Understanding the Genetic Diversity of the Invasive Callery pear, *Pyrus calleryana* Decne. in Native and Introduced Ranges of the U.S. using Microsatellite Loci

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Shiwani Sapkota

August 2021

DEDICATION

I dedicate this manuscript to my loving parents and my husband, Sujan Paudel. I am truly indebted for their unconditional love and support throughout my life.

ACKNOWLEDGMENTS

My special thanks to my advisor Dr. Marcin Nowicki for granting me with this opportunity of pursuing my Master's degree. I am beyond thankful for his immense support and guidance throughout the research project.

I would also like to thank my committee members, Dr. Robert N. Trigiano, Dr. William E. Klingeman, and Dr. David R. Coyle, for their guidance and feedbacks throughout my project.

My sincere thanks to Dr. Denita Hadziabdic and my lab manager, Ms. Sarah Boggess for their unconditional support and suggestions. A big thanks to Dr. Bode Olukolu and his PhD student, Mr. Ryan Kuster for their help and support throughout the project. I am also thankful to all my lab members and friends in the Entomology and Plant Pathology (EPP) Department.

A huge thanks to all herbaria and arboreta (Arnold Arboretum, Bartlett Tree Research Laboratory, Carnegie Herbarium, Harvard University Herbaria, Illinois Natural History Herbarium, Morris Arboretum, Morton Arboretum, and U.S. National Arboretum), collaborators, and sample collectors for providing us with the research samples. I would like to thank the United States Department of Agriculture (USDA) and University of Tennessee Institute of Agriculture (UTIA) for their financial support with the project. At last, but not the least, I am very thankful to Student-Faculty Research Award (SFRA) and Graduate Student Senate (GSS) for research and travel support, respectively.

ABSTRACT

Pyrus calleryana Decne. (Callery pear) is a deciduous tree native to China, Japan, Korea, and Taiwan. In the early 1900s, this species was initially brought in the U.S. to assist with disease resistance to fire blight-causing bacteria *Erwinia amylovora* Burrill. via hybridization with *P. communis* L. Since then, many popular ornamental hybrid cultivars of *P. calleryana* have been developed. 'Bradford' is the most well-known *P. calleryana* cultivar in the U.S. Today, *P. calleryana* has become an extremely common invasive tree species that is naturalized across the eastern U.S. Knowledge on genetic diversity and population structure of *P. calleryana* is very limited. In this study, we estimated the genetic diversity and population structure of *P. calleryana* across its native and introduced ranges. As the species is highly invasive in its introduced range and presents evidence of variation in morphological traits, we hypothesized that high genetic diversity, and the presence of population structure would be evident among *P. calleryana* collection distributed across its native and introduced ranges. For the first study, we developed and used 18 microsatellite loci that were used to analyze 147 *Pyrus* source samples and to articulate the status of genetic diversity within Asian *P. calleryana* and specimens from U.S. cultivars. For the second study, we used 15 microsatellite loci to determine the genetic diversity and population structure of 180 wild *P. calleryana* individuals collected across six naturally occurring sites in Tennessee, Georgia, and South Carolina. In both studies, our data revealed high genetic diversity (H_e of 0.81 and H_e of 0.74, respectively), high genetic differentiation (D_{est} of 0.42 and D_{est} of 0.21, respectively), high gene flow (N_m of 1.79 and N_m of 3.94, respectively), and presence of population structure in *P. calleryana*. Both of our studies supported China as the source of origin for *P. calleryana* cultivars of the U.S. These findings indicate the highly invasive capability of *P. calleryana* across its introduced range. Based on

these results, we suggest taking management actions to control invasive *P. calleryana*. We suggest homeowners consider planting native flowering deciduous tree species, including serviceberry (*Amelanchier* spp.), *Cornus florida*, or *Cercis canadensis*, as alternatives for *P. calleryana*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Pyrus is an Old-World genus of about 25 tree species originating from Asia, Europe, and the mountainous regions of northern Africa. *Pyrus* belongs to the subfamily Maloideae (Pomoideae) in the family Rosaceae. *Pyrus calleryana* Decne., commonly known as Callery pear, is a species of pear tree native to eastern and southern China, Taiwan, Korea, and Japan (Rehder, 1915). *Pyrus calleryana* is a popular ornamental tree predominately planted in commercial and residential areas, popular for its early spring blooms and fall color (Culley & Hardiman, 2009). The *Pyrus calleryana* genome is diploid with a haploid chromosome number of 17, 2n = 34 (Cuizhi & Spongberg, 2003), and flow cytometry estimated genome size of 588 Mbp/1C (Dickson et al., 1992).

Pyrus calleryana is a deciduous tree with a conical to rounded crown (Cuizhi & Spongberg, 2003). Flowering starts as early as three years of age (Bell & Zimmerman, 1990). In early spring, the flowers are produced before the full expansion of leaves, and they are mostly clustered together in approximately 6 to 12 per inflorescence. The flowers are about 2 to 2.5 cm in diameter with an unpleasant odor. Each flower contains 5 sepals, 5 petals, two sets of 10 anthers, and 2 to 5 carpels with 2 ovules per locule (Cuizhi & Spongberg, 2003). Individuals of *P. calleryana* cultivars are self-incompatible. Leaves are simple, alternate, and oval, 4 to 8 cm long. Leaves are shiny dark green in summer and various vivid colors in autumn including maroon, burgundy, yellow, orange, and red. Bark is characterized by a rough texture and green/brown color. The branches may bear sharp spur shoots that can be longer than 8 cm. The lifespan of *P. calleryana* is approximately 25 to 30 years (Dirr, 1990). The fruits are hard and small, 1 to 1.5 cm long. The fruits are green/brown, spherical to slightly oblong, each with 1 or 2 seeds (Vincent, 2005). The fruits serve as a secondary food source for birds resulting in seed dispersal in distant areas (Reichard & Hamilton, 1997). In the native ranges, *P. calleryana* trees

are mostly distributed in warm and humid regions (Liu et al., 2012). Furthermore, trees often remain quite small and are often widely scattered in nature (Culley & Hardiman, 2007).

In the early 1900s, Pacific Northwestern pear orchards were being decimated and wiped out by fire blight (caused by the bacterium *Erwinia amylovora* Burrill) generating great losses of the annual crop. Professor Frank C. Reimer from Southern Oregon Experiment Station found that *P. calleryana* could help combat the fire blight problem in *P. communis* using various the grafting technique. In 1916, Frank Meyer (USDA) collected and sent *P. calleryana* seeds in bulk to the U.S. Meyer died in China in 1918 but further collections of *P. calleryana* seeds were imported in the U.S. even after his death (Culley & Hardiman, 2007). Large numbers of *P. calleryana* seeds were planted at the Southern Oregon Experiment Station (Medford, OR) and the USDA Plant Introduction Station (Glenn Dale, MD). The initial research mainly focused on *E. amylovora* resistance and scion-rootstock compatibility (Whitehouse et al., 1963). Eventually, *P. calleryana* was used as a common rootstock for several cultivated *Pyrus* species. The hardiness of *P. calleryana* lent to the development and release of several intraspecific hybrid cultivars making it popular among local gardeners and landscapers (Culley & Hardiman, 2007). The cultivar 'Bradford' was developed as an ornamental street tree by grafting *P. calleryana* seedlings and a specific, vigorous, thornless selection tree at the USDA station (Whitehouse et al., 1963). By 1962, the tree was commercially released, and it ranked as one of the most widely planted avenue trees in U.S. urban areas (Culley & Hardiman, 2007). It is still the most widely planted and the most known cultivar throughout the U.S. (Swearingen et al., 2002). But, by the early 1980s, older 'Bradford' trees tended to break during windstorms and heavy snow. To overcome this problem, several other cultivars were developed within the next few decades as improved replacements for older 'Bradford' trees (Culley & Hardiman, 2007). Some of the

cultivars of *P. calleryana* developed for desirable ornamental characteristics include 'Aristocrat', 'Autumn Blaze', 'Capital', 'Chanticleer', 'Princess', 'Rancho', 'Redspire', 'Trinity', and 'Whitehouse' (Santamour & McArdle, 1983). With the increasing popularity of *P. calleryana*, more cultivars were developed and released by many nurseries.

Early indications of problems began to arise with the increasing popularity of 'Bradford' and other Callery pear cultivars. It is considered that *P. calleryana* escaped cultivation as early as 1964 in Arkansas and 1965 in Maryland (Vincent, 2005). But the species became widely noticed in the natural areas of the southern U.S. only by the 1990s. By the late 1990s, non-cultivated pears were abundant along the mid-Atlantic coast. By the early 2000s, several undetected young feral pear seedlings were growing among the roadside vegetation of the eastern and the southern U.S. Eventually, feral pears were noticed throughout the U.S. in marginal and disturbed areas with high light including but not limited to park boundaries, wetlands, and forests (Culley & Hardiman, 2007; Vincent, 2005). It is quite surprising to see the transformation of a popular landscaping tree into an invasive one within a mere decade (Culley, 2017).

Despite the highly invasive status of *P. calleryana* in many of the U.S. states, surprisingly little is known about the genetic diversity and spatial structure of the species. Hence, in this research project we assessed the molecular genetic diversity and spatial distribution of the species across its native and introduced ranges using microsatellite loci at two levels: i) samples of *P. calleryana* from the Asian native range and the U.S. cultivars were collected from herbaria and arboreta to evaluate the genetic diversity of the species and the origin of *P. calleryana* in the U.S. and ii) the non-cultivated *P. calleryana* samples were collected from three southeastern U.S. states to assess the fine-scale genetic diversity of the species across a small geographical area. Invasive tree species such as *Albizia lebbeck*, *Pueraria lobata*, etc. usually maintain high genetic

diversity within their populations (Hamrick et al., 1992; Hamrick & Godt, 1996; Dunphy & Hamrick, 2005; Pappert et al., 2000). Based on this information, we hypothesized that high genetic diversity would be present among collections of *P. calleryana* that include samples from across its Asian native range and introduced range in the U.S.

CHAPTER 1

MICROSATELLITE LOCI REVEAL GENETIC DIVERSITY OF ASIAN CALLERY PEAR (PYRUS CALLERYANA) IN THE SPECIES NATIVE RANGE AND IN THE NORTH AMERICAN CULTIVARS

*Chapter 1 has been published as a research article in the journal named "Life" under

https://doi.org/10.3390/life11060531

Abstract

Pyrus calleryana Decne. (Callery pear) includes cultivars that that are popular ornamental trees planted into commercial and residential landscapes in the U.S. In the last few decades, this species has increasingly naturalized across portions of the eastern and southern U.S. But the mechanisms behind this plant's spread are not well understood. The genetic relationship of present-day *P. calleryana* trees with their Asian *P. calleryana* forebears (native trees from China, Japan, and Korea) and the original specimens of United States cultivars are unknown. We developed and used 18 microsatellite markers to analyze 147 *Pyrus* source samples and to articulate the status of genetic diversity within Asian *P. calleryana* and specimens of cultivars. We hypothesized that Asian *P. calleryana* specimens and cultivars of *P. calleryana* would be genetically diverse and would show genetic relatedness. Our data revealed high genetic diversity, high gene flow, and presence of population structure in *P. calleryana*, potentially relating to the highly invasive capability of this species. Strong evidence for genetic relatedness between Asian *P. calleryana* specimens and cultivars of *P. calleryana* was also demonstrated. Our data suggest the source for *P. calleryana* that have become naturalized in United States was China rather than Korea or Japan. These results will help understand the genetic complexity of invasive *P. calleryana* individuals and may be used when developing management for escaped *P. calleryana* populations.

Introduction

Pyrus calleryana Decne., Callery pear, is a popular ornamental tree well-known for its early spring flowers, robust growth, and fall color display (Culley & Hardiman, 2009). 'Bradford' is the most widely planted and most well-known Callery pear cultivar in the U.S. (Culley, 2017; Swearingen et al., 2002). Flowers from the cultivars of *P. calleryana* are self-incompatible and cultivars are routinely vegetatively propagated by grafting and cuttings. The trees can produce

viable seeds, however, when flowers are cross-pollinated by other compatible *Pyrus* species or other *P. calleryana* cultivars (Culley & Hardiman, 2007). When this occurs, the fertile seeds can germinate and establish when dispersed in favorable environments (Swearingen et al., 2002). In addition, *P. calleryana* adapts well to soils of variable pH (acidic to alkaline soil) and trees are tolerant of drought (Culley & Hardiman, 2009).

Naturalized *Pyrus calleryana* trees can be found in 33 states (EDDMapS, 2021) and due to expanding populations of naturalized trees, *P. calleryana* has been listed, or watch-listed, as an invasive species in many U.S. states (Culley & Hardiman, 2007; Randall, 2002; Swearingen et al., 2002). Several traits, including gametophytic self-incompatibility, pathogen and herbivore resistance, and tolerance of various abiotic stresses have contributed to the spread and persistence of *P. calleryana* in a variety of environments (Culley & Hardiman, 2007). Although the underlying mechanisms and processes that have enabled the broad spread and invasiveness of the species are not well understood, intraspecific hybridization among the genetically distinct cultivars and interspecific hybridization with the other escaped *Pyrus* species could be among the possible reasons behind expansion of *P. calleryana* populations (Culley & Hardiman, 2009; Vincent, 2005). Additionally, several insects promote cross-pollination (Culley, 2017). The fruits on fertile naturalized trees become secondary foods for birds and other vertebrates that then disperse seeds into distant areas and contribute to the spread of the species (Reichard & Hamilton, 1997). *Pyrus calleryana* can potentially become one of the most problematic invasive species in the U.S. (Warrix & Marshall, 2018). Hence, there is need for understanding genetic contributions to invasiveness as a means to develop effective management plans and strategies that can mitigate or limit the spread of wild *P. calleryana* populations. To address this knowledge gap, we need to know more about the biology and genetics of this invasive species.

Genetic diversity and population structure of any species are important factors for determining the long-term survival of the species and their adaptation to environmental changes (Taberlet et al., 1998). The population genetic profile of a species can inform about the origins of sub-populations (Bossdorf et al., 2005; Goolsby et al., 2006), which provides greater understanding about the introduction history of the tree and its cultivars. Such knowledge can be used to formulate an effective management plan. Assessment of the existing genetic variation of any species requires widespread and intensive sampling from both native and introduced ranges (Schierenbeck & Ainouche, 2006). The assessment of the molecular population genetics preferably using co-dominant markers is necessary for understanding the dynamics of adaptation and the spread of any species (Purugganan, 2000). The use of simple sequence repeats (microsatellites; SSRs) in population genetics and phylogenetics provides a more reliable interpretation of genetic diversity than other arbitrary markers such as random amplified polymorphic DNAs (RAPDs) (Hadziabdic et al., 2012; Kamvar et al., 2014). SSRs, highly informative and multi-allele genetic markers, are experimentally reproducible and transferable among closely related species (Nowicki et al., 2019). The presence of multiple alleles at an SSR locus makes SSRs more informative than other molecular markers, including SNPs (Vieira et al., 2016).

Most of our knowledge on *P. calleryana* originates within its native range in Asia, where *P. calleryana* is an endangered species (Kato et al., 2013). Most of these studies have focused on identification and characterization of cultivars or *Pyrus* species using different types of DNA markers such as SSRs (Bao et al., 2007; Yamamoto et al., 2001; Yamamoto et al., 2002a; Yamamoto et al., 2002b), amplified fragment length polymorphisms (AFLPs) (Bao et al., 2008), RAPDs (Yuanwen et al., 2001), restriction fragment length polymorphisms (RFLPs) (Iketani et

al., 1998), and chloroplast DNA (cpDNA) (Kimura et al., 2003). Few studies, however, have investigated the genetic diversity and population structure of *P. calleryana*. Nuclear SSRs (nSSRs) and cpDNA were used in a study of genetic diversity of *P. calleryana* in Zhejiang province of China (Liu et al., 2012), which revealed the geographic distance as the major factor in shaping the population structure of *P. calleryana*. Additionally, an examination of the genetic diversity of *P. calleryana* var. *dimorphophylla* in the Tokai district of central Japan (Kato et al., 2013) found complex genetic structure of the species, probably originating from artificial propagation and introgression with other *Pyrus* species. There are limited studies investigating the genetic diversity and population structure of the invasive U.S. *P. calleryana* cultivars, and this issue urgently needs further investigation (Culley & Hardiman, 2009; Lalk et al., 2021). We assessed the genetic diversity and population structure of *P. calleryana* from both native (*i.e.*, Asian) and introduced ranges (the U.S.-introduced cultivars). We hypothesized that there would be high genetic diversity present within Asian *P. calleryana* populations, and that they would be genetically related to early U.S.-naturalized cultivars of *P. calleryana* in terms of genetic distance and their genetic clustering. Specifically, we aimed to (a) develop novel microsatellite markers for *P. calleryana*, (b) test the cross-species amplification of the developed genomic short sequence repeats (gSSRs) to other *Pyrus* and *Malus* species, and (c) use the most informative of these markers to evaluate the genetic diversity and genetic relatedness among Asian *P. calleryana* specimens and early U.S.-naturalized cultivars of *P. calleryana*.

Materials and Methods

Sample Collection

Leaf and flower bud samples of pears (*Pyrus*) and the original U.S. cultivar selections of *P. calleryana* were obtained from 10 herbaria and arboreta in the U.S. (Fig. 1.1). In total, 147

samples (80 samples of *P. calleryana* and 67 samples of other *Pyrus* and *Malus* species) were received. The geographical coordinates, country of origin, and year of collection were recorded for each sample and whenever possible for historical specimens (Kim et al., 2010). The collection was reduced to 90 specimens due to the unreliable/inconsistent amplification of samples. Out of those 90 samples, 57 were *P. calleryana* specimens (36 specimens of Asian *P. calleryana* and 21 samples of 7 U.S. commercial cultivars of *P. calleryana*) and 33 samples represented 14 different species of *Pyrus* and 2 samples of *Malus rockii* Rehder. These 33 *Pyrus* and *Malus* species samples were used to evaluate the potential for cross-amplification among gSSRs.

DNA Extraction

Approximately 100 mg of dried leaf samples or fresh flower buds were taken from each specimen and homogenized using a Bead mill 24 (Fisher Scientific, Pittsburgh, Pennsylvania, U.S.). Samples were homogenized four times and kept frozen in liquid nitrogen for at least 5 min between each homogenization step in order to improve the tissue homogenization. The gDNA was extracted using EZNA DNA DS Mini Kit (Omega Bio-Tek, Norcross, Georgia, U.S.) protocol. Nanodrop (Thermo Fisher Scientific, Wilmington, Delaware, U.S.) was used to evaluate the purity and measure the concentration of the isolated gDNA. Re-extraction was attempted following the CTAB protocol (Doyle & Doyle, 1987) for those samples that did not produce good quality and quantity of gDNA using EZNA DNA DS Mini Kit. gDNAs was re-extracted wherever enough of plant sample material was available.

Microsatellite Primers and Genotyping Conditions

Genomic SSRs (gSSRs) were developed by using closely related pear (*Pyrus* × *bretschneideri* Rehder) whole genome sequence data (GenBank number: JH994112)

(Wu et al., 2013). The development of gSSRs involved several steps including: (i) genome assembly using MaSuRCA (Zimin et al., 2013); and (ii) SSR mining using GMATA (Wang & Wang, 2016). The gSSRs of interest with a motif length of 2 to 6 bp and a minimum of 5 repeats were retained. Finally, 18 to 22 bp gSSR primers were designed using Primer3 (Untergasser et al., 2012) with 45 to 55% GC content, 58 to 62 °C melting temperature, and 100 to 500 bp of expected product sizes. The developed gSSRs were used for this *P. calleryana* study (IDT, Coralville, Iowa, U.S.). Genomic locations of the gSSRs used in this study were analyzed in six available genomes of related *Prunus* spp. using BLAST (Altschul et al., 1990) with default settings, owing to their relatively higher quality (contiguity, N score, and annotation) than the genome used for their development. Genome with most gSSRs mapping to it was used to visualize their locations. Additionally, for cross-species amplification analysis, gDNA samples of the other *Pyrus* species that failed to amplify with the gSSRs were further tested with Internal transcribed spacer (ITS) primers (White et al., 1990) and ribosomal protein S16 (*rps*16/cpDNA03) primers (Nowicki et al., 2018). Both ITS (nuclear) and *rps*16 (plastidic) primers were used in our study for more effective DNA quality controls.

Polymerase chain reaction (PCR) amplifications were performed in a 10 µl reaction mixture consisting of 4 ng gDNA, 1 µM final concentration of each primer, 5 µl of 2 × GoTaq® DNA Polymerase (Promega, Madison, Wisconsin, U.S.), and 0.5 µl of dimethyl sulfoxide (DMSO). In order to validate the data, the gDNA extracted from the *P. calleryana* herbarium specimen (Arnold Arboretum, catalog number: 156119) was used as a positive control, and sterile distilled water as a negative control, for every 18 gSSR tested. PCR amplification was performed using the touch-down protocol to ensure the specificity of the amplified fragments (Korbie & Mattick, 2008) using the following thermal profile: initial denaturation at 94 ºC for 3 min., followed by

10 cycles of denaturation at 94 ºC for 30 s, annealing at 65 ºC for 30 s with a touch-down of 0.7 ºC/cycle and an extension at 72 ºC for 30 s then, followed by 30 cycles of denaturation at 94 ºC for 30 s, annealing at 58 ºC for 30 s and an extension at 72 ºC for 30 s, with a final extension of 72 ºC for 4 min.

QIAxcel Capillary Electrophoresis System (QIAGEN, Germantown, Maryland, U.S.) was used to visualize the *P. calleryana* PCR products and determine allelic sizes using a 15/600 bp alignment marker and 25 to 500 bp DNA size marker. Allele sizing was performed for each of the 147 gDNA samples against each of the 18 gSSR markers. The PCR reaction was repeated twice for the samples that did not amplify, which were then declared missing data if still failed to amplify. Samples with missing data in more than 9 of 18 loci were discarded from the analyses resulting in the final dataset of 57 *P. calleryana* samples.

For cross-species amplification test, PCR amplification in the gDNA samples of the other *Pyrus* and *Malus* species was also performed using this PCR reaction mixture composition and following the touch-down protocols detailed above. In addition, PCR amplification for the samples that failed to amplify for each gSSR was further attempted using ITS (White et al., 1990) and *rps*16 primers (Nowicki et al., 2018) using a 10 µl reaction mixture: 10 ng gDNA, 5 µM final concentration of each primer (ITS and *rps*16, respectively), and 5 µl of AccuStart II PCR SuperMix (Quantabio, Beverly, Massachusetts, U.S.). *Pyrus calleryana* (Carnegie Herbarium, catalog number: 396078) DNA was used as a positive control and sterile distilled water was used as a negative control.

The optimized thermal profile for PCR with ITS1 and ITS4 primers included an initial denaturation at 94 °C for 2 min, 40 cycles of 95 °C for 30 s, 60 °C for 1 min, 72 °C for 90 s, and the final extension at 72 °C for 7 min. The thermal cycle for touchdown PCR involving *rps*16

primers included the following settings: 95 °C for 4 min, 95 °C for 20 sec, 58 °C for 30 sec (-0.3 °C/cycle), (72 °C for 90 sec) for 10 cycles; (95 °C for 20 sec, 55 °C for 30 sec, 72 °C for 90 sec) for 35 times, and 72 °C for 5 min. The products of the PCR reactions were electrophoresed using 2% w/v agarose gels with ethidium bromide stain at 100V/cm^2 for 1 h, visualized under UV light using UVP GelStudio PLUS (Analytikjena, Upland, California, U.S.) and documented using VisionWorks 8.22.18309.10577.

Data Analysis

The raw allele sizes were binned into statistically identical allelic classes using FlexiBin excel macro (Amos et al., 2007). The binned allelic data were used for further analyses. For each gSSRs, the *P. calleryana* dataset was transformed using the motif lengths to represent the number of repeats rather than the PCR allele sizes using PGDSpider (Lischer & Excoffier, 2012) version 2.1.1.5. Clone correction was completed using *poppr* (Kamvar et al., 2014) version 2.8.5 in RStudio version 1.2.5033 using R (Team, 2013) 3.6.2. There were no clones in the dataset (total n = 57). Based on their country of origin, the samples were divided into four population groups, denoted “China” (n = 20), “Japan” (n = 11), “Korea” (n = 5), and “U.S.” (n = 21).

Population Genetics of *P. calleryana*

Genetic Diversity

The genetic diversity indices across the 18 gSSRs and 4 population groups were calculated using R with packages *poppr* version 2.8.5 and *hierfstat* (Goudet, 2005) version 0.04-22, and SPAGeDi(O. Hardy & Vekemans, 2015) version 1.5. For each gSSR marker, the number of alleles amplified (N), observed heterozygosity (H_o), expected heterozygosity (H_e), Jost’s differentiation estimate (D_{est}), Stoddard and Taylor index (G), and allelic richness (A_r) were calculated using hierfstat. SPAGeDi was used to estimate the presence of private alleles (P_a), as

well as hierarchical fixation indices including inbreeding coefficients (F_{IS}, R_{IS}), allele fixation index (F_{ST}), and R_{ST} (F_{ST} analogue based on allele size) (Hardy & Vekemans, 2002; Slatkin, 1995). Gene flow ($N_m = ¼×[(1/F_{ST})-1]$) among populations was estimated using GenAlEx (Peakall & Smouse, 2006) version 6.5.

In order to assess the contribution of mutation rate to the population structure, SPAGeDi performed the permutation tests to determine the phylogenetic distance between individuals using 10,000 permutations among alleles within each locus and to determine the significance of inbreeding coefficients using 10,000 permutations among gene copies (Pons & Petit, 1996). For many samples, including many of the historical specimens, GPS coordinates were not available, shrinking the sample subset further, and consequently creating a possible bias in our study. The entire *P. calleryana* dataset (n = 57) was used to determine the phylogeographical signal within populations as R_{ST} does not depend on GPS coordinates, whereas only the individuals with known GPS coordinates of origin (total n = 28; “China”, “Japan”, “Korea” groups) were used to determine the phylogeographical signal among populations.

Analysis of Molecular Variance (AMOVA) was performed using the package *poppr* with 1,000 permutations in order to evaluate the molecular variance partitioning within and among the population groups. Linkage Disequilibrium (LD) of the used gSSRs was determined in *poppr*, with significance assessed using 1,000 permutations. The pairwise index of association ($\overline{r}_d$) was calculated using permutation approach to assess whether the loci were linked and to ensure that the observed pattern of LD is not due to a single pair of loci (Agapow & Burt, 2001).

Population Structure

Mantel Test for Isolation by Distance

Samples with known GPS coordinates (n = 28) were analyzed using Mantel and partial Mantel tests implemented in the package MASS (Ripley et al., 2013) version 7.3-50 using 1,000 permutations for the assessment of tests significance. These tests estimate isolation by distance (IBD) to determine the correlation between the genetic and the geographical distance matrices of the individuals. The Mantel test was also standardized using the year of sampling (partial Mantel test). Further, Mantel correlogram tests were performed to further examine the underlying correlative relationship using the packages *ade4* (Dray & Dufour, 2007) version 1.7-13 and *vegan* (Oksanen et al., 2013) version 2.5-3 using $\alpha = 0.05$.

Bayesian Clustering using STRUCTURE and DAPC

STRUCTURE (Pritchard et al., 2000) version 2.3.4 was used to analyze the population structure of the *P. calleryana* dataset utilizing a Bayesian clustering algorithm. Genetic clusters among the *P. calleryana* individuals were inferred using 30 independent Monte Carlo Markov Chains (MCMC) with 250,000 generations of burn-in period and 750,000 MCMC steps in the actual runs for each used number of clusters (K = 1 to 10). STRUCTURE results were then visualized with PopHelper (Francis, 2017) version 1.0.10 using the Evanno's method (Evanno et al., 2005). ObStruct (Gayevskiy et al., 2014) version 1.0 was used to determine the correlation of population structure of inferred ancestral profiles to that of predefined/sampled populations. The program uses the ad-hoc R^2 statistics whose values range from 0 (recent divergence of predefined populations or a lot of migration/admixture between populations) to 1 (strong diversification and/or population structure) (Gayevskiy et al., 2014). One-tailed t-tests were used

to accrue the significance of the differences when consecutively removing each one of the pre-defined populations or the inferred clusters, to assess this group's impact on R^2.

The population structure of the *P. calleryana* dataset was also analyzed using model-free multivariate clustering approach, Discriminant Analysis of Principal Components (DAPC) using the package *adegenet* (Jombart, 2008) version 2.1.1. The DAPC analysis was optimized and cross-checked utilizing 1,000 permutations over Principal Component Analysis (PCA) range from 2 to 109 (total number of alleles – 1). The analysis was further confirmed using a dendrogram of unrooted neighbor-joining tree of pairwise genetic distances (Nei, 1978) among *P. calleryana* inferred clusters. A separate DAPC analysis was also performed for U.S. cultivars only.

Results

gSSR Development and Selection

After gSSR mining with GMATA and execution of additional in-house scripts, a total of 115,838 SSRs were discovered from three *Pyrus* × *bretschneideri* draft genomes. Marker polymorphism was determined based on allelic variation within each genome or across the three genome assemblies. SSR mining and primer design resulted in 105,557 SSRs with acceptable primers and of these, 90,987 were dinucleotide, 10,913 were trinucleotide, and 2,891 were tetranucleotide repeats. Primers between 59 to 60 ºC with <1 ºC difference between forward and reverse primers melting temperatures resulted in the discovery of 15,269 single motifs monomorphic SSRs along with 306 single motifs polymorphic SSRs. Only single motifs polymorphic SSRs were considered for the further analysis. In the single motifs polymorphic SSRs collection, AG was the most frequent dinucleotide motif and AAG was the most frequent trinucleotide motif (Fig. 1.2a). For the study, 50 gSSRs were selected and tested using gDNA

samples from three locally escaped, naturalizing Callery pear trees (Third Creek Greenway, Knoxville, Tennessee, U.S.). Based on preliminary data and the resulting amplification robustness, polymorphic character, and agreement with the expected product sizes, 40 gSSRs were selected for further evaluations. Of these, based on the initial assessment, 18 gSSRs with high discriminating power for multi-locus genotypes of *P. calleryana* gDNA were included in the subsequent analyses. Genomic locations of those 18 gSSRs were analyzed using the seven available high-quality genomes of related *Prunus* spp. Among those, 17 gSSR mapped to *Prunus dulcis* (Mill.) D.A. Webb genome GCF_902201215.1 (Fig. 1.2b).

Cross-Amplification

Cross-amplification was performed using 18 gSSRs (Table 1.1). Within the genus *Pyrus*, all the gSSRs amplified in three *Pyrus* species: *P. communis* L., *P. longipes*/*cossonii* Rehder, and *P. pyrifolia* Burm. Likewise, the gSSRs cross-amplified at high rates in *P. pashia* L. (94%) and *P. amygdaliformis* Vill. (89%). The developed gSSRs performed well in *Malus rockii* Rehder (67%). Additionally, the gDNA samples that failed to amplify in *P. gharbiana*, *P. korshinskyi* Litv., *P. regelii,* and *P.* × *hybrid* ('Bartlett' × *P. salicifolia* Pall.) samples using our gSSRs were successfully amplified using both ITS and *rps*16 primers (Fig. 1.3).

Population Genetics of *P. calleryana*

Genetic Diversity

There were no clonal multi-locus genotypes (MLGs) present in the dataset. All 57 individual *P. calleryana* samples represented unique MLGs and their genotypic data were used for analyses. Throughout the dataset, there was about 10% missing data (Table 1.2). The locus and population with the highest missing data were PyC012 with 26%, and "China" with 15% missing data, respectively. The dataset deviated from Hardy-Weinberg equilibrium (HWE; Fig. 1.4),

which is a possible result from the relatively low sample number, from specimens that were sampled across a broad time range (approximately 1912 to 2019 AD), as well as from wide geographical origins.

Among the four population groups, the average Shannon-Wiener Index of MLG diversity (H) was 4.04, indicating high genetic diversity in the genotypic dataset (Table 1.2). The average effective number of alleles in the four population groups was 5, ranging from 2 in "Korea" to 6 in "China". This indicated high genetic diversity and variations of *P. calleryana* populations. Similarly, the overall A_r of 3.79 varied from 2.48 in "Korea" to 4.61 in "China". A total of 88 private alleles were found in the *P. calleryana* dataset, with "China" population having the most of private alleles (n = 37).

The gSSRs had high power in discriminating the MLGs requiring only 8 gSSRs to capture all the MLGs present (Fig. 1.5). The calculation of a small range of pairwise values of LD ($\overline{r}_d$ = 0 to 0.4, *P*-value = 0.113) indicates no linked loci and suggests distribution of gSSRs across the genome (Fig. 1.6).

Across the dataset, an average of about 12 alleles per locus (ranging from 5 to 20) were detected (Table 1.3). The mean allelic richness (A_r) calculated in the locus-wise manner was 4.70, ranging from 2.74 (PyC050) to 6.22 (PyC031), suggesting a high long-term adaptability and persistence potential of the *P. calleryana* populations. There was a moderate observed heterozygosity across 18 gSSRs overall (H_o = 0.34) ranging from 0.03 (PyC050) to 0.93 (PyC038). In addition, there was a high expected overall heterozygosity (H_e = 0.81) across all loci ranging from 0.41 (PyC050) to 0.92 (PyC031 and PyC017). The overall $H_o<H_e$ implied the presence of population structure within our *P. calleryana* collection. Furthermore, the dataset indicated high gene flow across all gSSRs with an overall value of 1.79.

For the assessment of phylogeographic signals within the *P. calleryana* dataset, SPAGeDi was implemented, using 10,000 permutations among alleles within each locus. The mean permuted R_{ST} over all loci was not statistically different from F_{ST} in accordance with the expectations of this test, and the observed R_{ST} was bigger than F_{ST} (Pobs>exp = 0) indicating the presence of phylogeographic signal within populations (data not shown). Furthermore, the slope test of pairwise R_{ST} was evaluated in both linear and logarithmic forms to assess the phylogeographic signal among populations. From this slope test, we found no evidence of phylogeographic signal among populations (Pobs>exp = 0.95). Only the samples with available GPS coordinates (n = 28) were used for the slope test, thus creating a possible bias in this analysis.

AMOVA was used to investigate the partitioning of the molecular variance. It suggested a low proportion of total molecular variance among populations (6.2%), with more than half of the proportion of molecular variance partitioned within individuals (63.5%) (Table 1.4). Significance of the test result ($P < 0.001$) indicated the existence of population structure within the genotyped collection of *P. calleryana*.

Population Structure

Mantel Test for Isolation by Distance

Several independent analyses were performed to determine the population structure in the *P. calleryana* collection. Isolation by distance analysis using the Mantel test yielded no evidence of correlation between genetic and geographic distances among the analyzed *P. calleryana* individuals (total n = 28; Mantel statistic (r) = 0.04, *P*-value = 0.30) (Fig. 1.7a). The partial Mantel test (standardized geographic distance matrix by the year of sample collection) did not change that result (Mantel statistic using year of sample collection (r') = 0.04, *P*-value = 0.33). Additionally, there was a non-linear relationship of the genetic and geographic distances across

the space (Fig. 1.7b). The amplitude of the Mantel's r scores in the correlogram was low (-0.1 to 0.05), indicating a low impact of spatial distancing on the population structure of *P. calleryana* dataset.

Bayesian Clustering using STRUCTURE and DAPC

Bayesian clustering using STRUCTURE indicated two genetically distinct clusters ($\Delta K = 2$) in the genotyped *P. calleryana* dataset (Fig. 1.8b). The admixture of both the inferred clusters was observed for all four population groups. Additionally, four genetically distinct clusters ($\Delta K = 4$) of the genotyped *P. calleryana* dataset (Fig. 1.8c) showed the admixture of all four inferred clusters in all population groups except "Korea". Negligible variability among 30 independent Markov chains of STRUCTURE was detected. The overall R^2 between predefined and inferred clusters of the *P. calleryana* collection (when K = 2) was 0.42 ± 0.07 suggesting moderate divergence among the predefined populations and STRUCTURE-derived genetic clusters within the dataset. The information on the contribution of sampled and inferred populations to the observed structure of *P. calleryana* was derived by changes to R^2 using iterative successive removal of the pre-defined populations and the inferred clusters using ObStruct (Table 1.5). Only minor changes in R^2 index value were evident when the predefined populations or the inferred clusters were removed sequentially. Removal of the "China" population caused decrease (*P*-value = 0.32) in R^2 indicating that this population contributed the most to the population structure from among those analyzed. Additionally, the removal of the "Japan" population caused increase (*P*-value = 0.20) in R^2 which suggests that this population was of mixed ancestries and contributed the least to the population structure. Furthermore, removal of the inferred clusters resulted in no major changes in R^2 indicating no major contribution of the inferred clusters to the structure within the data. In addition to this, the overall R^2 between

predefined and inferred clusters of the *P. calleryana* dataset (when K = 4) was 0.41 ± 0.02, again with no major changes when sequentially removing the pre-attributed or the inferred groups (Table 1.5).

Using DAPC, a multivariate analysis, the *P. calleryana* dataset showed the clustering different from STRUCTURE. DAPC indicated 4 clusters for the given *P. calleryana* dataset. The DAPC result indicated the possibility of the "China" population being ancestral to the bulk of the species with diverged "Japan" and "Korea" populations (Fig. 1.9). This was also supported by the genetic distance dendrogram and in congruence with the results of the Bayesian clustering. In addition, DAPC analysis of U.S. cultivars alone showed that the cultivars with the same name same from the same or different source institutions were not identical (Fig. 1.10).

Discussion

We investigated the genetic diversity of *P. calleryana* in the collection of original non-cultivated Asian specimens and early developed U.S. cultivars. Available historical records and our results support the hypothesis for high genetic diversity within Asian pear specimens and their genetic relatedness with the early developed U.S. cultivars. Sample collection from herbaria and arboreta demonstrate their great value for research studies such as ours, or when sampling in native environment is hindered. We found high levels of genetic diversity within *P. calleryana* populations supporting the fact that wild populations of forest tree species maintain high levels of genetic diversity (Hamrick & Godt, 1996). Our findings on *P. calleryana* diversity are supported by other studies of woody forest trees; for example, North America-native *Cercis canadensis* (Ony et al., 2020) and Asia-native *Cornus kousa* (Nowicki et al., 2020).

We developed and used 18 gSSRs discriminating individual MLGs of *P. calleryana*. These novel gSSRs also cross-amplified to conserved sequences of DNA extracted from several *Pyrus*

species. But these gSSRs did not perform well, with fewer gSSRs amplifying informative sequences from the more distantly related *Malus* species. Wide genetic variation among *Pyrus* species exists, and the failure of gSSRs to cross-amplify some *Pyrus* species could be due to speciation and changing genomic landscape (Liu et al., 2012). Those non-amplified *Pyrus* species are related to *Pyrus bretschneideri* and *Pyrus calleryana* but are highly admixed likely due to inter-species hybridization and genetic admixture common in *Pyrus* (Wu et al., 2018). Nevertheless, the overall cross-amplification potential of our gSSRs makes these novel markers useful for future studies of several *Pyrus* species.

The genetic diversity, allelic richness, and gene flow patterns of any plant species are related to their ecology and evolution(Caballero et al., 2010; Leimu et al., 2006; Sakai et al., 2001). In an outcrossing and widespread species, genetic diversity mainly exists within populations (Hamrick et al., 1992). Our study revealed a high level of genetic diversity and allelic richness, and the presence of population structure in *P. calleryana*. A similar high level of genetic diversity was found in *P. calleryana* using 14 nSSRs to study 77 individuals (Liu et al., 2012) and in *Ambrosia artemisiifolia* using 13 gSSRs and 13 expressed sequence tag SSRs (EST-SSRs) to study 321 individuals (Meyer et al., 2017). We found high genetic diversity for *P. calleryana* populations compared to other invasive species such as southwestern Puerto Rico's invasive tree *Albizia lebbeck* (Dunphy & Hamrick, 2005) or *Pueraria lobata,* which is invasive in the southeastern U.S. (Pappert et al., 2000).

We recorded high gene flow among *P. calleryana* populations across the species' native range. Irrespective of the geographical barriers, high gene flow could be a possible reason for this species' invasiveness because gene flow promotes evolution through the spread of new genes or mixture of genes throughout the range of species (Slatkin, 1987). The high gene flow rate we

observed is consistent with other invasive species (Dunphy & Hamrick, 2005; Gaskin et al., 2014) such as *A. lebbeck* and *Fallopia* spp. Such a high gene flow rate promotes exchange of genes among populations, especially in self-incompatible species by providing seeds for more population growth and colonization of new habitats (Dunphy & Hamrick, 2005). As an outcrossing species, *P. calleryana* has been able to maintain a population structure with high genetic diversity. The individuals of a particular cultivar have the same self-incompatibility genotype and cannot produce fruits. But, if the rootstock is allowed to sprout then that rootstock can cross with genetically different scion resulting in fruit set with mixed cultivar types (Culley & Hardiman, 2007). Also, our study found a low impact of spatial distance in the population structure of the *P. calleryana* dataset which could support the seeds or pollen as the far-reaching mechanism of dispersal.

Pyrus calleryana flowers are indiscriminately visited by various generalist pollinator species and the pollen is often carried among neighboring cultivars, resulting in intraspecific hybridization (Culley & Hardiman, 2009). Seeds can be dispersed over long distances as a result of frequent frugivorous animal activities (Liu et al., 2012). In a previous study, most of the *P. calleryana* seedlings were found with almost no genetic similarity to nearby mature trees, implying that those seedlings might have originated from foraging birds' defecation (Culley & Hardiman, 2007). In addition, there could be an intraspecific hybridization among the cultivars and an interspecific hybridization with the other escaped *Pyrus* species (Vincent, 2005), as implied by our cross-amplification data. These characteristics may help *P. calleryana* maintain high genetic diversity and high gene flow rate, aiding in the continued spread of the species by providing environment-specific genotypes needed to adapt to a varied environmental condition (Dunphy & Hamrick, 2005).

High genetic differentiation found in our *P. calleryana* collection suggested the existence of population structure. Pears are expected to undergo random mating, but an unexpected positive result for the inbreeding coefficient (F_{IS}) values was found in our study indicating the probability of alleles coming from a common ancestor. A similar positive F_{IS} result was found in a study in *P. calleryana* in China (Liu et al., 2012), where the species is considered threatened. One of the possible explanations for such positive F_{IS} result could be that the sampled individuals might have experienced some human interferences, such as selection, propagation, and intentional transportation resulting in the escape of cultivation in the U.S. Other explanations could be the insect pollination with limited pollen flow creating local population structure, or the selection of loci under negative selection by chance, where any mutation would be lethal. Studies testing this phenomenon in the escaped *P. calleryana* populations are underway.

Our study did not indicate geographic distance as the major factor contributing to the genetic structure of the populations. A non-significant correlation between genetic and geographic distance could be a result of the small size of our *P. calleryana* dataset collection. Our result is unlike the Mantel test result obtained for wild *P. calleryana* from China (Liu et al., 2012). In addition, our study partitioned 57 individuals from 4 population groups into 2 major genetic clusters. But, considering the biology of *P. calleryana*, DAPC result, unrooted neighbor-joining tree, and bias of STRUCTURE (Cullingham et al., 2020; Lombaert et al., 2018) towards k = 2, we assumed k = 4 as the best clustering for our *P. calleryana* dataset. Almost all *P. calleryana* individuals throughout all population groups showed extensive genetic admixture, in agreement with the detected high gene flow among populations. The individuals from “China” and “U.S.” population groups had relatively higher assignment probability to the major genetic clusters. Furthermore, genetic distances of “China” and “U.S.” populations placed them close to each

other. Thus, our results support the historical import records of *P. calleryana* from China to the U.S. (Culley & Hardiman, 2007) with the "China" population being ancestral to the "U.S." population.

Historical events suggest the development of *P. calleryana* cultivars using *P. calleryana* as a common rootstock. We expected at least some clones within the "U.S." *P. calleryana* individuals as such cultivars were multiplicated clonally. However, no clones within the "U.S." *P. calleryana* individuals were found, as each "U.S." *P. calleryana* individual was unique with no shared MLGs within this population. Our study also found no identical cultivars although named the same from the source institutions. In such case, 'Trinity' cultivar of one source is genetically different from 'Trinity' cultivar of another source. Such a great genetic composition and diversity with no clones in our "U.S." *Pyrus calleryana* individuals signifies the great invasive potential of the species, and alarms about the level of the previously noted cultivar mislabeling and mishandling (Culley & Hardiman, 2009; Santamour & Demuth, 1980). Our result is in stark contrast to those of others, who used 2 SSRs from *P. pyrifolia* and 7 from *Malus* × *domestica* on 14 *P. calleryana* cultivars (Culley & Hardiman, 2009). They detected clonal MLGs in agreement with the cultivar description, confirming genetic identity of 'Chanticleer', 'Cleveland Select', and 'Stonehill'- in accord with their history, and origin from the same source (Culley & Hardiman, 2009).

The high level of genetic diversity within the investigated *P. calleryana* collection suggests the high potential of the species for evolving resilience and ability to thrive in a variety of environmental conditions. In addition, the high gene flow among *P. calleryana* populations creates the complex genetic admixture, adding more challenge to their control. For the development of measures for improved control, it is necessary to study in detail the genetic

composition and diversity, and to infer the evolutionary potential of the species across the introduced ranges. Considering the ultimate goal of well-informed control measures, we are already taking the next steps in analyses of naturalized *P. calleryana*: the fine-scale study of *P. calleryana* in localized area of the U.S. and the broad-scale study of *P. calleryana* across the U.S. Thus, *P. calleryana* populations across the U.S. provide a great prospect for further research and study, to better understand the spread mechanism of this invasive species.

Appendix: Tables and Figures

Table 1. 1. Cross species amplification of the studied gSSRs

Species	PyC006	PyC008	PyC009	PyC012	PyC013	PyC014	PyC015	PyC017	PyC018	PyC020	PyC031	PyC032	PyC035	PyC038	PyC041	PyC042	PyC047	PyC050
P. calleryana Decne.	384-425*	273-347	417-465	394-430	177-214	176-218	193-274	389-453	439-480	275-304	294-356	415-434	363-389	158-179	272-289	332-350	356-375	370-386
P. amygdaliformis Vill.	420-421	308-318	439	-	184-209	-	198-202	387-388	472-507	278-294	310-342	419	341-367	158-174	273-275	332-333	378-385	376-384
P. pashia Linnaeus	420	301-312	441-450	402	188-207	192-196	204-240	397-402	441-462	263-290	-	423	369	158-180	274-284	337-338	357-363	372-373
P. communis L.	416-421	304-313	418-461	403-430	180-213	173-196	200-227	384-415	455-479	272-304	306-348	414-431	339-378	158-175	118-280	330-334	359-379	362-380
P. pyrifolia (Burm.) Nak.	421-423	305-311	464	413	201-208	187	212	400	461-473	276-287	309-343	416	368-376	161-176	277	332	366	375
M. rockii Rehder	423	-	-	-	176-185	169-180	201-211	-	-	265-285	282-313	-	365	178-179	200	334	360	381-391
P. longipes/*cossonii*	415-429	298-311	433-439	428-430	186-211	168	196-199	385-388	463-475	279-303	340-348	422	338-362	159-174	274-275	334	375-384	380-387
P. gharbiana	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P. korshinskyi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P. regelii	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bartlett × P. *salicifolia*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Species: name of the *Pyrus* species used for cross-amplification study; Remaining each column in the table represent the respective gSSRs name and the allele sizes for each respective *Pyrus* species. *: Allele sizes of the species in base pair (bp); -: species with no amplification in the given locus.

Table 1. 2. Genetic diversity indices of *P. calleryana* dataset based on population groups using 18 microsatellite loci

Population	N	% Missing	# Alleles	NAe	H	G	F_{IS}	A_r	λ	H_e	H_o	$\overline{r}_d$	P_a
China	20	15.28	8.06	6.40	3.00	20	0.636***	4.61	0.95	0.80	0.30	0.10^{ns}	37
Japan	11	5.56	5.72	4.74	2.40	11	0.511***	4.02	0.91	0.74	0.37	0.16***	19
Korea	5	7.78	2.56	2.41	1.61	5	0.569***	2.48	0.80	0.51	0.24	0.03^{ns}	10
U.S.	21	7.94	7.22	4.62	3.04	21	0.397***	4.06	0.95	0.72	0.44	0.11^{ns}	22
Total	57	10.04	11.78	4.54	4.04	57	0.552***	3.79	0.98	0.80	0.36	0.12^{ns}	88

N: Number of individual samples used; % Missing: percent of genotypes missing; # Alleles: Number of alleles present in the given population group (calculated using SPAGeDi); NAe: Effective number of alleles present in the given population group (Nielsen et al., 2003); H: Shannon-Wiener Index of MLG diversity (Shannon, 2001); G: Stoddart and Taylor's Index of MLG diversity (Stoddart & Taylor, 1988); F_{IS}: Individual inbreeding coefficient (*** = $P < 0.0001$); A_r: Allelic richness giving the expected number of alleles among 8 gene copies; λ: Simpson's Index (Simpson, 1949); H_e: Nei's unbiased gene diversity (Nei, 1978); H_o: Observed heterozygosity; $\overline{r}_d$: Standardized index of association taking into account for the number of loci sampled (Kamvar et al., 2014); P_a: Number of private alleles in each population group. Significance was assessed by using 10,000 randomizations of gene copies among all individuals, *** = $P < 0.0001$ and ^{ns} = $P > 0.05$ at 10,000 permutations.

Table 1. 3. Population genetics indices of *P. calleryana* dataset based on microsatellite loci

gSSR Locus	Forward and Reverse Primers (5'-3')	Motif	# Alleles	% Missing	A_r	H_o	H_e	D_{est}	G	F_{ST}	F_{IS}	R_{ST}	R_{IS}	N_m
PyC006	F: ACAGGGTGAACTCGATGAGC R: GTGTACGGAATTGCTGCTGC	TA	10	12.28	3.89	0.13	0.72	0.2	3.5	0.08^{ns}	0.74****	0.18*	-0.08^{ns}	5.6
PyC008	F: AAACCTGTGCGGAAACATCC R: GAAAGCAAGGCTCATCCACG	CT	16	3.51	4.86	0.36	0.87	0.82	7.4	0.16****	0.38****	0.47****	0.69***	0.9
PyC009	F: ACGTGGTTTAGCATCCTGGG R: CTAAACGGAATAAGGCGGCG	CT	15	12.28	5.39	0.25	0.85	-0.14	6.1	-0.02^{ns}	0.72****	0.18*	0.70****	NA
PyC012	F: CCTTGACACGCTTCACATGC R: GTTCTTGCGGTTTAGGCTGC	CT	13	26.32	4.73	0.08	0.89	0.78	8.3	0.14***	0.85****	0.17ns	0.85****	1.1
PyC013	F: GACGTGTTTATGCACCACCG R: AACCCAGCTAGCTGACAACC	GA	14	0	4.34	0.43	0.84	0.5	5.9	0.11***	0.46****	0.11*	0.36*	1.2
PyC014	F: CAAATGACTCGCAGCAGACG R: GTTTACCGATACATGCCGCG	TC	14	7.02	5.45	0.21	0.86	0.72	6.8	0.16****	0.70****	0.50****	0.67****	0.9
PyC015	F: ACATGTGACAGAGGCATCCG R: TTTGTGGATTCCCCTCCTCG	TC	18	5.26	5.4	0.12	0.87	0.85	7	0.20****	0.85****	0.74****	0.87****	0.7
PyC017	F: AATGAAGCTGCAGCAATGGC R: GGCTGTGTGCATGTAAAGCG	TC	20	21.05	6.48	0.6	0.92	0.46	11	0.06**	0.19**	0.18*	0.37*	3.3
PyC018	F: TGTCTGTGGAAACGAGAGGC R: GAGCGTTGCCGTTTTAACCC	TC	16	8.77	5.89	0.65	0.91	0.51	10	0.05**	0.17**	-0.04^{ns}	0.09^{ns}	3.2
PyC020	F: ATTCTTGGGGTTGGTTGGGG R: CTTGCACTGAACGAGGTTGC	AG	13	0	4.89	0.39	0.8	0.2	4.8	0.03^{ns}	0.49****	0.07^{ns}	0.34*	5.5
PyC031	F: TTCAACCTGGTACGTCGTCG R: TCAGCTCAACCGAAAACAGC	TTC	20	8.77	6.22	0.84	0.92	0.57	12	0.04*	0.13*	0.03^{ns}	-0.38**	6.9
PyC032	F: TACAACGTCACACTGCTCCC R: TTCCCCTCCCTTTCATTGGC	AAT	6	19.3	4.24	0.09	0.74	0.58	3.8	0.22****	0.86****	0.37***	0.73****	0.9
PyC035	F: ATCTGATGGAGACGGCAACC R: CGCACCATGTAAGTTGCTCC	GGA	13	1.75	4.77	0.41	0.85	0.31	6.5	0.03^{ns}	0.49****	0.14*	0.17^{ns}	7.9
PyC038	F: TTCCTGGGTTGGAACTTGGG R: AGAATCGTCCTTGGAAGGCG	CAC	6	0	4.06	0.93	0.76	-0.04	4.1	-0.01^{ns}	0.23***	0^{ns}	-0.75****	NA
PyC041	F: GCTCATGCAATAGGACAAAGGC R: AAGGTGTACAGGAGACGTGC	TCT	8	14.04	4.08	0.29	0.81	0.04	5	0.01^{ns}	0.65****	0.08^{ns}	0.22****	7
PyC042	F: AACGGTCTGTGGAAAGCTCC R: TCAGCATCATTCCACACCCC	AGA	7	15.79	3.57	0.06	0.81	0.7	5.1	0.23****	0.89****	0.62****	0.82****	0.6
PyC047	F: ACTTGGACTTGGAACCCAGC R: GGACCAATGAACCCTTTGCG	AAAG	6	12.28	3.64	0.18	0.76	0.36	4	0.10*	0.61****	0.19*	0.75****	2.4
PyC050	F: TCGTCTGTCGGGTCAAATCG R: TGTACAGGTTAGTTGCCCGC	AGTG	5	12.28	2.74	0.03	0.41	0.07	1.7	0.10^{ns}	0.90****	0.09^{ns}	0.92****	2.7
Average			12.2	10.04	4.7	0.34	0.81	0.42	6.3	0.09****	0.52****	0.33****	0.31****	1.8

Alleles: Number of alleles detected; % Missing: % of samples that failed to amplify in the given locus; A_r: Allelic richness; H_o: Observed heterozygosity (Frequency of heterozygous individuals per locus averaged over the number of sampled loci); H_e: Expected heterozygosity (Nei's unbiased gene diversity; Nei 1978); Dest: Jost's differentiation estimate (Jost, 2008); G: Stoddard and Taylor index (Stoddard & Taylor, 1988); F_{ST}: a measure of sub-population genetic structure; F_{IS}: a measure of deviations from Hardy-Weinberg equilibrium in terms of heterozygote deficiency if <0 or homozygote excess if >0; R_{ST} and R_{IS}: Analogues of F_{ST} and F_{IS} based on allele sizes (Slatkin, 1995); N_m: Gene flow estimated as $N_m = ¼×[(1/F_{ST})-1]$. Significance of the dataset was assessed by 10,000 permutations using **** = P < 0.0001; *** = P < 0.001; ** = P < 0.01; * = P < 0.05; ns = P > 0.05. NA: Not applicable as a negative value do not represent any gene flow data.

Table 1. 4. Analysis of Molecular Variance of *P. calleryana* dataset

Source of variation	df	Sum of squares	Mean squares	Sigma	% Variance	Φ
Variations among populations	3	36.92	12.31	0.27**	6.23	0.37
Variations within populations	53	280.15	5.29	1.29**	30.25	0.32
Variations within individuals	57	154.34	2.71	2.71**	63.53	0.06
Total variations	113	471.41	4.17	4.26**	100	

Df: Degree of freedom given by size of sample - 1; Sum of squares: sum of squares of the deviations of all the observations from their mean; Mean squares: sample variance obtained by dividing the sum of squares by the respective df; Sigma: variance for each hierarchical level; % Variance: percent of the total variance for each hierarchical level; Φ: statistics calculated by the test; Significance was assessed by using 1,000 permutations of the dataset, ** = $P < 0.001$ at 1,000 permutations.

Table 1. 5. R^2 value using Obstruct for *P. calleryana* dataset

	When k = 2		When k = 4
Overall R^2 for the dataset	0.418±0.072***	Overall R^2 for the dataset	0.409±0.023***
R^2 without predefined population "Japan"	0.295±0.139^{ns}	R^2 without predefined population "Japan"	0.362±0.016^{ns}
R^2 without predefined population "Korea"	0.374±0.053^{ns}	R^2 without predefined population "Korea"	0.301±0.018^{ns}
R^2 without predefined population "U.S."	0.355±0.066^{ns}	R^2 without predefined population "U.S."	0.380±0.038^{ns}
R^2 without predefined population "China"	0.523±0.086^{ns}	R^2 without predefined population "China"	0.457±0.027^{ns}
R^2 without inferred population GREEN	0.418±0.072^{ns}	R^2 without inferred population YELLOW	0.403±0.070^{ns}
R^2 without inferred population ORANGE	0.418±0.072^{ns}	R^2 without inferred population GREEN	0.411±0.076^{ns}
		R^2 without inferred population ORANGE	0.407±0.051^{ns}
		R^2 without inferred population PURPLE	0.393±0.078^{ns}

***: significant at $P < 0.0001$; Inferred population colors (GREEN, ORANGE, YELLOW, and PURPLE) represent the clusters inferred with STRUCTURE for k = 2 or k = 4 across 30 independent Markov chains.

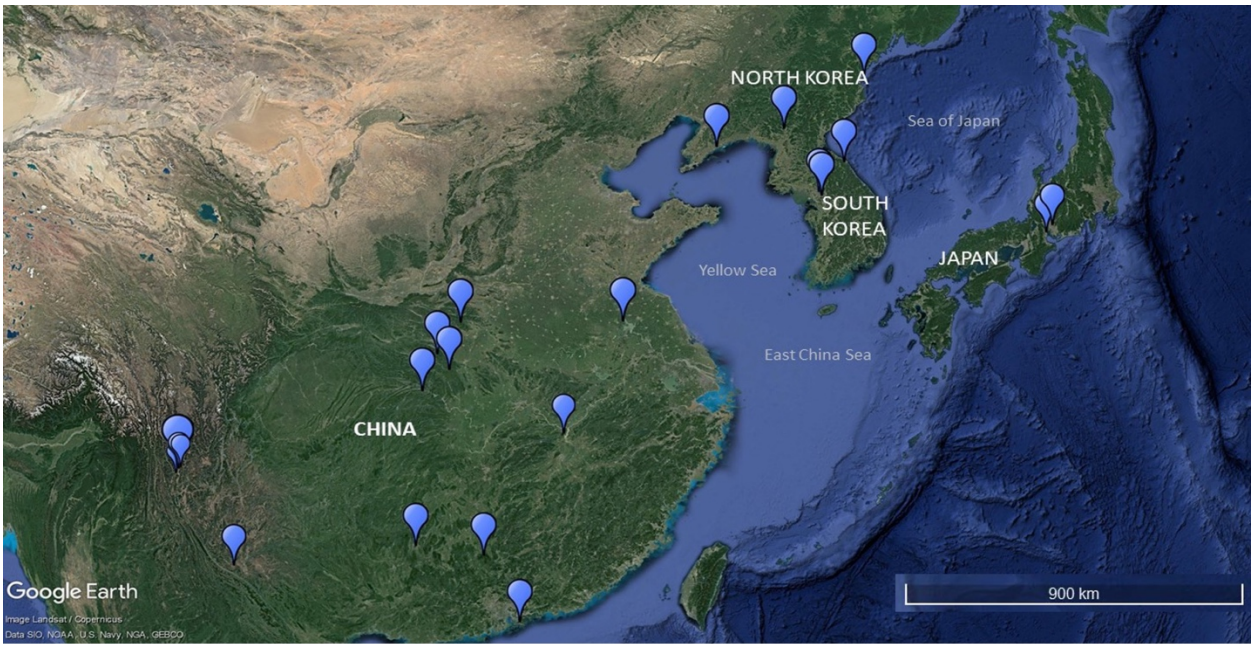

Figure 1. 1. Map showing collection sites of P. calleryana samples used for the study across East Asia. The blue marker points represent the sample collection location. The scale line indicates the ground-level distance of 900 km. The map was generated using Google Earth Pro version 7.3.

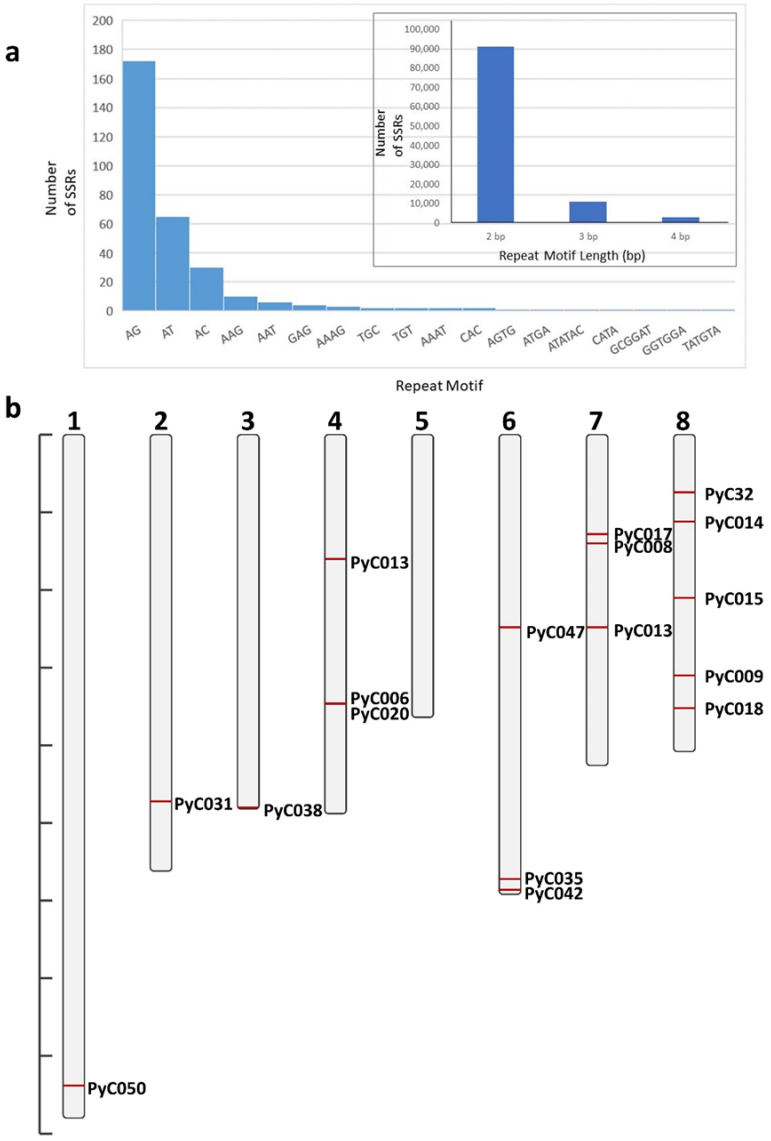

Figure 1. 2. The gSSRs used for the study. (a) Frequency of di-, tri-, tetra-, and hexa-nucleotide motif repeats discovered in 306 single motifs polymorphic gSSRs collection. The indicated motifs and their redundant iterations are grouped together. Insert: Frequency of di-, tri-, and tetra-nucleotide motif repeat gSSRs in the discovered 105,557 gSSRs collection; bp = base pair. (b) Locations of the gSRRs used in this study were investigated using BLAST and visualized on *Prunus dulcis* Mill. genome.

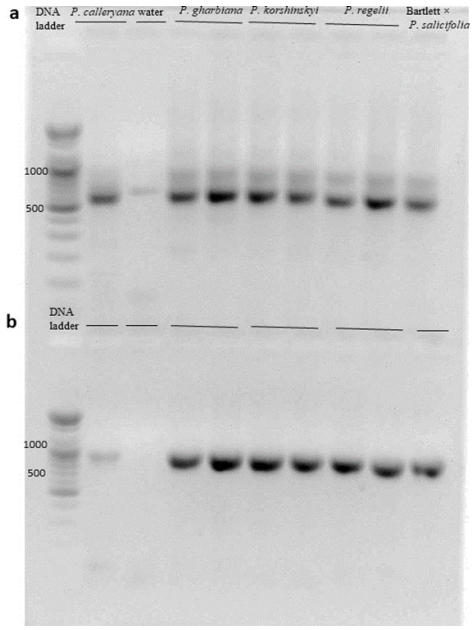

Figure 1. 3. Gel image of cross-species amplification using ITS and *rps*16. (**a**) The uppermost row in gel image represents the amplification done using ITS primers; (**b**) The lowermost row in gel image represents the amplification done using *rps*16 primers Nowicki et al., 2018). Order of samples in (a) and (b) is identical, and from left to right includes: DNA ladder (DNA molecular marker of 100 bp; BIONEER, Catalog No.:D-1030; Oakland, California, U.S.); positive control (*Pyrus communis*; PC_A_019); negative control (water); *Pyrus gharbiana* (20_PC_AO_29 and 20_PC_AO_33); *Pyrus korshinskyi* (20_PC_AO_14 and 20_PC_AO_15); *Pyrus regelii* (20_PC_AO_16 and 20_PC_AO_21); *Pyrus hybrid* ('Bartlett' × *Pyrus salicifolia*; 20_PC_AO_08). The expected size of PCR products for ITS was 678 bp and for *rps*16 was 911 bp.

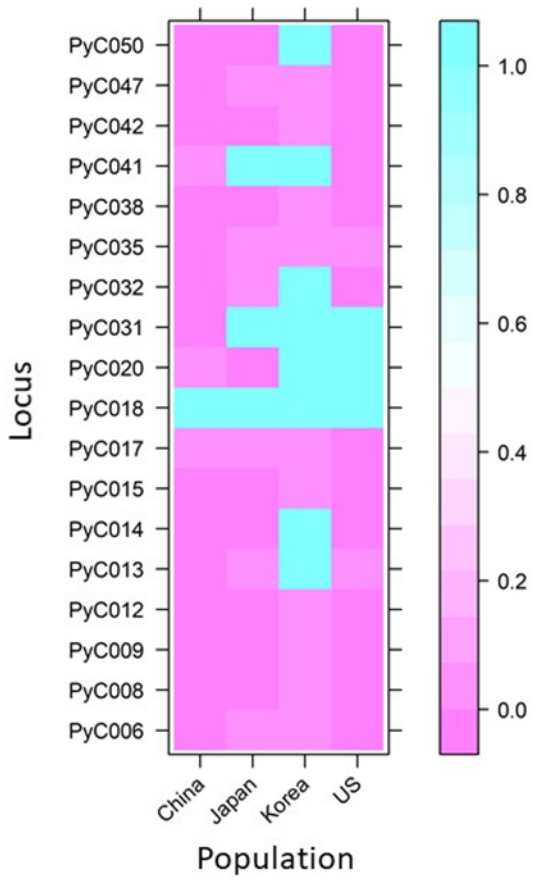

Figure 1. 4. Hardy-Weinberg Equilibrium (HWE) observed for samples included within the *P. calleryana* dataset. The rows represent loci and columns represent populations. The legend represents the probability of loci to follow HWE. The loci in pink are suspected of not being in HWE with $P \leq 0.05$.

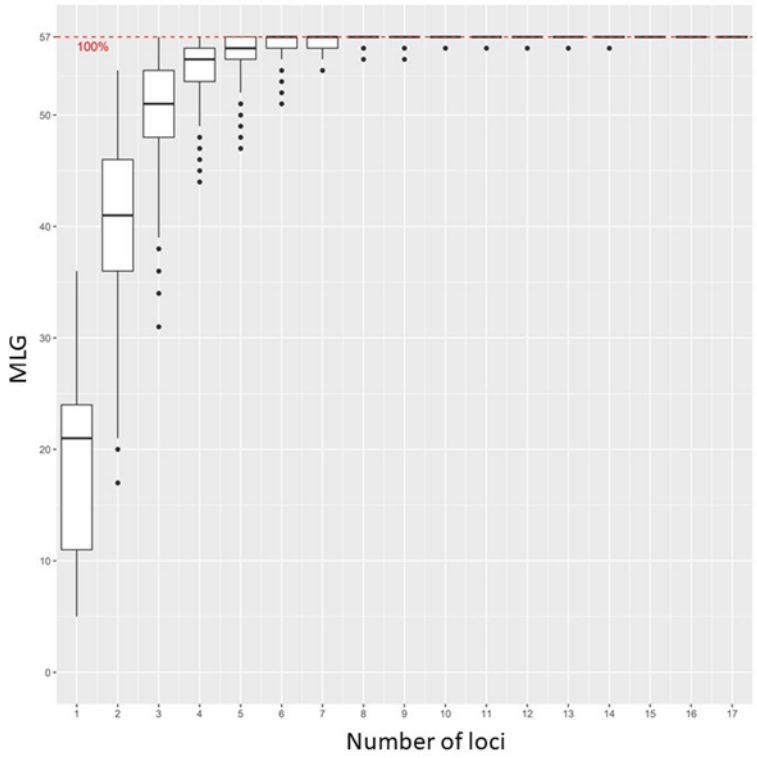

Figure 1. 5. Genotype Accumulation Curve (GAC) for samples included within the *P. calleryana* dataset. It represents the number of MLG detected (Y-axis) in relation to the number of loci (X-axis) used for genotyping.

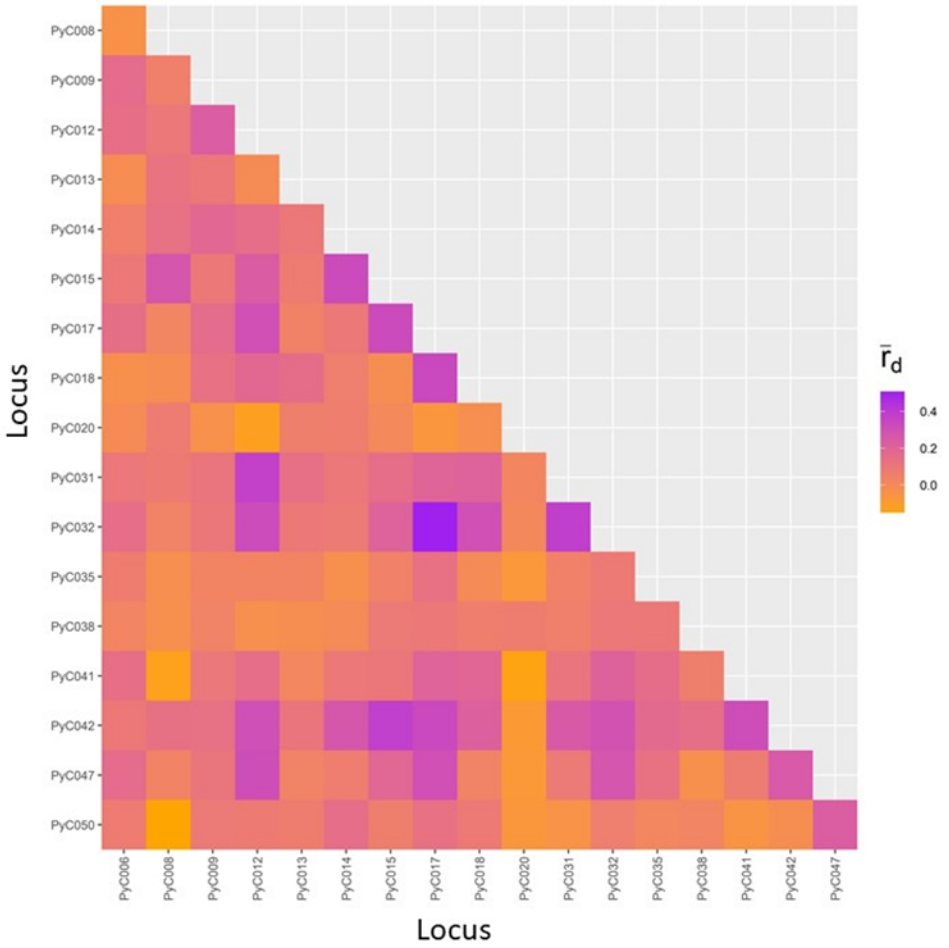

Figure 1. 6. Pairwise linkage disequilibrium (LD) among the 18 gSSRs for samples included within the *P. calleryana* dataset. Pairwise linkage disequilibrium (LD) among the 18 gSSRs for samples included within the *P. calleryana* dataset. LD is expressed as standardized index of association ($\bar{r}_d$). Legend values coincide with hues that explain the strength of the linkage calculated between each pair of the markers.

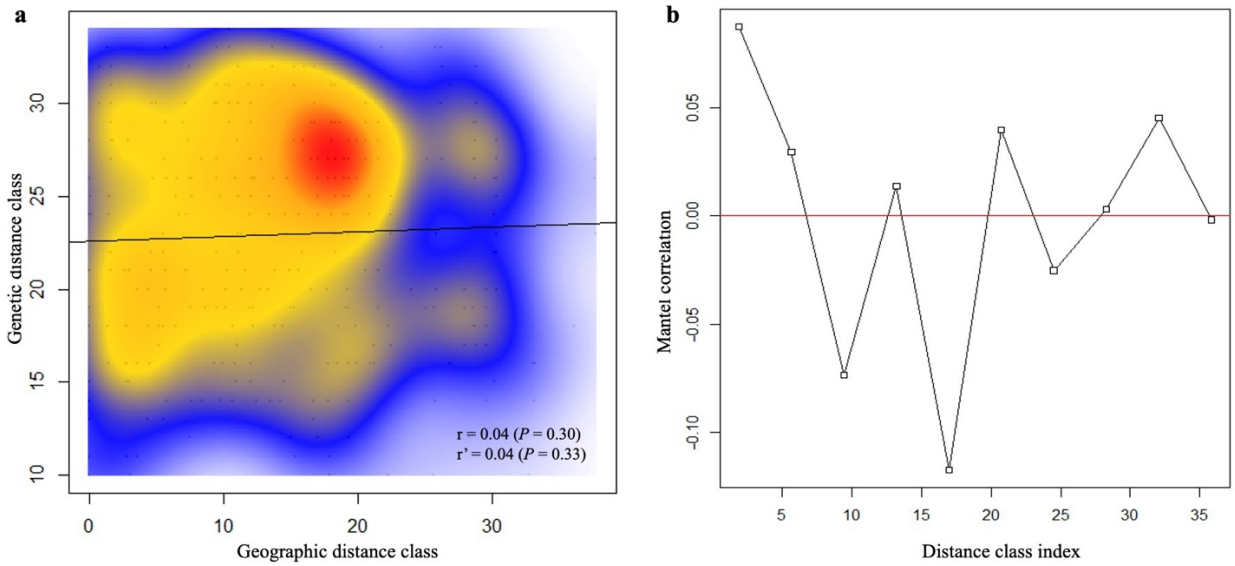

Figure 1. 7. Mantel test of *P. calleryana* dataset. Mantel test (a) with Mantel correlogram (b) for isolation-by-distance analyses of samples within the *Pyrus calleryana* dataset. The correlation between geographic and genetic distance for the dataset was determined using 1,000 permutations. Distance class indices (in 100s of km) indicates that the maximum linear distance between samples was 3,600 km. Significance (alpha = 0.05) is reported for the Mantel index (r) and the Mantel index standardized by the year of sampling (r').

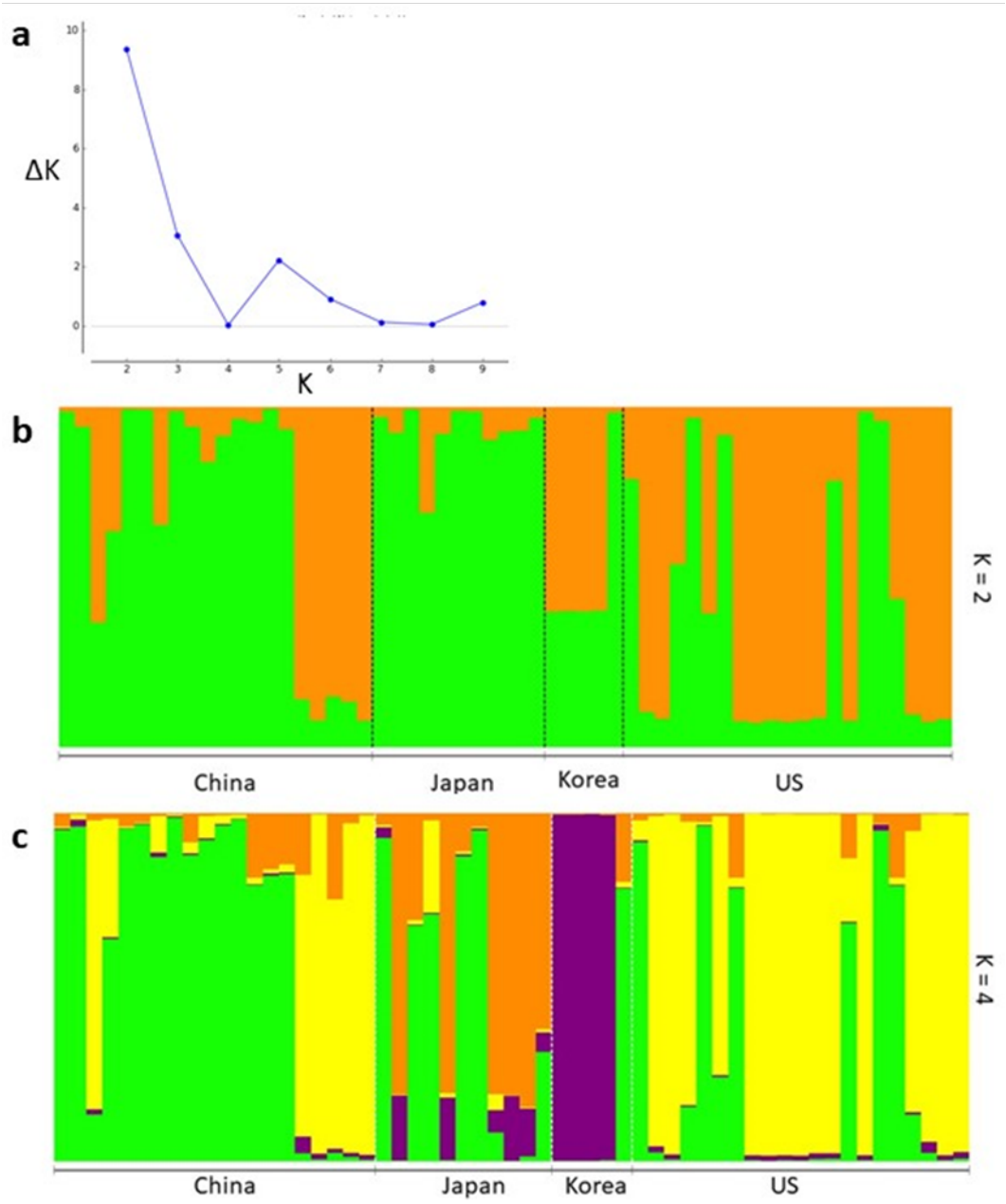

Figure 1. 8. STRUCTURE Bayesian clustering of *P. calleryana*. STRUCTURE Bayesian clustering analyzed with (a) the Evanno method and visualized using (b) 2 genetic clusters and (c) 4 genetic clusters. Each vertical bar represents an individual sample, and the bar color indicates the probability of an individual to get assigned to one of clusters identified.

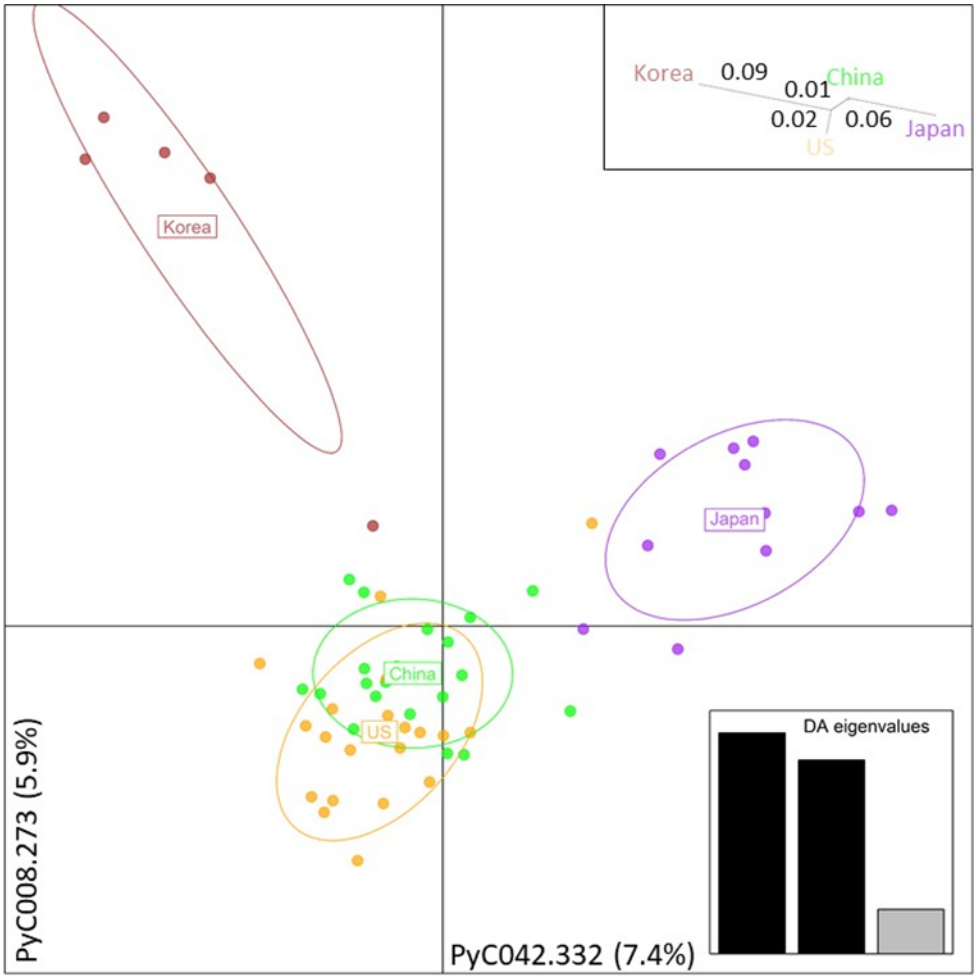

Figure 1. 9. DAPC of *P. calleryana* dataset. DAPC for determining the molecular variance partitioning projected using 15 Principal Components cross-checked and optimized with 1,000 permutations. Eigenvalues (Insert bar graph, bottom right) represent the factor by which eigenvectors are scaled, which expresses the spatial relationship among populations at different spatial scales. The two respective axes are indicated by the alleles explaining the most of variance within the sampled population. Insert genetic distance tree, top right: Unrooted neighbor-joining tree of pairwise genetic distances (F_{ST}) among the sampled *P. calleryana* specimens.

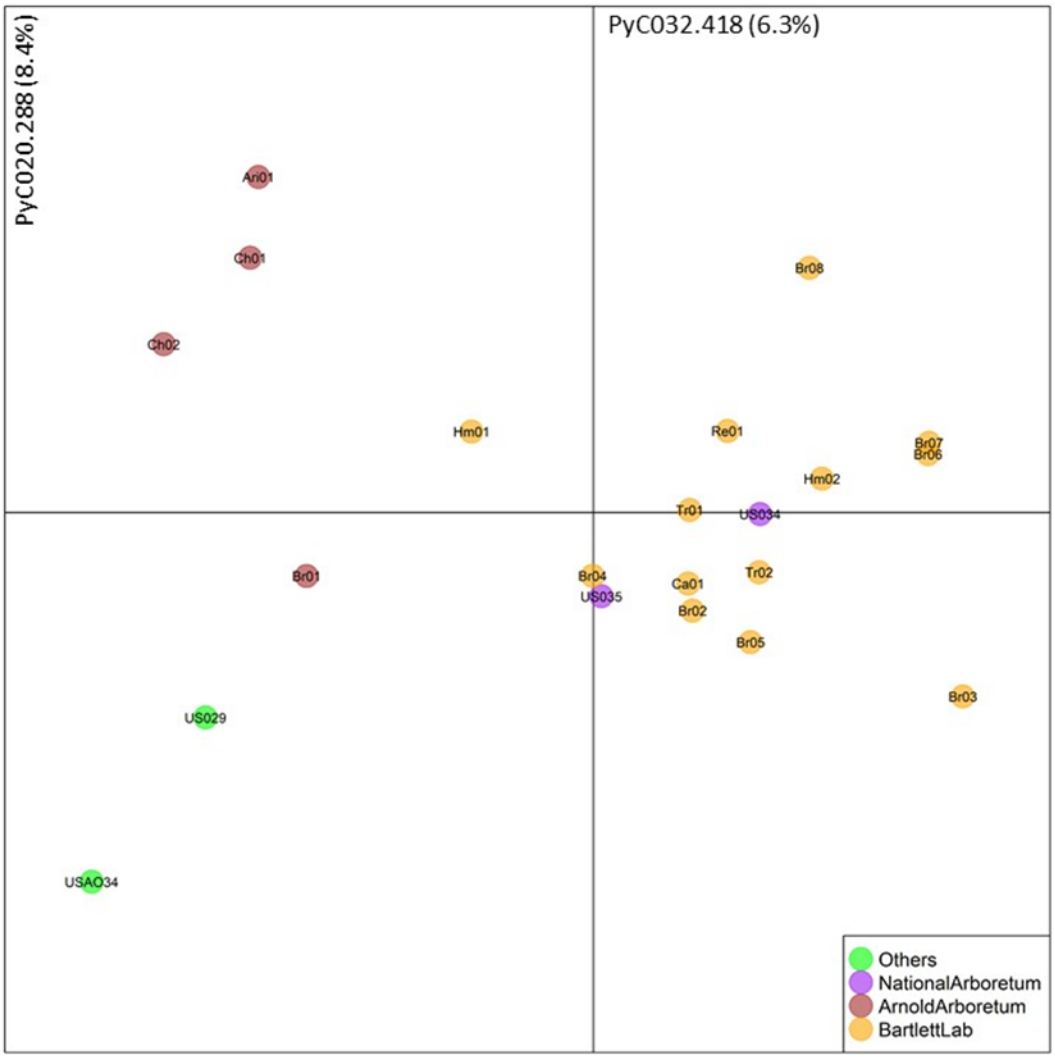

Figure 1. 10. DAPC of *P. calleryana* U.S. cultivars from different source institutions. DAPC for determining the molecular variance partitioning projected using 15 Principal Components cross-checked and optimized with 1,000 permutations. The two respective axes are indicated by the alleles explaining the most of variance within the sampled population. Each different color represents the different source institutions. Each colored dot represents *P. calleryana* U.S. cultivar individuals. The names in the figure indicate the following *P. calleryana* cultivars: Ari01: 'Aristocrat'; Br0 through Br08: 'Bradford'; Ca01: 'Cambridge'; Ch01 and Ch02: 'Chanticleer'; Hm01 and Hm02: 'Holmford'; Re01: 'Redspire'; Tr01 and Tr02: 'Trinity'; and US034 and US035: Unknown cultivars.

CHAPTER 2

UNDERSTANDING THE GENETIC DIVERSITY OF THE INVASIVE CALLERY PEAR, *PYRUS CALLERYANA* IN THE SOUTHEASTERN U.S.

Abstract

Pyrus calleryana Decne. (Callery pear) is a deciduous pear tree native to China, Japan, Korea, and Taiwan. *Pyrus calleryana* is a popular ornamental tree with early spring blooms and beautiful fall color. 'Bradford' is the most well-known *P. calleryana* cultivar in the U.S. *Pyrus calleryana* has become one of the most common invasive tree species across the eastern U.S. The formulation of better management practices for invasive *P. calleryana* require the knowledge of genetic diversity and biology of the species. As little is known about the genetic variation within the species, we investigated the genetic diversity and population structure of invasive *P. calleryana* populations collected within a small geographic area of 110 miles radius in the southeastern U.S. Since *P. calleryana* possesses a wide range of morphological variation with a great invasive potential, we hypothesized a high genetic diversity among *P. calleryana* collected across a small geographic area. We used 15 robust microsatellite loci for genotyping DNA samples from 180 non-cultivated *P. calleryana* individuals collected across six naturally occurring sites in Tennessee, Georgia, and South Carolina. The data and analyses supported revealed the presence of a population structure with high genetic diversity, high gene flow, and high genetic differentiation among individuals at the collection sites. AMOVA attributed the significantly low amount of molecular variance among populations, which validated the presence of population structure in the dataset. STRUCTURE and DAPC congruently revealed two genetic clusters present in our *P. calleryana* collection. DIYABC analyses revealed the studied *P. calleryana* populations differentiated shortly after introduction to the U.S. from 'Unsampled' population(s), which in turn descended from specimens imported from China, as per historical records and our prior research. The recent differentiation and transformation of recently introduced *P. calleryana* specimens from China into an invasive species highlight the invasive

potential of the species. Our data suggests *P. calleryana* has high genetic diversity with a great evolutionary potential creating a great threat to economy and native biodiversity at the invaded areas. Based on our results, we suggest considering proper management practices for controlling invasive *P. calleryana* trees such as using other alternative trees (example: Eastern redbud, Flowering dogwood), and developing sterile *P. calleryana* trees through various breeding techniques.

Introduction

Invasive species cause a great economic, social, and ecological threats to both natural and managed ecosystems (Pimentel et al., 2000, 2005). Invasive species are the second major cause of endangerment and extinction of native species, and their annual economic cost to society is estimated to about $120 billion (Crowl et al., 2008, Pimentel et al., 2000, 2005). A successful plant invasion includes an introduction, establishment of the species in a new area, and a lag phase followed by a colonization of additional areas (Sakai et al., 2001). Not all introduced plant species become invasive, but the reasoning behind this failure is not fully understood. In most cases, a species is introduced in a new area either through natural spread, accidentally, or deliberately (Culley & Hardiman, 2007). Many species introduced deliberately for horticultural purposes escape cultivation, which imposes a high economic cost for their control (Pimentel et al., 2005; Reichard & White, 2001).

Pyrus calleryana Decne. (also known as Callery pear) is one example of such an introduced species brought for horticultural purposes which later became invasive in the U.S. and a potential environmental weed in Australia (Cusrhes & Edwards, 1998). *Pyrus calleryana* is native to China, Taiwan, Korea, Vietnam, and Japan (Bell & Zimmerman, 1990; Cuizhi & Spongberg, 2003; Rubtsov, 1944). ‘Bradford’ is the most widely planted and commonly known ornamental

cultivar of *P. calleryana* in the U.S. In the landscape, *P. calleryana* and its cultivars are used as deciduous landscape trees prized for their shade and bright white flowers. The species typically boasts a rounded to conical crown and sharp spurs on the branches. Young plants start flowering as early as 3 years (Bell & Zimmerman, 1990). The flowers are about 2 to 2.5 cm in diameter with an unpleasant odor (Cuizhi & Spongberg, 2003). Leaves are simple, alternate, and oval, 4 to 8 cm long, shining dark green in summer, and various vivid colors in autumn. The fruits are hard and small, measuring 1 to 1.5 cm long. The fruits are green/brown, spherical to slightly oblong, each with 1 or 2 seeds (Vincent, 2005), and as a secondary food source for birds that results in seed dispersal. The species is relatively short-lived, lasting approximately 25 to 30 years (Dirr, 1990).

In the early 1900s, *P. calleryana* accessions were imported to the U.S. to introduce resistance to fire blight causing bacteria, *Erwinia amylovora* Burrill in *P. communis* L. Several *P. calleryana* selections exhibited desirable horticultural traits for urban use, which led to the development and release of many hybrid cultivars with other *Pyrus* species (Culley & Hardiman, 2007). By 1962, 'Bradford' was available commercially in the U.S. This cultivar was not expected to escape cultivation as *P. calleryana* is self-incompatible, propagated by vegetative methods, and in the native range produces very few seeded fruits (Culley & Hardiman, 2007; Gilman & Watson, 1994). Furthermore, *P. calleryana* trees often grew quite small, and widely scattered across its native ranges of China (Culley & Hardiman, 2007). However, soon after the release of commercial ornamental cultivars in the U.S., the species was noticed in several natural habitats and the first reported escapes were identified in 1964 in eastern Arkansas (Vincent, 2005). Since then, the escaped *P. calleryana* were increasingly found in the natural areas of the eastern U.S. (Swearingen et al., 2002; Vincent, 2005). 'Bradford' and related cultivars were identified to have

invasive potential in 1994 and after about 10 years, the naturalized non-cultivated *P. calleryana* were found in natural areas of at least 26 states (Vincent, 2005). Invasive *P. calleryana* are reported to be distributed now in 33 states (EDDMapS, 2021). Additionally, *P. calleryana* is predicted to have the potential of becoming one of the most problematic invasive plant species in the U.S. (Sapkota et al., 2021; Warrix & Marshall, 2018).

Management of invasive species such as *P. calleryana* is challenging because the control techniques are difficult to implement (Warrix & Marshall, 2018). Some of the management practices identified so far to control the species are the complete removal of *P. calleryana* trees (Culley & Hardiman, 2007; Swearingen et al., 2002), herbicide applications (Vogt et al., 2020), and prescribed burns of occupied areas (Warrix & Marshall, 2018). These control strategies such as complete removal of *P. calleryana* trees is effective but expensive; herbicide application techniques are effective and range in cost, and the technique of prescribed fire is cheap but not effective as new sprouts appear even after the treatment of fire. We currently have no great control options for invasive *P. calleryana,* and the available control options are general management practices rather than the targeted ones. In addition, knowing more about the plant may help us figure out a better control option than we currently have. Invasive plant management and control strategies should be developed based on their genetic and biological characteristics (Allendorf & Lundquist, 2003). In most invasive species, a great variability in their respective genetic diversity, gene flow pattern, resistance/tolerance to control methods, and modes of reproduction are expected, collectively requiring a targeted management practices for each such species (Gaskin et al., 2014). Hence, it is imperative to understand the genetic diversity and biology of the species so that effective targeted management strategies can be formulated accordingly.

Genetic diversity assessment of a plant species using loci under neutral selection can be used to help understand the adaptation dynamics and the spread of that species (Purugganan, 2000), while assessing the genetic diversity studies over different spatial scales can assist in detecting and characterizing the spatial distribution of genetic structure and genetic variation within a species' population. In invasive species, a large-scale genetic diversity study across a wide geographic area may help infer the mass (human-mediated) dispersal, whereas a fine-scale study across a narrow geographic area may help infer the natural dispersal through pollen, seed, or root sprouts of unknown genetic composition (Darling & Folino-Rorem, 2009). Even in invasive species populations with low genetic structure, the fine-scale patterns observed across portions of the total geographical area occupied by the species may help reveal important evolutionary impacts (Short & Petren, 2011).

Most research studies in *Pyrus* are limited to the identification and characterization of cultivars/species using various DNA markers (Bao et al., 2007; Bao et al., 2008; Iketani et al, 1998; Yamamoto et al., 2001; Yamamoto et al., 2002a; Yamamoto et al., 2002b; Yuanwen et al., 2001). A few population studies have been conducted within native range (Kato et al., 2013; Liu et al., 2012; Sapkota et al., 2021) with a limited study of non-cultivated populations in the U.S. (Culley & Hardiman, 2009). Despite the invasive status of *P. calleryana* in many U.S. states, very little is known about the genetic diversity of the escaped populations within affected regions.

The goal of this study was to assess the genetic diversity and population structure of invasive *P. calleryana* populations within a narrow geographic range using microsatellite loci. From our previous *P. calleryana* study across the native ranges, we found high genetic diversity maintained among *P. calleryana* trees (Sapkota et al., 2021). Based on that information, we

hypothesized there would be high genetic diversity among *P. calleryana* trees that escaped cultivation distributed across a relatively small area. The previously developed microsatellite loci (Sapkota et al., 2021) were used to test this hypothesis by addressing the following specific objectives within a narrow geographical portion of Tennessee, Georgia, and South Carolina: (a) to evaluate the fine-scale genetic diversity present within open-pollinated *P. calleryana* populations; (b) to investigate fine-scale patterns in spatial distribution and gene flow within *P. calleryana* trees in these locations; and (c) to infer the evolutionary history of *P. calleryana* in relation to China as a source of origin using Approximate Bayesian Computation.

Materials and Methods

Sample Collection

Leaf samples of *P. calleryana* were collected from individual trees growing in selected regions in parts of Tennessee, Georgia, and South Carolina, U.S. (Fig 2.1). Healthy, fully developed leaves (6-8 leaves per tree) were collected from 30 wild/non-cultivated location of *P. calleryana* trees per collection site (considered a subpopulation for this study). There were 180 samples collected for the study, out of which, 90 were collected in eastern Tennessee, and 90 from northeastern Georgia/northwestern South Carolina. The geographical coordinates of the collected samples were recorded. Based on the site of sample collection, samples were grouped into the following major subdivisions: (i) North Group (n = 90) and (ii) South Group (n = 90). Each of the North and South groups contained three subpopulations including North Group A (n = 30); North Group B (n = 30); North Group C (n = 30), and South Group A (n = 30); South Group B (n = 30); South Group C (n = 30) respectively. We used some samples from Sapkota et al. (2021) to gain access to the genotyping data from the originating population ‘China’ (n = 20).

gDNA Extraction

Approximately 100 mg of air-dried leaves per sample (tree) was homogenized using a Bead mill 24 (Fisher Scientific, Pittsburgh, Pennsylvania, U.S.) and used for genomic DNA (gDNA) extraction using EZNA DNA DS Mini Kit (Omega Bio-Tek, Norcross, Georgia, U.S.), according to the manufacturer's protocol. Nanodrop was used to measure the concentration and purity of the extracted gDNA samples (Thermo Fisher Scientific, Wilmington, Delaware, U.S.).

Microsatellite Primers and Genotyping Conditions

Genomic Short Sequence Repeats (gSSRs) were developed in our previous *P. calleryana* study (Sapkota et al., 2021) using genome sequence data of a closely related pear, *Pyrus* × *bretschneideri* Rehder (GenBank number: JH994112) (Wu et al., 2013). Considering high cross-species amplification rates, polymorphic character, and agreement with the expected PCR product size, a total of 15 robust gSSRs with were selected from our previous study and used here.

The gSSR loci were amplified using polymerase chain reaction (PCR) in a 10 µl reaction mixture consisting of the following: 4 ng gDNA, 5 µl of 2 × GoTaq® DNA Polymerase (Promega, Madison, Wisconsin, U.S.), 1 µM final concentration of each primer, and 0.5 µl of dimethyl sulfoxide (DMSO). For the data validation, DNA sample from *P. calleryana* var. *dimorphophylla* (collected from Japan in 1933 and maintained by Morton Arboretum) was used as a positive control, and sterile water was used as a negative control for each used primer pair. The touch-down protocol was used for PCR amplification to increase the specificity of the amplified products (Korbie & Mattick, 2008). The following thermal profile was used for PCR amplification: initial denaturation at 94 ºC for 3 min., followed by 10 cycles of denaturation at 94 ºC for 30 s, annealing at 65 ºC for 30 s with a touch-down of 0.7 ºC/cycle and

an extension at 72 ºC for 30 s then, followed by 30 cycles of denaturation at 94 ºC for 30 s, annealing at 58 ºC for 30 s and an extension at 72 ºC for 30 s, with a final extension of 72 ºC for 4 min.

The amplified PCR products were visualized using QIAxcel Capillary Electrophoresis System (QIAGEN, Germantown, Maryland, U.S.) and analyzed with a 15/600 bp alignment marker and 25 to 500 bp DNA size marker. Samples that failed to amplify even after two subsequent PCR attempts were referred to as missing data; those that failed to amplify in more than 5 loci were excluded from the study, making the final *P. calleryana* dataset 176 tree samples.

Data Analysis

The MS Excel macro, FlexiBin (Amos et al., 2007), was used to bin the raw allelic sizes into the statistically identical allelic classes using the nucleotide repeat motif size information. The binned allelic dataset was used for further data analyses. For each gSSR, PGDSpider (Lischer & Excoffier, 2012) version 2.1.1.5 was used to transform binned PCR allele sizes into repeat numbers. Clone correction of the dataset was performed using *poppr* (Kamvar, Tabima, & Grünwald, 2014) version 2.8.5 in RStudio version 1.2.5033 using R (Team, 2013) version 3.6.2. No clonal multi-locus genotypes (MLGs) were found in North/South groups and 6 subpopulations, and 176 unique samples were used for subsequent analyses.

Population Genetics of *P. calleryana*

Genetic Diversity

For each of the 15 gSSRs, the genetic diversity indices the number of alleles amplified (N), rarefied allelic richness (A_r), observed heterozygosity (H_o), Nei's unbiased expected heterozygosity (H_e) (Nei, 1978), Jost's differentiation estimate (D_{est}) (Jost, 2008), and Stoddart and Taylor index of MLG diversity (G) (Stoddart & Taylor, 1988) were calculated in R using the

packages *poppr* and *hierfstat* (Goudet, 2005) version 0.04-22. GenAlEx (Peakall & Smouse, 2006) version 6.5 was used to estimate gene flow ($N_m = ¼×[(1/F_{ST})-1]$) and the presence of private alleles (P_a) among subpopulations. SPAGeDi (Hardy & Vekemans, 2015) was used to calculate hierarchical fixation indices including inbreeding coefficients (F_{IS}), allele fixation index (F_{ST}), and their respective analogues based on allele size (R_{IS} and R_{ST}) (Hardy & Vekemans, 2002; Slatkin, 1995). The analyses were performed for both dataset subdivisions (North/South groups and 6 subpopulations). To determine the significance of hierarchical indices, 10,000 permutations were performed among gene copies in SPAGeDi (Pons & Petit, 1996).

SPAGeDi was used to investigate the contribution of mutation rate in the population structure of *P. calleryana* dataset using 10,000 permutations among alleles within each locus. Additionally, we used SPAGeDi to determine the phylogenetic patterns in *P. calleryana* dataset using 10,000 permutations of gene copies among individuals within populations or individuals among all populations.

Analysis of Molecular Variance (AMOVA) was performed for both dataset subdivisions (North/South groups and 6 subpopulations) to estimate the molecular variance distribution among and within subpopulations using the package *poppr* with 1,000 permutations. Linkage Disequilibrium (LD) of the used gSSRs was assessed in *poppr* using 1,000 permutations. The pairwise index of association ($\overline{r}_d$) was used for assessing the linkage among loci to identify the possible bias of the pattern of LD due to a single or few pair(s) of loci.

Population Structure

Mantel Test for Isolation by Distance

The Mantel test was performed to estimate isolation by distance (IBD) using package MASS (Ripley et al., 2013) version 7.3-50 in R with 1,000 permutations. Mantel test results were used

to determine the correlation between genetic and geographical distance matrices of the studied individuals. The underlying correlative relationship between genetic and geographical distance matrices was further confirmed with the Mantel correlogram test using package *ade4* (Dray & Dufour, 2007) version 1.7-13 and *vegan* (Oksanen et al., 2013) version 2.5-3 in R, at $\alpha = 0.05$.

STRUCTURE and DAPC

The population structure of the *P. calleryana* dataset was analyzed using the Bayesian approach in STRUCTURE (Pritchard et al., 2000) version 2.3.4. Thirty independent Monte Carlo Markov Chains (MCMC) were used with 250,000 generations of burn-in period and 750,000 MCMC repetitions for each number of clusters (K = 1 to 10). PopHelper (Francis, 2017) version 1.0.10 with the Evanno method (Evanno et al., 2005) was then used to analyze and visualize the STRUCTURE results. ObStruct (Gayevskiy et al., 2014) version 1.0 was used to determine the correlation between the population structure of STRUCTURE-inferred ancestral profiles and predefined subpopulations. It uses an ad-hoc R^2 statistic whose value ranges from 0 (admixture between populations/recent divergence) to 1 (population structure/complete divergence).

A model-free multivariate clustering approach, Discriminant Analysis of Principal Components (DAPC), was also used to investigate the genetic structure of the *P. calleryana* dataset using the package *adegenet* (Jombart, 2008) version 2.1.1 in R. A principal component analysis (PCA) was performed and the PCA vectors explaining the majority of variance but minimizing the over-fit of DAPC were selected. The number of PCAs selected in this manner was then used to optimize and cross-check the DAPC analysis using 1,000 permutations. The result was further confirmed using a dendrogram of the unrooted neighbor-joining tree of pairwise genetic distances (Nei, 1978) using *ape* (Paradis et al., 2004) version 5.5 in R.

Population Demography

BOTTLENECK

BOTTLENECK (Cornuet & Luikart, 1996) version 1.2.02 was used to investigate the evidence for an evolutionary recent bottleneck of *P. calleryana* populations for both dataset subdivisions (North/South groups and 6 subpopulations). We used a stepwise-mutation model (S.M.M.) and two-phase mutational model (T.P.M.) to test for a recent bottleneck or expansion of the *P. calleryana* populations. The variance of geometric distribution for T.P.M. was set to 12 together with 95% of S.M.M. Significance of the test under either of these models was evaluated using sign test, standardized differences test, and Wilcoxon sign-rank test with 10,000 simulations. The heterozygosity excess and heterozygosity deficiency of each group was assessed using Wilcoxon sign-rank test, whereas the results of all three basic tests *i.e.*, sign, standardized differences, and Wilcoxon sign rank test, were used to assess the mode-shift in population size. BOTTLENECK outputs a graph as either an L-shaped graph indicating a stable population or a mode-shift graph indicating a population that experienced a bottleneck.

Approximate Bayesian Computation

The population history of *P. calleryana* was investigated with the Approximate Bayesian Computation approach using the program DIYABC (Cornuet et al., 2014) version 2.1. For this analysis, *P. calleryana* samples from China (Sapkota et al., 2021) were used as the originating population as evidenced and supported by that previous study. Based on samples' geographical location, the collected samples were divided into 3 major population groups *i.e.*, 'North Group', 'South Group', and 'Origin Group' (*P. calleryana* samples from China (Sapkota et al., 2021), n = 20). This population grouping was also supported by both STRUCTURE and DAPC analysis.

To assess the parameter values for the main DIYABC run, the initial run that used the entire dataset as one population was done with the following parameters under uniform distribution: population size (min: 10; max: 10000); time (min: 10; max: 10000), and a generalized S.M.M. with a mean mutation rate of 5×10^{-4} (min: 10^{-4}; max: 10^{-3}) mutations per generation per locus. In the subsequent run, the *P. calleryana* dataset was tested using 6 possible scenarios considering population divergence, admixture, and the presence of a 'Unsampled' population (population sample group not included in our study). For each scenario, 1 million pseudo-observed datasets (PODs) were simulated from the dataset. For the main run with 6 million simulations (1 million simulations per each regarded scenario), we set the following parameters under uniform distribution: population size (min: 100; max: 100000), and time (min: 1; max: 40 with t2 >= t1, considering the introduction time of *P. calleryana* in the U.S. and the species biology of flowering as early as 3 years of age). We assumed a uniform prior distribution and a generalized S.M.M. with a mean mutation rate of 5×10^{-4} (ranging from 1.12×10^{-5} to 9.53×10^{-2} mutations per generation per locus). The following summary statistics for each scenario were calculated by the DIYABC program: the mean number of alleles, mean genetic diversity, mean population size variance, classification index, pairwise F_{ST}, and distance between pairs of populations $(d\mu)^2$. The effective population size (given in number of individuals) for each population groups are the constructs but not the exact values. For each scenario, summary statistics of the simulated datasets were compared with that of the observed dataset of genotyped *P. calleryana*.

The logistic regression analysis on 1% of the simulated datasets closest to the observed dataset was used to infer the relative posterior probabilities of the scenarios under comparison (Cornuet et al., 2010). Two scenarios having the highest posterior probabilities from the main DIYABC run were used for subsequent comparative analyses. The bias and precision analysis on

parameter estimations was computed on the dataset for the best-supported scenario. The "Confidence in scenario choice" program option was used to test the goodness of fit for top 2 considered scenarios by estimating the proportion of type I and type II errors based on 1,000 pseudo-observed datasets.

Results

Population Genetics of *P. calleryana*

Genetic Diversity

All 176 *P. calleryana* individuals in both dataset subdivisions (North/South groups and 6 subpopulations) represented unique multi-locus genotypes (MLGs) and were used for analyses. Overall, 2.8% of missing data were detected across the entire dataset. The locus PyC006 had the highest missing data of 8.5% and the subpopulation South Group B had 5.3% of missing data. Both North and South groups did not follow the Hardy-Weinberg equilibrium (HWE), but the North group deviated comparatively more from HWE than the South group. This suggests that some changes might be occurring in the population structure across time and that our results should be interpreted cautiously (Fig 2.2).

We found overall about 12 alleles per locus and about 5 effective alleles per locus for both North and South groups (Table 2.1). The overall observed heterozygosity ($H_o = 0.32$) was lower than the overall expected heterozygosity ($H_e = 0.74$) indicating high genetic diversity of *P. calleryana* subpopulations and the presence of population structure in *P. calleryana* across the sampled areas. The only two private alleles recovered were found in the North group. The overall allelic richness (A_r) was 6.71, ranging from 5.35 in North Group C to 7.87 in North Group A, suggesting high allelic richness with the long-term adaptability potential of *P. calleryana* individuals. The standardized index of association ($\overline{r}_d$) among 6 subpopulations was 0.32 ($P <$

0.001). Furthermore, a positive inbreeding coefficient (F_{IS} = 0.56, P < 0.001) was detected suggesting that a substantial level of homozygosity possibly resulting from inbreeding within the *P. calleryana* dataset.

The gSSRs chosen for the study were powerful in discriminating MLGs as only 6 gSSRs were able to discern all the MLGs present in the dataset (Fig 2.3). The values of pairwise LD ($\bar{r}_d$) ranged from 0 to 0.4 (P = 0.002), which indicated the absence of linked loci and the genome-wide distribution of the gSSRs (Fig 2.4). Overall, an average of about 12 alleles per locus (ranging from 8 to 19) were detected across the dataset (Table 2.2). Our dataset suggested a high overall genetic differentiation (D_{est} = 0.21), indicating the presence of population structure in the tested *P. calleryana* dataset. Furthermore, our data detected the presence of high gene flow (N_m = 3.94).

The phylogeographic signals within the *P. calleryana* dataset assessed using SPAGeDi indicated statistically similar results for F_{ST} and the mean permuted R_{ST} over all loci (data not shown). The mean of permuted R_{ST} and observed F_{ST} values were statistically comparable (Pobs>exp = 0.48) suggesting the absence of a phylogeographic signal within populations. To further evaluate presence of a phylogeographic signal among populations, the slope test (b-log values) of pairwise R_{ST} was evaluated. No evidence of phylogeographic signal was demonstrated among populations (Pobs>exp = 0.95).

AMOVA was used to assess the proportion of molecular variance present within the *P. calleryana* dataset. A low proportion of molecular variance was present among populations (6.80%) and the major portions of molecular variance were attributed to within individuals (43.38%) and within populations (49.82%) (Table 2.3). The significant result of this test (P < 0.001) demonstrated the existence of population structure within the *P. calleryana* dataset.

Population Structure

Mantel Test for Isolation by Distance

Isolation by distance using the Mantel test was used to determine the correlation between genetic and geographic distances among *P. calleryana* individuals (Fig 2.5). We found a positive correlation between genetic and geographic distances (Mantel's r = 0.24, $P = 0.001$), indicating that about 6% of the genetic variance is explained by the geographic distance. The maximum linear distance between samples was found as approximately 260 km. Across space, there was a non-linear relationship between genetic and geographic distances of *P. calleryana* individuals. This suggested that the increased geographic distance between *P. calleryana* individuals does not necessarily mean increased genetic dissimilarity between them. The amplitude of the Mantel's r scores in the correlogram ranged between -0.20 and 0.15, indicating a relatively a low impact of spatial distancing on the population structure of the *P. calleryana* dataset.

STRUCTURE and DAPC

Bayesian clustering analysis using STRUCTURE indicated an optimum of $\Delta K = 2$ suggesting the presence of two genetically distinct clusters among the studied subpopulations of *P. calleryana* (Fig 2.6). The result comprised of 2 genetic clusters consisting of 3 subpopulations from each of North and South groups respectively with a little admixture among them.

The overall R^2 between the predefined populations and inferred clusters under K = 2 was 0.88 ± 2.28E-16 suggesting strong divergence among the predefined populations and STRUCTURE-derived genetic clusters within the dataset (Table 2.4). We found only negligible changes in R^2 when the predefined populations were removed sequentially. In addition, there was no change in R^2 when the inferred clusters were removed, suggesting no major contribution of the inferred clusters to the structure of the *P. calleryana* dataset. As such, the results of successive removal of

populations/clusters imply that our *P. calleryana* populations might be a part of an even bigger community of *P. calleryana*.

A multivariate analysis, DAPC, for the *P. calleryana* dataset showed a similar clustering pattern to STRUCTURE (Fig 2.7). The *P. calleryana* dataset was divided into two major clusters similar to their geographical location. This result was further supported by an unrooted neighbor-joining tree of pairwise genetic distances among the sampled *P. calleryana* individuals (Fig 2.7, Insert topright).

Population Demography

BOTTLENECK

Using the Wilcoxon test in BOTTLENECK, significant signals were found under the T.P.M. and S.M.M. mutation models for a possible bottleneck (heterozygosity deficiency) in both North and South groups. The mode shift in the population size was analyzed using the information from all three basic tests *i.e.*, sign, standardized differences, and Wilcoxon sign-rank tests. In the cumulative mode shift test, a normal L-shaped distribution was detected in each North/South groups signifying that no recent bottleneck events were evident within the data (Table 2.5).

Approximate Bayesian Computation

Following initial runs for estimation of parameters, six hypothetical evolutionary scenarios were evaluated (Fig 2.8). The highest support for evolutionary scenario reconstruction was found for scenario 2 followed by scenario 5. In scenario 2, 'Unsampled' population was derived from the originating population (Origin Group: DNA samples extracted from China-sourced leaf tissues, n = 20) shortly after *P. calleryana* introduction to the U.S. (Fig 2.9). The 'Unsampled' population later split into 'North group' and 'South group' subpopulations shortly after *P. calleryana* introduction to the U.S. In scenario 5, North Group population was derived from the Origin

Group population ‘China’ and South Group population was in turn derived from North Group population. Scenario 2 was accepted as the most likely evolutionary scenario for the analyzed *P. calleryana* dataset as it had the highest relative posterior probability, highest support by logistic regression, and the range of 95% CI did not overlap with the CI ranges of other models. The related posterior parameter estimates under scenario 2 suggested the high mutation rate of 0.053 per locus per generation as our expectation as gSSRs are expected to have high mutation rates (Table 2.6). In accordance with our assumptions, these data indicated that the North and South groups *P. calleryana* evolved recently from the unsampled escape population characterized by a great evolutionary potential. The relative mean absolute deviation for the *P. calleryana* dataset derived using prior and posterior distributions was 4.91 (95% coverage: 0.90) and 0.51 (95% coverage: 0.99) (Table 2.7). The confidence prior type I error in scenario 2 for the genotyped *P. calleryana* dataset using the direct and logistic approach was 0.431 and 0.471, respectively, whereas the confidence prior type II error in scenario choice using the direct and logistic approach was 0.537 and 0.482, respectively (Table 2.8).

Discussion

We investigated the genetic diversity and population structure of the invasive *P. calleryana* collected from a narrow geographical area of the southeastern U.S. Our study showed high genetic diversity (H_e = 0.74) for the *P. calleryana* dataset which was similar to genetic diversity reported for *Malus orientalis* Uglitzk. (H_e = 0.76) in Iran using nine SSRs (Farrokhi et al., 2011) and somewhat higher than that reported in *P. calleryana* (H_e = 0.64) using 14 nuclear SSRs (nSSRs) in China (Liu et al., 2012). The genetic diversity statistics for *P. calleryana* were also higher compared to other invasive tree species such as *Albizia lebbeck* (L.) Benth. and *Pueraria lobata* (Willd.) (Dunphy & Hamrick, 2005; Pappert et al., 2000) Our study results

support the hypothesis of *P. calleryana* has high genetic diversity. Such a high level of genetic diversity within *P. calleryana* could be the result of high gene flow, high genetic differentiation, multiple introductions of different rootstocks and cultivars into landscapes across time (Culley et al., 2011), great evolutionary potential, and widespread distribution of the species through various dispersal mechanisms via birds, insects, and human (Dlugosch & Parker, 2008; Ellstrand & Schierenbeck, 2000; Hamrick et al., 1992; Sexton et al., 2002). Our study also supports the contention that outcrossing species tend to have higher levels of within-population genetic diversity and lower levels of among-population genetic diversity (Hamrick & Godt, 1996). Between the North and South groups, only the North group had two private alleles which were unique only for North group. The low number of private alleles in *P. calleryana* population indicates the presence of a few unique alleles within populations suggesting a possible admixture among the populations. Such genetic admixture in *P. calleryana* populations could be the result of high gene flow among *P. calleryana* individuals, different *P. calleryana* cultivars in nearby locations cross-pollinating with each other, and high seeds and pollen dispersals to distant and nearby locations via birds and insect pollinators. The genetic diversity of the North group (H_e = 0.71) was not significantly higher compared to that of the South group (H_e = 0.69). Other genetic diversity indices were similar for both North and South groups. Additionally, DAPC, unrooted neighbor-joining genetic tree, and STRUCTURE partitioned our North and South groups into two major genetic clusters with a little admixture between them, in agreement with the geographical locations of the collected *P. calleryana* samples.

Isolation by distance indicated a positive correlation between genetic distance and geographic distance (Mantel r = 0.24, P = 0.001), implicating geographic distance as one of the factors in determining the genetic structure of the *P. calleryana* dataset, albeit with a low effect. This result

was also supported by our SPAGeDi data. Compared to our study, a higher positive correlation between genetic and geographic distances (Mantel r = 0.59, $P = 0.008$) was obtained for wild *P. calleryana* in China, where most of *P. calleryana* populations grow fragmented and isolated (Liu et al., 2012). We found high gene flow estimates among *P. calleryana* populations, which is consistent with our previous study of a native Asian collection and U.S. cultivars (Sapkota et al., 2021) and other studies in invasive species such as *A. lebbeck* (Dunphy & Hamrick, 2005) and *Fallopia* species (Gaskin et al., 2014). Such a high gene flow rate helps exchange the alleles among populations and ensures abundant fruit crops in self-incompatible species contributing the seeds for population growth and colonization of new areas (Dunphy & Hamrick, 2005). Compared to other trees, fruits of *P. calleryana* stay on the trees for a longer period, thereby making the fruits an emergency food for birds during winter months when other food sources are scarce (Culley, 2017). This ultimately leads to the dispersal of seeds to distant areas making it more successful as an invasive species. *Pyrus calleryana* trees are visited by various pollinators such as honeybees and frugivorous animals leading to short and long-distance dispersal of both pollen and seed (Culley & Hardiman, 2007, 2009; Liu et al., 2012). In addition, there is high human-mediated dispersal of *P. calleryana* trees as a result of which disturbed land has become the hotspot for fruiting wild *P. calleryana* trees.

Our dataset displayed a high level of genetic differentiation among populations as reported in other studies of *P. calleryana* (Liu et al., 2012; Sapkota et al., 2021) and other hardwood species such as *Cornus florida* and *Cornus kousa* (Nowicki et al., 2020) that were able to maintain high level of genetic diversity. Despite the high gene flow between *P. calleryana* individuals and the evident founder effect, the species has been able to maintain a high level of genetic differentiation which could be the result of various dispersal mechanisms of *P. calleryana* trees.

This high level of genetic differentiation suggests the presence of population structure in the *P. calleryana* dataset. We found low value of standardized index of association ($\overline{r}_d$) for our dataset but the positive inbreeding coefficient (F_{IS}) was found in contrast to *P. calleryana* biology. Such positive F_{IS} for our dataset could be the result of human interference (selection, propagation, and intentional transportation of *P. calleryana* leading to the escape of species), insect pollination limiting the long-distance pollen flow, and the founder effect. This positive result agrees with other studies in *P. calleryana* across the native ranges (Liu et al., 2012; Sapkota et al., 2021). Furthermore, we found no significant differences between the permuted values of R_{ST} and F_{ST}, indicating the absence of phylogeographic patterns within our *P. calleryana* dataset and the mutation rate as negligible when compared to migration rate of the species (Hardy & Vekemans, 2015; Nowicki et al., 2020).

The evolutionary history of invasive *P. calleryana* has not yet been studied extensively in the U.S. We used DIYABC to reconstruct the evolutionary scenario of the sampled subpopulations using DNA samples from China to construct the originating population (Sapkota et al., 2021). Our DIYABC analyses indicated that a generated originating population (about 100 derived individuals) diverged into an 'Unsampled' population represented by about 106 derived individuals and the 'Unsampled' population later diverged into the North group (about 488 derived individuals) and South group (about 1050 derived individuals) shortly after introduction to the U.S. in the early 20th century. Furthermore, the Obstruct analyses showed strong differentiation among populations evolved within a short time (equivalent to about 40 generations (Vincent 2005)). These findings indicate how contemporary *P. calleryana* individuals would be capable of evolving from a small number of individuals into an established invasive population on a continental scale in a short time frame. Considering how these

P. calleryana individuals diverged recently and how they have already become invasive in their introduced ranges, *P. calleryana* individuals possess extremely high evolutionary potential presenting a greater threat to native biodiversity. However, if we compare the present status of *P. calleryana* across its native and introduced ranges, then it is reported to be fragmented and nearing extinction in China and Japan due to human overexploitation (Kato et al., 2013; Liu et al., 2012) while it is invasive in the U.S. (Culley et al., 2011).

Our study showed how *P. calleryana* populations in smaller geographic areas have been able to maintain high levels of genetic diversity making trees more adaptable to the given environment, which is similar to their native habitat. The species' high gene flow pattern creating genetically admixed *P. calleryana* populations creates a greater challenge for targeted management options for *P. calleryana*. This study helps broaden the existing knowledge on genetic diversity patterns, origins, and evolutionary potential of this species, already deemed invasive and quite broadly distributed across the U.S.

Appendix: Tables and Figures

Table 2. 1. Genetic diversity indices of the genotyped *P. calleryana* dataset for North/South groups and six subpopulations using fifteen microsatellite loci

Subpopulations	N	% Missing	# Alleles	NAe	$\overline{r}_d$	H_e	H_o	A_r	F_{IS}	P_a
North Group A	30	0.20	8	5	0.01***	0.74****	0.41	8.04	0.45****	2
North Group B	30	0.40	7	4	0.03***	0.69****	0.35	6.99	0.50****	0
North Group C	30	3.30	6	3	0.04***	0.63****	0.27	5.50	0.58****	0
South Group A	29	2.30	7	4	0.06***	0.68****	0.29	6.75	0.57****	0
South Group B	29	5.30	7	4	0.02***	0.66****	0.32	6.32	0.52****	0
South Group C	28	5.20	8	5	0.01***	0.71****	0.31	7.34	0.57****	0
Σ/Overall	176[a]	2.80[b]	12[b]	4[b]	0.03[b]***	0.74[b]****	0.32[b]	8.24[b]	0.56[b]****	2[a]
Groups	N	% Missing	# Alleles	NAe	$\overline{r}_d$	H_e	H_o	A_r	F_{IS}	P_a
North Group	90	1.30	10	4	0.03***	0.71****	0.34	9.86	0.52****	2
South Group	86	4.30	9	4	0.03***	0.69****	0.31	8.93	0.56****	0
Σ/Overall	176[a]	2.80[b]	12[b]	5[b]	0.03[b]***	0.74[b]****	0.32[b]	10.43[b]	0.56[b]****	2[a]

[a]: Summation (Σ); [b]: Overall; N: Number of samples used for the study in each population group; % missing: % of data missing in the given population group; # Alleles: Number of alleles present; NAe: Effective number of alleles; $\overline{r}_d$: Standardized index of association considering the number of loci sampled (Kamvar et al., 2014); H_e: Nei's gene diversity corrected for sample size (Nei, 1978); H_o: Observed heterozygosity; A_r: Allelic richness; F_{IS}: Individual inbreeding coefficient; P_a: Number of private alleles in each population group. Significance of the dataset was assessed by 10,000 permutations using $P < 0.0001$ = ****; $P < 0.001$ = ***; $P < 0.01$ = **; $P < 0.05$ = *; $P > 0.05$ = [ns].

Table 2. 2. Genetic diversity indices for six subpopulations of genotyped *P. calleryana* dataset based on 15 microsatellite loci

SSR Locus	# Alleles	% Missing	H_o	H_e	R_{ST}	R_{IS}	D_{est}	N_m
PyC006	13	8.50	0.06	0.81****	0.31****	0^{ns}	0.57	0.86
PyC008	15	0.60	0.45	0.82****	0.38****	0.29*	0.52	1.14
PyC013	10	0.00	0.49	0.70*	-0.01^{ns}	0.10^{ns}	0.04	6.99
PyC014	16	0.60	0.25	0.86***	0.07**	0.67****	0.18	4.68
PyC015	13	3.40	0.24	0.84****	0.04ns	0.84****	0.37	2.11
PyC017	19	5.10	0.54	0.90^{ns}	0.03*	0.18*	0.09	8.16
PyC018	13	1.70	0.48	0.84^{ns}	0.03*	0.05^{ns}	0.06	8.08
PyC020	14	0.00	0.34	0.76****	0^{ns}	0.27**	0.16	3.33
PyC031	13	5.70	0.58	0.78****	0.02^{ns}	-0.25^{ns}	0.28	2.17
PyC032	8	10.20	0.03	0.55**	0.04*	0.83****	0.06	3.29
PyC035	12	0.60	0.53	0.86****	0.05**	0.04^{ns}	0.19	4.80
PyC041	9	0.60	0.46	0.7**	-0.01^{ns}	0.55****	0.07	5.07
PyC042	10	0.60	0.06	0.4*	0.08***	0.61****	0.11	2.45
PyC047	12	2.80	0.28	0.84****	0.27****	0.77****	0.47	1.53
PyC050	8	1.10	0.09	0.42^{ns}	0.14****	0.32**	0.03	4.41
Σ/Overall	185^a	2.80^b	0.32^b	0.74^b****	0.08^b****	0.31^b****	0.21^b	3.94^b

[a]: Summation (Σ); [b]: Overall; # Alleles: Number of alleles identified; H_o: Observed heterozygosity; H_e: Expected heterozygosity (Nei's unbiased gene diversity; Nei, 1978); R_{ST} and R_{IS} are complementary measures of F_{ST} (fixation index) and F_{IS} (inbreeding coefficient) respectively; D_{est}: Jost's differentiation estimate (Jost, 2008); Nm: Gene flow given as N_m = ¼×[(1/F_{ST})-1]. Significance of the dataset was assessed by 10,000 permutations using $P < 0.0001$ = ****; $P < 0.001$ = ***; $P < 0.01$ = **; $P < 0.05$ = *; $P > 0.05$ = [ns].

Table 2. 3. AMOVA of genotyped *P. calleryana* dataset using six subpopulations and North/South groups

Subpopulations						
Source of Variation	df	Sum of Squares	Mean Squares	Sigma	% Variance	Φ
Variations among populations	5	229.64	45.93	0.58**	6.80	0.57
Variations within populations	170	2060.25	12.12	4.22**	49.82	0.53
Variations within individuals	176	646.93	3.68	3.68**	43.38	0.06
Total Variations	351	2936.82	8.37	8.37**	100.00	
North and South Groups						
Source of Variation	df	Sum of Squares	Mean Squares	Sigma	% Variance	Φ
Variations among populations	1	131.89	131.89	0.68**	7.79	0.58
Variations within populations	174	2158.00	12.40	4.36**	50.05	0.54
Variations within individuals	176	646.93	3.68	3.68**	42.16	0.08
Total Variations	351	2936.82	8.37	8.72**	100.00	

Df: Degree of freedom (sample size - 1); Sum of Squares: Sum of squares of deviations of the observations from mean; Mean Squares: Sample variance as given by the sum of squares divided by the respective df; Sigma: Variance given for each hierarchical level; % Variance: Total variance percent for each hierarchical level; Φ: Statistics given by the test; Significance of the test was assessed using 1,000 permutations; ** = $P < 0.001$.

Table 2. 4. Obstruct analysis for genotyped *P. calleryana* dataset

	When k = 2
Overall R^2 for the dataset	0.88±2.28E-16***
R^2 without predefined population "North Group A"	0.86±2.28E-16***
R^2 without predefined population "North Group B"	0.88±2.28E-16***
R^2 without predefined population "North Group C"	0.85±1.14E-16***
R^2 without predefined population "South Group A"	0.87±2.28E-16***
R^2 without predefined population "South Group B"	0.91±1.14E-16***
R^2 without predefined population "South Group C"	0.88±2.28E-16***
R^2 without inferred population RED	0.88±2.28E-16***
R^2 without inferred population PURPLE	0.88±2.28E-16***

***: Significant at $P < 0.0001$; RED and PURPLE are the inferred population colors indicating the inferred clusters using STRUCTURE for k = 2 across 20 independent Markov chains.

Table 2. 5. BOTTLENECK analyses for North and South groups of the genotyped *P. calleryana* dataset

			T.P.M.			S.M.M.		
Parameter	**Group**		**Result**	**Ho**	***P***	**Result**	**Ho**	***P***
Sign Test	North		14/1***	8.82	0.00004	14/1***	8.80	0.00004
	South		11/4*	8.84	0.01	12/3**	8.83	0.003
Standardized Differences Test	North		***	-6.30	0	***	-9.06	0
	South		***	-6.37	0	***	-8.95	0
Wilcoxon Test	North	P (1T Hdef):	***	na	0.00008	***	na	0.00003
		P (1T Hexc):	ns	na	0.99995	ns	na	0.99998
		P (2T Hdef&exc):	**	na	0.00015	***	na	0.00006
	South	P (1T Hdef):	**	na	0.00513	**	na	0.0005
		P (1T Hexc):	ns	na	0.99582	ns	na	0.99962
		P (2T Hdef&exc):	*	na	0.01025	**	na	0.00101
Mode Shift	North	Not shifted (normal L-shaped distribution)						
	South	Not shifted (normal L-shaped distribution)						

T.P.M.: Two-phase mutation model; S.M.M.: Stepwise mutation model; Significance of heterozygosity excess/deficiency for each group and mutation model was assessed using 10,000 permutations with $P < 0.0001$. Shift in the population size (model shift) considers all three basic tests (sign, standardized differences, and Wilcoxon sign-rank tests). na: not applicable.

Table 2. 6. DIYABC analyses of the genotyped *P. calleryana* dataset

Parameter	mean	median	mode	q025	q050	q250	q750	q950	q975
$N_{North\ Group}$	2140.00	488.00	113.00	112.00	122.00	218.00	1480.00	8000.00	15700.00
$N_{South\ Group}$	4360.00	1050.00	131.00	133.00	158.00	409.00	3330.00	19700.00	34600.00
$N_{Origin\ Group}$	129.00	100.00	100.00	100.00	100.00	100.00	108.00	150.00	227.00
$N_{Unsampled}$	143.00	106.00	100.00	100.00	100.00	100.00	120.00	243.00	372.00
t_1	39.00	40.00	40.00	35.60	37.50	39.60	40.00	40.00	40.00
t_2	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00
μmic_1	0.05	0.05	0.05	0.01	0.01	0.04	0.07	40.00	0.09

N_X: effective size of the given population; t_X: estimated time since split (in generations); μmic_1: overall mutation rate (in mutations per locus per generation).

Table 2. 7. Bias and precision on parameter estimation analysis of the genotyped *P. calleryana* dataset

		$N_{North\ Group}$	$N_{South\ Group}$	$N_{Origin\ Group}$	$N_{Unsampled}$	t_1	t_2	μmic_1
RMeanAD	Drawn from prior distributions	1.68	1.52	1.16	1.69	1.40	0.60	4.91
95% Coverage		0.96	0.96	0.92	0.96	0.97	0.95	0.90
RMeanAD	Drawn from posterior distributions	15.89	16.81	1.94	430.01	0.49	0.09	0.51
95% Coverage		0.81	0.84	0.85	0.00	0.27	1.00	0.99

Bias and precision on parameter estimation analysis of the *P. calleryana* dataset using DIYABC. RMeanAD (Relative mean absolute deviation) and 95% Coverage gives information about how much % of it was contained in 95% dataset. Only the genetic data generated by the analyses have been reported in the given table. N_X: effective size of the given population; t_X: estimated time since split (in generations); μmic_1: overall mutation rate (in mutations per locus per generation).

Table 2. 8. Confidence in scenario choice of the given *P. calleryana* dataset using DIYABC

	Direct approach	Logistic approach
Confidence global	0.53	0.47
Confidence prior global	0.80	0.82
Confidence prior type I error	0.43	0.47
Confidence prior type II error	0.54	0.48

Evaluation of confidence in scenario choice of the given *P. calleryana* dataset using DIYABC. Confidence global: posterior based error computation over all scenarios; Confidence prior global: prior based error computation over all scenarios; Confidence prior type I error: prior based error computation for scenario 2, discriminated with scenario 5; Confidence prior type II error: prior based error computation for scenario 5, discriminated with scenario 2.

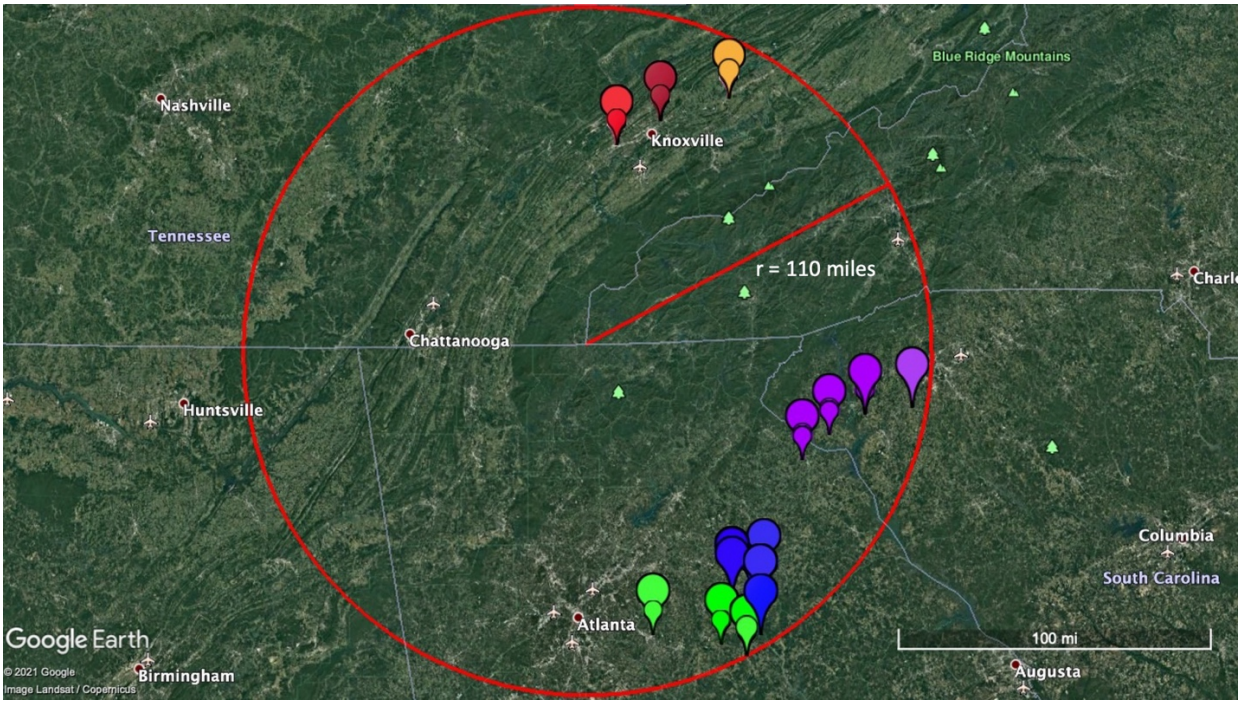

Figure 2. 1. Map showing the collection sites of the open-pollinated tree samples collected within the radius of 110 miles used for the *P. calleryana* study. Each colored symbol represents individual samples taken from 10 different trees. Leaf samples from trees were collected from Tennessee, Georgia, and South Carolina. Each of the six colors represent six subpopulations (Brown: North Group A, Red: North Group B, Light Orange: North Group C, Blue: South Group A, Light Green: South Group B, and Purple: South group C). The scale line indicates the ground-level distance of 100 miles. The map was generated using Google Earth Pro version 7.3.

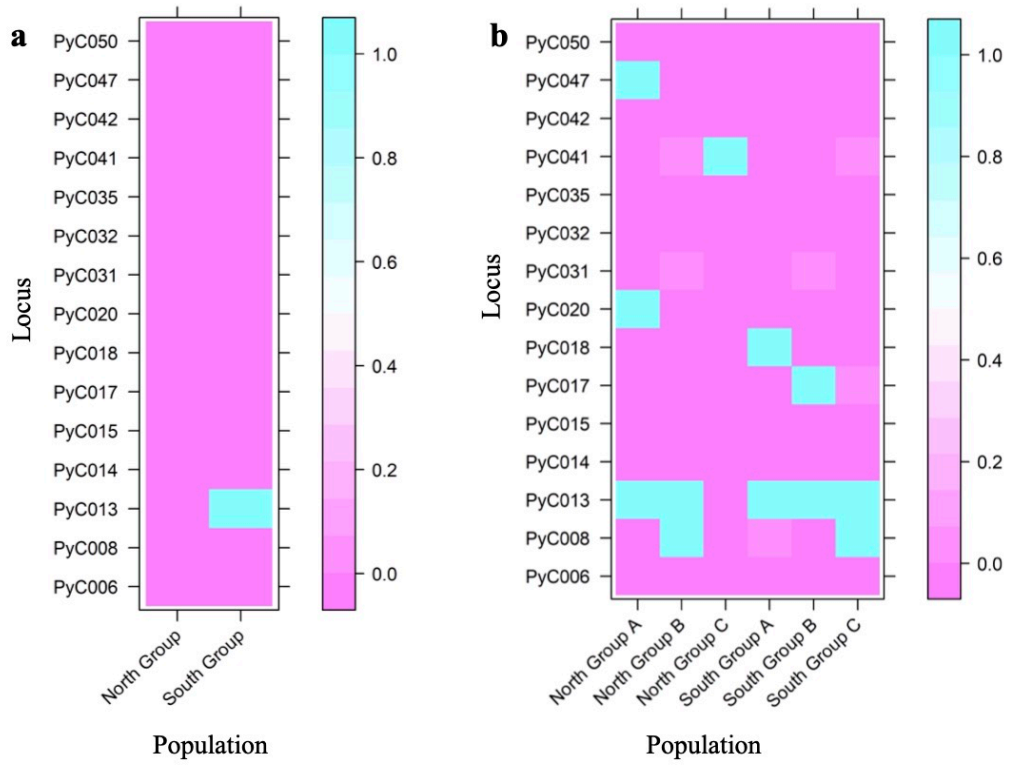

Figure 2. 2. Hardy-Weinberg Equilibrium (HWE) for 6 subpopulations and loci of the *P. calleryana* dataset. (a) HWE for North and South groups and loci, and (b) HWE for all 6 subpopulations and loci. The rows represent the loci, and the columns represent the sample populations used for the study. The probability of the given loci following HWE is shown in the legend. The pink color represents the loci not in HWE at $P \leq 0.05$.

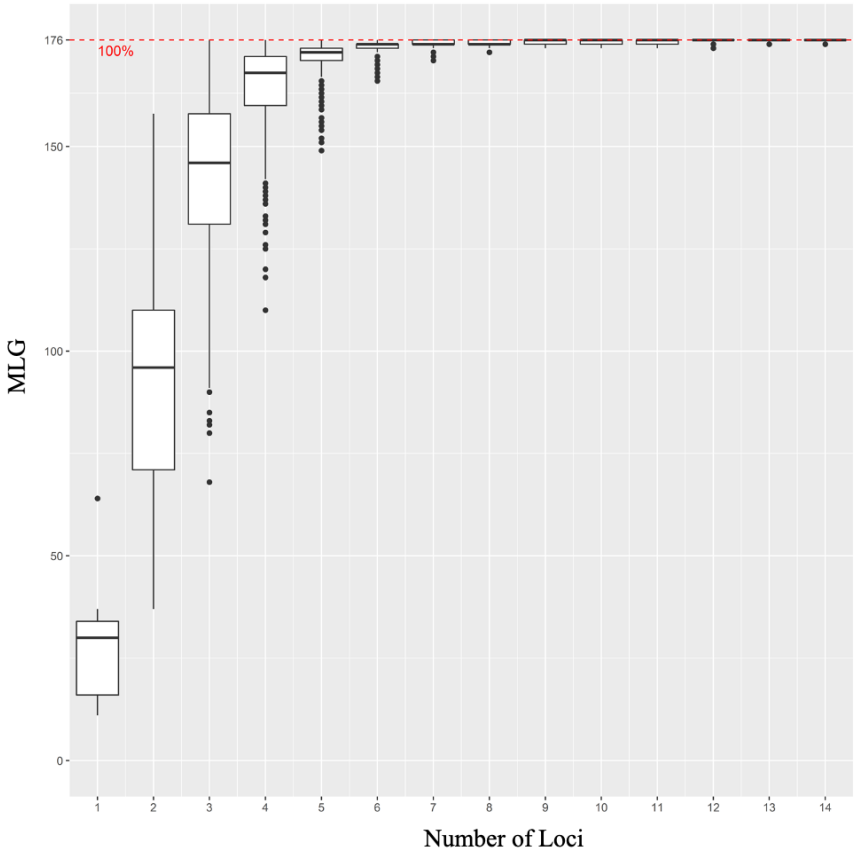

Figure 2. 3. Genotype Accumulation Curve (GAC) for the *P. calleryana* dataset. The X-axis represents the number of loci used for the study and the Y-axis represents the number of multi-locus genotype (MLG) detected.

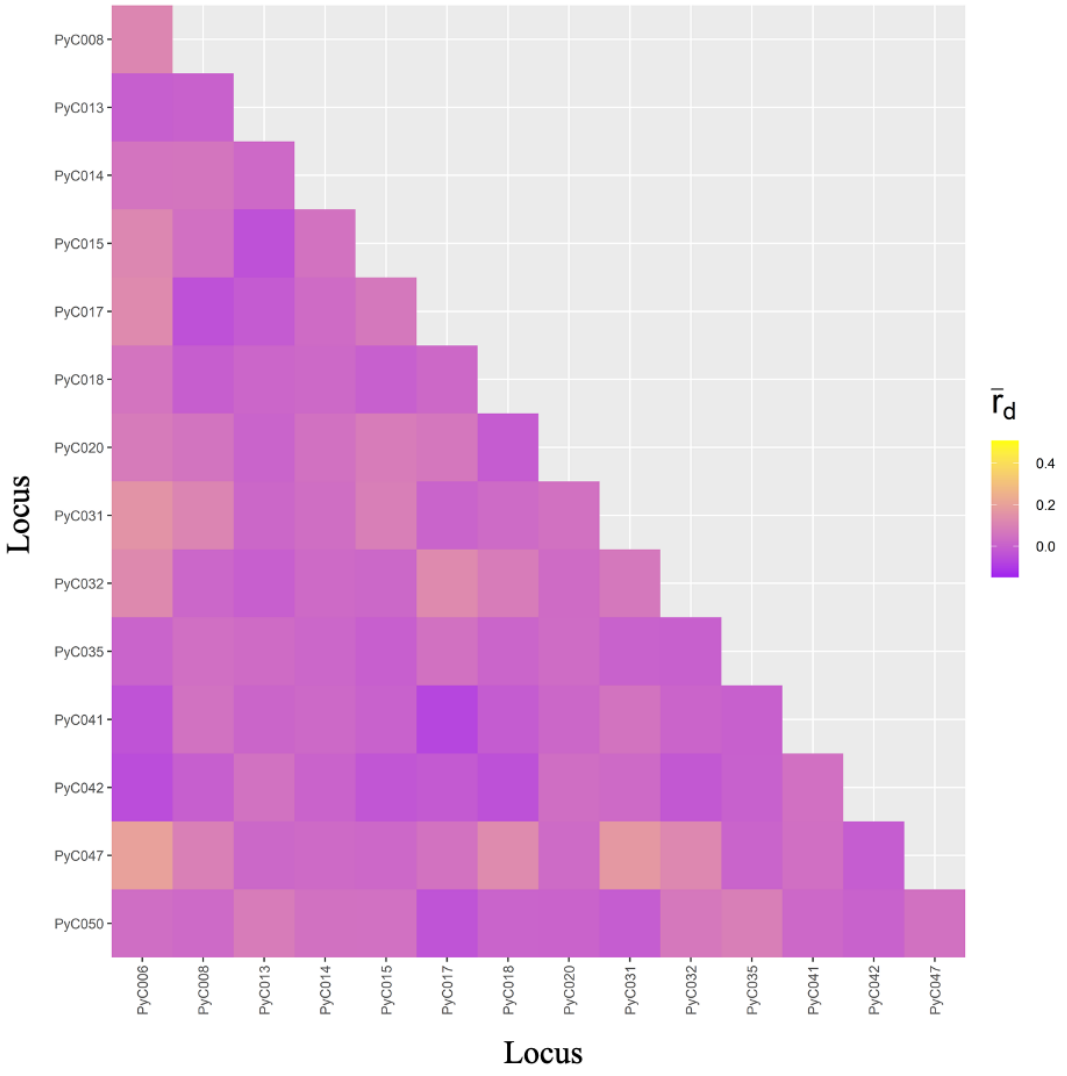

Figure 2. 4. Pairwise Linkage Disequilibrium (LD) among the studied 15 loci included in the *P. calleryana* dataset. The linkage strength between pairs of loci is represented by hues depicted in the legend that are expressed in relation to a standardized index of association ($\bar{r}_d$).

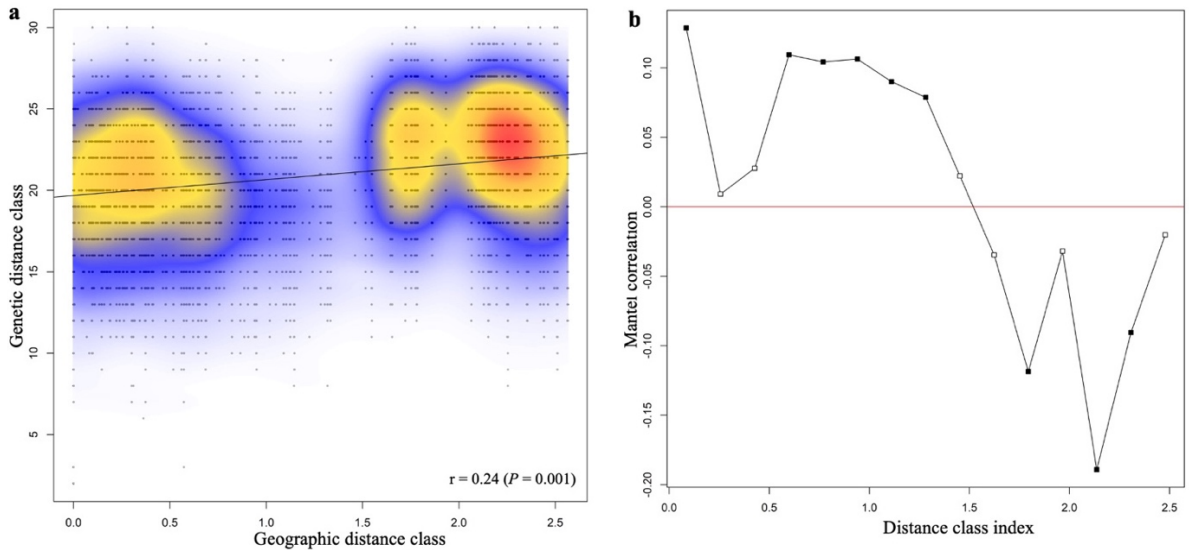

Figure 2. 5. Mantel test of the studied *P. calleryana* dataset. Mantel test (**a**) with Mantel correlogram at $\alpha = 0.05$ (**b**) using isolation-by-distance correlogram for samples included within the *P. calleryana* dataset using 1,000 permutations. Distance class index (in 100s of km) represents the maximum linear distance between samples *i.e.*, 260 km. Correlograms marked with solid black symbols are significant at $P < 0.001$.

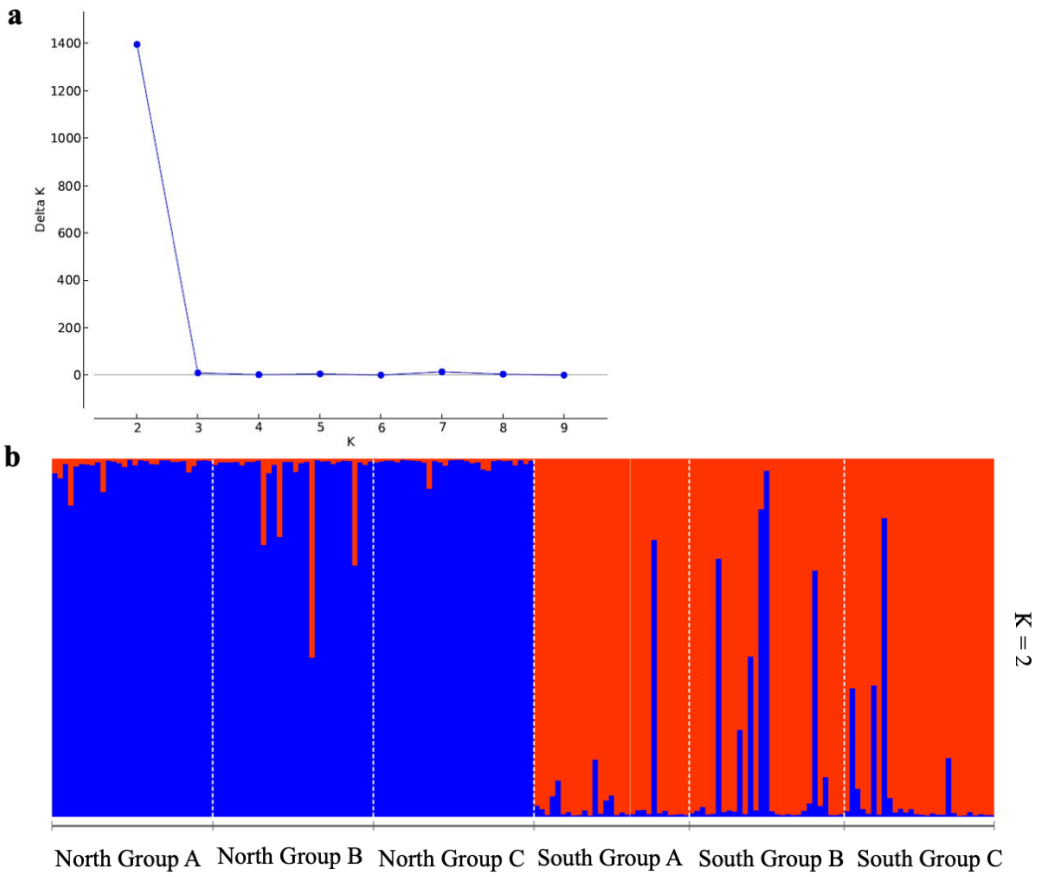

Figure 2. 6. Bayesian clustering using STRUCTURE for the *P. calleryana* dataset. The results were analyzed using (**a**) the Evanno method and visualized using (**b**) 2 inferred genetic clusters. An individual sample is represented by each vertical bar and an individual probability to fall under the identified cluster is represented by the bar color.

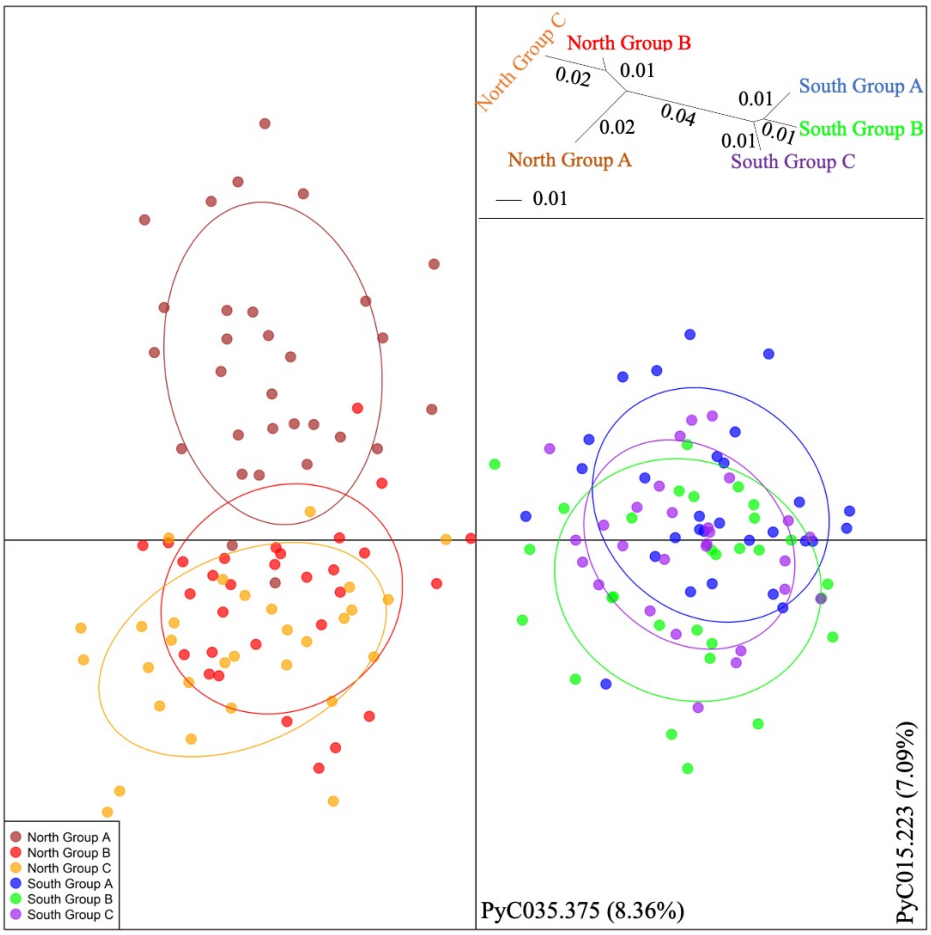

Figure 2. 7. Discriminant Analysis of Principal Component (DAPC) of the tested *P. calleryana* dataset. The alleles that explained the most of variance within the sampled populations (and their contributions) are indicated in both X and Y axes. The genetic distance tree (Insert top right) represents the unrooted neighbor-joining tree of pairwise genetic distances (Nei, 1978) among the sampled 176 *P. calleryana* individuals.

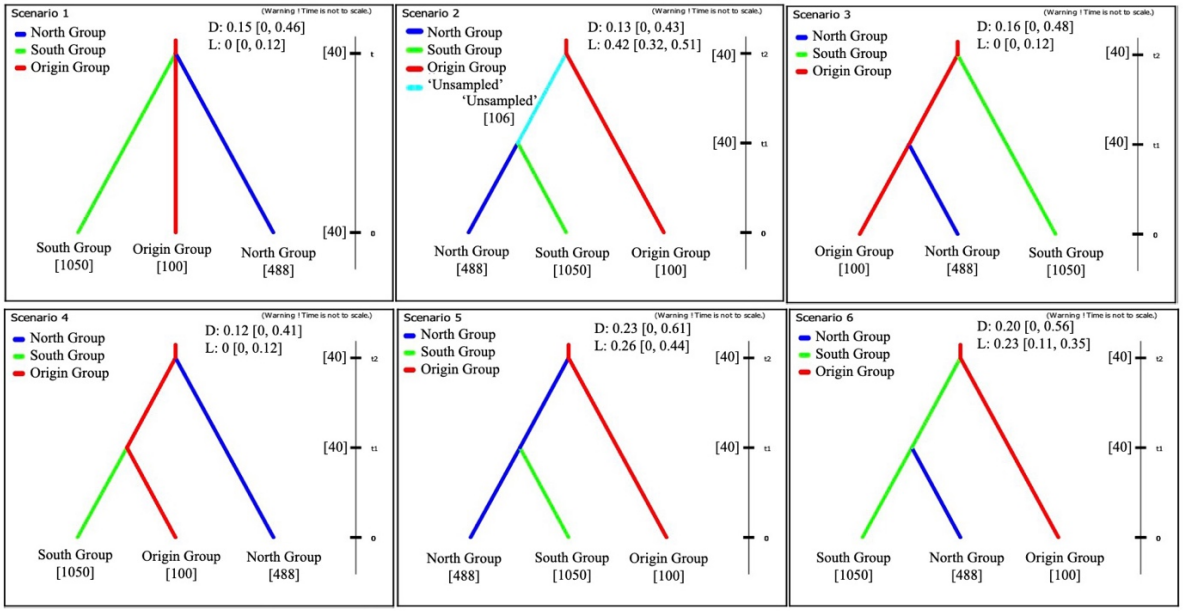

Figure 2. 8. All 6 scenarios tested for the *P. calleryana* study using DIYABC. In total, 6 hypothetical evolutionary scenarios were considered and tested with North Group, South Group, Origin Group, and 'Unsampled' population. In the figure, the numbers (construct number of individuals generated by DIYABC) given below each population name represent the effective population size for each population groups. For each of the tested scenarios, D and L indicate the values derived from direct and logistic regression approaches, respectively, with their probability values of 95% confidence intervals given in [].

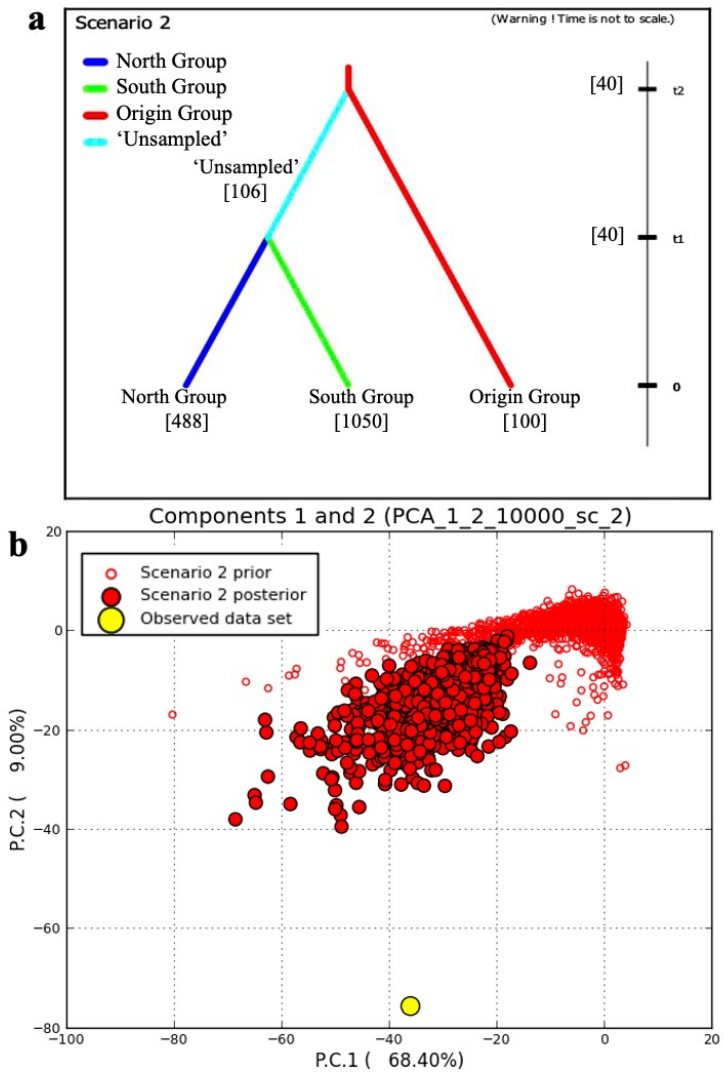

Figure 2. 9. Scenario 2: The best-supported scenario by DIYABC for the *P. calleryana* dataset. (a) Scenario 2 had the highest support; here the 'Unsampled' population (effective population size of about 106 generated individuals) split from the Origin Group (effective population size of about 100 individuals) at about 40 generations into the coalescent, which shortly later split into North Group (about 488 individuals of effective population size) and South Group (about 1050 individuals of effective population size) subpopulations. (b) Model-checking of the closest 1% simulated prior and posterior datasets was performed using PCA.

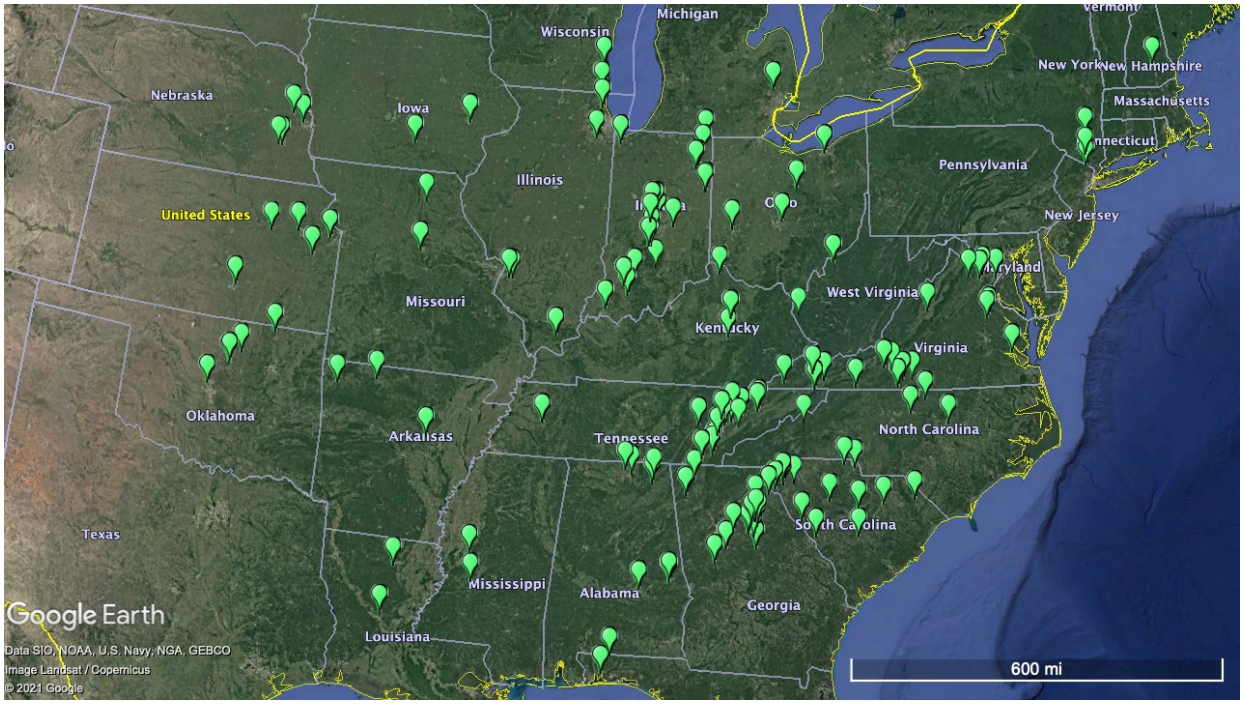

Figure 2. 10. Map of *P. calleryana* collection sites for the broad scale study. The bright-green marker points represent the locations of collection sites. The geographical coordinates for each collection site were provided by respective collectors. The scale line indicates the ground-level distance of 600 miles. The map was generated using Google Earth Pro version 7.3.

CONCLUSION

There is very limited information available about the genetic diversity and population structure of invasive *P. calleryana* across the U.S. This study acts as the first attempt to comprehensively address the genetic diversity and population structure of *P. calleryana* across its native and introduced (U.S.) ranges. Our first study across the species native ranges and the U.S. cultivars indicated the prevalence of high genetic diversity, high genetic differentiation, and high gene flow for *P. calleryana* populations. That study revealed China as the source of origin for invasive *P. calleryana* trees of the U.S, in agreement with the historical records available. Our second study across the introduced ranges of the southeastern U.S. indicated similar results, suggesting genetically diverse *P. calleryana* populations across the small geographic area. This study further supported China as the source of origin for *P. calleryana* populations of the U.S. that were diverged very shortly after introduction, leveraging the species' great evolutionary potential.

As *P. calleryana* is genetically diverse and has high invasive potential across its introduced ranges, it is important to ensure their control and management on time.

The management actions for invasive *P. calleryana* trees could include planting alternative trees such as *Viburnum prunifolium* (Blackhaw Viburnum), *Prunus americana* (Wild plum), *Ostrya virginiana* (Ironwood), *Cercis canadensis* (Eastern redbud), *Amelanchier arborea* (Serviceberry), *Carpinus caroliniana* (American hornbeam), *Cornus florida* (Flowering dogwood), *Nyssa sylvatica* (Black gum), *Cladrastis kentukea* (Yellowood), *Prunus virginiana* (Chokeberry), *Betula nigra* (River birch), *Asimina triloba* (Pawpaw), and *Halesia tetraptera* (Silverbell) which have great ornamental properties as of *P. calleryana*. Rootstock of *P. calleryana* produces many sprouts which in turn can cross-pollinate with the scion of the same plant. In such case, the management practices could include the rootstock manipulations for grafted scion so that rootstock plant cannot cross-pollinate with the scion. The production of

sterile *P. calleryana* trees through the breeding technique could also be a good option. The *P. calleryana* cultivar identification could also be helpful so that a particular cultivar be maintained in an area to avoid cross-pollination between different cultivars. Furthermore, multiple re-introductions of the plants from the native or other areas should be avoided to reduce the genetic diversity within invasive *P. calleryana* populations.

Our study provides a great prospect for future research on invasive *P. calleryana*. This study helped us understand and visualize the genetic diversity and population structure of the invasive *P. calleryana* in its native range and a small, localized area of the southeastern U.S. But considering our sample size and sample location, our study might be limited to provide the complete picture of *P. calleryana* distributed across the U.S. Hence, our data encouraged a broad-scale *P. calleryana* study covering a wide invasive geographical range. With this goal, we have already started working on a broad-scale study of *P. calleryana* using samples from 1,312 individual open-pollinated, wild-type trees that were collected across 125 collection sites of the U.S. (Fig 2.10). Isolated DNA samples from that collection were submitted for Reduced-Representation Sequencing. This approach will help us understand some other aspects of *P. calleryana* which were not covered in the two microsatellite-based studies such as thorniness of the species. Furthermore, this approach will help us compare the genomic characteristics of *P. calleryana* to our present findings and will help us better understand the invasive character of the species.

REFERENCES

Allendorf, F. W., & Lundquist, L. L. (2003). Introduction: population biology, evolution, and control of invasive species. *Conservation Biology, 17*(1), 24-30.

Agapow, P. M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes, 1*(1-2), 101-102.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Amos, W., Hoffman, J., Frodsham, A., Zhang, L., Best, S., & Hill, A. (2007). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes, 7*(1), 10-14.

Bao, L., Chen, K., Zhang, D., Cao, Y., Yamamoto, T., & Teng, Y. (2007). Genetic diversity and similarity of pear (Pyrus L.) cultivars native to East Asia revealed by SSR (simple sequence repeat) markers. *Genetic Resources and Crop Evolution, 54*(5), 959-971.

Bao, L., Chen, K., Zhang, D., Li, X., & Teng, Y. (2008). An assessment of genetic variability and relationships within Asian pears based on AFLP (amplified fragment length polymorphism) markers. *Scientia Horticulturae, 116*(4), 374-380.

Bell, R. L., & Zimmerman, R. H. (1990). Combining ability analysis of juvenile period in pear. *HortScience, 25*(11), 1425-1427.

Bossdorf, O., Auge, H., Lafuma, L., Rogers, W. E., Siemann, E., & Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia, 144*(1), 1-11.

Caballero, A., Rodríguez-Ramilo, S., Avila, V., & Fernández, J. (2010). Management of genetic diversity of subdivided populations in conservation programmes. *Conservation Genetics, 11*(2), 409-419.

Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics, 144*(4), 2001-2014.

Cornuet, J.-M., Ravigné, V., & Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1. 0). *BMC Bioinformatics, 11*(1), 1-11.

Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Estoup, A. (2014). DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics, 30*(8), 1187-1189.

Crowl, L. A., Crist, T. O., Parmenter, R. R., Belovsky, G., & Lugo, A. E. (2008). The spread of invasive species and infectious disease as drivers of ecosystem change. *Frontiers in Ecology and the Environment*, *6*(5), 238-246.

Cusrhes S., & Edwards, R. (1998). Potential Environmental Weeds in Australia. *Canberra (Australia): Queensland Department of Natural Resources*.

Cuizhi, G., & Spongberg, S. (2003). Pyrus. *Flora of China, 9*, 173-179.

Culley, T. M. & Hardiman N. A. (2007). The beginning of a new invasive plant: a history of the ornamental Callery pear in the United States. *BioScience, 57*(11), 956-964.

Culley, T. M. (2017). The rise and fall of the ornamental Callery pear tree. *Arnoldia, 74*(3), 2-11.

Culley, T. M., & Hardiman, N. A. (2009). The role of intraspecific hybridization in the evolution of invasiveness: a case study of the ornamental pear tree *Pyrus calleryana*. *Biological Invasions, 11*(5), 1107-1119.

Culley, T. M., Hardiman, N. A., & Hawks, J. (2011). The role of horticulture in plant invasions: how grafting in cultivars of Callery pear (*Pyrus calleryana*) can facilitate spread into natural areas. *Biological Invasions*, *13*(3), 739-746.

Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct K value using the ΔK method: When does K= 2? *Molecular Ecology, 29*(5), 862-869.

Darling, J. A., & folino-rorem, N. C. (2009). Genetic analysis across different spatial scales reveals multiple dispersal mechanisms for the invasive hydrozoan Cordylophora in the Great Lakes. *Molecular Ecology, 18*(23), 4827-4840.

Dickson, E., Arumuganathan, K., Kresovich, S., & Doyle, J. (1992). Nuclear DNA content variation within the Rosaceae. *American Journal of Botany, 79*(9), 1081-1086.

Dirr, M. A. (1990). *Manual of woody landscape plants: their identification, ornamental characteristics, culture, propagation and uses*: Stipes Publishing Co.

Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology, 17*(1), 431-449.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, 19*, 11-15.

Dray, S., & Dufour, A. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software, 22*(4), 1-20.

Dunphy, B., & Hamrick, J. (2005). Gene flow among established Puerto Rican populations of the exotic tree species, *Albizia lebbeck*. *Heredity, 94*(4), 418-425.

EDDMapS. (2019). Early Detection & Distribution Mapping System. *The University of Georgia - Center for Invasive Species and Ecosystem Health.* Retrieved from http://www.eddmaps.org/.

EDDMapS. (2021). Early Detection & Distribution Mapping System. *The University of Georgia - Center for Invasive Species and Ecosystem Health.* Retrieved from http://www.eddmaps.org/.

Ellstrand, N. C., & Schierenbeck, K. A. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences, 97*(13), 7043-7050.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology, 14*(8), 2611-2620.

Farrokhi, J., Darvishzadeh, R., Naseri, L., Azar, M. M., & Maleki, H. H. (2011). Evaluation of Genetic Diversity Among Iranian Apple (*'Malus domestica*'Borkh.) Cultivars and Landraces Using Simple Sequence Repeat Markers. *Australian Journal of Crop Science, 5*(7), 815-821.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources, 17*(1), 27-32.

Gaskin, J. F., Schwarzländer, M., Grevstad, F. S., Haverhals, M. A., Bourchier, R. S., & Miller, T. W. (2014). Extreme differences in population structure and genetic diversity for three invasive congeners: knotweeds in western North America. *Biological Invasions, 16*(10), 2127-2136.

Gayevskiy, V., Klaere, S., Knight, S., & Goddard, M. R. (2014). ObStruct: a method to objectively analyse factors driving population structure using Bayesian ancestry profiles. *PLoS One, 9*(1), e85196.

Gilman, E. F., & Watson, D. G. (1994). Pyrus calleryana 'Bradford':'Bradford'Callery Pear. Gainesville: Environmental Horticulture Department, Florida Cooperative Extensive Service, Institute of Food and Agricultural Sciences, University of Florida. Fact Sheet ST-537.

Goolsby, J. A., De Barro, P. J., Makinson, J. R., Pemberton, R. W., Hartley, D. M., & Frohlich, D. R. (2006). Matching the origin of an invasive weed for selection of a herbivore haplotype for a biological control programme. *Molecular Ecology, 15*(1), 287-297.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes, 5*(1), 184-186.

Hadziabdic, D., Wadl, P. A., Vito, L. M., Boggess, S. L., Scheffler, B. E., Windham, M. T., & Trigiano, R. N. (2012). Development and characterization of sixteen microsatellite loci for *Geosmithia morbida*, the causal agent of thousand canker disease in black walnut (*Juglans nigra*). *Conservation Genetics Resources, 4*(2), 287-289.

Hamrick, J. L., Godt, M. J. W., & Sherman-Broyles, S. L. (1992). Factors influencing levels of genetic diversity in woody plant species. In *Population Genetics of Forest Trees* (pp. 95-124): Springer.

Hamrick, J. L., & Godt, M. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351*(1345), 1291-1298.

Hardy, O., & Vekemans, X. (2015). SPAGeDi 1.5. *A program for Spatial Pattern Analysis of Genetic Diversity. User's manual http://ebe. ulb. ac. be/ebe/SPAGeDi_files/SPAGeDi_1. 5_Manual. pdf. Université Libre de Bruxelles, Brussells, Belgium*.

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes, 2*(4), 618-620.

Iketani, H., Manabe, T., Matsuta, N., Akihama, T., & Hayashi, T. (1998). Incongruence between RFLPs of chloroplast DNA and morphological classification in east Asian pear (*Pyrus* spp.). *Genetic Resources and Crop Evolution, 45*(6), 533-539.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405.

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology, 17*(18), 4015-4026.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *2*, e281.

Kato, S., Imai, A., Rie, N., & Mukai, Y. (2013). Population genetic structure in a threatened tree, *Pyrus calleryana* var. *dimorphophylla* revealed by chloroplast DNA and nuclear SSR locus polymorphisms. *Conservation Genetics Resources, 14*(5), 983-996.

Kim, H., Chang, K. S., & Chang, C.-S. (2010). EH Wilson's expedition to Korea from 1917 to 1919: resolving place names of his collections. *Journal of Japanese Botany, 85*(2), 99-117.

Kimura, T., Iketani, H., Kotobuki, K., Matsuta, N., Ban, Y., Hayashi, T., & Yamamoto, T. (2003). Genetic characterization of pear varieties revealed by chloroplast DNA sequences. *The Journal of Horticultural Science and Biotechnology, 78*(2), 241-247.

Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols, 3*(9), 1452.

Lalk, S., Hartshorn, J., & Coyle, D. R. (2021). Invasive woody plants and their effects on arthropods in the United States: challenges and opportunities. *Annals of the Entomological Society of America*.

Leger, E. A., & Espeland, E. K. (2010). Perspective: coevolution between native and invasive plant competitors: implications for invasive species management. *Evolutionary Applications, 3*(2), 169-178.

Leimu, R., Mutikainen, P., Koricheva, J., & Fischer, M. (2006). How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology, 94*(5), 942-952.

Lischer, H. E., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics, 28*(2), 298-299.

Liu, J., Zheng, X., Potter, D., Hu, C., & Teng, Y. J. B. s. (2012). Genetic diversity and population structure of Pyrus calleryana (Rosaceae) in Zhejiang province, China. *Biochemical Systematics Ecology, 45*, 69-78.

Lombaert, E., Guillemaud, T., & Deleury, E. (2018). Biases of STRUCTURE software when exploring introduction routes of invasive species. *Heredity, 120*(6), 485-499.

Meyer, L., Causse, R., Pernin, F., Scalone, R., Bailly, G., Chauvel, B., Le Corre, V. (2017). New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive Ambrosia species. *PLoS One, 12*(5), e0176197.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics, 89*(3), 583-590.

Nielsen, R., Tarpy, D. R., & Reeve, H. K. (2003). Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology, 12*(11), 3157-3164.

Nowicki, M., Boggess, S. L., Saxton, A. M., Hadziabdic, D., Xiang, Q.-Y. J., Molnar, T., Trigiano, R. N. (2018). Haplotyping of *Cornus florida* and *C. kousa* chloroplasts: Insights into species-level differences and patterns of plastic DNA variation in cultivars. *PLoS ONE, 13*(10), e0205407.

Nowicki, M., Houston, L. C., Boggess, S. L., Aiello, A. S., Payá-Milans, M., Staton, M. E., Trigiano, R. N. (2020). Species diversity and phylogeography of *Cornus kousa* (Asian dogwood) captured by genomic and genic microsatellites. *Ecology and Evolution, 10*(15), 8299-8312.

Nowicki, M., Zhao, Y., Boggess, S. L., Fluess, H., Payá-Milans, M., Staton, M. E., Trigiano, R. N. (2019). *Taraxacum kok-saghyz* (rubber dandelion) genomic microsatellite loci reveal modest genetic diversity and cross-amplify broadly to related species. *Scientific Reports, 9*(1), 1-17.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Wagner, H. (2013). Package 'vegan'. *Community Ecology Package, Version, 2*(9), 1-295.

Ony, M. A., Nowicki, M., Boggess, S. L., Klingeman, W. E., Zobel, J. M., Trigiano, R. N., & Hadziabdic, D. (2020). Habitat fragmentation influences genetic diversity and differentiation: Fine-scale population structure of *Cercis canadensis* (eastern redbud). *Ecology and Evolution, 10*(8), 3655-3670.

Pappert, R. A., Hamrick, J., & Donovan, L. A. (2000). Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany, 87*(9), 1240-1245.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 20*(2), 289-290.

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*(1), 288-295.

Pimentel, D., Lach, L., Zuniga, R., & Morrison, D. (2000). Environmental and economic costs of nonindigenous species in the United States. *BioScience, 50*(1), 53-65.

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics, 52*(3), 273-288.

Pons, O., & Petit, R. (1996). Measwring and testing genetic differentiation with ordered versus unordered alleles. *Genetics, 144*(3), 1237-1245.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Purugganan, M. D. J. M. E. (2000). The molecular population genetics of regulatory genes. *9*(10), 1451-1461.

Randall, R. P. (2002). A global compendium of weeds. *R.G. and F.J. Richardson, Victoria, Australia.*

Rehder, A. (1915). *Synopsis of the Chinese species of Pyrus.* Paper presented at the Proceedings of the American Academy of Arts and Sciences.

Reichard, S. H., & Hamilton, C. W. (1997). Predicting invasions of woody plants introduced into North America: Predicción de Invasiones de Plantas Leñosas Introducidas a Norteamérica. *Conservation Biology, 11*(1), 193-203.

Reichard, S. H., & White, P. (2001). Horticulture as a pathway of invasive plant introductions in the United States: most invasive plants have been introduced for horticultural use by nurseries, botanical gardens, and individuals. *BioScience, 51*(2), 103-113.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran R, 538*.

Rubtsov, G. (1944). Geographical distribution of the genus *Pyrus* and trends and factors in its evolution. *The American Naturalist, 78*(777), 358-366.

Sapkota, S., Boggess, S. L., Trigiano, R. N., Klingeman, W. E., Hadziabdic, D., Coyle, D. R., & Nowicki, M. (2021). Microsatellite loci reveal genetic diversity of Asian Callery pear (*Pyrus calleryana*) in the species native range and in the North American cultivars. *Life*, *11*(6), 531.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., Ellstrand, N. C. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics, 32*(1), 305-332.

Santamour JR, F. S., & Demuth, P. (1980). Identification of Callery pear cultivars by peroxidase isozyme patterns. *Journal of Heredity, 71*(6), 447-449.

Santamour Jr, F. S., & McArdle, A. J. (1983). Checklist of cultivars of Callery pear (*Pyrus calleryana*). *Journal of Arboriculture*.

Schierenbeck, K. A., & Ainouche, M. L. (2006). The role of evolutionary genetiocs in studies of plant invasions. In *Conceptual ecology and invasion biology: reciprocal approaches to nature* (pp. 193-221): Springer.

Sexton, J. P., McKay, J. K., & Sala, A. (2002). Plasticity and genetic diversity may allow saltcedar to invade cold climates in North America. *Ecological Applications, 12*(6), 1652-1660.

Shannon, C. E. (2001). A mathematical theory of communication. ACM SIGMOBILE Mob. *Computing Communications Review, 5*(1), 3-55.

Short, K. H., & Petren, K. (2011). Fine-scale genetic structure arises during range expansion of an invasive gecko. *PLoS One, 6*(10), e26258.

Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*(4148), 688-688.

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

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics, 139*(1), 457-462.

Stoddart, J. A., & Taylor, J. F. (1988). Genotypic diversity: estimation and prediction in samples. *Genetics, 118*(4), 705-711.

Swearingen, J., Reshetiloff, K., Slattery, B., & Zwicker, S. (2002). Plant invaders of mid-Atlantic natural areas. In: National Park Service and U.S. Fish and Wildlife Service. Washington, DC. 168pp.

Taberlet, P., Fumagalli, L., WUST-SAUCY, A. G., & COSSON, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology, 7*(4), 453-464.

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

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research, 40*(15), e115-e115.

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

Vincent, M. A. (2005). On the spread and current distribution of *Pyrus calleryana* in the United States. *70*(1), 20-32.

Vogt, J. T., Coyle, D. R., Jenkins, D., Barnes, C., Crowe, C., Horn, S., & Roesch, F. A. (2020). Efficacy of five herbicide treatments for control of *Pyrus calleryana*. *Invasive Plant Science and Management*, *13*(4), 252-257.

Wang, X., & Wang, L. (2016). GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science, 7*, 1350.

Warrix, A. R., & Marshall, J. M. (2018). Callery pear (*Pyrus calleryana*) response to fire in a managed prairie ecosystem. *Invasive Plant Science and Management, 11*(1), 27-32.

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications, 18*(1), 315-322.

Whitehouse, W. E., Creech, J., & Seaton, G. (1963). Bradford ornamental pear—a promising shade tree. *American Nurseryman, 117*, 7-8.

Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., Postman, J. D. (2018). Diversification and independent domestication of Asian and European pears. *Genome Biology, 19*(1), 1-16.

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Wang, H. (2013). The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research, 23*(2), 396-408.

Yamamoto, T., Kimura, T., Sawamura, Y., Kotobuki, K., Ban, Y., Hayashi, T., & Matsuta, N. (2001). SSRs isolated from apple can identify polymorphism and genetic diversity in pear. *Theoretical and Applied Genetics, 102*(6-7), 865-870.

Yamamoto, T., Kimura, T., Sawamura, Y., Manabe, T., Kotobuki, K., Hayashi, T., Matsuta, N. (2002). Simple sequence repeats for genetic analysis in pear. *Euphytica, 124*(1), 129.

Yamamoto, T., Kimura, T., Shoda, M., Ban, Y., Hayashi, T., & Matsuta, N. (2002). Development of microsatellite markers in the Japanese pear (*Pyrus pyrifolia* Nakai). *Molecular Ecology Notes, 2*(1), 14-16.

Yuanwen Teng, Y., Tanabe, K., Tamura, F., & Itai, A. (2001). Genetic relationships of pear cultivars in Xinjiang, China, as measured by RAPD markers. *The Journal of Horticultural Science and Biotechnology, 76*(6), 771-779.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics, 29*(21), 2669-2677.

VITA

Shiwani Sapkota was born in Chitwan, Nepal on July 08, 1994. She had her Bachelor of Science (BSc) in Agriculture from Agriculture and Forestry University (AFU), Chitwan, Nepal in December 2017. After her BSc completion, she worked as an instructor in a technical school of Nepal for six months. She then moved to the U.S. and joined the University of Tennessee for her Master's degree in bioinformatics and Genomics in the Department of Entomology and Plant Pathology under the supervision of Dr. Marcin Nowicki.