistent with present-day east-west clustering of *A. triloba* populations, assuming most parsimonious post-glacial expansion routes (Swenson & Howard 2005). An inland area of high habitat suitability found in present day southwest Georgia may have allowed for admixture between the two refugia and may explain the high levels of genetic diversity observed in the middle latitudes of the species range (Petit et al. 2003).

Twenty populations of putative anthropogenic origin (i.e., those occurring on or adjacent to known Native American settlements) were also characterized using neutral nuclear genetic markers. Levels and partitioning of genetic variation were compared between anthropogenic and 62 wild populations (i.e., those not located adjacent to known Native American settlements). Seven Six population genetic measures (MLGs, A_T, A_E, P_A, A_{RARE}, and F_{IS}) revealed significantly higher genetic diversity in wild populations than anthropogenic populations. Interestingly observed heterozygosity (H_O) was significantly lower in wild populations. These data are consistent with human-mediated dispersal and manipulation of this highly prized tree species. The patterns observed in anthropogenic populations are likely to have resulted from admixing of seeds from wild populations followed by loss of alleles associated with successive founder events and perhaps strict sense genetic drift. The number of rare alleles (i.e., those occurring in < 10 % of populations) for all 82 populations decrease significantly with increasing latitude, though rare alleles were significantly less frequent in anthropogenic populations. The high levels of genetic structure among the northernmost anthropogenic populations suggest that subsequent to their colonization, there was little gene flow among them. These anthropogenic populations showed high levels of genetic identity and low G_{ST} values with geographically distant wild populations, which suggests long-distance dispersal and the possibility that these wild

populations served as the sources of propagules that founded the anthropogenic populations. These data support the hypothesis that pre-Columbian humans may have been responsible for the expansion of *A. triloba* to the northern limit of its distribution.

Our data further illustrate that clonal growth is very important to the within-population expansion and persistence of this understory tree species, particularly in more marginal habitats. Clonal growth patterns are not uniform across the broad geographic range of *A. triloba* and the degree of clonality is likely shaped by abiotic factors, with populations at the edges of the species' range engaging in little to no sexual reproduction, relying instead on clonal growth for long-term persistence. Fruit-set among all 82 sampled populations of this self-incompatible species was strongly dependent on number of multilocus genotypes (MLGs). Additionally, populations of *A. triloba* occurring on or adjacent to pre-Columbian settlements are characterized by significantly fewer MLGs and lower evenness of clones (Fager 1972). Spatial autocorrelation analyses revealed that genetic structure of *A. triloba* populations is strongly influenced by the production of ramets; when clones were excluded from analyses, populations possessed little to no significant spatial genetic structure.

LITERATURE CITED

Fager, EW (1972) Diversity: a sampling study. *The American Naturalist* 106: 293 – 310.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405: 907 – 913.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563 – 1565.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist* 166: 581 – 591.