Summer 2024

Genetic Diversity of Paxistima canbyi A. Gray: Conserving A Rare Plant Species Endemic to the Eastern United States

Isaac Buabeng
ikb001@bucknell.edu

Genetic Diversity of *Paxistima canbyi* A. Gray: Conserving A Rare Plant Species Endemic to the Eastern United States

By

Isaac K. Buabeng

A Thesis

Presented to the Faculty of Bucknell University

In partial fulfillment of the requirements for the Degree of Master of Science in Biology

July 2024

Approved by:

Prof. Christopher T. Martine, **Thesis Advisor** …  ……..

Dr. Scott Schuette, **Committee Member** …  ……

Dr. Tanisha M. Williams, **Committee Member** …  ……

Dr. Melody P. Sain, **Committee Member** ……  ……

Prof. Mark Haussmann, **Chair, Department of Biology** ……  ……

Acknowledgments

First, I want to thank Prof. Chris Martine for your guidance and support over the past two years. Thank you for the encouragement and always being present. I would also like to thank Dr. Melody Sain and Dr. Tanisha Williams, for their immense support and openness in teaching me. I have gained so many valuable skills through your guidance. A special thank you to Dr. Scott Schuette for pushing me to look more closely into natural history, as well as for being an incredibly supportive fieldwork partner.

To Prof. Sarah Lower, Prof. Ken Field, and Dr. Terry Campbell in the Bucknell Department of Biology, I say thank you for challenging me and for providing me with the necessary skills and knowledge needed for this project. To Mike Harvey and Jeremy Dreese at BisonNet thanks for your support and helpful discussions on HPC related questions and issues, and to Beth Williams and Janine Glathar in Bucknell GIS and Digital Pedagogy, thank you very much for the help with my maps and GIS. I appreciate it.

My sincerest gratitude to the Martine Lab, especially P. J. Newhart and Beth Williams for being wonderful companions during field work and extractions. And to all the other lab members, I will always cherish our hearty conversations.

Thank you to Dr. Scott Schuette with the Western Pennsylvania Conservancy and Pennsylvania Natural Heritage Program for providing information regarding locations of populations and bridging the connections to the people connected to the populations and sites. Many thanks to Rachel Goad of the Pennsylvania Natural Heritage Program, Mike Powell of The Nature Conservancy - West Virginia, John Townsend of the Virginia Department of Conservation and Recreation, Dr. Melody Sain, Dr. Scott Schuette, and Prof. Chris Martine for assistance during fieldwork and sample collection. A special thank you to Tara Littlefield for

help with sample collection and the valuable insight provided on the Natural History of this species.

I am also grateful to my colleague graduate students and friends in the biology department for their companionship over the past two years. I am thankful to Derick Adu, and my family: Sarah and Abbie for their immense support, love, and encouragement. Me da mo nyinaa ase! Nyame nhyira mo pii!

Funding for this project came from the PA DCNR Wild Resources Conservation Program (project 17571), with additional support from the David Burpee Endowment, Bucknell Biology Graduate Program and the Bucknell Biology Department. This work was also funded in part by travel grants awarded to I. K. B by the Association of Plant Taxonomists.

Table of Contents

List of Tables

List of Figures

Abstract

Premise: The glacial cycles of the Pleistocene period likely had a profound impact on the genetic diversity and phylogeographic patterns of numerous plant taxa in North America. However, the genetic diversity of rare, limestone-endemic woody species has not been highlighted. This study assesses genetic diversity in *Paxistima canbyi* A. Gray (Canby's mountain lover or rat stripper), a rare limestone endemic small woody plant.

Methods: We investigated the genetic diversity of *P. canbyi* using genomic Single Nucleotide Polymorphism (SNP) data from 165 individuals from 14 populations sampled from across the entire range distribution. We first; d*e novo* assembled the SNP dataset using ipyrad and then estimated Ploidy levels of *P. canbyi* using the observed heterozygosity and the ratio of DNA sequences through a Bayesian assignment method-based R package, gbs2ploidy. We subsequently use a hard-filtered SNP dataset containing only biallelic SNPs to make inferences on the genetic diversity of the extant *P. canbyi* populations.

Results: Our results reveal *P. canbyi* is a polyploid species. Also, genetic diversity within the extant populations seems to be moderately high with observed heterozygosity higher than expected. Also, there appears to be a significant level of differentiation between populations with the overall F_{ST} calculated to be 0.390. Except for two populations in West Virginia who (even with large numbers of individuals**)** appear to be inbred, most *P. canbyi* populations are not inbred as heterozygotes were observed to be abundant (F_{IS} = -0.273). Using a cross-entropy criterion in a population structure analysis, we also identified fourteen genetic clusters with all populations except four, which form two unique genetic clusters.

Conclusion: Almost all *P. canbyi* populations are distinct genetic units that share some ancestral genetic variation even under genetic isolation. Despite promising inferences regarding genetic health of the species, some populations may not persist without control of the invasive scale insect. While all populations have high genetic diversity and are not yet inbred, the genetic isolation observed across all the populations within the disjunct distributions indicate that these populations may be at risk of future inbreeding depression especially within the Smoke Hole and Blue Rock populations in West Virginia.

Chapter 1: Background on *Paxistima canbyi* and Issues on the Conservation of the Species

Introduction

As articulated by Soulé (1985), conservation biology entails leveraging insights derived from biological sciences to address challenges faced by species, communities, and ecosystems disrupted by natural and anthropogenic activities. In addressing these challenges, species are often classified into various 'statuses' that use data collected on the existing populations of such species (International Union for Conservation of Nature [IUCN], 2012; NatureServe, 2022). Species within Latin America, North America, and Canada are classified based on statutes generated by NatureServe, which provides such ranks in conformity with internationally accepted standards (NatureServe, 2022). These ranks as defined by NatureServe are meant to assess the risk of extinction for species at the global, national, and sub-national scales. They are based on a scale of 1 to 5 and are assigned by NatureServe scientists or a designated lead office in the NatureServe network. These ranks or statuses and their interpretations can be found in Table 1.

Driven by the scientific community's commitment to addressing the escalating threats posed by global environmental changes, biodiversity conservation has undergone a transformative shift. This evolution emphasizes the conservation of populations as distinct units with evolutionary potential, departing from the traditional focus on safeguarding habitats (Allendorf et al., 2013; Chokheli et al., 2020). Central to this paradigm shift is the integration of genetic information into conservation efforts, marking a new era in biodiversity conservation (Theissinger et al., 2023). Recognizing the vital role of genomic data obtained through modern sequencing methods, practitioners aim to inform policies for managing species populations,

particularly those facing rarity challenges (Delgado et al., 2007; Laikre et al., 2010; Allendorf et al., 2010; Frankham, 2010; Allendorf et al., 2013; Theissinger et al., 2023).

In the United States, there are approximately 8,840 plant taxa formally recognized as rare (USFS, 2024), a considerable number of which face ongoing threats resulting from reduced genetic variation within small populations (Falk & Holsinger, 1991; Barrett & Kohn, 1991; Godt et al., 1996; Honnay & Jacquemyn, 2007). This situation underscores the urgency of strategically utilizing genomic data in conservation efforts. Unraveling complexities such as genetic bottlenecks, founder effects, inbreeding, and genetic drift not only inform targeted interventions; but it also contributes to mitigating these challenges and promoting the long-term viability of rare plant species because not addressing these challenges can reduce genetic variation within their populations.

Paxistima canbyi A. Gray (Celastraceae), also referred to as rat stripper or Canby's mountain lover, is a clonally-growing shrub of eastern North America currently ranked as globally-imperiled (G2) because it is at an elevated risk of extinction or collapse due to its restricted range, few populations or occurrences, and steep population declines from severe threats posed by insect herbivory and deer grazing. The species is distributed across two major subpopulations (Fig. 1) on limestone outcrops and shale barrens: One group of central Appalachian Mountains populations found in western Virginia, eastern West Virginia, western Maryland, and southern Pennsylvania; and the other set occurring on the western edges of the Allegheny - Cumberland plateaus of central Kentucky, southern Ohio, and central Tennessee (Fig. 2) (Littlefield, 2021; NatureServe, 2022). The occurrence of disjunct populations in the range of a species is ostensibly linked to patterns of colonization and vicariance (Fryxell, 1966; Watson et al., 2002; Mohn et al., 2021), which suggests that the disjunction observed in the

range of *P. canbyi* may have resulted from either the colonization of new regions through long-distance dispersal or from a vicariance event that may have partitioned a previously continuous distribution into two or more subpopulations.

Eastern North America is recognized as a global hotspot for rare plants and species of conservation concern (Murdock, 1995; Marcinko, 2007; Crain et al., 2011; Fortney et al., 2015; Allen & Lendemer, 2016). Climatological and geologic activities during the Pleistocene are thought to have had significant impacts on the region (Delcourt, 1998; Soltis et al., 2006; Hyseni & Garrick, 2019, Mohn et al., 2021), due largely to environmental changes brought on by the Last Glacial Maximum (LGM), approximately 22,000 years ago. The Pleistocene is said to have contributed to the contraction and isolation of plant populations into refugia across eastern North America. These refugial contractions enabled species populations to persist despite the harsh environmental and climatological conditions of the period, which later allowed for their post-LGM expansion, and recolonization of suitable habitats (Soltis et al., 2006; Lyman & Edwards, 2022).

Because of the interplay of LGM climate challenges and refugial retreat, the Pleistocene is frequently characterized as the origin of various phylogeographic discontinuities in eastern North America, typically linked to significant geomorphological features such as the Appalachian Mountains, Apalachicola River, Mississippi river, and others, as detailed by Soltis et al. (2006). These discontinuities, as further reviewed by Lyman & Edwards (2022), include (a) the maritime-Atlantic vs. Gulf Coast discontinuity; (b) the Apalachicola River; (c) the Tombigbee River discontinuity; (d) the Mississippi River discontinuity; (e) the Apalachicola River and Mississippi River discontinuity; and (f) the Appalachian Mountains discontinuity. However, other studies have argued that some discontinuities may predate the Pleistocene, with

topographic features like the Mississippi River and the Appalachian Mountains acting as historical barriers to gene flow (Burbrink et al., 2008; Herman and Bouzat, 2016; Wang et al., 2023). For instance, the Appalachian Mountain system formed over 480 million years ago and underwent significant geologic cycles, with the last occurring in the mid-Miocene, approximately 16 million years before the LGM (Wang et al., 2023). This challenges the notion that Pleistocene events solely influenced species lineages in the Appalachian Mountain discontinuity.

Population genetics research assessing the historical processes responsible for the present-day geographic distributions of rare species within eastern North America has historically explored such questions with traditional genetic markers such as DNA microsatellites, amplified fragment length polymorphisms (AFLPs), and other older methods that are known to operate under different sets of evolutionary assumptions (Table 2) (Andrews et al., 2016; Peterson et al., 2012). For instance, DNA microsatellite data was used to detect a high level of differentiation in Carolina hemlock (*Tsuga caroliniana,* Engelm.), an imperiled southern Appalachian endemic conifer (Potter et al., 2017). Also, microsatellite data was used to discover phylogeographic domains across the populations of the North American common ragweed (*Ambrosia artemisiifolia* L.) (Kočiš Tubić et al., 2015). However, the application of Next Generation Sequencing (NGS) to assess the genetic diversity and structure of the populations of rare species and species of conservation concern is fast becoming a mainstay in biodiversity conservation (e.g., Capblancq et al., 2020; Hayes, 2021; McDonell et al., 2021; Moore et al., in review). This is largely because, the adoption and use of high-quality long-read and short-read sequencing, and the associated bioinformatic technologies used to analyze those data, facilitate

genome sequencing and assembly for any species and are effective tools in characterizing biodiversity (Theissinger et al., 2023).

To investigate the genetic diversity of *P. canbyi*, I used data generated via the Genotype-By-Sequencing (GBS) approach (Elshire et al., 2011). GBS makes use of restriction enzymes in cutting DNA into smaller fragments, focusing on reducing the complexity of the organism's genome by generating genome-wide high-throughput sequencing data and obtaining a large number of genetic polymorphism tag sequences to fully represent the whole genome information of the species (Elshire et al., 2011). These regions often harbor relevant genetic variations like single nucleotide polymorphisms (SNPs) that can generate useful genetic information about the species such as the specific genotypes within a population. GBS has been proven to be valuable for assessing the population genomics of non-model organisms particularly rare species (e.g., Rowe et al., 2011; Liu et al., 2020; McDonell et al., 2021; Ye et al., 2021; Moore et al., in review), and has been found to yield more data that is more representative of the entire genome than microsatellites, AFLPs, and other traditional sequencing approaches (Andrews et al., 2016; Peterson et al., 2012). Using NGS methods such as GBS, to assess geographical distribution patterns and to inform the conservation of rare plants, is valuable for establishing best management practices.

Therefore, by ascertaining the genetic structure underlying the existing populations of *P. canbyi* using genotypes from SNP data generated via the GBS sequencing method, I aim to investigate the genetic diversity among the existing populations of *P. canbyi.* I hypothesize that: **i)** the sampled populations exhibit some genetic structure even under limited gene flow, and **ii)** gene flow among the current populations is low due to the existing geographical barriers and the clonal nature of the species.

Background of Species

Taxonomy and Species Description

Paxistima canbyi A. Gray is an evergreen shrub of the family Celastraceae (Staff-tree or Bittersweet family) (Ma et al., 2016). It is one of only two species within the genus *Paxistima*, with the other being *Paxistima myrsinites* (Pursh) Raf. (Ma et al., 2016) of western North America. *Paxistima canbyi* is a low, clonally growing, decumbent shrub with opposite evergreen leaves that measure 5–20 × 2–4 mm (about 0.16 in) with a blunt apex. The leaves of the plant are sessile, and leaf margins are revolute, obscuring the widely spaced tiny teeth that are sometimes absent. Plants can grow to 4 dm (Ma et al., 2016), but it has been observed to achieve this only in cultivation. Wild plants are usually if not always shorter (Stoutamire, 1991; Goad et al., 2019).

Paxistima canbyi flowers from late March through April, and buds begin formation in the early parts of the summer (Stoutmire, 1991). The flowers occur on short peduncles either in clusters of up to five or singly on short pedicels. The sepals and petals of the flower are 4-merous, no longer than 2.5 mm (about 0.1 in). The petals are yellowish-green in color and sometimes partly or wholly maroon. Fleshy nectaries occur between the four stamens in each flower (Weakley et al., 2012; Ma et al., 2016). In cultivation, *P. canbyi* flowers have been observed to have a faint unpleasant odor that attracts sciarid flies, which are thought to be possible pollinators (Stoutmire, 1991). However, observations of this phenomenon in the wild have not been reported.

Paxistima canbyi has been observed to take up to two years to flower from seedling in cultivation; and when it does; it typically forms fruits with usually a low seed set which reportedly fall off within 24 hours of fruit set (Stoutamire, 1991; Goad et al., 2019). This

phenomenon of having a low seed set in cultivation has been largely attributed to two possible phenomena: 1) Low pollination efficiency, where the stigmas in the flowers of *P. canbyi* have been observed to be relatively dry and devoid of pollen oil droplets (pollenkitt), which together limit the attachment of pollen to pollinators and the stigmatic surface (Stoutamire, 1991); and 2) The lack of appropriate mates within populations due to the inability to produce zygotes after self-pollination in fertile co-sexual individuals (Stoutamire 1991; White & Williams, 2017; Goad et al., 2019) where most of the existing populations are expected to contain only a small number of genetically different individuals (Goad et al., 2019; White & Williams, 2017). For example, in an initial attempt to assess the genetic diversity of some *P. canbyi* populations in Ohio and West Virginia, 119 individuals across 9 populations were sampled across both states. A microsatellite analysis revealed only 35 unique multi-locus genotypes (MLGs) across the 119 individuals sampled, indicating that levels of genetic diversity are quite low in the 9 sampled populations (White & Williams, 2017).

Threats

The biggest threat to the extant populations of *Paxistima canbyi* is the threat posed by suspected low genetic variation within the extant populations. This comes off the back of an initial assessment made on *P. canbyi* populations in Ohio and West Virginia where only a few MLGs are identified (White & Williams, 2017). *Paxistima canbyi* also faces a severe threat of herbivory from the euonymus scale insect (*Unapsis euonymi*). The euonymus scale insect is a sexually dimorphic (i.e., sexes of the same species exhibit different morphological structures), invasive insect native to Japan and China that attacks vines and evergreen shrubs. It colonizes plants by first attacking the leaves, causing yellow spots to appear on them, and later moving on to other parts of the plant. Scale insects feed by inserting microscopic, thread-like mouthparts

into the plant and sucking out plant sap. Euonymus scale insects usually have two or three generations per year, where tiny crawlers hatch and emerge from the mother's shell in April, May, and June. A second brood hatches in the late summer, and a partial third brood may emerge later (Cockfield & Potter, 1990; Mathada et al., 2003).

Although this scale is small, infestations are often dense and visible, particularly with heavy populations where males usually greatly outnumber females (Cockfield & Potter, 1990; Mathada et al., 2003). In dense infestations, clusters of females typically settle on stems and secrete a grey test while males infest both stems and leaves and produce a white, cottony test. Males are typically more abundant than females during infestations. The development of the adult male or female takes between three and four weeks from the moment the eggs hatch. Before depositing eggs, the females survive for at least 4 to 5 more weeks with an extra 4 to 6 weeks (about 1 and a half months) passing during the ovipositional stage (Cockfield & Potter, 1990). Female scale insects are brown and shaped like tiny oyster shells, growing up to 1/16-inch in length. Male scale insects are white and grow up to 1/32 inches long. Young crawlers are yellow and very tiny before they create the protective scale covering. Scales are quite difficult to manage due to their reproductive phenology, but horticultural oils have proven effective in management (Goad et al., 2019).

Other members of the Celastraceae such as *Euonymous* spp., and *Celastrus orbiculatus* along with other problematic invasive species like *Berberis thunbergii* often co-occur with *P. canbyi.* The presence of these species is contributing to the spread of the scale. The availability and presence of alternative hosts such as these plants may be helping to keep the scale insect populations still active all year round, even though the scale insect is noted to have two generations in a year (Pellitteri, 2012). Serving as intermediary hosts for the insect, these plants

can provide alternative nutrition and accommodate the scale insects at periods of the year when most other plants may have lost their leaves or branches. The evergreen nature of *P. canbyi,* can provide shelter and protection all year for various insects, particularly in their overwintering stages. This can create favorable conditions for their populations to thrive (R. Goad, personal communication, September 9, 2022).

As a result of these threats and its limited range, *P. canbyi* is classified as globally imperiled (G2) by NatureServe. It is also listed as critically imperiled (S1) in all of the states it inhabits (i.e., in Pennsylvania, Maryland, Kentucky, Ohio, and Tennessee), except for Virginia and West Virginia, where it is classified as imperiled (S2) due to the viability of the populations in both states. To ensure the continued existence and preservation of the extant *P. canbyi* populations, all its known populations are monitored by state Natural Heritage Programs in charge of managing data on species deemed to be essential elements of biodiversity (Natural Heritage Programs). A critical component of maintaining up-to-date conservation ranks is having current and accurate data regarding population sizes and habitat conditions. This information is gathered during field surveys that are a core function of heritage programs. As part of my thesis research, surveys were conducted with botanists from the Pennsylvania Natural Heritage Program and West Virginia Nature Conservancy, where we performed NatureServe vegetation assessments of the habitats and made tissue collections for population genomics work.

Species Range

As described above, the range of *P. canbyi* is divided into two main disjunct distributions, one in the interior low plateaus (i.e., the Allegheny and Cumberland plateaus), and the other in the central Appalachians (Fig. 2) (NatureServe, 2022). *Paxistima canbyi* is said to have a total ground cover of 20,000 - 200,000 square km across its distribution, with

approximately 67 historical occurrences (64 extant, 31 with good viability). The northernmost and southernmost extents of the distribution are represented by occurrences in south-central Pennsylvania and northern Tennessee, respectively. Information on what might have historically caused the restrictions in the distribution of these populations is unknown (NatureServe, 2022).

Species Habitat

The habitat of *P. canbyi* is thought to be a remnant of Arcto-Tertiary woods (i.e. a hypothesized floral assemblage that once covered the Northern Hemisphere, from roughly the late Mesozoic to mid-Cenozoic Eras), persisting in locations where geo-climatic conditions have allowed its continuity—often referred to as refugia (Navaro & Blackwell, 1990; Littlefield, 2008; Goad et al., 2019). *Paxistima canbyi* is naturally found exclusively on calcareous bedrocks such as limestone, dolomite, or calcareous shale, typically thriving in shallow soils covering these substrates (Stoutamire, 1991). Within these environments, plant growth is primarily constrained by edaphic factors, and habitat choice or preference may also be influenced by fire (Goad et al., 2019).

Although sometimes cultivated as a ground cover in loamy soil, *P. canbyi* also does well in rocky-sandy soil and tolerates shade. It can also grow in open or partially shaded places (NatureServe, 2022). Some populations also exist in what can be referred to as mid-successional habitats such as open canopied woodlands and glades that may have been maintained historically by fire or grazing by herbivores (Goad et al., 2019). The nature of the habitats of *P. canbyi* as being in their early- to mid-successional state conceals how little tolerance this species has for disturbances such as overgrazing, herbivory, pests, and other invasive species (Stoutamire, 1991; Ma et al., 2016).

Significance and Applications of Study

Insights from the results of my project will directly inform the nationwide community of conservation practitioners of the genetic health of the existing *P. canbyi* populations. It will provide information about genetic variability and gene flow between populations, and also provide information about the genetic delineation of *P. canbyi* populations. This work would contribute to the precedent already set for genetic work which enhances the protection and management of endangered species populations in this region (e.g., McDonnell et al., 2021; Hayes, 2021; Moore et al., in review).

Another main application will be reassessing the number of Elements of Occurrences (EOs) [i.e. an area of land and/or water in which a species or natural community is, or was, present] in the various Natural Heritage program documents. Natural Heritage programs as part of their duties are obligated to make a comprehensive documentation of their states' ecological resources, especially rare and endangered species. Additionally, natural history observations such as herbivory impacts from site visits to the existing populations can also be used in the NatureServe Conservation Rank Calculator and Climate Change Vulnerability Index used by the various Heritage Programs (Master et al., 2012; Tuberville et al., 2015).

Additionally, data from my work will determine how genetic populations of *Paxistima canbyi* exist in each state across its range, which could lead to an enhanced process of management for the rare species. Data from my work will enhance species management at the state level and globally through NatureServe, a bigger international network of natural heritage programs sharing the same methods of data collection and management. NatureServe then consolidates and manages all the data and information collected by the various Natural Heritage

Programs and affiliates to ensure the effective management of species on a global scale (Regan et al., 2004)

References

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., & Hohenlohe, P.A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81–92.

Allen, J. L., & Lendemer, J. C. (2016). Climate change impacts on endemic, high-elevation lichens in a biodiversity hotspot. Biodiversity and Conservation, 25, 555-568.

Allendorf, F. W., Funk, W. C., Aitken, S. N., Byrne, M., & Luikart, G. (2013). Conservation and the genomics of populations. Oxford University Press.

Barnard-Kubow, K. B., Debban, C. L., & Galloway, L. F. (2015). Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. *American Journal of Botany*, *102*(11), 1842-1853. https://doi.org/10.3732/ajb.1500267

Barrett, K., Nibbelink, N. P., & Maerz, J. C. (2014). Identifying priority species and conservation opportunities under future climate scenarios: amphibians in a biodiversity hotspot. Journal of Fish and Wildlife Management, 5(2), 282-297.

Barrett, S. C., & Kohn, J. R. (1991). Genetic and evolutionary consequences of small population size in plants: implications for conservation. Genetics and conservation of rare plants, 3-30.

Brant, S. V., & Ortí, G. (2003). Phylogeography of the northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): Past fragmentation and postglacial recolonization. Molecular Ecology, 12(6), 1435-1449. https://doi.org/10.1046/j.1365-294X.2003.01789.x

Capblancq, T., Butnor, J. R., Deyoung, S., Thibault, E., Munson, H., Nelson, D. M., Fitzpatrick, M. C., & Keller, S. R. (2020). Whole-exome sequencing reveals a long-term decline in effective population size of red spruce (*Picea rubens*). Evolutionary Applications, 13(9), 2190-2205. https://doi.org/10.1111/eva.12985

CBD, U.N. (1992). United Nations Convention on Biological Diversity. Abgerufen am, 5, 17.

Caterino, M. S., & Recuero, E. (2023) Shedding light on dark taxa in sky-island Appalachian leaf litter: Assessing patterns of endemicity using large-scale, voucher-based barcoding. Insect Conservation and Diversity. https://doi.org/10.1111/icad.12697

Chokheli, V. A., Dmitriev, P. A., Rajput, V. D., Bakulin, S. D., Azarov, A. S., Varduni, T. V., Stepanenko, V. V., et al. (2020). Recent development in micropropagation techniques for rare plant species. Plants, 9(12), 1733. MDPI AG. Retrieved from http://dx.doi.org/10.3390/plants9121733

Cockfield S.D, Potter D.A., (1990). Euonymus scale patterns of damage to woody plants. Journal of Arboriculture 16:239–241.

Crain, B. J., White, J. W., & Steinberg, S. J. (2011). Geographic discrepancies between global and local rarity richness patterns and the implications for conservation. Biodiversity and Conservation, 20, 3489-3500.

Delcourt, P. A., & Delcourt, H. R. (1998). Paleoecological insights on conservation of biodiversity: A focus on species, ecosystems, and landscapes. Ecological Applications, 8(4), 921-934.

Delgado, P., L.E. Eguiarte, F. Molina-Freaner, E.R. Alvarez-Buylla, & D. Piñero (2008). Using phylogenetic, genetic and demographic evidence for setting conservation priorities for Mexican rare pines. Biodiversity and Conservation 17:121–137.

Erlandson, S. K. (2018). History, dispersal limitation, and environment shape the current and future ranges of forest herbs of the Southern Appalachians (Doctoral dissertation, University of Minnesota).

Erlandson, S. K., Bellemare, J., & Moeller, D. A. (2021). Limited range-filling among endemic forest herbs of eastern North America and its implications for conservation with climate change. Frontiers in Ecology and Evolution, 9, 751728.

Elliott, K. J., & Hewitt, D. (1997). Forest species diversity in upper elevation hardwood forests in the southern Appalachian Mountains. Castanea, 32-42.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one, 6(5), e19379.

Falk, D. A., & Holsinger, K. E. (Eds.). (1991). Genetics and conservation of rare plants. Oxford University Press, USA.

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation 143:1919–1927.

Fortney, R. H., Stephenson, S. L., & Rentch, J. S. (2015). Rare plant communities in Canaan Valley, West Virginia. Southeastern Naturalist, 14(sp7), 121-135.

Fryxell, P. A. (1966). The interpretation of disjunct distributions. Taxon, 16(4), 316-324. https://doi.org/10.2307/1216382

Goad R.K., Grund, S.P., & Tracey, C., (2019). Resource recovery plan for Canby's mountain lover (*Paxistima canbyi)* in Pennsylvania. Pennsylvania Natural Heritage Program. Pittsburgh, Pennsylvania.

Godt, M. J. W., Johnson, B. R., & Hamrick, J. L. (1996). Genetic diversity and population size in four rare southern Appalachian plant species. Conservation biology, 10(3), 796-805.

Graham, S. P., Steen, D. A., Nelson, K. T., Durso, A. M., & Maerz, J. C. (2010). An overlooked hotspot? Rapid biodiversity assessment reveals a region of exceptional herpetofaunal richness in the southeastern United States. Southeastern Naturalist, 9(1), 19-34.

Hayes, J. D. (2021). Genetic diversity & connectivity of *Chasmanthium latifolium* (poaceae) in pennsylvania & the effect on conservation status of a rare species. Bucknell University, Honors Thesis.

Hodkinson, B. P. (2010). A first assessment of lichen diversity for one of North America's 'biodiversity hotspots' in the southern Appalachians of Virginia. Castanea, 75(1), 126-133.

Herman, T. A., & Bouzat, J.L. (2016). Range-wide phylogeography of the four-toed salamander: Out of Appalachia and into the glacial aftermath. Journal of Biogeography, 43, 666–678. https://doi.org/10.1111/jbi.12679

Hyseni, C., & Garrick, R. C. (2019). The role of glacial-interglacial climate change in shaping the genetic structure of eastern subterranean termites in the southern Appalachian Mountains, USA. Ecology and Evolution, 9(8), 4621-4636. https://doi.org/10.1002/ece3.5065

Kočiš Tubić, N., Djan, M., Veličković, N., Anačkov, G., & Obreht, D. (2015). Microsatellite DNA variation within and among invasive populations of *Ambrosia artemisiifolia* from the southern Pannonian Plain. Weed Research, 55(3), 268-277.

Laikre, L. (2010). Genetic diversity is overlooked in international conservation policy implementation. Conservation Genetics, 11, 349-354.

Levins, R., (1970). Extinction. Some mathematical questions in biology.

Littlefield T., (2008). Chronicles of a glacial relict. Naturally Kentucky: 2–3.

Ma J., Ball P.W., & Levin G.A., (2016). Celastraceae R. Brown. Page 111 Flora of North America: North of Mexico. Oxford University Press, New York, Oxford.

McAllister, C. A., & Miller, A. J. (2016). Single nucleotide polymorphism discovery via genotyping by sequencing to assess population genetic structure and recurrent polyploidization in Andropogon gerardii. American Journal of Botany, 103(7), 1314-1325.

Marcinko, S. E. (2007). Pattern and process in rare plant conservation: an assessment of the Southeastern US.

Master L, Faber-Langendoen D, Bittman R, Hammerson GA, Heidel B, Ramsay L, Snow K, Teucher A, & Tomaino A. (2012). NatureServe Conservation Status Assessments: Factors for Evaluating Species and Ecosystem Risk. NatureServe, Arlington, VA.

Matadha, D., Hamilton, G. C., Hughes, M. G., & Lashomb, J. H. (2003). Distribution of natural enemies of euonymus scale, *unaspis euonymi* (comstock) (homoptera: diaspididae), in New Jersey. Environmental Entomology, 32(3), 602-607. https://doi.org/10.1603/0046-225X-32.3.602

McDonnell, A.J., Moore, C.L., Schuette, S., & Martine, C.T. (2021). Population genomics and conservation of *Erigenia bulbosa* (Apiaceae), an edge-of-range species in Pennsylvania. International Journal of Plant Sciences. 182(5): 344-355.

McPherson, J.I., (2013). Conservation assessment of calcareous ecosystems in Pennsylvania. Page 152 pg. Report to Wild Resources Conservation Program, Grant #10391—Pennsylvania Natural Heritage Program at Western Pennsylvania Conservancy, Pittsburgh, PA.

Merker, S. A., & Chandler, R. B. (2020). Identifying global hotspots of avian trailing-edge population diversity. Global Ecology and Conservation, 22, e00915.

Milanovich, J. R., Peterman, W. E., Nibbelink, N. P., & Maerz, J. C. (2010). Projected loss of a salamander diversity hotspot as a consequence of projected global climate change. PloS one, 5(8), e12189.

Mills, M. H., & Schwartz, M. W. (2005). Rare plants at the extremes of distribution: broadly and narrowly distributed rare species. Biodiversity & Conservation, 14, 1401-1420.

Moore, C. L., McDonnell A. J., Schuette, S., & Martine, C.T., (In review). Population genomics of a rare riparian plant species. Natural Areas Journal.

Mohn, R. A., Oleas, N. H., Smith, A. B., Swift, J. F., Yatskievych, G. A., & Edwards, C. E. (2021). The phylogeographic history of a range disjunction in eastern North America: The role of post-glacial expansion into newly suitable habitat. American Journal of Botany, 108(6), 1042-1057. https://doi.org/10.1002/ajb2.1686

Moura, A. E., Pilot, M., Tehrani, A., Best, P. B., Thornton, M., Plön, S., De Bruyn, P. J., Worley, K. C., Gibbs, R. A., Dahlheim, M. E., & Hoelzel, A. R. (2014). Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. Molecular Biology and Evolution, 31(5), 1121-1131. https://doi.org/10.1093/molbev/msu058

Murdock, N. A. (1995). Rare and endangered plants and animals of southern Appalachian wetlands. Wetlands of the interior southeastern United States, 189-209.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403(6772), 853-858. https://doi.org/10.1038/35002501

Navaro, A.M., & Blackwell, W.H., (1990). A revision of paxistima (celastraceae). Sida, Contributions to Botany 14:231–249.

NatureServe (2022). NatureServe Explorer: Conservation ranks and statuses [web application]. Version 7.1. Available from https://explorer.natureserve.org/AboutTheData/DataTypes/ConservationStatusCategories (accessed November 1, 2022)

Noss, R. F., Platt, W. J., Sorrie, B. A., Weakley, A. S., Means, D. B., Costanza, J., & Peet, R. K. (2015). How global biodiversity hotspots may go unrecognized: lessons from the North American coastal plain. Diversity and Distributions, 21(2), 236-244.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., & Hoekstra, H.E. (2012). Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7: e37135.

Poag, C. W., & Sevon, W. D. (1989). A record of Appalachian denudation in post-rift Mesozoic and Cenozoic sedimentary deposits of the US middle Atlantic continental margin. Geomorphology, 2(1-3), 119-157.

Soulé, M. E. (1985). What is conservation biology? BioScience, 35(11), 727-734.

Stoutamire, W. P., (1991). Sexual reproduction in *Paxistima canbyi*: Final report on grant #910533. Ohio Department of Natural Resources, Division of Natural Areas and Preserves

Ramírez-Barahona, S., & Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics: the case of cloud forests during the Last Glacial Maximum. Ecology and evolution, 3(3), 725-738.

Regan, T. J., Master, L. L., & Hammerson, G. A. (2004). Capturing expert knowledge for threatened species assessments: a case study using NatureServe conservation status ranks. Acta Oecologica, 26(2), 95-107.

Rowe, H.C., Renaut, S., & Guggisberg, A. (2011). RAD in the realm of next-generation sequencing technologies. Molecular Ecology 20:499-3502.

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G. R., Godoy, J. A., Jentoft, S., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Secomandi, S., & Zammit, G. (2023). How genomics can help biodiversity conservation. Trends in Genetics, *39*(7), 545-559. https://doi.org/10.1016/j.tig.2023.01.005

Thesing, B. D., Noyes, R. D., Starkey, D. E., & Shepard, D. B. (2016). Pleistocene climatic fluctuations explain the disjunct distribution and complex phylogeographic structure of the Southern Red-backed Salamander, *Plethodon serratus*. Evolutionary Ecology, 30, 89 –104. https://doi.org/10.1007/s10682 -015 -9794-3

Tuberville, T. D., Andrews, K. M., Sperry, J. H., & Grosse, A. M. (2015). Use of the NatureServe climate change vulnerability index as an assessment tool for reptiles and amphibians: Lessons learned. Environmental management, 56, 822-834.

USFS (2024), How many rare plants are there? https://www.fs.usda.gov/wildflowers/Rare_Plants/howmany.shtml

Wang, C., Yap, Z. Y., Wan, P., Chen, K., Folk, R. A., Damrel, D. Z., & Li, P. (2023). Molecular phylogeography and historical demography of a widespread herbaceous species from eastern North America, *Podophyllum peltatum*. American Journal of Botany, 110(11), e16254.

Watson, L. E., Kornkven, A. B., Miller, C. R., Allison, J. R., McCarty, N. B., & Unwin, M. M. (2002). Morphometric and genetic variation in eriocaulon koernickianum Van Heurck & Muller-Argoviensis (Eriocaulaceae): A disjunct plant species of the southeastern United States. Castanea, 67(4), 416-426. https://doi.org/4034135

Weakly A.S., Ludwig C.J., & Townsend J.F., (2012). Flora of Virginia. Foundation of the Flora of Virginia Project Inc., Richmond. Fort Worth: Botanical Research Institute of Texas Press.

White A., & Williams E.W., (2017). Genetic diversity in isolated populations of *Paxistima canbyi* (Canby's Mountain Lover). Prepared for the National Forest Genetics Laboratory. NFGEL Project #329. Atlanta, GA.

Wroblewska, A. (2012). The role of disjunction and postglacial population expansion on phylogeographical history and genetic diversity of the circumboreal plant *Chamaedaphne calyculata*. Biological Journal of the Linnean Society, 105(4), 761-775.

Tables and Figures

Table 1: *Table showing ranks and their definitions adapted from NatureServe, 2022.*

NATIONAL / SUB - NATIONAL RANK	DEFINITION	GLOBAL RANK	DEFINITION
NX / SX	**Presumed Extirpated** — An environment or species is thought to have disappeared from the jurisdiction (i.e., the country, a state, or a province). Despite extensive searches of historical places and other suitable environments, it has not been found, and there is almost no chance that it will ever be. [corresponding to "Regionally Extinct" in terms of the IUCN Red List]	**GX**	**Presumed Extinct (species)** — Despite extensive searches, not been found, and almost minimal chance of rediscovery. **Presumed Collapsed (ecosystem)** — collapsed over its range as a result of the eradication of places and ecological processes critical to the type's survival, as well as the loss of key dominant and distinctive species.
NH / SH	**Possibly Extirpated** – Only known from historical records, yet there is still a chance of rediscovery. Although there isn't enough evidence to say for sure, there is some evidence that the species or ecosystem may no longer exist in the jurisdiction. Examples of this evidence include the following: (1) the fact that a species has not been observed in human-dominated landscapes for 20–40 years despite some searching and/or some evidence of significant habitat loss or degradation; and (2) the fact that a species or ecosystem has been searched for but not thoroughly enough to assume it is no longer	**GH**	**Possibly Extinct (species) or Possibly Collapsed (ecosystem)** — only from historical incidents, yet there is still a chance of rediscovery. Examples of evidence include (1) the fact that a species has not been observed in human-dominated landscapes for 20 to 40 years despite some searching and/or some evidence of severe habitat loss or degradation, (2) the fact that a species or ecosystem has been searched for unsuccessfully but not thoroughly enough to assume it is extinct or has collapsed across its range.

	present in the jurisdiction.		
N1 / S1	**Critically Imperiled** — Due to a very small range, few individuals or occurrences, extremely rapid losses, serious threats, or other circumstances, the species is at a very high risk of extinction in the jurisdiction.	**G1**	**Critically Imperiled** — Due to a very small range, a small number of individuals or occurrences, a very steep decline, a very serious threat, or other circumstances, the species is extremely vulnerable to extinction or collapse.
N2 / S2	**Imperiled** — At high risk of extirpation in the jurisdiction due to restricted range, few populations or occurrences, steep declines, severe threats, or other factors.	**G2**	**Imperiled** — At high risk of extinction or collapse due to restricted range, few populations or occurrences, steep declines, severe threats, or other factors.
N3 / S3	**Vulnerable** — At moderate risk of extirpation in the jurisdiction due to a fairly restricted range, relatively few populations or occurrences, recent and widespread declines, threats, or other factors.	**G3**	**Vulnerable** — At moderate risk of extinction or collapse due to a fairly restricted range, relatively few populations or occurrences, recent and widespread declines, threats, or other factors.
N4 / S4	**Apparently Secure —** At a fairly low risk of extirpation in the jurisdiction due to an extensive range and/or many populations or occurrences, but with possible cause for some concern as a result of recent local declines, threats, or other factors.	**G4**	**Apparently Secure** — At a fairly lower risk of extinction or collapse due to an extensive range and/or many populations or occurrences, but with possible cause for some concern as a result of recent local declines, threats, or other factors.
N5 / S5	**Secure** — At very low or no risk of extirpation in the jurisdiction due to a very extensive range, abundant populations or occurrences, with little to no concern from declines or threats.	**G5**	**Secure** — At very low risk of extinction or collapse due to a very extensive range, abundant populations, or occurrences, and little to no concern from declines or threats.

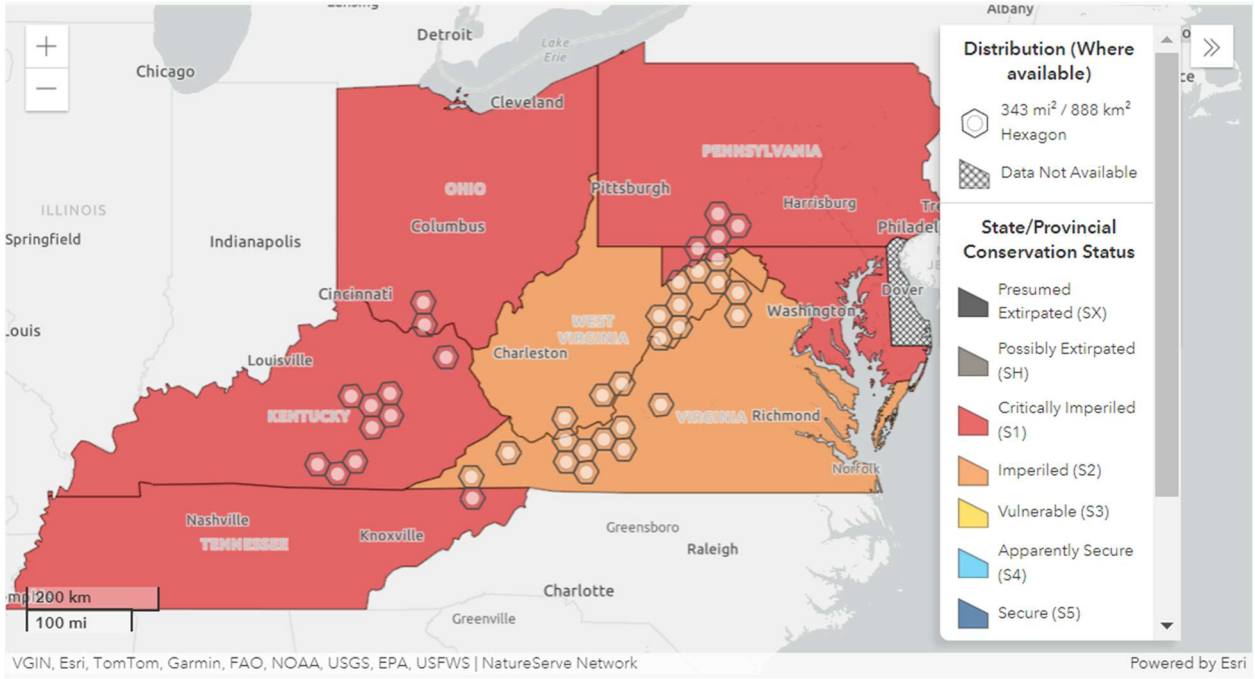


Figure *1*: A map showing the distribution of all *P. canbyi* populations across its range. Map adapted from NatureServe, 2022. States in red indicate states where *P. canbyi* is ranked as critically imperiled and those in orange indicate that populations are imperiled.

Chapter Two: Assessing the Genetic Diversity, Gene Flow, and Connectivity amongst the Range-Wide Populations of *Paxistima canbyi*: The Use of An Effective Reduced Representation Sequencing Method, Genotype-By-Sequencing (GBS).

Introduction

The use and adoption of genomic data in managing rare species and species of conservation concern are fast becoming a mainstay in conservation biology (e.g., Melville et al., 2007; Gebremedhin et al., 2009; Coates et al., 2016; McDonnell et al., 2021; Moore et al., in review). This trend has come off the back of several persuasive but pragmatic arguments and advocacy by researchers and conservation biologists where researchers argue the importance of genomics in conservation biology such as the detection of population substructure, assessing genetic connectivity, identifying and predicting potential risks associated with demographic change and inbreeding. Researchers argue that these advantages provided by using genomic tools, provide evidence for making scientifically informed conservation management decisions (Shafer et al., 2015; Taylor et al., 2017; Theissinger et al., 2023; Hogg, 2023). These arguments highlight the need for, and the continued integration of genomic data generated from current sequencing methods and tools in the formation of policies to manage species populations (Delgado et al., 2007; Laikre et al., 2009; Allendorf et al., 2010; Frankham, 2010).

Integrating genomic data into conservation plans and policies improves our ability to maintain conservation units large enough to ensure species continuity (Theissinger et al., 2023). This is critical for maintaining healthy populations and preventing inbreeding depression and helps delineate distinct conservation units within species (Frankham, 2003; Reed & Frankham, 2003). It also enhances our understanding of the genetic diversity of the species, and their populations by providing a detailed understanding of the genetic structure of species and

populations. This includes identifying genetic variations, such as alleles and genotypes, which are crucial for assessing the overall genetic diversity within a species (Frankham, 2003; Reed & Frankham, 2003; DeWoody et al., 2021). Furthermore, understanding the dynamics occurring between the spatial distribution of species (both past and present) and their relation to genotypes found within populations of a species' is vital to the formulation of successful conservation strategies and the maintenance of conservation management units as this allows for the maintenance of adaptive genetic diversity and the resilience of populations facing environmental challenges (Waits et al., 1998; Lee et al., 2006; Funk et al., 2012; Médail & Baumel, 2018).

Current population genomics studies that have informed conservation efforts often apply sequencing techniques that utilize Next Generation Sequencing (NGS) platforms in generating genomic data for their studies (e.g., Zavodna et al., 2013; Purahong et al., 2019; Tan et al., 2019; Teixeira et al., 2021; Marinček et al., 2021; McDonnell et al., 2021; Moore et al., in review). Sequencing techniques such as; restriction site-associated DNA sequencing (RADseq), Genotype-By-Sequencing (GBS), and other similar reduced-representation sequencing methods remain the most used NGS methods, for generating genomic data on species of conservation concern (Morin et al., 2004; Matz, 2018; Brandeis et al., 2019; Wold et al., 2021). This is because, more often than not, rare species or species of conservation concern tend to be non-model organisms that lack well-characterized genomes. Therefore, understanding the structure of such genomes by characterizing them requires methods that do not only break down the complexity of the genome for studies, but also methods that generate large enough amounts of genomic data that can used for exploring within-species diversity, constructing haplotype maps, and performing genome-wide association studies (GWAS) (Elshire et al., 2011).

To assess the genetic diversity and gene flow amongst the range-wide populations of *Paxistima canbyi* A. Gray, I used genomic data generated from a sequencing method that approaches sequencing via a reduced-representation sequencing technique known as Genotype-By-Sequencing (GBS). GBS makes use of restriction enzymes in cutting DNA into smaller fragments, focusing on reducing the complexity of the organism's genome by generating genome-wide high-throughput sequencing data and obtaining a large number of genetic polymorphism tag sequences to make a reduced representation of the whole genome information of the species (Elshire et al., 2011). These regions often harbor relevant genetic variations like single nucleotide polymorphisms (SNPs); SNPs provide a powerful tool for studying the evolutionary history of populations. By analyzing SNP patterns, researchers can reconstruct past migration events, detect signatures of founder effects and population bottlenecks, assess the impact of genetic drift, and identify regions of the genome under selective pressure. SNPs through nucleotide sequence information of alleles can also be used to determine the specific genotypes (*AA, AB, or BB*) for each individual in a population.

The use of SNP data from the GBS method has also proven to be more informative than data produced from other traditional genetic markers such as: microsatellites, restriction fragment length polymorphisms (RFLPs), and amplified fragment length polymorphisms (AFLPs), and other methods that operate under a different set of evolutionary assumptions (Table 2) as SNPs arise due to mutations at single nucleotide positions in the DNA sequence and are the most common type of genetic variation in the genome (Andrews et al., 2016; Peterson et al., 2012). As a result, GBS-generated SNP data has been used in several recent studies looking to resolve issues on the genetic diversity of plant species of conservation concern. For example, Martin et al. (2016) explored fine-scale population genetic structure across all the populations of

the North American common ragweed (*Ambrosia artemisiifolia* L.) using data generated from diallelic SNP loci generated through GBS. To demonstrate the ability of data generated from GBS to resolve a reportedly weak genetic structure within the highly admixed species, the study further corroborated previously discovered phylogeographic domains which had been previously discovered by Tubić et al. (2015) using microsatellite data.

My work addresses questions on the genetic diversity of *Paxistima canbyi*, Canby's mountain lover (Celastraceae), an evergreen clonally growing shrub, that is usually found growing only in habitats with calcareous bedrocks like limestone, dolomite, or calcareous shale across eastern North America (Stoutamire, 1991). All across its range, *P. canbyi* appears to form a clustered distribution of smaller subpopulations that come together to form two disjunct distributions occupying the eastern ridge of the central Appalachian Mountains which comprises western Virginia, eastern West Virginia, western Maryland, and southern Pennsylvania, and the western edges of the Allegheny-Cumberland interior low plateaus which comprises central Kentucky, southern Ohio, and central Tennessee.

Paxistima canbyi is at an elevated risk of extirpation globally, due to its restricted range, very few populations, steep population declines, and severe threats from insect herbivory and the invasion of the species' habitat by invasive weed species in all states within which the species' populations are found. As a result, *P. canbyi* is classified as Imperiled (G2) on the global scale by NatureServe (NatureServe, 2022). On the state level, it is considered Critically Imperiled (S1) in most states except Virginia and West Virginia where it is classified as Imperiled (S2) (NatureServe, 2022). My study was designed to provide support to conservation agencies such as NatureServe, the Pennsylvania Department of Conservation and Natural Resources, and the Pennsylvania Natural Heritage Program in crafting future management plans for this globally

imperiled species. Due to the species' conservation status, each sub-population recorded for each state is well-documented and re-evaluated every ten years by the various Natural Heritage Programs through NatureServe's protocols (NatureServe, 2022).

Across the species range, *P. canbyi* populations are distributed into two disjoined and distinct (disjunct) distributions that are grouped based on their relative geographic positions and features (Fig. 2). These disjunct distributions in the populations of *P. canbyi* form two major subpopulations that are geographically isolated by major vicariances such as mountains and rivers. For example, *Paxistima canbyi* populations in Bedford County, Pennsylvania, and Grant County, West Virginia are isolated by streams, rivers, and a row of mountains and hills, which deepen their isolation. This makes the two subpopulations an ideal system for studying the genetic structure and genetic diversity in rare plant species populations that are influenced by the soils on which they grow.

As some *P. canbyi* populations are at the edge of the species range (e.g., *P. canbyi* populations in Pennsylvania and southern West Virginia), my study includes a thorough sampling across the two major disjunct distributions of *P. canbyi.* This effort is essential to detect unique multilocus genotypes (Frankhams et al., 1996; Spielman et al., 2004). Rare plant species populations found at the edges of their range, often possess higher genetic diversity than central populations due to gene flow from other edge populations within the species range (Cook et al., 1961; Franks et al. 2014; Tomáš and Zuzana, 2022). Species populations at the edge of their range are also useful in inferring levels of genetic traits across species ranges as such populations tend to undergo genetic bottlenecks and drift (Sexton et al., 2011). Therefore, it is possible to infer the levels of these genetic traits in edge-of-range populations, which are expected to have higher genetic diversity than central populations that may have gone through a

bottleneck or experienced genetic drift, by having a better understanding of the genetic structure and connectivity of edge-of-range populations (Sexton et al., 2009; Sexton et al., 2011).

While geographic isolation and edge-of-range effects (i.e. the changes in population or community characteristics that occur at the boundaries of a species' geographic distribution) remain the principal threats to the genetic integrity of the existing *P. canbyi* populations, the activities of the invasive euonymus scale insect (*Unaspis euonymi*), may also be contributing to the reduction in the size of *P. canbyi* subpopulations and the further genetic isolation of the extant populations. Invasive insect species are known to reduce diversity and alter population density in plant communities (Myers & Safraz, 2017) and herbivorous insects can be a threat to rare plants that have restricted habitat preferences, narrow geographic ranges, low genetic diversity, and smaller populations (Bevill et al., 1999; Myers & Safraz, 2017). The impact of insect herbivory on rare and endangered plants can differ across populations, environmental gradients, and various habitats, including those with distinct disturbance histories (Myers & Safraz, 2017; Moore et al., 2021). Even though the conservation efforts made by the various Natural Heritage Programs and the location of the habitats of the existing *P. canbyi* populations protect them from major disturbance activities such as mining, farming, etc.; the introduction of invasive host plant species whose life cycles synchronize with the breeding activities of euonymus scale insect (e.g., *Celastrus orbiculatus*) may be detrimental to the evergreen shrub.

Additionally, the reproductive biology of *Paxistima canbyi* may pose a threat to its persistence. *Paxistima canbyi* is a potentially polyploid, clonally reproducing species (Stoutamire, 1991; Simmons et al., 2001). Polyploidy, the condition of having more than two complete sets of chromosomes, can arise through mechanisms like genome duplication or hybridization between distinct species. This results in individuals with multiple sets of

chromosomes, leading to greater genomic complexity and potential for novel gene interactions compared to their diploid counterparts. The increased chromosome number in polyploids can result in altered allele frequencies and genotypic ratios, challenging the assumptions of the Hardy-Weinberg theorem, which include random mating, no selection, no mutation, no migration, and infinite population size, where allele and genotype frequencies remain constant from generation to generation (Hartl & Clark, 2007).

Further exaggerating the challenge raised by polyploidy in *P. canbyi* is the issue of clonality, where offspring of the species are produced asexually and are thought to be genetically identical to a single parent, complicates the genetic structure of the extant *P. canbyi* populations (Ellstrand & Roose, 1987; White & Williams, 2017). This mode of reproduction bypasses the genetic recombination typical of sexual reproduction, resulting in populations consisting of many genetically identical individuals, or clones. In clonal populations, the assumptions of random mating and independent assortment of alleles are violated, leading to departures from the Hardy-Weinberg equilibrium (Arnaus-Haond et al., 2007). The presence of clones can create genetic clusters with reduced genetic diversity and non-random allele frequencies, as certain genotypes proliferate more than others due to their clonal nature (Eckert et al., 2010). Consequently, the genetic structure of the potentially clonal and polyploid populations like that of *P. canbyi* can be markedly different from that predicted by Hardy-Weinberg equilibrium. The combined effects of polyploidy and clonality can lead to significant deviations in allele frequencies and genotypic distributions, posing challenges not only to the analysis of data produced in from this project but also for the continuation and conservation of the species.

I used a reduced representation sequencing method - Genotype-By-Sequencing (GBS), a particularly useful method for understanding the population genomics of non-model organisms

such as *P. canbyi* (e.g., Elshire et al., 2011; Peterson et al., 2012; Hayes, 2021; Liu et al., 2020; McDonnell et al., 2021; Qiao et al., 2021; Ye et al., 2021; Moore et al., in review). This is because this sequencing approach allows a targeted fraction of the *P. canbyi* genome (a reduced representation library) to be sequenced with next-generation technology (in this case, Illumina NovaSeq 6000) rather than the entire genome. This allows for the identification of SNPs which would then be used to assess genetic diversity across the extant populations of *P. canbyi.* The GBS approach is again useful for *P. canbyi* as this species has little or no previous genomic information (reference genome) which would have allowed for the easy identification of polymorphic changes across the *P. canbyi* genome.

I utilized the single nucleotide polymorphisms (SNP) dataset generated to calculate population genetic metrics including F_{ST} (genetic variance), F_{IS} (inbreeding coefficient), and heterozygosity. I also performed clustering analyses to deduce the group affiliation of individuals. This involved techniques such as discriminant analysis of principal components (DAPC) and the determination of individual admixture coefficients using methods like SNMF. I then applied these Inferences to two hypotheses I intended testing including **i)** the sampled populations exhibit a high genetic structure even under limited gene flow, and **ii)** gene flow throughout the range including the two disjunct distributions is low due to the existing geographical barriers and the clonal nature of the species.

Materials and Methods

Study species

The genus *Paxistima* is endemic to North America and consists of only two species occupying the eastern and western edges of continental North America (NatureServe, 2022; Weakley and Southern Flora Team, 2023). *Paxistima canbyi*, the study species, ranges along the eastern coast of the continent, dividing into two disjunct distributions which separates into two subpopulations (Fig. 1); where one with six sub-populations occupies the Cumberland-Allegheny plateaus of central Kentucky and southern Ohio, and the other with fifteen subpopulations in the central Appalachians of Southern Pennsylvania, western Maryland, western Virginia, eastern West Virginia, and Tennessee (NatureServe, 2022). *Paxistima canbyi* is generally considered a woodland understory species, typically occurring on calcareous bedrocks like limestone, dolomite, or calcareous shale but has been found growing in open woodlands (Ma et al., 2016; Goad et al., 2019). Although this has yet to be well studied, *P. canbyi* is thought to be a polyploid because of its membership in the Celastraceae, a family where species tend to be polyploids (Stoutamire, 1991; Simmons et al., 2001; Lunardi et al., 2004).

Sample Collection

Leaf tissue samples were collected from 185 individuals across 21 populations spanning the range-wide distribution of *P. canbyi* with the number of individuals per population ranging between 1 to 16 between the spring and summer of 2023 (Table 3). Where, healthy, young, and actively growing leaf tissue were collected from spatially distant individuals across the population and then placed in separate coin envelopes (2 ¼ x 3 ½ inches) containing a small amount of silica gel by sample, properly labeled with sample name, date, species, collection location, and collector's name (Gostel et al., 2016).

The collections included 16 sites along the central Appalachian distribution and a collection representing the interior low plateaus from the Atlanta Botanical Garden of 5 Kentucky populations that are currently lost or missing (Fig. 2). Access to the site of the only *P. canbyi* population in southern Ohio was denied, due to the 'precarious' nature of the population. Due to the species' clonal nature, sampling was conducted on phenotypically representative individuals spaced 1-2 meters away. Also, three voucher specimens collected from three different populations in Virginia (Collection ID: CTM – 5232), West Virginia (Collection ID: IKB - 021), and Pennsylvania (Collection ID: IKB - 012) were deposited at the Wayne E. Manning Herbarium (BUPL) at Bucknell University.

DNA Extraction and Sequencing

Total Genomic DNA was extracted from leaf tissue dried on silica using the FastDNA® Spin Kit (MP Biomedicals, Santa Ana, California) following a modified DNA extraction protocol for *P. canbyi* where, 10 – 25 mg of Leaf tissue was weighed out into lysing matrix tubes and then dry ground using the Geno/Grinder 2010 [Spex ® Sample Prep (Cole – Palmer ®), Metuchen, New Jersey] instead of a mortar and pestle. Also, a SPIN™ column filter was used to capture the suspended silica matrix instead of multiple attempts to dry the silica-bound pellets. To assess the total quality of the DNA material extracted, the concentration of the DNA extracted from the samples were measured with a Qubit dsDNA BR assay kit using a Qubit v.4.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts). 3µl of the extracted DNA was mixed with 1µl 6X loading dye and then run on 1% agarose gels at 60V for 120 minutes. The gels were then viewed in a dark room under UV light, to assess the integrity and size of the extracted DNA material through the bands produced after gel electrophoresis. Prepared samples were sent to the University of Wisconsin Biotechnology Center

(https://dnaseq.biotech.wisc.edu/gbs/) where they underwent digestion enzyme optimization, library preparation, and GBS sequencing.

During sequencing, fragment analyses revealed the restriction enzyme *ApeKI*, to have exhibited the highest activity in our samples. Consequently, we adopted a single-enzyme genotyping by sequencing approach, following the method outlined by Elshire et al., (2011). DNA from two plates was digested using *ApeKI*, and subsequent steps involved library preparation, quantification, pooling, and 150bp paired end read sequencing on a NovaSeq 6000 instrument with six million reads requested (Illumina, San Diego, California). Distinct barcodes were also assigned to each sample on both plates to allow for their identification during sequence assembly.

Sequence processing

The open-source, python-written, tool for assembling and analyzing restriction site-associated DNA sequence datasets, ipyrad [v.0.9.90], was used in processing sequenced raw read data (Eaton & Overcast, 2020). The sequenced raw read data were then sorted and de novo assembled in ipyrad, with the following parameter settings: datatype=pairgbs, max_low_qual_bases=5, clust_threshold=0.85, max_barcodes_mismatch=0, filter_adapters=2, max_alleles_consens=4, min_samples_locus=4, and max_SNPs_locus=20. To begin the assembly process, sequenced reads were sorted into separate files for each sample in a sequenced run (Demultiplexing). After each plate was demultiplexed, the two plates were then merged using the merge function in ipyrad. After this, reads underwent quality-based filtering to exclude samples with Phred scores lower than 20 ($Q < 20$). Subsequently, clustering within samples was performed, grouping reads into an assembly of de novo loci with an 85% similarity threshold and processing into an assembly where all loci were shared among four or more samples. Final

output files were then generated including a variant call format (vcf) file which was used in further downstream analyses.

Inferring Ploidy Levels

Paxistima canbyi is often considered an autopolyploid, most probably a tetraploid, due to its affiliation to the family Celastraceae (Stoutamire, 1991; Simmons et al., 2001). However, the actual levels of ploidy within the genus *Paxistima* are unknown and must be determined to bring further resolution to further downstream data analyses. Using the Bayesian assignment method-based r package, gbs2ploidy (Gompert & Mock, 2017), observed heterozygosity and the ratio of DNA sequences in the assembled GBS dataset were used to estimate ploidy levels. However, it must be noted that this method can be limiting to situations where the data has low coverage and instances where there are no reference genomes for the assembled genomic dataset (Gompert & Mock, 2017).

To begin with, the vcf (.vcf) output files for 14 populations from ipyrad were converted into data frame objects in the Heterozygous Allele Depth (HAD) format using the *vcf2had* function in the r package, vcf2ploidy (https://github.com/dandewaters/vcf2ploidy) (DeWaters, 2020; Oyundelger et al., 2022). Using the *'estprops'* function in gbs2ploidy, the proportion of heterozygous GBS SNPs with different allelic ratios was obtained through Bayesian estimates of allelic proportions using Markov chain Monte Carlo from the data frame objects. The converted data frame objects were then used in assigning individuals to different cytotype groups using principal component analysis and discriminant analysis through the function '*estploidy'* in gbs2ploidy.

Data Preparation and Filtering

Due to the inability of most tools used in the downstream analyses of population genetic data to handle multiallelic and potentially polyploid data; the assembled variant call format (vcf) output file was split into two datasets i.e., biallelic and multiallelic datasets using BCFtools. The biallelic dataset was used for genetic diversity analyses, multivariate analyses, and sparse-nonnegative matrix factorization (SNMF) analyses to assess population structure and admixture.

The resulting assembled vcf output file from ipyrad was further filtered using VCFTools v. 0.1.16 (Danecek et al. 2011), where SNPs were filtered to remove sites not genotyped in 40% of all individuals within the dataset and sites located within 10 base pairs of each other, this was to ensure that linked loci and possible confounders were removed from the dataset (Danecek et al., 2011). The resulting SNPs were then further filtered in PLINK v1.90b7 to remove invariant sites based on minor allele counts and minor allele frequencies thresholds of 1 and 0.01, respectively. This was done to remove invariant sites and alleles only found once or rare variants (singletons). Genetically linked loci or loci in linkage disequilibrium were also filtered out based on squared coefficient of correlation values (R^2), R^2 values are used to analyze how differences in one variable can be explained by a difference in a second variable, in this case, the relationship between loci close to each other and the probability of they being in linkage disequilibrium. Therefore, loci with R^2 scores >= 0.8 were pruned from the dataset to further resolve population (individual) structure and association. The resulting assembly from the filtering process was then imported as a .vcf file and converted to adegenet (.raw) and structure (.strct) files for further downstream analysis using various population genomics packages in the R (R Core Team 2020).

Assessing Genetic Diversity

Genetic diversity is defined as any measure that quantifies the magnitude of genetic variability within a population (Hughes et al., 2008). Genetic diversity can be measured using population genetic parameters like F_{IS} (Inbreeding Coefficient), the probability that two alleles will be identical and derived from the same ancestor. F_{IS} assesses global variation in individuals, relative to the variation in their subpopulation; F_{ST} (Fixation Index) is also defined as the variance of allele frequencies among populations at a given locus. It assesses the variation in the subpopulations relative to that in the total population; H_E (Expected Heterozygosity) which is the probability that an individual will be heterozygous at a given locus or over the assayed loci for a multi-locus system, and H_O (Observed Heterozygosity) which is defined as the percentage of loci that are heterozygous and is determined by dividing the total number of heterozygotes by the sample size.

To calculate or estimate these measures, the vcf file generated from the ipyrad assembly process was converted into a 'genind' object in R using the adegenet package. The genind object was again converted into a hierfstat object using the 'genind2hierfstat' function in adegenet to be used by the Hierfstat package in R for calculating the statistic on genetic diversity (Jombart, 2008; Goudet, 2005). The parameters were then subsequently calculated using the 'basic.stats' function in Hierfstat (Goudet, 2005). To estimate a pairwise genetic distance comparison among all populations, the function 'gene.dist' in the package Hierfstat was applied to the hierfstat object using the Weir and Cockerham (1984) method with 100 bootstraps.

Population structure

To infer population structure, a genetic dissimilarity test was performed through Principal Coordinate Analysis (PCoA) using the R package poppr (Kamvar, et al., 2015). A PCoA utilizes a genetic distance matrix calculated between each pair of samples which is mapped as the relative similarities or differences between samples onto a two-dimensional plane for visualization by selecting the first two axes that best preserve the original distribution of distances for data representation (Mohammadi et al., 2003; Sankaran & Holmes, 2019). In performing the PCoA, the vcf file from the filtered dataset was first converted into a genind object using the vcfR package and then suffixed with population and hierarchical information. A Euclidean dissimilarity distance matrix was then computed from the genind object using the 'bitwise.dist' function in poppr. The generated matrix was then log-transformed and then used for the principal coordinate analysis and plotting of a simple ordination plot.

To further assess the genetic structure of the species' populations and corroborate results from the PCoA, a Discriminant Analysis of Principal Components (DAPC) was used to identify and describe clusters of genetically related individuals. A DAPC uses a multivariate statistical approach in detecting genetic variation which is later partitioned into two components; i.e., variation between and within groups. DAPC emphasizes and maximizes the variation observed within groups by utilizing a discriminant analysis (DA) to provide membership probabilities (Jombart et al., 2010).

In performing the DAPC, the 'find.clusters' function in the adegenet package in R was used to transform and subject genetic data from the filtered dataset to a Principal Component Analysis (PCA) and then to a clustering algorithm (using calculated means) that considers a range of potential K values to group the data into separate genetic clusters (K). This was done to

assess the number of genetic clusters present within our dataset. The goodness of fit of each K value is indicated by a Bayesian Information Criterion (BIC) score, where the lowest BIC indicates the best fit. The PCA transformation aims to summarize the overall variability among individuals, including both the variation between groups (i.e., structured genetic variability) and the variation occurring within groups ('random' genetic variability) (Jombart et al., 2010). The principal components of the PCA which are a component of the chosen best fit K value were then submitted to a Linear Discriminant Analysis (LDA) using the 'dapc' function in the adegenet package to assess the relationships between the different genetic clusters.

Additionally, the LEA (Landscape and Ecological Association) package in R was used to assess population structure and admixture (Frichot et al., 2014; Frichot and François, 2015) within the filtered dataset. Unlike the DAPC, the LEA package uses a cross-entropy criterion (based on the prediction of masked genotypes to evaluate the fit of a model with a specific number of populations) to estimate the number of ancestral populations or the best run for a fixed value of the number of genetic clusters (K). Therefore, the number of genetic clusters (K) was first estimated using a cross-entropy criterion, where one thousand iterations were run 10 times (because we were using 21,696 SNPs, a general rule of thumb is to run a lesser alpha for >10,000 SNPs) for 14 different K values (K = 1 – 14) using the 'snmf' function in the LEA package. A STRUCTURE-like ancestry coefficient proportions plot was generated from the results obtained from the iteration process. This plot comprises colored bars representing all individuals in our dataset grouped into specific genetic clusters (depicted by distinct colors that represent the different genetic clusters), and an x-axis with estimations of the membership probability of each individual in the dataset to a genetic cluster.

Finally, to determine if there are any patterns of isolation by distance within my data, the DartR package was used to perform a Mantel test which measured the correlation between genetic distance and geographic distance (spatial autocorrelation). Isolation by distance refers to the accrual of local genetic variation under limited geographical dispersal. It is used to determine the distribution of gene frequencies over a geographic region. To do this, I first transformed the filtered dataset which had been converted from a vcf file into a genlight object using the vcfR package in R into a Euclidean genetic distance matrix using the 'gl.dist.pop' function in the DartR package. I also converted the geo-coordinates of all the populations sampled into a geo-coordinate matrix using the 'dist' function in the DartR package. To perform the Mantel test with both matrices, the function 'gl.ibd' in DartR was called, for which each test had 999 permutations summarized. Results from the test were plotted to visualize the spatial autocorrelation tested.

Results

Sequence processing

The sequencing process generated a total of 684,171,375 150bp paired-end raw read sequences i.e., approximately 3,698,221.49 reads as an average number of reads per individual; each with an average Phred score of 39 and 0 - 0.01% Ns (below average quality reads). These raw reads were then demultiplexed, assembled, and filtered in a seven-stage process in ipyrad, which yielded 2,164,011 SNPs from 165 samples across 14 populations from an initial 185 samples from across 21 populations. 20 samples from 7 populations were dropped due to their small sample sizes and the low number of reads recovered after the filtering and formatting output files stage of the assembly process. This was done to avoid bias and the production of false results in further downstream analyses. The assembled dataset was then filtered for only biallelic SNPs using BCFtools which yielded 1,897,199 biallelic SNPs. This was done to reduce the complexities associated with the handling of multiallelic SNP data with current population genomics tools as many do not support multiallelic marker data. The biallelic SNP data was hard filtered in VCFtools and PLINK retaining 21,696 SNPs after filtering (Table. 5).

Inferring Ploidy Levels

Ploidy levels were inferred from estimated allelic proportions of the assembled datasets of 14 *P. canbyi* populations using the *'estprops'* function of the gbs2ploidy package in R. Results from the estimated ratioed allelic proportions indicate ploidy levels to be between 2n (diploids), 3n (triploids), and 4n (tetraploids) within mostly 4 allelic ratios across all the sampled populations i.e., 1:1 (0.5), 1:2 or 2:1 (0.33 or 0.66), and 1:3, 2:2, or 3:1 (0.25, 0.5, or 0.75) (Fig. 6; Fig. 8).

Genetic diversity

Between populations, a relatively high level of differentiation was observed with global F_{ST} reaching 0.390 at a 95% confidence interval. F_{ST} values range from 0 to 1, where values ≥ 0.15 indicate a significantly high level of genetic differentiation between or within populations (Wright 1978). On the other hand, very low levels of inbreeding were detected across almost all populations except for two i.e., Blue Rock and Smoke Hole populations (with global F_{IS} = -0.273 with values for individual populations falling within the range of -0.8 to 0.0057) (Table 7). F_{IS} values range from -1 to 1, with values closer to -1 signifying outbreeding and those closer to 1 signifying inbreeding (Wright 1978). Overall, heterozygosity was observed to be higher than expected, which was confirmed to be significantly different by a Bartlett test (Bartlett's K-squared = 425.13, df = 1, p-value < 2.2e-16) except for Blue Rock and Smoke Hole populations where heterozygosity observed were lower than expected (Table 7).

Population structure

Results from the PCoA did not show individuals clustering into hypothesized two main sub-populations i.e., Central Appalachian Mountains and Cumberland - Allegheny Interior Low Plateaus sub-populations (Fig. 9). The DAPC supports 14 genetic clusters (K=14), however all the populations cluster together except for five populations i.e., Saint Clairsville, Lutzville, Powers Hollow, Sweet Lilly Ridge and SELU (Fig. 7). Also, the DAPC does not support the distinction between the two sub-populations (Fig. 7).

The clustering and cross-entropy criterion in the SNMF analysis suggested K = 14 to be the optimal number of genetic clusters (Fig. 4). The resulting analysis indicated high levels of admixture between populations as indicated by the mixing of the colored bars (indicating individuals) from the different genetic clusters (Fig. 4). However, individuals from the various

populations seem to cluster together except for two clusters that are composed of individuals from four different populations i.e., Smoke Hole (WV) and Brush Creek Falls (WV) populations cluster together in one cluster with SELU (VA) and Carter Caves (KY) clustering together in another. Results from F_{ST} estimates support the genetic differentiation of populations from each other (Fig. 4).

A Mantel test aimed at assessing the relationship between the geographic and genetic distances of samples and populations revealed no significant support for the tested relationship (Fig. 8: 999, replicates; Mantel statistic r: 0.06225, p-value < 0.412), indicating no significant pattern of Isolation by Distance. This is also supported by the results from the PCoA, with all the populations grouping together except for a few individuals from Powers Hollow and Blue Rock (Fig. 9). The largest proportion of genetic variation detected through an Analysis of Molecular Variance (AMOVA) is attributed to variation within samples, indicating that there is considerable genetic diversity within the studied populations (71.15%) rather than between populations (11.61%) and the two disjunct distributions (17.23%). This suggests that the genetic structure within the existing *P. canbyi* populations is not defined by geographic structure.

Discussion

The impact of insect herbivory and other threats, on rare and endangered plants such as *Paxistima canbyi*, can differ across populations, environmental gradients, and various habitats (Myers & Safraz, 2017; Moore et al., 2021). Here, questions on the population genomics of a rare glacial relict species are addressed by using a reduced representation sequencing method, Genotype-By-Sequencing (GBS) (Elshire et al., 2011; Peterson et al., 2012). With the life history and reproductive biology of *P. canbyi* being thought of as the prime influencing factors in the genetic differentiation of the extant populations of *P. canbyi*, this project involved the sampling of leaf tissue from across the range-wide distribution of the species including the two major disjunct distributions i.e., the Central Appalachian Mountains and Cumberland-Allegheny Interior Low Plateaus (Fig. 3).

As indicated by the overall calculated global F_{ST} value (0.390), the 14 sampled *P. canbyi* populations appear to be significantly genetically isolated. This observation is thought to be a common phenomenon in species limited in gene flow by the geographic distances between populations, as shown, for example in species such as *Festuca pratensis* (Poaceae), *Brassica insularis* (Brassicaceae), *Centaurea corymbosa* (Asteraceae), and *Meles meles* (Mustelidae) where low levels of gene flow were reported to occur between populations due to limited gene flow (Rognli et al. 2000, Petit et al., 2001; Pope et al., 2006; Sexton et al., 2014). Low levels of gene flow observed between *P. canbyi* populations may be exacerbated by low pollination efficiency; given that stigmas in its flowers have been observed to be relatively dry and devoid of pollen oil droplets (pollenkitt), which limits the attachment of pollen to pollinators and the stigmatic surface, thus seemingly limiting the efficiency of pollen transfer, especially over long distances (Stoutamire, 1991).

Also, the assumed lack of appropriate mates within populations due to the inability of *P. canbyi* to produce zygotes after self-pollination in fertile co-sexual individuals was expected to have significantly influenced genetic diversity and caused increased levels of inbreeding across *P. canbyi* populations. However, our results suggest an excess of heterozygotes where heterozygosity was observed to be higher than expected, with low levels of inbreeding as global F_{IS} was calculated to be -0.273 across all populations. However, two West Virginia populations (Blue Rock and Smoke Hole) both recorded significantly positive F_{IS} values and lower than expected heterozygosity (Table 7). This may explain the reason for which only 35 unique multilocus genotypes were identified in an initial attempt to assess the genetic diversity of some *P. canbyi* populations in Ohio and West Virginia, where 119 individuals across 9 populations were sampled across both states indicating that levels of genetic diversity were quite low in the 9 sampled populations (White & Williams, 2017). Lastly, total genetic diversity (Ht) which comprises heterozygosity between and within the extant *P. canbyi* populations was calculated to be 0.195 indicating substantial genetic diversity.

Results from the PCoA, DAPC, and the STRUCTURE-like plot produced from the SNMF analysis did not show individuals clustering into the hypothesized two main subpopulations i.e., Central Appalachian Mountains and Cumberland - Allegheny Interior Low Plateaus subpopulations. On the other hand, almost all the populations cluster together indicating some genetic similarity except a few in the DAPC (Fig. 7) and the PCoA biplot (Fig. 9); these populations, Lutzville and Saint Clairesville in Pennsylvania, and Powers Hollow and Sweet Lilly Ridge in West Virginia and Kentucky, respectively, are within the north-central distribution of the species. In the STRUCTURE-like plot, Smoke Hole (WV) and Blue Rock (WV) populations appear to cluster together forming a single genetic unit, while SELU (VA) and

Carter Caves (KY) also cluster together forming another unit. The four populations comprise some of the centrally distributed populations of *P. canbyi.* These results suggest some recent admixture between the extant populations of *P. canbyi* (Fig. 4).

This observed admixture may not be explicable given the constraints observed with respect to the reproductive biology of the species and the current geographic constraints set by the mountains, rivers, and roads separating these populations which limit gene flow between populations and isolate them as gleaned from our F_{ST} results where populations are genetically isolated. However, these results may be likely when the species have undergone some recent adaptation through selection on pre-existing genetic variation which may have existed in their initial gene pool (Standing Genetic Variation). Adaptation is likely to occur faster from standing genetic variation because beneficial alleles are immediately available and at higher frequencies so they may have been pre-tested by selection in past environments, in another part of the species' range, or even in another species with which the population has exchanged genes (Barrett & Schulter, 2008). So, in the case of *P. canbyi*, I propose two hypotheses: **i.)** That the Last Glacial Maximum being the last known genetic drift event which occurred approximately 20,000 years ago, led to the split up of the initial gene pool of *P. canbyi* populations present during the period into smaller subpopulations which have managed to adapt to their present-day habitats through selection on pre-existing genetic variation from the initial gene pool. **ii.)** That the Last Glacial Maximum led to break away of the subpopulations on the Cumberland - Allegheny interior low plateaus from the initial gene pool populations initially established around the Central Appalachian Mountains which have managed to adapt to their present-day habitats through selection on pre-existing genetic variation from the initial gene pool.

To investigate these hypotheses, I propose a phylogeographic molecular clock estimation of the divergence times and dates of the ancestral *P. canbyi* populations. The study may want to consider fossil records of taxa within the family Celastraceae present in and around the Teays river valley system and the Central Appalachian Mountains before the last glacial maximum as outgroups in a phylogenetic reconstruction of the divergence dates and times of the extant populations. The use of taxa present before the last glacial maximum will serve as a strong standpoint for making divergence time comparisons. Ultimately, this study will further enhance our understanding of the genetic structure of the extant *P. canbyi* populations and will explain the source of the standing genetic variation observed within the extant populations.

Overall, the sampled populations would be considered moderately genetically "healthy". As the F_{IS} and heterozygosity are adequate, with no significant loss of heterozygosity due to inbreeding. However, observed heterozygosity (H_O) and expected Heterozygosity (H_E) vary significantly across most populations with H_O greater than H_E for all populations. The general Bartlett test supports this (Table 7). However, the threat posed by the herbivorous activities of the euonymus scale leaves both small and large populations of *P. canbyi* in a very precarious state. My attempt to sample the last known population of *P. canbyi* in Ohio was denied due to the delicate state of the populations caused by the effects of the herbivory activities of the scale insect. Other populations of *P. canbyi* were once present in Ohio but have become extirpated, likely due to the effects of the herbivory activities of the scale insect. A similar pattern could be followed by other *P. canbyi* populations, especially ones smaller in size, if proper conservation management strategies are not adopted (especially when inbreeding is occurring). Management strategies focused on increasing overall genetic diversity in *P. canbyi* may be critical, as it appears that a significant level of inbreeding occurs in certain large populations, such as those at

Smoke Hole and Blue Rock, where there appear to be many individuals but a significant level of inbreeding.

Conclusions

To conclude, almost all *P. canbyi* populations are distinct genetic units that share some ancestral genetic variation even under isolation except for four populations which form two distinct genetic units. Our findings indicate that *P. canbyi* populations on the edge of the species' range are genetically healthy. Although the majority of the populations I studied have high genetic diversity and are not yet inbred, the genetic isolation observed across all the populations within the disjunct distributions indicates that these populations may be at risk of future inbreeding depression especially with the Smoke Hole and Blue Rock populations in West Virginia. Therefore, local (state-by-state) ex-situ conservation efforts must be encouraged, where stem cuttings from the maternal lines of the extant populations can be grown and made to flower and produce viable seeds which can be stored in an established seed bank for future conservation efforts.

References

AAT Bioquest, Inc. (2024). Quest database™ restriction enzymes cut sites reference Table. AAT Bioquest. https://www.aatbio.com/data-sets/restriction-enzymes-cut-sites-reference-table.

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., & Lander, E. S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature, 407(6803), 513-516.

Arnaud-haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. Molecular Ecology, 16(24), 5115-5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Avise, J. C. (2000). Phylogeography: the history and formation of species. Harvard University Press.

Bevill R, Louda S, Stanforth L. (1999). Protection from natural enemies in managing rare plant species. Conserv. Biol. 13:1323–31

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics, 32(3), 314.

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. Genes, 10(11), 846.

Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. Frontiers in Ecology and Evolution, 6, 165.

Cook, L. M., Judd, S., Hodkinson, I., Smith, F. A., Minckley, W. L., Deacon, J. E., Berry, R. J., Power, D. M., Case, T. J., Gill, D. E., ENDLER, J. A., Bantock, C. R., Price, D. J., Bush, G. L., Soule, M., Berry, R. J., Soule, M., Carlquist, S., Grant, P. R., & Berry, R. J. (1961). The edge effect in population genetics. The American Naturalist, 95(884), 295-307. https://doi.org/2458653

DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. Molecular ecology, 30(17), 4147-4154.

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2010). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Molecular Ecology, 17(5), 1170-1188.

Ellegren, H. (2004). "Microsatellites: Simple sequences with complex evolution." *Nature Reviews Genetics*, 5(6), 435-445.

Ellstrand, N. C., & Roose, M. L. (1987). Patterns of genotypic diversity in clonal plant species. The American Naturalist, 129(5), 709-732.

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics, 36(8), 2592-2594

Frankham, R. (2003). Genetics and conservation biology. Comptes rendus. Biologies, 326(S1), 22-29.

Gebremedhin, B., Ficetola, G. F., Naderi, S., Rezaei, R., Maudet, C., Rioux, D., Luikart, G., Flagstad, Ø., Thuiller, W., & Taberlet, P. (2009). Combining genetic and ecological data to assess the conservation status of the endangered Ethiopian *walia ibex*. Animal Conservation, 12(2), 89-100. https://doi.org/10.1111/j.1469-1795.2009.00238.x

Gostel M.R., Kelloff C., Wallick K., & Funk V.A. (2016) A workflow to preserve genome-quality tissue samples from plants in botanical gardens and arboreta. Appl Plant Sci. 2016 Sep 19;4(9): apps.1600039. doi: 10.3732/apps.1600039. PMID: 27672517; PMCID: PMC5033361.

Hartl, D. L., & Clark, A. G. (2007). Principles of population genetics (4th ed.). Sinauer Associates.

Hayes, J. D. (2021). Genetic diversity & connectivity of *chasmanthium latifolium* (poaceae) in pennsylvania & the effect on conservation status of a rare species.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics, 6(2), 95-108.

Husband, B. C., & Schemske, D. W. (2000). Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. Journal of Ecology, 88(4), 689-701.

Hogg, C. J. (2023). Translating genomic advances into biodiversity conservation. Nature Reviews Genetics, 1-12. https://doi.org/10.1038/s41576-023-00671-0

Kaiser, V. B., & Charlesworth, B. (2009). The effects of deleterious mutations on evolution in non-recombining genomes. Trends in Genetics, 25(1), 9-12.

Kreitman, M., & Wayne, M. L. (1994). Organization of genetic variation at the molecular level: lessons from Drosophila. *M*olecular ecology and evolution: approaches and applications, 157-183.

Kwok, P. Y. (2001). Methods for genotyping single nucleotide polymorphisms. *A*nnual review of genomics and human genetics, 2(1), 235-258.

Lee, S. L., Ng, K. K., Saw, L. G., Lee, C. T., Muhammad, N., Tani, N., Tsumura, Y., & Koskela, J. (2006). Linking the gaps between conservation research and conservation management of rare dipterocarps: A case study of *Shorea lumutensis*. Biological Conservation, 131(1), 72-92. https://doi.org/10.1016/j.biocon.2006.02.005

Liu, Y., Yi, F., Yang, G., Wang, Y., Pubu, C., He, R., & Ma, W. (2020). Geographic population genetic structure and diversity of *Sophora moorcroftiana* based on genotyping-by-sequencing (GBS). PeerJ, 8, e9609.

Lunardi, M. P. M., Schifino-Wittmann, M. T., & de Barros, I. B. (2004). Chromosome number variability in the south american medicinal plant Maytenus ilicifolia Mart. ex Reiss (Celastraceae). Cytologia, 69(4), 439-445.

Martin, M. D., Olsen, M. T., Samaniego, J. A., Zimmer, E. A., &, Gilbert, M. T. (2016). The population genomic basis of geographic differentiation in North American common ragweed (*Ambrosia artemisiifolia* L.). Ecology and Evolution, 6(11), 3760-3771. https://doi.org/10.1002/ece3.2143

Marinček, P., Wagner, N. D., & Tomasello, S. (2021). Using herbarium samples for NGS methods–a methodological comparison. bioRxiv, 2021-08.

McDonnell, A.J., Moore, C.L., Schuette, S., Martine, C.T. (2021). Population genomics and conservation of *Erigenia bulbosa* (Apiaceae), an edge-of-range species in Pennsylvania. International Journal of Plant Sciences. 182(5): 344-355.

Médail, F., & Baumel, A. (2018). Using phylogeography to define conservation priorities: The case of narrow endemic plants in the Mediterranean basin hotspot. Biological Conservation, 224, 258-266. https://doi.org/10.1016/j.biocon.2018.05.028

Melville, J., Goebel, S., Starr, C. et al. (2007) Conservation genetics and species status of an endangered Australian dragon, *Tympanocryptis pinguicolla* (Reptilia: Agamidae). Conserv Genet 8, 185–195. https://doi.org/10.1007/s10592-006-9161-6

Mohammadi, S.A., & Prasanna, B.M. (2003) Review and interpretation analysis of genetic diversity in crop plants —salient statistical tools. Crop Science, 2003, 43, 1235-1248.

Moore, C.L., McDonnell, A.J., Schuette, S., & Martine, C.T., (2021). Lepidopteran granivory reduces seed counts in a rare species of riparian scour prairies. Natural Areas Journal 41:47-54.

Moore, C.L., McDonnell, A.J., Schuette, S., & Martine., C.T. In review. Population genomics of a rare riparian plant species. Natural Areas Journal.

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution, and conservation. Trends in Ecology & Evolution, 19(4), 208-216. https://doi.org/10.1016/j.tree.2004.01.009

Moritz, C., Patton, J. L., Schneider, C. J., & Smith, T. B. (2003). Diversification of rainforest faunas: An integrated molecular approach. https://doi.org/10.1146/annurev.ecolsys.31.1.533

Nielsen, R. (2005). Molecular signatures of natural selection. Annual Review of Genetics, 39, 197-218.

Oyundelger, K., Harpke, D., Herklotz, V., Troeva, E., Zheng, Z., Li, Z., Oyuntsetseg, B., Wagner, V., Wesche, K., & Ritz, C. M. (2022). Phylogeography of *Artemisia frigida* (Anthemideae, Asteraceae) based on genotyping-by-sequencing and plastid DNA data: Migration through Beringia. Journal of Evolutionary Biology, 35(1), 64-80. https://doi.org/10.1111/jeb.13960

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., & Hoekstra, H.E. (2012). Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135.

Peterson, G. W., Dong, Y., Horbach, C., & Fu, Y. B. (2014). Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity, 6(4), 665-680.

Petit, C., Fréville, H., Mignot, A., Colas, B., Riba, M., Imbert, E., Hurtrez-Boussés, S., Virevaire, M., & Olivieri, I. (2001). Gene flow and local adaptation in two endemic plant species. Biological Conservation, 100(1), 21-34. https://doi.org/10.1016/S0006-3207(00)00204-4

Pope, L. C., Domingo-Roura, X., Erven, K., & Burke, T. (2006). Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. Molecular Ecology, 15(2), 371-386.

Purahong, W., Wubet, T., Krüger, D., & Buscot, F. (2019). Application of next-generation sequencing technologies to conservation of wood-inhabiting fungi. Conservation Biology, 33(3), 716-724.

Qiao, Y., Guo, F., Huo, N., Zhan, L., Sun, J., Zuo, X., & Liu, Y. (2021). Genotyping-by-sequencing to determine the genetic structure of a Tibetan medicinal plant *Swertia mussotii* Franch. Genetic Resources and Crop Evolution, 68, 469-484.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. Conservation biology, 17(1), 230-237.

Rognli, O., Nilsson, N.O., & Nurminiemi, M. (2000) Effects of distance and pollen competition on gene flow in the wind-pollinated grass *Festuca pratensis* Huds. Heredity 85, 550–560. https://doi.org/10.1046/j.1365-2540.2000.00789.x

Rose, J. P., & Sytsma, K. J. (2023). Phylogeography and genetic variation in Western Jacob's ladder (Polemonium occidentale) provide insights into the origin and conservation of rare species in the Great Lakes region. Molecular Ecology, 32(1), 79-94. https://doi.org/10.1111/mec.16730

Sankaran, K., & Holmes, S. P. (2019) Multitable methods for microbiome data integration. Frontiers in genetics, 2019 10, 627.

Schlotterer, C. (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma, 109(6), 365-371.

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution, 68(1), 1-15.

Simmons, M. P., Clevinger, C. C., Savolainen, V., Archer, R. H., Mathews, S., & Doyle, J. J. (2001). Phylogeny of the Celastraceae inferred from phytochrome B gene sequence and morphology. American Journal of Botany, 88(2), 313-325. https://doi.org/10.2307/2657021

Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., & Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. Trends in ecology & evolution, 30(2), 78-87.

Tan, M. P., Wong, L. L., Razali, S. A., Afiqah-Aleng, N., Mohd Nor, S. A., Sung, Y. Y., & Danish-Daniel, M. (2019). Applications of next-generation sequencing technologies and computational tools in molecular evolution and aquatic animals conservation studies: A short review. Evolutionary Bioinformatics, 15, 1176934319892284.

Taylor, H. R., Dussex, N., & van Heezik, Y. (2017). Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. Global Ecology and Conservation, 10, 231-242.

Teixeira, T. M., & Nazareno, A. G. (2021). One step away from extinction: a population genomic analysis of a narrow endemic, tropical plant species. Frontiers in Plant Science, 12, 730258.

Tishkoff, S. A., & Verrelli, B. C. (2003). "Patterns of human genetic diversity: Implications for human evolutionary history and disease." Annual Review of Genomics and Human Genetics, 4, 293-340.

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G. R., Godoy, J. A., Jentoft, S., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Secomandi, S., & Zammit, G. (2023). How genomics can help biodiversity conservation. Trends in Genetics, *39*(7), 545-559. https://doi.org/10.1016/j.tig.2023.01.005

Tubić, N. K., Djan, M., Veličković, N., Anačkov, G., & Obreht, D. (2015). Microsatellite DNA variation within and among invasive populations of Ambrosia artemisiifolia from the southern Pannonian Plain. Weed Research, 55(3), 268-277. https://doi.org/10.1111/wre.12139

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. V. D., Hornes, M., & Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. Nucleic acids research, 23(21), 4407-4414.

Waits, L. P., Talbot, S. L., Ward, R. H., & Shields, G. F. (1998). Mitochondrial DNA phylogeography of the north american brown bear and implications for conservation. Conservation Biology, 12(2), 408-417. https://doi.org/10.1111/j.1523-1739.1998.96351.x

Weakley, A.S., and Southeastern Flora Team. (2023). Flora of the Southeastern United States Web App. University of North Carolina Herbarium, North Carolina Botanical Garden, Chapel Hill, U.S.A. fsus.ncbg.unc.edu Accessed Dec 10, 2023.

Wold, J., Koepfli, P., Galla, S. J., Eccles, D., Hogg, C. J., Le Lec, M. F., Guhlin, J., Santure, A. W., & Steeves, T. E. (2021). Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. Molecular Ecology, 30(23), 5949-5965. https://doi.org/10.1111/mec.16141

Ye, M., Wang, X. I. A. O. P. E. N. G., Zhou, Y. U. L. I., Huang, S., & Liu, A. I. R. O. N. G. (2021). Genetic diversity and population structure of cultivated *Dendrobium huoshanense* (CZ Tang et SJ Cheng) using SNP markers generated from GBS analysis. Pak. J. Bot, 53(5), 1683-1690.

Zavodna, M., Grueber, C. E., & Gemmell, N. J. (2013). Parallel tagged next-generation sequencing on pooled samples–a new approach for population genetics in ecology and conservation. PLoS One, 8(4), e61471.

Tables and Figures

Table *2*: List of genetic Marker types and attributes under mutation and variation, selection, genetic drift and their use.

Marker	Mutation and Variation	Selection	Genetic Drift	Population History	Other Uses (e.g., Genetic Mapping, Phylogenetics)
SNPs (Single Nucleotide Polymorphisms)	Arise from mutations at single nucleotide positions (Altshuler et al., 2000; Kwok, 2001).	Can be under selective pressure (Nielsen, 2005).	Prone to genetic drift, particularly in small populations (Kaiser et al., 1997).	Reveal information about population history, migrations, bottlenecks, and founder effects (Tishkoff and Verrelli, 2003).	Used extensively in genetic association studies and population genetics (Hirschhorn and Daly, 2005).
Microsatellites	Mutate rapidly due to slipped-strand mispairing (Ellegren, 2004; Schlotterer, 2000).	Highly sensitive to genetic drift (Ellegren, 2004).	Useful for assessing population structure (Selkoe and Toonen, 2006).	Used to infer phylogenetic relationships and genetic differentiation (Selkoe and Toonen, 2006).	Commonly used in forensic genetics and studies of kinship (Schlotterer, 2000).
RFLPs (Restriction Fragment Length Polymorphisms)	Result from variations in DNA sequences affecting restriction enzyme recognition sites (Botstein et al., 1980; Kreitman et al., 1994).	May be under selective pressure (Botstein et al., 1980).	Provide insights into historical population dynamics (Avise, 2000).	Historically important in genetic mapping studies (Botstein et al., 1980).	Early markers in genetic diversity and mapping studies (Avise, 2000).
AFLPs (Amplified Fragment	Detect polymorphisms based on variations in restriction enzyme recognition sites	Less affected by genetic drift due to dominant nature	Reveal patterns of genetic differentiati	Useful for reconstructing phylogenetic relationships	Widely used in population genetics and evolutionary

Length Polymorphisms)	combined with PCR amplification (Vos et al., 1995; Bonin et al., 2004).	(presence/absence) (Vos et al., 1995).	on among populations (Bonin et al., 2004).	(Vos et al., 1995).	studies (Bonin et al., 2004).

Table *3*: A table showing range-wide population information on sampled P. canbyi populations.

State	County Name	Number of Populations
West Virginia	Grant	6
Pennsylvania	Bedford	3
Kentucky	Pulaski	1
Kentucky	Madison	1
Kentucky	Carter	1
Kentucky	Jessamine	1
Kentucky	Breckinridge	1
Virginia	Mercer	1
Virginia	Craig	1
Virginia	Russell	1
Virginia	Page	1
Virginia	Scott	1
Maryland	Howard	1
Maryland	Allegany	1
		Total: 21

Table *4*: A table showing range-wide population information on sampled *P. canbyi* populations with the number of individuals collected from those populations.

Population Name	**State**	**Number of individuals**
Sweet Lilly Ridge **(SLR)**	KY	5
Carter Caves **(CC)**	KY	7
Jessamine Creek **(JC)**	KY	6
Berea College **(BC)**	KY	2
Sinking Creek **(WC)**	KY	2
Lutzville **(LE)**	PA	15
Saint Clairesville **(SC)**	PA	15
Shawnee State Park **(SP)**	PA	15
Smoke Hole **(SH)**	WV	15
Brush Creek Falls 1 **(BF-1)**	WV	5
Brush Creek Falls 3 **(BF-3)**	WV	3
False Castle Rock **(FCR)**	WV	15
Blue Rock **(BR)**	WV	16
Powers Hollow **(PH)**	WV	15
Webbs Mill **(WM)**	WV	2
Little River Dolomite Bluffs **(SELU)**	WV	10
Pinnacle **(PIN)**	VA	9
Knobbs Mountain **(KM)**	VA	15
Natural Tunnel Limestone Slopes **(NTLS)**	VA	5
Elk Ridge Land Trust **(EG)**	MD	1
Town Creek **(TC)**	MD	8

Table 5: Table contains a list of parameters by which the dataset was filtered and the settings and thresholds by which they were filtered.

Filtering Parameter	Tool Used	Threshold/Settings	Number of SNPs filtered	Number of SNPs left after filtering
Filter Biallelic SNPs	BCFtools/SAMTools	bcftools view -m2 -M2 -v snps	2,164,011	1,897,199
Site thinning (thin) and Site Missingness (max-missing)	VCFTools	10 and 0.4 respectively	1,851,774	45425
Minor Allele Frequency (MAF) and Minor Allele Count (MAC)	VCFTools	0.01 and 1 respectively	23729	21,696
Linkage Disequilibrium	PLINK	R2 >= 0.8	21,696	21,696

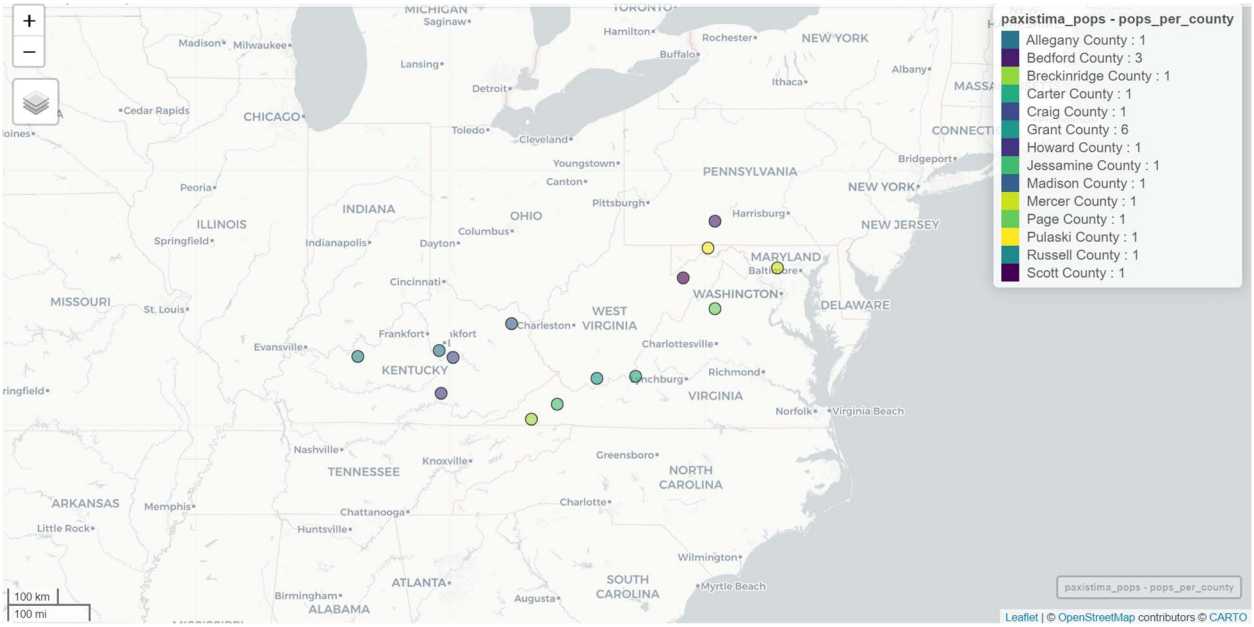


Figure 2: A map showing the disjunct distribution of the extant *Paxistima canbyi* populations on the county level. Legend has county names and number of populations.

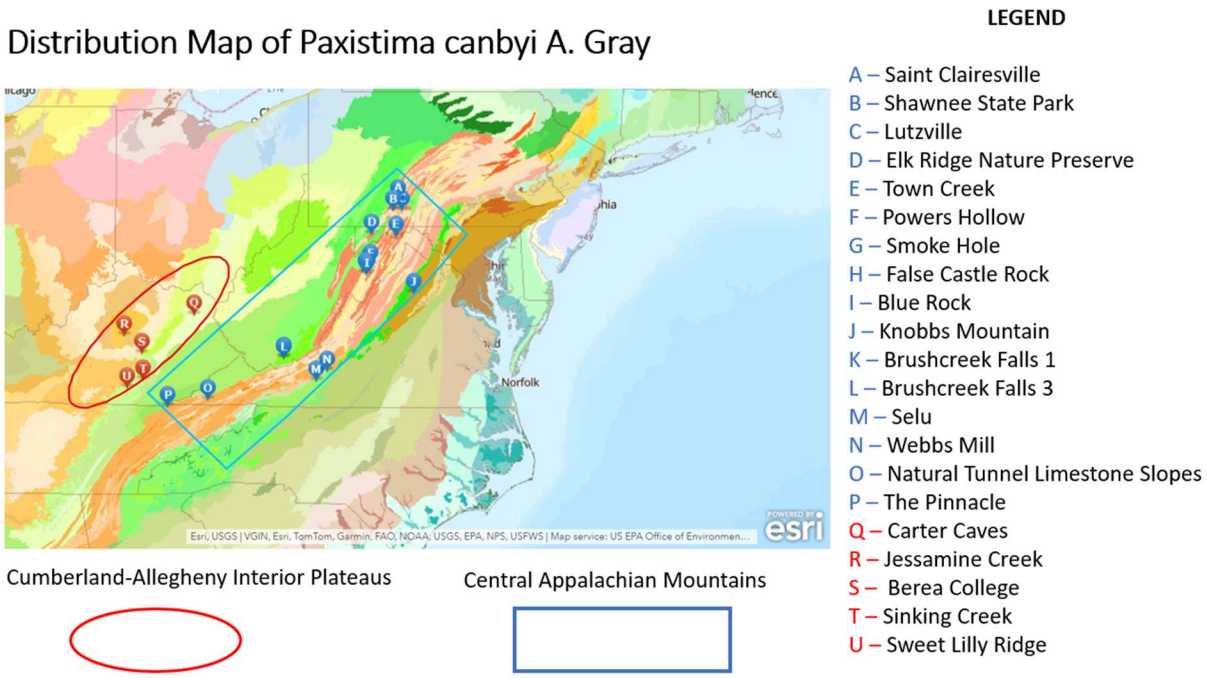


Figure 3: A map showing the distribution of all sampled *P. canbyi* populations across its range considering all the distinct terrestrial ecoregions these populations occupy. A portion of these populations spread along the northern and southern sandstone ridges ecoregion which make up the Central Appalachian Mountains; this is depicted by populations in the blue-colored rectangle. While another occupies group occupies the Cumberland-Allegheny Interior Low Plateaus shown by populations in the red-colored circle. The rectangle and square also show the disjunction in the distribution of *P. canbyi* across its range.

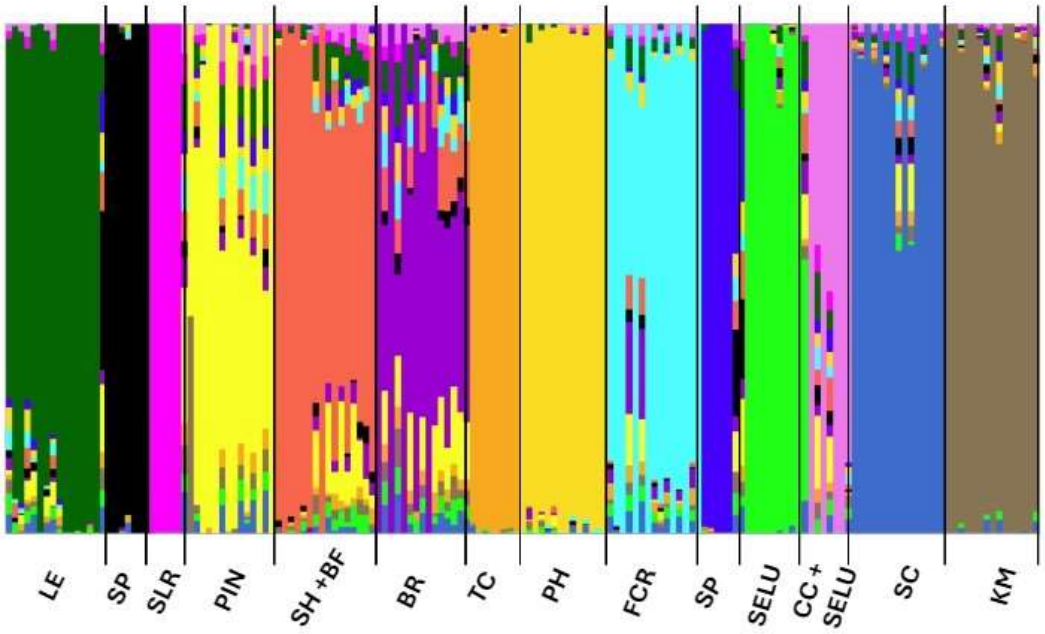


Figure *4*: Plot of sNMF ancestry coefficient proportions derived from ancestry plot for 14 sampled populations of *Paxistima canbyi* A. Gray with K=14. Populations appear to be distinctive genetic clusters by themselves except for 4 populations (Smoke Hole, Brush creek falls, Carter Caves, and SELU) which cluster into two genetic clusters. Abbreviations in this figure match those reported for sampled populations in table 4.

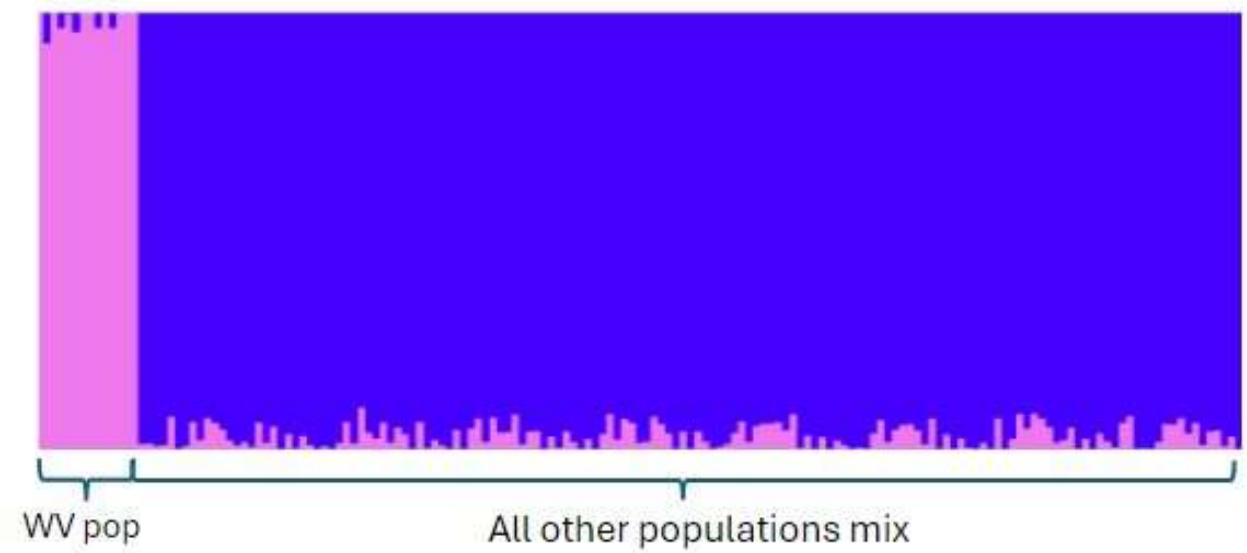


Figure *5*: Plot of SNMF ancestry coefficient proportions derived from ancestry plot for 14 sampled populations of *Paxistima canbyi* A. Gray with K=2. Populations do not cluster by their known disjunct distribution. Blue-colored cluster shows individuals from all the other populations mixing with a lot of admixtures from Powers Hollow (colored pink) a West Virginia population.

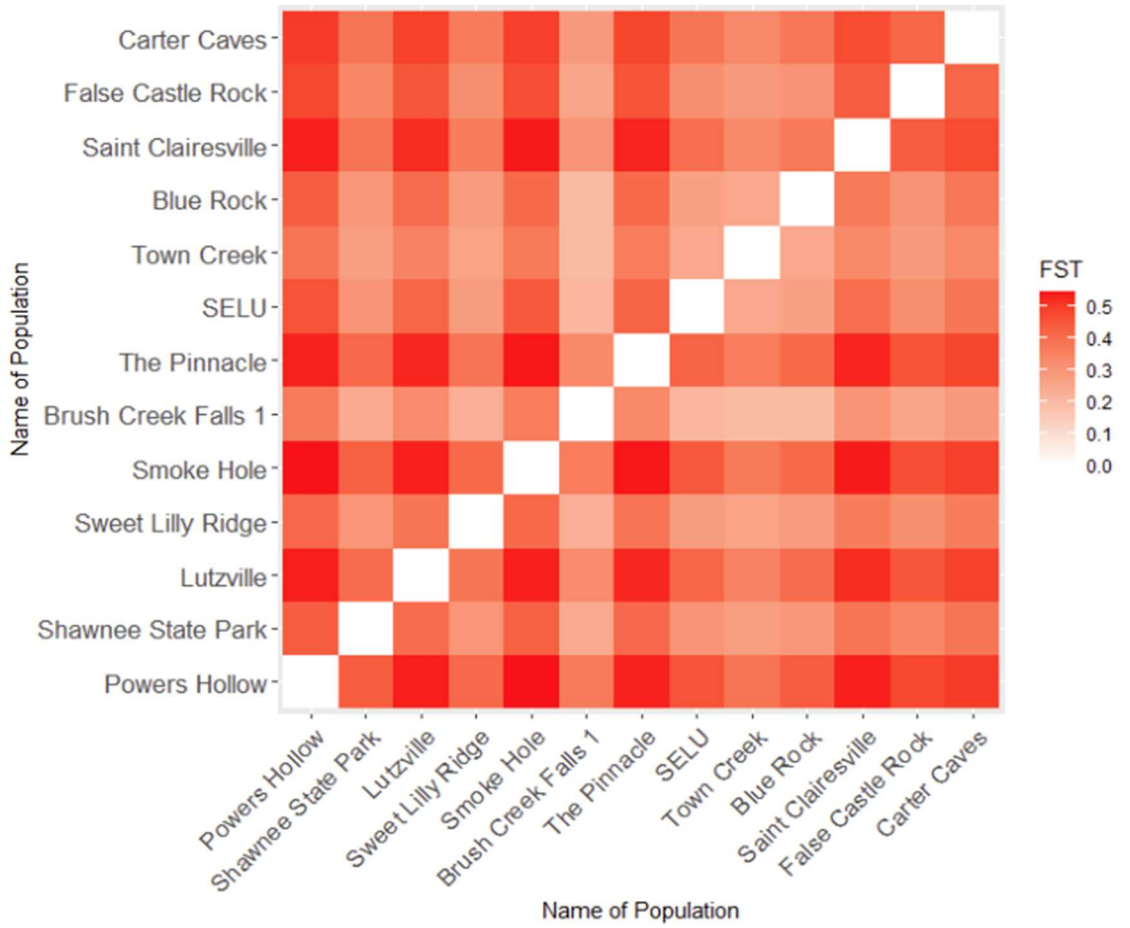


Figure *6*: Heatmap of pairwise FST values. A significant level of genetic differentiation was observed between populations (FST > 0.15).

Supplemental Tables and Figures

Figure *7*: Discriminant analysis of principal components (DAPC) scatterplot of the first two linear discriminant axes showing the spatial relationship between populations of *Paxistima canbyi*. Eigenvalues for the first 60 principal components (PCs) are shown on the bottom right corner of the plot. The plot shows that Sweet Lilly ridge, SELU and Powers Hollow have the greatest variance between each other. The central cluster is made up of a mix of populations from the central Appalachian Mountains and the Cumberland - Allegheny interior low plateaus demonstrating that populations do not cluster by their geographic distribution.

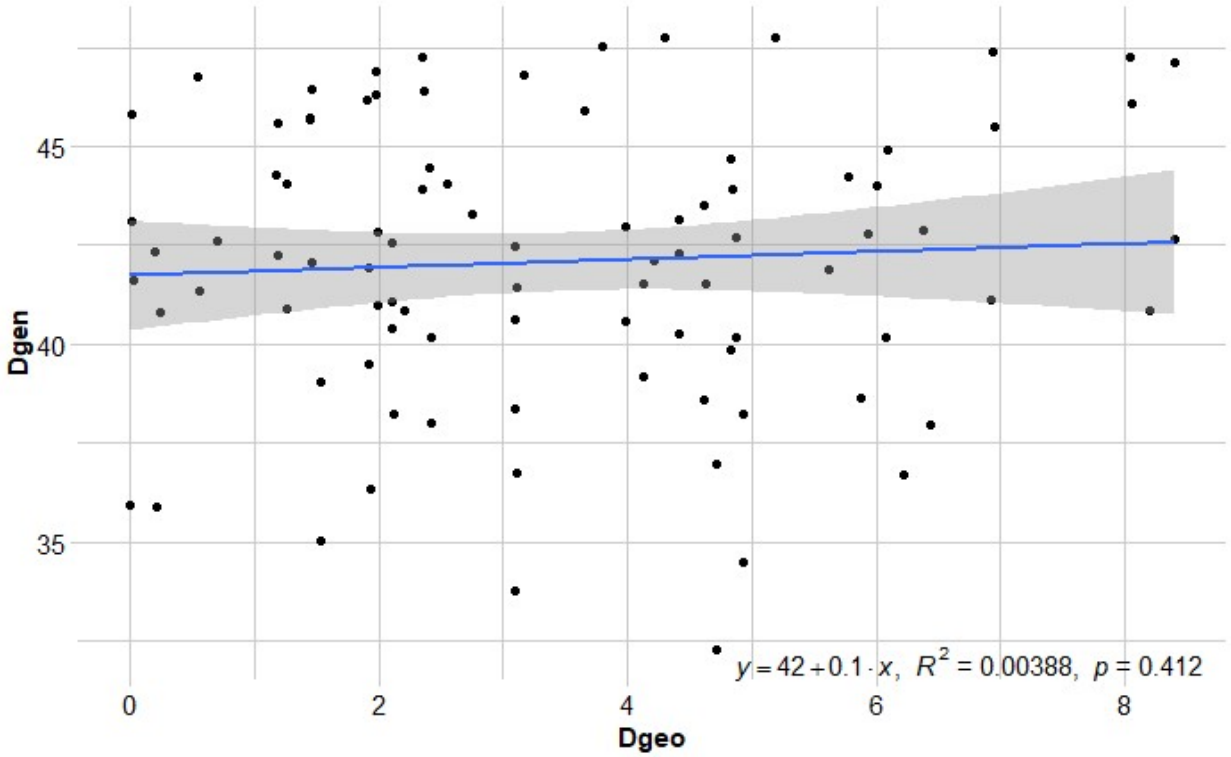
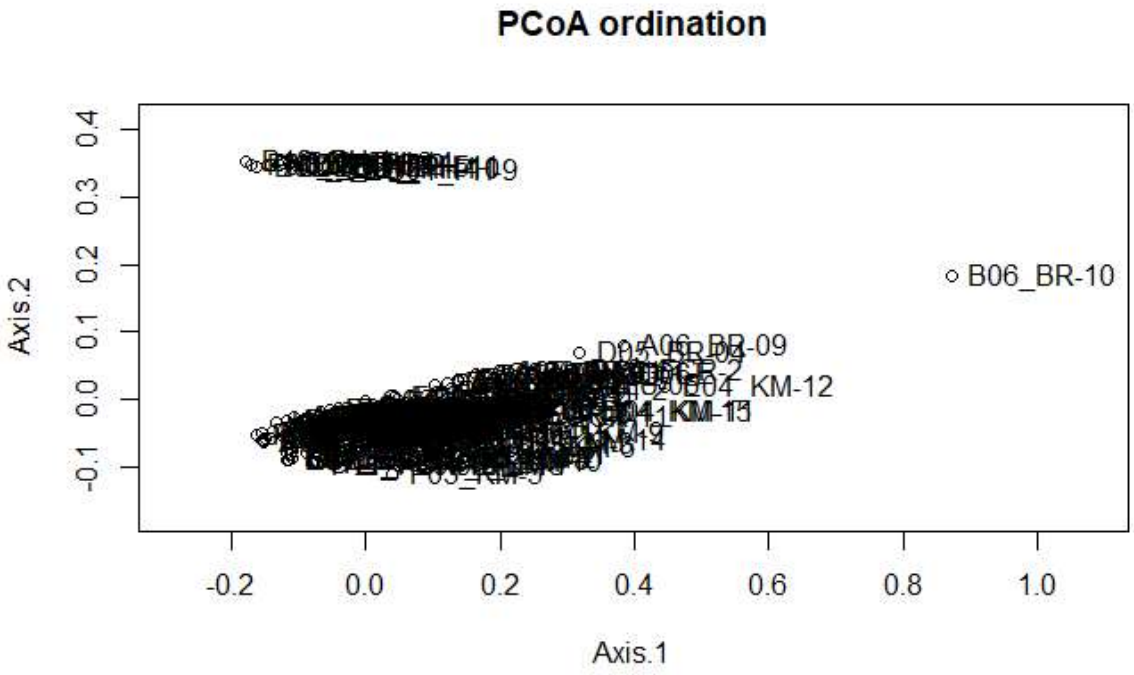
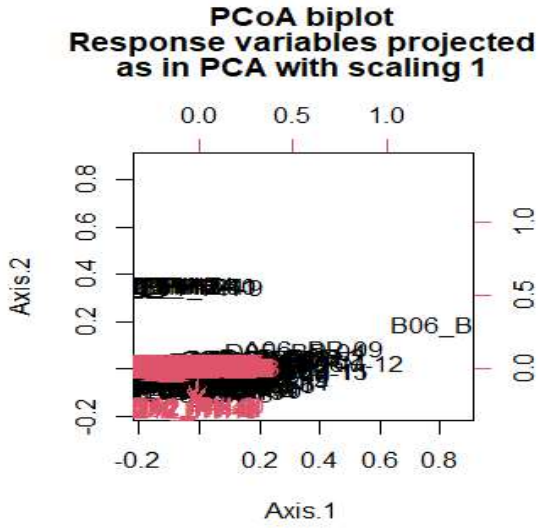
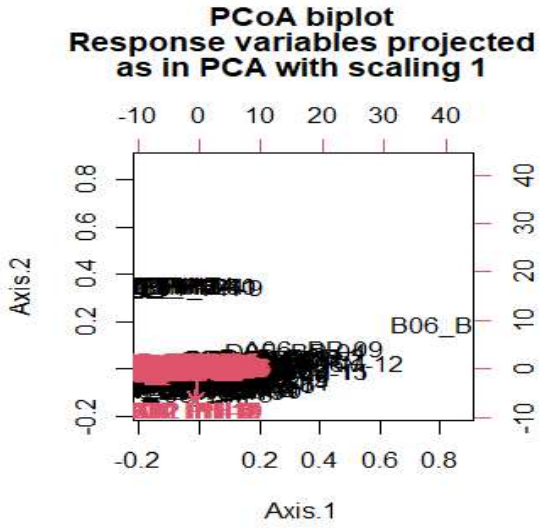


Figure *8*: Mantel Test showing no significant isolation by distance for all populations across the range of *P. canbyi* with genetic distance on the y-axis and geographic distance on the x-axis.


PCoA ordination
Axis.2
Axis.1
B06_BR-10
A06_BR-09
KM-12
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
-0.1
0.1
0.3


PCoA biplot
Response variables projected
as in PCA with scaling 1
Axis.2
Axis.1
B06_B
-0.2
0.2
0.4
0.6
0.8
0.0
0.5
1.0


PCoA biplot
Response variables projected
as in PCA with scaling 1
Axis.2
Axis.1
B06_B
-10
0
10
20
30
40
-0.2
0.2
0.4
0.6
0.8

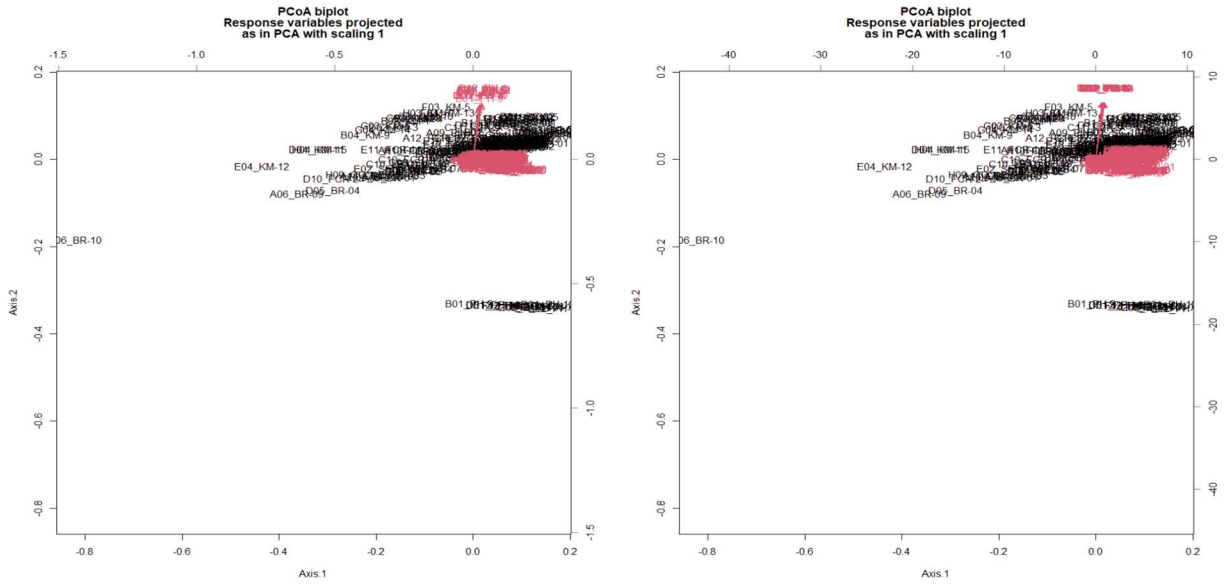


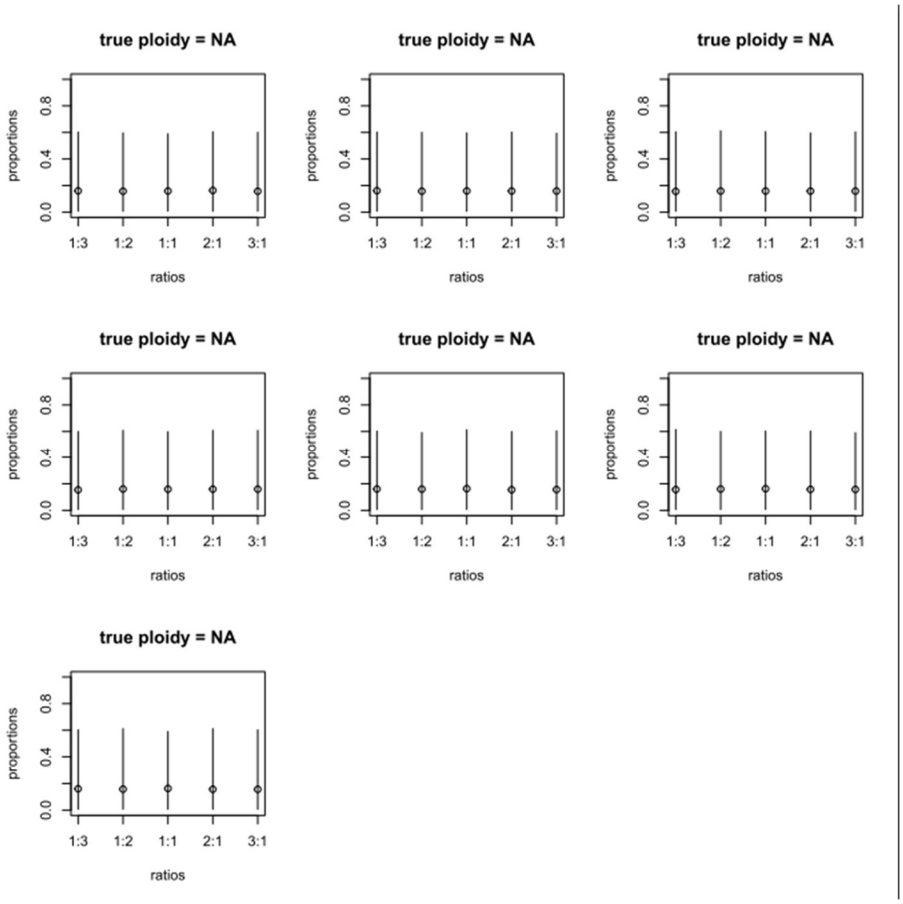
Figure *9*: PCoA ordination plots showing projected samples from *P.canbyi* populations across the range of the species. The first axis indicates the axis with the most variance while the second has the second most variance. Populations are not clustering by their geographic distribution, however, individuals from Powers Hollow (colored red) and Blue Rock seem to cluster away from the main cluster

Table *6*: Table of pairwise FST values (Weir and Cockerham, 1984).

	Powers Hollow	Shawnee State Park	Lutzville	Sweet Lilly Ridge	Smoke Hole	Brush Creek Falls 1	The Pinnacle	SELU	Town Creek	Blue Rock	Saint Clairesville	False Castle Rock	Carter Caves	Shawnee State Park
Powers Hollow	0	0.50398193	0.53881810	0.41350091	0.54611814	0.36825691	0.53450127	0.46094907	0.38709663	0.44049767	0.53703509	0.48400684	0.50057741	0.4393824107
Shawnee State Park	0.50398193	0	0.48697390	0.36057713	0.49720011	0.32630348	0.48165107	0.38648848	0.35728522	0.38482703	0.47081196	0.42567744	0.44869456	0.3911838599
Lutzville	0.53881810	0.48697390	0	0.38191626	0.53609508	0.32994706	0.52893188	0.42269819	0.35036364	0.40646946	0.52221444	0.45419217	0.49179604	0.4069193725
Sweet Lilly Ridge	0.41350091	0.36057713	0.38191626	0	0.40907727	0.23362365	0.38466308	0.28238609	0.26501117	0.28602706	0.36689103	0.32098611	0.36531245	0.3030056975
Smoke Hole	0.54611814	0.49720011	0.53609508	0.40907727	0	0.36327516	0.54401842	0.44955824	0.37408396	0.40938479	0.53965659	0.46819172	0.49499759	0.4303231589
Brush Creek Falls 1	0.36825691	0.32630348	0.32994706	0.23362365	0.36327516	0	0.33608972	0.21303724	0.19905818	0.19749704	0.30608351	0.25953762	0.28769633	0.2452685995
The Pinnacle	0.53450127	0.48165107	0.52893188	0.38466308	0.54401842	0.33608972	0	0.42418272	0.36276387	0.40901987	0.53260417	0.45944402	0.48646902	0.4123937612
SELU	0.46094907	0.38648848	0.42269819	0.28238609	0.44955824	0.21303724	0.42418272	0	0.25113958	0.27042412	0.40258870	0.32138832	0.38380293	0.303799671
Town Creek	0.38709663	0.35728522	0.35036364	0.26501117	0.37408396	0.19905818	0.36276387	0.25113958	0	0.25067692	0.33614277	0.28784748	0.33533454	0.2795666382
Blue Rock	0.44049767	0.38482703	0.40646946	0.28602706	0.40938479	0.19749704	0.40901987	0.27042412	0.25067692	0	0.37224034	0.30735432	0.37743363	0.2978402775
Saint Clairesville	0.53703509	0.47081196	0.52221444	0.36689103	0.53965659	0.30608351	0.53260417	0.40258870	0.33614277	0.37224034	0	0.43880050	0.47664270	0.3870001181
False Castle Rock	0.48400684	0.42567744	0.45419217	0.32098611	0.46819172	0.25953762	0.45944402	0.32138832	0.28784748	0.30735432	0.43880050	0	0.42199450	0.3401649041
Carter Caves	0.50057741	0.44869456	0.49179604	0.36531245	0.49499759	0.28769633	0.48646902	0.38380293	0.33533454	0.37743363	0.47664270	0.42199450	0	0.3830565602
Shawnee State Park	0.43938241	0.39118385	0.40691937	0.30300569	0.43032315	0.24526859	0.41239376	0.30379967	0.27956663	0.29784027	0.38700011	0.34016490	0.38305656	0

Table *7*: Barlett's K2 and p value indicating if there is a significant difference between observed and expected heterozygosity. Bartlett's K2 test statistic which compares that variances between Ho and He, the p values which are not significant. Inbreeding coefficient (F_{IS}) and expected and observed heterozygosity (He and Ho) as calculated by "hierfstat."

Pop_Name	No. of indv.	Ho	He	Bartlett's K-squared	p-value	F_{IS}
Powers Hollow	14	0.1618341	0.09172102	2011.5	< 2.2e-16	-0.7922
Brush Creek Falls1	5	0.1469706	0.13650947	92.968	< 2.2e-16	-0.1254
Carter Caves	7	0.172879	0.12881525	1639.5	< 2.2e-16	-0.3995
Bluerock	16	0.1450255	0.16267403	368.8	< 2.2e-16	0.1027
Sweet Lilly Ridge	5	0.1654352	0.09059953	427.33	< 2.2e-16	-0.8843
The Pinnacle	9	0.1454777	0.14378781	522.12	< 2.2e-16	-0.0491
SELU	10	0.1517348	0.0933997	1464.1	< 2.2e-16	-0.6732
Knobbs Mountain	15	0.1566519	0.08920196	1581.9	< 2.2e-16	-0.7703
Lutzville	15	0.1428226	0.1335061	544.63	< 2.2e-16	-0.0763
Saint Clairesville	15	0.1472206	0.1088433	1418.5	< 2.2e-16	-0.36
Shawnee State Park	15	0.1553514	0.14228932	1578.7	< 2.2e-16	-0.0976
Town Creek	8	0.1603291	0.0897952	847.73	< 2.2e-16	-0.8224
Smoke Hole	16	0.1493377	0.15080949	427.33	< 2.2e-16	0.0057
False Castle Rock	15	0.1466444	0.10705365	1533.8	< 2.2e-16	-0.3874


true ploidy = NA
proportions
0.0
0.4
0.8
1:3
1:2
1:1
2:1
3:1
ratios

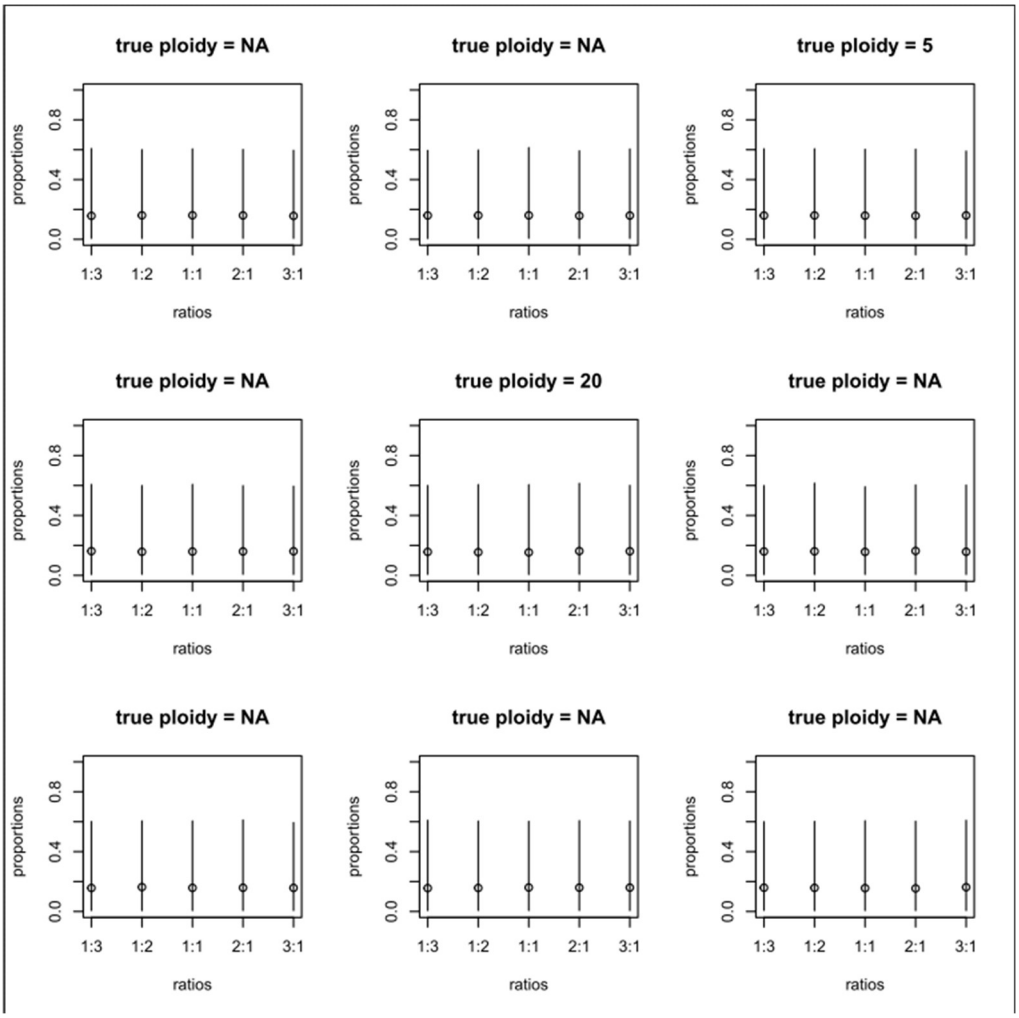


Figure *10*: Plots showing Allelic proportions and ratios across 14 populations. On the x-axis is the estimated allelic proportions of the population and on the y-axis are the ratios within which these of these allelic proportions are.