Spring 2021

Genetic Diversity & Connectivity of *Chasmanthium latifolium* (Poaceae) in Pennsylvania & the Effect on Conservation Status of a Rare Species

Jonathan D. Hayes
Bucknell University, jdh037@bucknell.edu

GENETIC DIVERSITY & CONNECTIVITY OF *CHASMANTHIUM LATIFOLIUM* (POACEAE) IN PENNSYLVANIA & THE EFFECT ON CONSERVATION STATUS OF A RARE SPECIES

Jonathan D. Hayes

Presented to the Faculty of Bucknell University
In Partial fulfillment of the Requirements
for Honors in the Department of Biology

May 2021

Approved by:

Official Approval Obtained via Examination Approval Form
Christopher T. Martine
Thesis Advisor

Official Approval Obtained via Examination Approval Form
Tanisha M. Williams
Thesis Co-Advisor

Official Approval Obtained via Examination Approval Form
Brian Smith
Committee Member

Official Approval Obtained via Examination Approval Form
Rachel Goad
Committee Member

Official Approval Obtained via Examination Approval Form
Ken A. Field
Chair, Department of Biology

Acknowledgements

Firstly, I would like to thank my advisor Dr. Chris Martine for his support and guidance. He went above-and-beyond to make it possible for me to join his lab team in the summer of 2019 and has extended the same level of support throughout the past two years. His willingness to push me out of my comfort zone to apply for grants, present at conferences, develop better science communication skills, and complete this project has made me a better scientist and academic. Your service as a mentor, a scientist, and a friend has made my time at Bucknell exceptionally special. Your love for botany and ability to effectively share that passion to everyone has kept me motivated and taught me many skills that I will use for the rest of my life.

I would also like to thank my co-advisor, Dr. Tanisha Williams, for her never-ending support and willingness to teach me. Your encouragement, flexibility, and insightful feedback has been an important factor in my success. Your assistance in genome analyses has also had a huge impact on the success of this study. Seeing your love of learning, ability to create social change, and spread your love of botany and science in general is truly inspirational and has pushed me to be a better scientist and person.

Additionally, I want to recognize several former members of the Martine Lab and collaborators for their support throughout this study. Thank you, Ariel Antoine for your assistance with DNA extractions and teaching me many of the laboratory techniques that were necessary for this project. Thank you, Cheyenne Moore, and Angela McDonnell, for the guidance that both of you provided with this work and refining the pipeline used for analyses. I would also like to thank Rachel Goad and Scott Schuette for their expertise on the study and assistance with tissue collections.

Furthermore, I would like to express my gratitude towards my family, and many of my close friends for their unending support throughout the years. This thesis would not have been possible without their continual encouragement and help.

Finally, I would like to recognize and thank the funding sources for this research. This project was made possible from the funding granted by the PA DCNR Wild Resource Conservation Program (project 17571) as well as support from the David Burpee Endowment and Wayne E. Manning Internship Fund. Grants awarded to J.D.H. from the Botanical Society of America, Sigma Xi, the Torrey Botanical Society, and the American Society of Plant Taxonomists also helped to fund this research.

Table of Contents

List of Tables

List of Figures

Abstract

Chasmanthium latifolium (Michx.) Yates (Poaceae) is a loosely colonial, rhizomatous, perennial grass species that lives in riparian habitats, making it fittingly referred to as river oats. Native to the southern Midwest and the eastern half of the United States, *C. latifolium* reaches the northeastern edge of its range in Pennsylvania. Within Pennsylvania, eleven extant *C. latifolium* populations are found along four waterways: the Monongahela River, the Susquehanna River, and two tributaries to the Susquehanna River. This limited state distribution exhibits an east-west disjunct distribution, where western populations are largely separated from eastern populations with one centrally located population. Between the limited distribution and number of remaining populations as well as habitat threats, *C. latifolium* is considered critically imperiled (S1) at the state-level by the Pennsylvania Natural Heritage Program. While western populations appear contiguous with the core distribution, central and eastern populations are separated by the Allegheny Mountain range with large distances between populations along the Susquehanna River. Because of these conservation concerns, a better understanding of the natural history and genetics of *C. latifolium* should prove useful for conservation practitioners. My research aims to investigate the genetic diversity and connectivity of the critically imperiled taxon to better understand the natural history of the species and develop scientifically informed conservation practices. This work utilizes a genotyping by sequencing (GBS) approach to generate genomic data for use in population genetics analyses. I found that all populations appear to be genetically healthy, with high levels of heterozygosity and no inbreeding. Western populations appear as one genetic unit with some sub-structuring, while central-eastern populations are genetically different from western populations and other populations along the Susquehanna River system. Although there is currently no evidence of inbreeding, given the

genetic isolation seen within the Susquehanna River populations, inbreeding may be of concern in the future. My research provides an updated, scientifically-informed conservation status assessment of *C. latifolium* in Pennsylvania. This project combines rare plant surveys done by the Pennsylvania Natural Heritage Program and Western Pennsylvania Conservancy with genetic work done at Bucknell University to address broad conservation questions.

Chapter 1: Background on Pennsylvania plants and relevant geography.

Background

Pennsylvania is home to approximately 3,400 plant species, of which almost 2,300 are classified as native or naturalized (Rhoads & Block, 2007). Of the 2300, 582 species are classified as native by the Pennsylvania Department of Conservation and Natural Resources (DCNR), 60% (349) are considered rare, threatened, or endangered (PA DCNR, 2017). The geologic history of the state is linked to impressive levels of plant diversity, including a significant number of rare species that are associated with substrates like serpentinite, limestone, and peat (Rhoads & Block, 2007). My honors thesis research focuses on a rare grass species of conservation concern within Pennsylvania, *Chasmanthium latifolium* (Michx.) Yates (Poaceae). The geographic distribution of *C. latifolium* extends from the southern Midwest and along the eastern half of the United States, extending as far northeast as Pennsylvania (Figure 1; PNHP, 2019a). Although the species is considered globally secure (G5), within Pennsylvania *C. latifolium* is listed as critically imperiled (S1) by the Pennsylvania Natural Heritage Program (PNHP) and has a tentatively undetermined status by the DCNR (PNHP, 2019a; PA DCNR, 2017). The tentatively undetermined classification is selected because this species is believed to be in threat of decline but cannot be included in another classification due to insufficient data (PNHP, 2019b).

The remaining known populations of *C. latifolium* are found along four waterways within the state: the Monongahela River, the Susquehanna River, and two tributaries to the Susquehanna River (Conewago Creek and the Raystown Branch Juniata River). The distribution of extant populations exhibits a large geographic disjunction between eastern populations that occur along the Susquehanna River and Conewago Creek, and western populations that occur

along the Monongahela River, with a centrally located population along the Raystown Branch Juniata River (Table 1; Figure 2). In recent history, there has been a decline in native *C. latifolium* occurrences due to elimination of much of the floodplain habitats that populations once inhabited (PNHP, 2019a). Recent growth in industry, agriculture, housing, and the damming of rivers and altering of flood patterns have all contributed to the current, limited distribution found throughout the state (PNHP, 2019a).

While the current S1 status accounts for the limited distribution and declining habitat, there is limited knowledge on the genetic stability of the Pennsylvania populations. To better understand the status of *Chasmanthium latifolium* in the state, my research utilizes next-generation sequencing technology which will provide insight into the population genetics of the species. The leading hypothesis for this species is that populations are genetically structured by an east-west disjunction, where eastern populations from the Susquehanna River and its tributaries are genetically distinct from the western Monongahela River populations. While this type of isolation would not cause significant concern, further isolation within each side of the state could have significant impacts on isolated sites. By gaining a better understanding of the population genetics of *C. latifolium*, I hope to develop a more informed conservation assessment of the species and ensure the conservation of remaining occurrences in the state.

Geographic considerations

Pennsylvania has a total land area of approximately 45,000 square miles. Elevation ranges from sea level along the Delaware River to over 3,200 feet above sea level at Mount Davis. Pennsylvania consists of six physiographic provinces: Central Lowlands, Appalachian Plateaus, Ridge and Valley, New England, Piedmont, and Atlantic Coastal Plains (PA DCNR,

2021). The three largest provinces, which account for 98% of the land, are the Appalachian Plateaus, Ridge and Valley, and Piedmont (PA DCNR, 2021). State forest account for 58% of the total land area, crops another 14%, and the remaining 28% split between pasture, developed, rural, and federal land use (Widmann, 2016; PASDC, 2019). Pennsylvania is also known for its vast river basin system; the state is covered by 1,100 square miles of water. The watersheds are divided into five major and two minor river basins, of which the three largest basins are the Susquehanna River, Ohio River, and Delaware River (Fayette County Conservation District, 2016; Conservation Voters of Pennsylvania, 2020). The climate of the region is, generally, considered humid continental type, having significant seasonal oscillations with hot summers and cold winters (NCDC, 2009). Temperatures ranging between zero to 100 °F, in the northern and central portions of the state temperature averages 47 °F and 57 °F in the southern region (NCDC, 2009). Precipitation is spread evenly throughout the year with yearly totals ranging between 35–54 inches (NCDC, 2009).

The complex geological and ecological systems (e.g., river basins) found throughout Pennsylvania can have significant impacts on species distribution patterns and gene flow. The sheer distance between eastern and western *C. latifolium* populations presents a clear disjunction hypothesis, however the connection between the central population and other populations is less obvious. While there could be uni- or bidirectional gene flow (Figure 3) between the central-eastern populations and/or central-western populations, there could also be gene flow with only one side of the state, or the central population could be completely isolated (McDonnell et al., 2021; Moore, 2020). The Allegheny Mountain range is a significant geographic barrier that could limit gene flow between the central Raystown Branch Juniata River and the western Monongahela River populations (Li et al., 2019). Both the distance between central and western

populations, as well as the harsh terrain of the Allegheny Mountains can greatly limit the potential for gene flow via wind dispersal or animal dispersal between these populations. On the other hand, the Raystown Branch Juniata River is a tributary of the Susquehanna River, which could facilitate gene flow to the eastern populations. While cross pollination (by wind) between eastern populations and the central population is highly unlikely due to the distance between sites, these populations could be connected by seed dispersal via waterflow, wind, or animal dispersal. More likely, the large distance and geographic landscape between populations could make the central population relatively isolated from other native populations.

Taxon description

Chasmanthium latifolium is a loosely colonial, rhizomatous, perennial grass (Poaceae) species that occurs in a variety of shady habitats from dry shaly cliffs to moist lowlands. Most commonly however, *C. latifolium* is found along waterways, making it fittingly referred to as river oats (Yates, 1966; PNHP, 2019a). This species is easily identifiable by its large, flattened, and drooping spikelets arranged in open panicles, which has made it a desirable ornamental grass (Figures 4A & 4B; Yates, 1966). *Chasmanthium latifolium* is monecious and produces florets that are able to undergo sexual reproduction via wind pollination, as well as florets that exhibit self-pollination (Yates, 1966). Spikelets separate from their pedicels when ripe, thus allowing *C. latifolium* to be ‘self-seeding’ (Davis, 2001). The combination of rhizomatous root growth and self-seeding dispersal result in large population sizes, even in areas like Pennsylvania where there is limited distribution (Figure 4C; Keck et al., 2014).

Of ecological significance, *C. latifolium* is one of two host species in Pennsylvania for the Pepper and Salt Skipper (*Amblyscirtes hegon*) - providing cover from predation and acting as

a larval food source (Lotts et al., 2020; Bess, 2005). Seeds are also a minor food source for birds and rodents while the foliage provides cover for other insects (Neill, n.d.). As a rhizomatous species, the root system aids in the prevention of soil erosion in shaded areas, thus improving water quality (Neill, n.d.).

Chapter 2: Population genomics of Pennsylvania *Chasmanthium latifolium* & the implications on conservation

Introduction

The development of next-generation sequencing (NGS) technology has allowed for genetic studies to be conducted on non-model organisms, which has extended the use of genetic sequencing to be used in many more species (Unamba et al., 2015). In particular, NGS has made it possible to conduct population genetics studies for use in conservation biology (Hunter et al., 2018). These techniques allow us to understand the genetic health of species of concern and how populations are related and connected to one another (see McDonnell et al., 2021; Hohenlohe et al., 2021). In this study, I use NGS to conduct a population genetics study to assess the genetic diversity and population structure of a Pennsylvania state critically imperiled grass (Poaceae) species, *Chasmanthium latifolium*.

Chasmanthium latifolium (Michx.) Yates is a rhizomatous perennial species that is endemic to the southern Midwest and along the eastern half of the United States, extending as far northeast as Pennsylvania (Figures 1 & 3). *Chasmanthium latifolium* is a loosely colonial, rhizomatous, perennial grass (Poaceae) species that occurs in a variety of shady habitats from dry shaly cliffs to moist lowlands. Most commonly, *C. latifolium* is found along waterways, making it fittingly referred to as river oats (Yates, 1966; PNHP, 2019a). This species is easily identifiable by its large, flattened, and drooping spikelets arranged in open panicles, which has made it a desirable ornamental grass (Figures 4A & 4B; Yates, 1966). *Chasmanthium latifolium* is monoecious and produces both chasmogamous and cleistogamous flowers (Yates, 1966). Chasmogamous florets are able to undergo sexual reproduction via wind pollination, however, cleistogamous florets only exhibit self-pollination (Yates, 1966). Spikelets separate from their

pedicels when ripe, thus allowing *C. latifolium* to be 'self-seeding' (Davis, 2001). The combination of rhizomatous root growth and self-seeding dispersal result in large clonal population sizes, even in areas of limited distribution, such as Pennsylvania (Figure 4C; Keck et al., 2014).

Although the species is globally classified as a secure (G5) species, within Pennsylvania, *C. latifolium* is ranked as critically imperiled (S1) by the Pennsylvania Natural Heritage Program (PNHP) and has a tentatively undetermined status by the state (PNHP, 2019a; PA DCNR, 2017). *Chasmanthium latifolium* populations are found along four rivers in the state: the Monongahela River on the western side of the state, the Susquehanna River and one of its tributaries, Conewago Creek, on the eastern side of the state, and the Raystown Branch Juniata River (another Susquehanna River tributary) which is centrally located between eastern and western populations (Table 1; Figure 2). Pennsylvania populations of *C. latifolium* appear to be declining due to relatively recent growth in industry, agriculture, and housing (PNHP, 2019a). Many floodplain areas *C. latifolium* once inhabited have been eliminated and much of the remaining habitat has been impacted by damming of rivers, altering of flood patterns, timber harvesting, and invasive species (PNHP, 2019a).

Chasmanthium latifolium is of conservation concern because the species is rare and is at the northeastern edge of its distribution here in Pennsylvania. Understanding the ecological and evolutionary processes that determine species distributions, although an old idea in science (Darwin, 1859; MacArthur, 1972), is still an important concept continuing to be explored with more data and new techniques (Sexton et al., 2009; Sexton et al., 2011). Given the impacts of anthropogenic climate change, there has been a new vigor in trying to understand what determines a species' range limit. Populations located at the edge of a species distribution are

often adapted to the highly complex and dynamic environments (Sexton et al., 2009; Gaston, 2003). While gene flow between the edge and central populations could promote increased genetic diversity by reducing inbreeding depression, this type of gene flow could also decrease fitness by swamping edge populations with maladaptive traits that are less suited for the harsher environments edge populations often inhabit (Sexton et al., 2011). Alternatively, if the edge populations are isolated, there is the potential of increased inbreeding events leading to an increase in homozygous deleterious genes, and populations could also be more vulnerable to genetic drift (Dolgin et al., 2007; Frankham, 2010). Gene flow between edge populations can increase genetic diversity and reduce inbreeding, while maintaining adaptive traits, in certain environments (Sexton et al. 2009; Sexton et al., 2011). To fully understand range limit dynamics, we must understand the species' spatial and temporal variation, evolutionary history, as well as abiotic and biotic interactions leading to current distribution. By using populations genomics methods and expert botanical knowledge, I can start to illuminate the complex and dynamic landscape that determines a species' distribution and range limits.

I hypothesized support for one of three hypotheses regarding Pennsylvanian *C. latifolium* populations. Due to the Allegheny Mountain range acting as a potential barrier of gene flow between the two waterways that endemic populations inhabit, as well as water dispersal as a mechanism to connect populations along a river system, the leading hypothesis for my work was: central-eastern and western populations are isolated from each other, but populations within each region will have high levels of gene flow. Alternatively, if long distance gene flow of *C. latifolium* between rivers is better mediated than expected by the proposed mechanisms of wind dispersal and zoochory, there may be one statewide metapopulation with gene flow among all localities (i.e., no population structure). Another possibility is that gene flow via water dispersal,

wind dissemination, and zoochory is very limited within Pennsylvania populations and cleistogamous self-pollination may be a predominant reproductive method; if this is the case, it is expected that there are eleven distinct populations where all populations are genetically isolated and there is very little gene flow among them.

These hypotheses are rooted in our understanding of the tight link between the reproductive biology of *C. latifolium* and consequent opportunities for gene flow. *Chasmanthium latifolium* spikelets separate from their pedicels when ripe and naturally fall to the ground, which by itself offers limited opportunity for seed dispersal. However, the riparian environment that many populations inhabit could aid in gene flow via the unidirectional flow of water (Honnay et al., 2010). Flood damage and heavy rains could wash stems, rhizomes, and spikelets downstream, which has the potential to result in a downstream accumulation of genetic diversity, termed the 'unidirectional dispersal hypothesis' (Figure 3; Honnay et al., 2010; Yan et al., 2016). Upstream gene flow is still possible in some species through both biotic and abiotic mechanisms, resulting in bidirectional dispersal (Figure 3); however, this is much more likely to occur in taxa that have insect-mediated seed and pollen dissemination, compared to wind-pollinated and wind-dispersed grass (Honnay et al., 2010; Yan et al., 2016). Although the potential for biotic mediated gene flow is more limited in grasses, upstream dispersion could still occur through zoochory (e.g., epizoochory on birds and mammals or possible endozoochory by waterfowl and fish) (Yan et al., 2016; Lovas-Kiss et al., 2020; Pollux et al., 2006). As it pertains to *C. latifolium*, we expected that seeds are dispersed primarily by waterflow, which would likely result in a gene flow by the unidirectional dispersal model. Thus, we expect to see a greater similarity between populations along the Raystown Branch Juniata River and the Susquehanna River populations due to the connection of these waterways. Meanwhile, the Raystown Branch

Juniata River tributary and the Susquehanna River are not connected to the Monongahela River which prevents the opportunity of water dispersion.

Here I utilize next-generation sequencing to better understand the genetic diversity and connectivity of the extant populations of *C. latifolium* in Pennsylvania to provide a scientifically-informed conservation assessment and better manage this rare species.

Methods

Sampling and sequencing

Sampling was conducted in collaboration with the Pennsylvania Natural Heritage Program and Western Pennsylvania Conservancy at all eleven extant locations within Pennsylvania, including 5 sites along the Monongahela River and 6 sites along the Susquehanna River and its tributaries (Figure 2). At each site, between 7 and 16 tissue samples were collected from both leaves and seed pods and dried using silica. In total, 133 individuals were collected across the 11 remaining known sites within the state. DNA was extracted from the silica-dried tissue samples using the FastDNA kits (MP Biomedicals, Santa Ana, California). Extracted DNA was quantified using a Qubit dsDNA BR assay kit on a Qubit v2.0 fluorometer (ThermoFisher Scientific, Waltham, Massachusetts). DNA quality was assessed by visualizing 2-5 µL samples on a 1% agarose gel run at 100V for 1.5 hours. Restriction enzyme cleavage was checked on approximately 10% of the samples using EcoR1-HF (New England BioLabs, Ipswich, Massachusetts) and successful cleavage was assessed via gel electrophoresis on 1% agarose at 100V for 1.5 hours. Following quality and quantity assessments, samples were shipped to the University of Wisconsin-Madison Biotechnology Center

(https://www.biotech.wisc.edu/services/dnaseq) for additional enzyme testing as well as library preparation and sequencing.

A genotype-by-sequencing (GBS) method was selected because similar methods have been used for population genetics studies on other species in the Martine lab, which yielded promising results (McDonnell et al., 2021; Moore, 2020). Fragment analyses indicated that the restriction enzyme, ApeK1, and the restriction enzyme pair, PstI/MspI, showed the greatest activity with our samples. GBS using a two-enzyme approach has been shown to decrease complexity and generate a more uniform library compared to single-enzyme sequencing; therefore, a two-enzyme GBS approach was used (Poland et al., 2012). Following MstI/PstI digestion of plates, libraries were prepared, quantified, and pooled, and 150bp paired-end sequencing was preformed using a NovaSeq 6000 instrument (Illumina, San Diego, California).

Raw sequencing reads were filtered and assembled following the UNEAK assembly pipeline in TASSEL version 3.0.174 (Lu et al., 2012; Lu et al., 2013). The resulting dataset contained 133 native individuals across 11 populations, and 999 single nucleotide polymorphisms (SNPs). The filtered SNPs were analyzed using various packages in R version 3.6.3 (R Core Team, 2020).

Genetic diversity & population structure

The R package, vcfR (Knaus & Grünwald, 2017) was used to convert the vcf output file generated from TASSEL, into a hierfstat format, which can be used by the hierfstat package (Goudet, 2005). To better understand genetic variation within and among populations, hierfstat was used to calculate *F*-statistics, including the inbreeding coefficient (F_{IS}), and F_{ST}, which gives the proportion of genetic variance observed in a population relative to the total genetic variance

observed across all collected individuals (Holsinger and Weir, 2009). Hierfstat was also used to calculate the observed heterozygosity (H_O) and expected heterozygosity (H_E) which can provide useful insight into genetic stability of populations. A Bartlett's test of homogeneity of variances (Bartlett, 1937) was also performed, using base R, to assess if the difference between H_O and H_E was significant.

A principal component analysis (PCA) was performed using the package adegenet (Jombart, 2008). Adegenet was also used to conduct a Discriminant Analysis of Principal Components (DAPC) which uses discriminant analysis to assign membership probabilities for analyzing principal components (Jombart & Collins, 2015). The LEA or "Landscape and Ecological Association" was used to determine the number of ancestral populations (K) through a comparison of cross-entropy values (K=1-11 was tested), and generate a STRUCTURE plot, which was used to assess population structure and admixture (Frichot & François, 2015).

Results

GBS data

Genetic sequencing yielded 315.0 million raw reads and an average of 3 million raw reads per individual (lowest: 184; highest: 4.6 million). After assembly of the dataset and hard filtering (filtering for read quality and depth, missingness per site, missingness per individual, allelic frequency, and linkage disequilibrium), 999 SNPs were used for analyses. These data were from 133 Pennsylvania native individuals collected from 11 populations.

Genetic diversity & population structure

The Pennsylvania native populations had a global F_{ST} of 0.1130, showing moderate differentiation across populations (Wright, 1978). Overall, there was no inbreeding observed among these individuals (F_{IS} = -0.6219). Globally, the observed heterozygosity was greater than the expected heterozygosity (global H_O = 0.6590; global H_E = 0.3969), and a Bartlett's test confirmed significant differentiation (Bartlett's K^2 = 1209.3, df = 1, p-value < $2.2e^{-16}$). Population-level statistics reflected the global statistics - across all populations, there was a significantly greater observed heterozygosity than expected heterozygosity and no inbreeding detected (Table 2). Although the global F_{ST} was moderately high, a pair-wise F_{ST} test showed high levels of gene flow among populations along the Monongahela River (C1N, C1S, C2, FH1, FH2), while all of the Susquehanna River populations (central population: RB; eastern populations: H, CR, SFR, NFR, EC) showed high genetic differentiation from western populations, but also differentiation from other Susquehanna River populations (Figure 5).

The PCA showed that populations on the eastern side of the state cluster together, while western populations cluster together separately, with the central population (RB) clustered intermediately between the eastern and western populations (Figure 6). The STRUCTURE analysis supported K=5 as the best supported number of ancestral populations, however, K=4 to K=7 showed low cross-entropy as well. STRUCTURE analysis showed genetic diversity within populations, yet a clear separation between some populations (Figure 7). Eastern populations are generally genetically different from each other, the central population, and very different from western populations. The central population, RB, showed the most similarity to the eastern, EC, population, while western populations appear as one genetic unit.

Discriminant analysis was able to assign group membership at a rate of 81% which was due to the amount of admixture within populations, as shown by the STRUCTURE analysis. The DAPC scatterplot shows that most populations cluster together, while three eastern Susquehanna River populations (H, NFR, CR) cluster independently (Figure 8). Looking at the STRUCTURE-like plot based on DAPC analyses, all Susquehanna River & tributary (eastern and central) populations are genetically distinct from each other and western populations (Figure 9). The western populations generally cluster together, but have subdivisions within the cluster, where C1S and C2 show genetic difference.

Discussion

Existing at the species range edge, Pennsylvania populations of *Chasmanthium latifolium* may be impacted by several factors that have been identified previously. Edge-of-range taxa have been found to frequently inhabit ecologically marginalized sites (Abeli et al., 2014), have a decrease in seed production (Jump & Woodward, 2003), and experience a greater impact from climate change than populations that are located more centrally within the global distribution (Rehm et al., 2015). The central marginal hypothesis predicts edge-of-range species will exhibit low genetic diversity and show genetic differentiation due to historical genetic drift, founder, inbreeding, and/or bottleneck events (Eckert et al., 2008; Antonovics et al., 2002).

The life history and biology of *C. latifolium* may also influence inbreeding and genetic differentiation of Pennsylvania populations. As part of the Poaceae family, *C. latifolium* is wind pollinated, which has traditionally been assumed to limit the efficiency of pollen transfer, especially over long distances (Friedman & Barrett, 2009; Osborne & Free, 2003). Research done in *Festuca pratensis* (Poaceae) showed that beyond 75 meters gene flow was significantly

limited (Rognli et al. 2000). Thus, over the large geographic distances between many Pennsylvania populations, especially along the Susquehanna River, we might expect to see the sort of genetic isolation found in our F_{ST} , DAPC, and STRUCTURE results. The likelihood of inbreeding would be thought to be relatively high within *C. latifolium* due to the presence of cleistogamous florets and potential limited long-distance dispersal of pollen and seeds, which also aligns with what is expected by the central marginal hypothesis (Eckert et al., 2008; Antonovics et al., 2002). Contrary to the central marginal hypothesis, our results suggest that *C. latifolium* populations show no evidence of inbreeding and genetic diversity is high, despite significant genetic isolation between the two waterways and among the populations along the Susquehanna River and its tributaries.

The PCA indicates that Susquehanna River populations cluster together while the western populations cluster separately. Additionally, the STRUCTURE analysis and DAPC STRUCTURE-like plot showed that western populations are genetically different from all populations to the eastern side of the state. Given that Susquehanna River populations cluster together and appear separate from Monongahela River populations yet show a differentiating genetic structure within the Susquehanna River, *C. latifolium* populations along the eastern waterway may have diverged from each other in more recent history than the differentiation from western populations. Considering the geographic barrier that the Allegheny Mountain range poses between the two waterways, which limits gene flow, this makes sense.

When looking at the distance between populations, the distance between sites along the Monongahela River is much less than the populations along the Susquehanna River system, which are more spread apart (~1-7 vs. ~1-120 miles). These results indicate that distance may

limit gene flow between populations, which would align with previous findings within the Poaceae family (Rognli et al., 2000).

Along the Susquehanna River, significant genetic isolation between populations is observed, most notably, the PCA, STRUCTURE analysis, and DAPC STRUCTURE-like plot all showed genetic distinction within the centrally located RB population. The genetic isolation observed in RB may be due to a founder event, where little gene flow has occurred since. Alternatively, there may have been connecting populations intermediary to RB and eastern populations that have since been extirpated. The isolation within eastern populations could also be due to a founder effect, but because of the closer proximity among populations and long history of disturbance in the Susquehanna River Valley, a genetic bottleneck caused by habitat alteration is a more likely explanation. Another important implication from the genetic isolation observed among Susquehanna River populations, is that long-distance mechanisms of gene flow appear very limited within *C. latifolium*, even via water-dispersal. As observed in other systems, unidirectional down-stream gene flow through water-dispersion would be observed through genetic similarity and connectivity between sites along a river, with populations further downstream having increased genetic diversity (measured by heterozygosity). However, all *C. latifolium* populations along the Susquehanna River were shown to be genetically isolated, indicating that there is very limited down-stream gene flow within this system (Love et al., 2013).

Relating to conservation, these results indicate that Monongahela River populations to the west appear to be of less concern - they are genetically diverse, have no inbreeding, and experience gene flow. Populations occurring around the Susquehanna River, however, may be of greater concern when accounting for genetics. Although these sites are genetically diverse and

not yet inbred, there is very limited gene flow between populations. Due to the genetic isolation observed along the waterway, inbreeding depression may be of concern for these populations in the future. The potential negative effects of genetic drift could also thus have a greater impact on the populations along the Susquehanna River and its tributaries. While crossbreeding that may occur between cultivars and native individuals could limit the potential for inbreeding, it could also inundate native populations with traits maladapted for the harsh Pennsylvania winters. Thus, facilitated gene flow via. seedings from other Pennsylvania sites may be an effective way to maintain adaptive genetic diversity and limit the potential for inbreeding.

Conclusion

The main finding from this work is that populations of *Chasmanthium latifolium* in Pennsylvania are composed of one genetic unit along the Monongahela River with some sub-structuring, and several genetically distinct groups along the Susquehanna River and its tributaries. Our findings indicate that Pennsylvania populations of *C. latifolium* appear genetically healthy as of now. While all populations have high genetic diversity and are not yet inbred, the genetic isolation observed across eastern and central populations within the Susquehanna basin indicates that these populations may be at risk of future inbreeding depression. Western populations, on the other hand, show genetic connectivity within the Monongahela waterway which indicates that these genetically healthy populations also have a greater genetic stability and are less susceptible to genetic drift. In terms of conservation practices, we should continue to conserve all native populations due to the limited number of occurrences throughout the state. However, these results also highlight eastern populations as being the most vulnerable. While all populations appear to be genetically healthy, this population

genetics research revealed that populations along the Susquehanna River and its tributaries may experience greater affects from habitat alterations and other threats to this rare species than Monongahela River populations.

Tables

Table 1. Sample sites where *Chasmanthium latifolium* tissue was collected and information about location and sampling. Exact location information is redacted due to the species' PNHP critically imperiled status. * indicates the water way is a Susquehanna River tributary.

Region	Water Way	Site Name	Abbrev.	Collector	Tissue Collection Date	Plants Sampled	County
East	Susquehanna River	Haines	H	C.T. Martine	9/13/2018	15	Lancaster Co.
East	Susquehanna River	North of Fisherman Run	NFR	C.T. Martine	9/13/2018	16	Lancaster Co.
East	Susquehanna River	South of Fisherman Run	SFR	C.T. Martine	9/13/2018	12	Lancaster Co.
East	Susquehanna River	Chickies Ridge	CR	C.T. Martine	9/13/2018	11	Lancaster Co.
East	Conewago Creek*	Erney Creek	EC	T.M. Williams	9/5/2019	12	York Co.
Central	Raystown Branch Juniata River*	Raystown Branch	RB	S. Schuette	9/24/2019	15	Montour Co.
West	Monongahela River	Cheat River 1N	C1N	G. Malone	9/27/2018	8	Fayette Co.
West	Monongahela River	Cheat River 1S	C1S	G. Malone	9/27/2018	7	Fayette Co.
West	Monongahela River	Cheat River 2	C2	S. Schuette	9/27/2018	8	Fayette Co.
West	Monongahela River	Friendship Hill 1	FH1	G. Malone	9/28/2018	14	Fayette Co.
West	Monongahela River	Friendship Hill 2	FH2	G. Malone	9/28/2018	15	Fayette Co.

Table 2. Expected and observed heterozygosity (H_O and H_E) and inbreeding coefficient (F_{IS}) as calculated by hierfstat. Significant difference between expected and observed heterozygosity was observed within all populations as assessed by a Bartlett's test. As well, all populations showed significant F_{IS} values as assessed using a 95% confidence interval. All populations show a greater than expected genetic diversity and no inbreeding.

Region	Population	H_O	H_E	Bartlett's K^2 $*p < 2.2e^{-16}$	F_{IS}	F_{IS} [95% CI]
East	Haines	0.7103	0.3837	421.6*	-0.8512	[-0.8615, -0.8408]
	North of Fisherman Run	0.7355	0.3944	457.0*	-0.8649	[-0.8754, -0.8526]
	South of Fisherman Run	0.5154	0.4097	372.1*	-0.2580	[-0.2905, -0.2238]
	Chickies Ridge	0.7388	0.3926	454.6*	-0.8817	[-0.8911, -0.8713]
	Erney Creek	0.7400	0.3985	482.0*	-0.8569	[-0.8694, -0.8434]
Central	Raystown Branch Juniata River	0.6440	0.4047	401.3*	-0.5915	[-0.6087, -0.5746]
West	Cheat River 1N	0.6969	0.3958	423.1*	-0.7608	[-0.7789, -0.7412]
	Cheat River 1S	0.6212	0.4147	475.5*	-0.4978	[-0.5246, -0.4692]
	Cheat River 2	0.5507	0.3863	355.5*	-0.4258	[-0.4577, -0.3963]
	Friendship Hill 1	0.6847	0.3909	396.4*	-0.7517	[-0.7648, -0.7387]
	Friendship Hill 2	0.6119	0.4035	355.8*	-0.5166	[-0.5360, -0.4976]

Figures

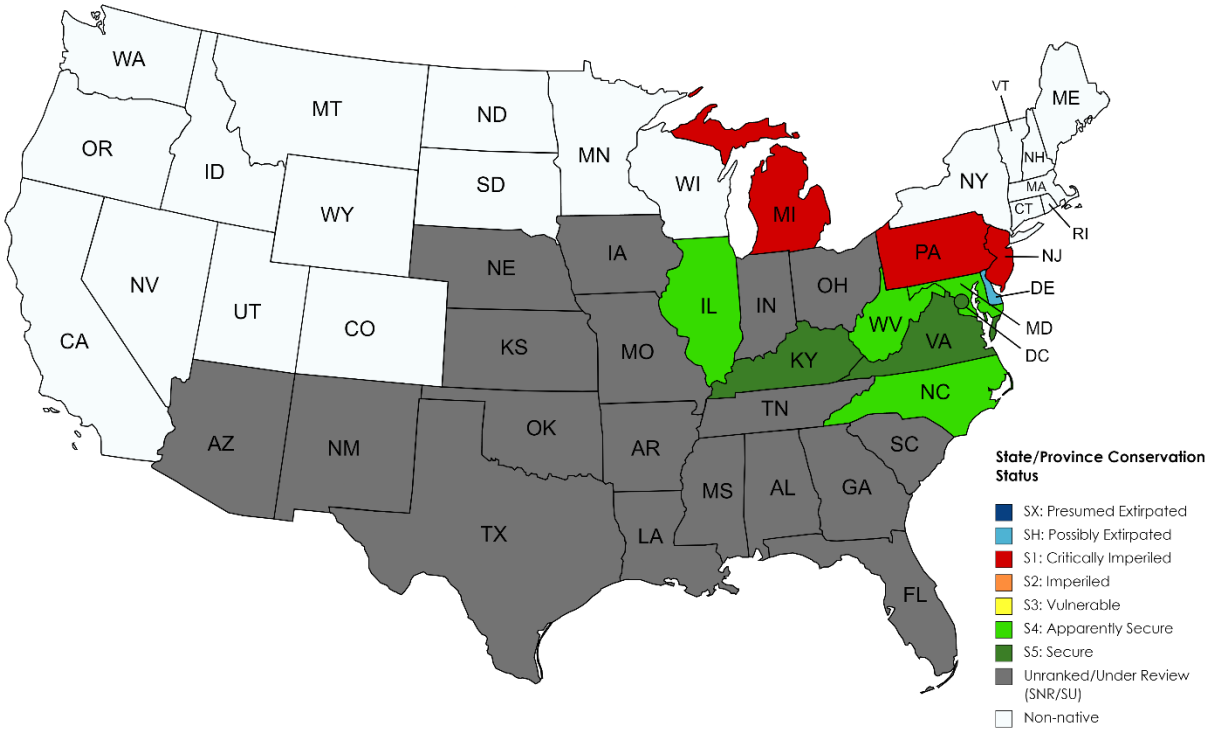

Figure 1. Range map for *Chasmanthium latifolium* by state with state conservation status (NatureServe, 2021).

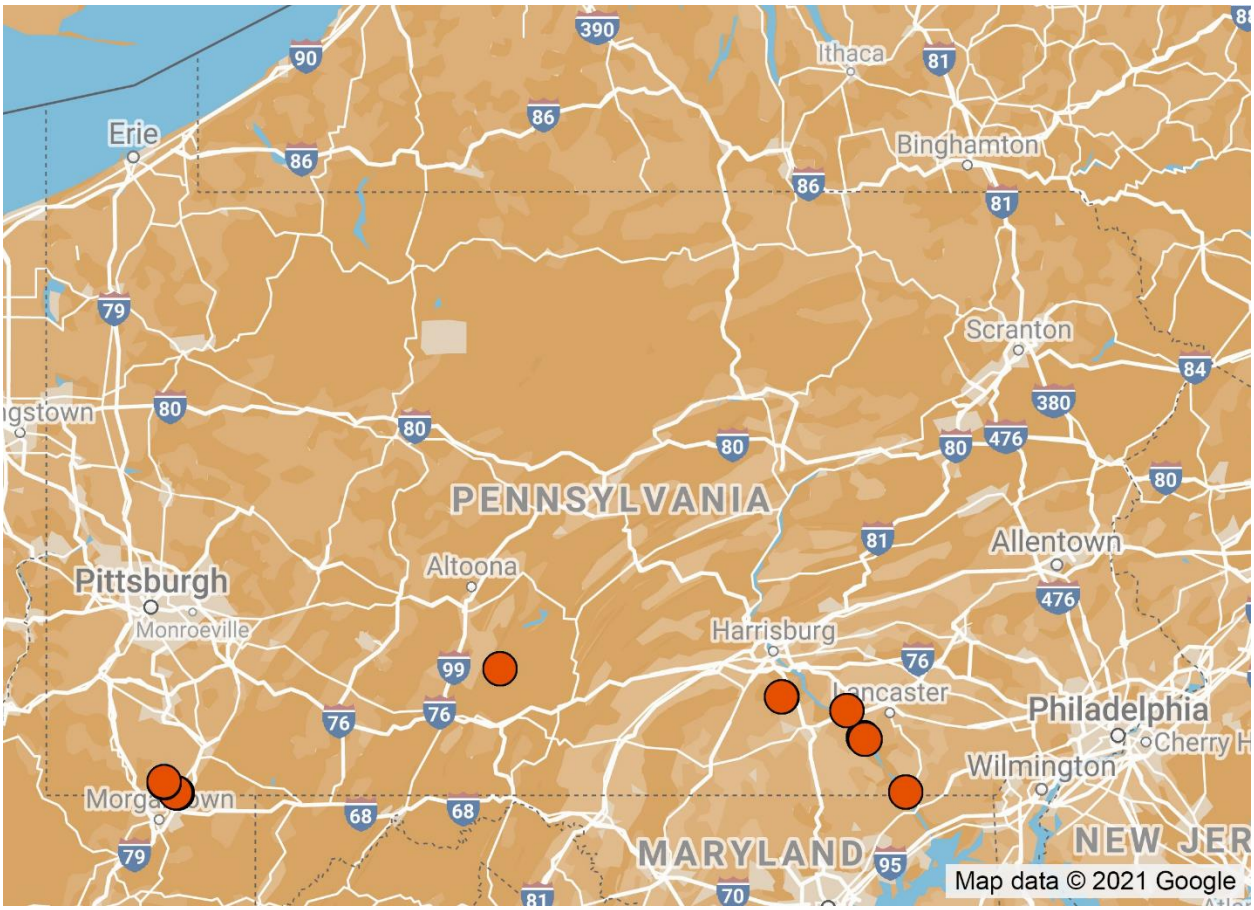

Figure 2. Map of all Pennsylvania sites where *C. latifolium* was collected. Map generated using Google Maps.

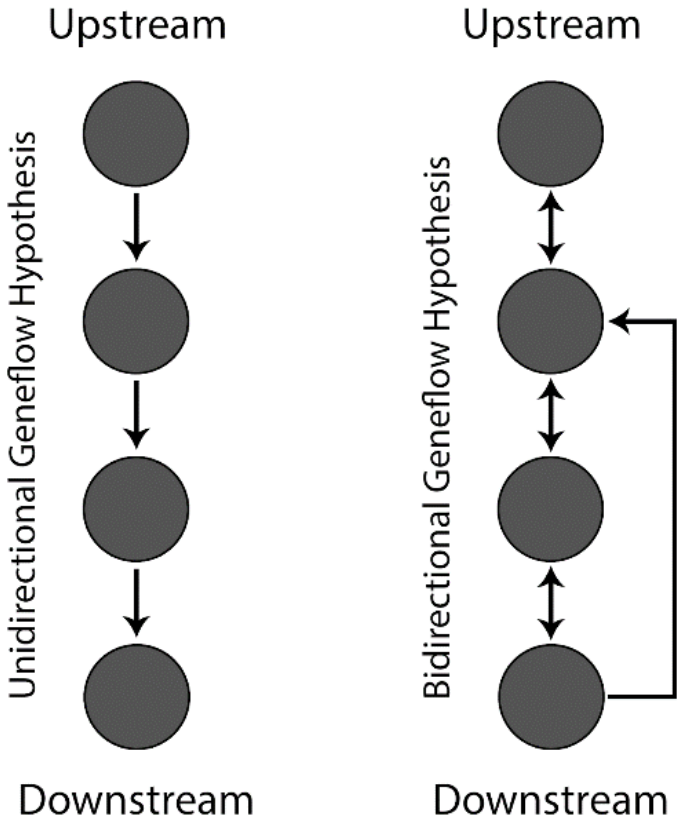

Figure 3 Representation of the unidirectional and bidirectional dispersion hypotheses (Honnay et al., 2010; Yan et al., 2016).

Figure 4. Photos of *Chasmanthium latifolium* characteristic spikelet (A & B). Photos: J. Hayes. *C. latifolium* growing in bluffs habitat above the Susquehanna River (C). Photo: C. Martine

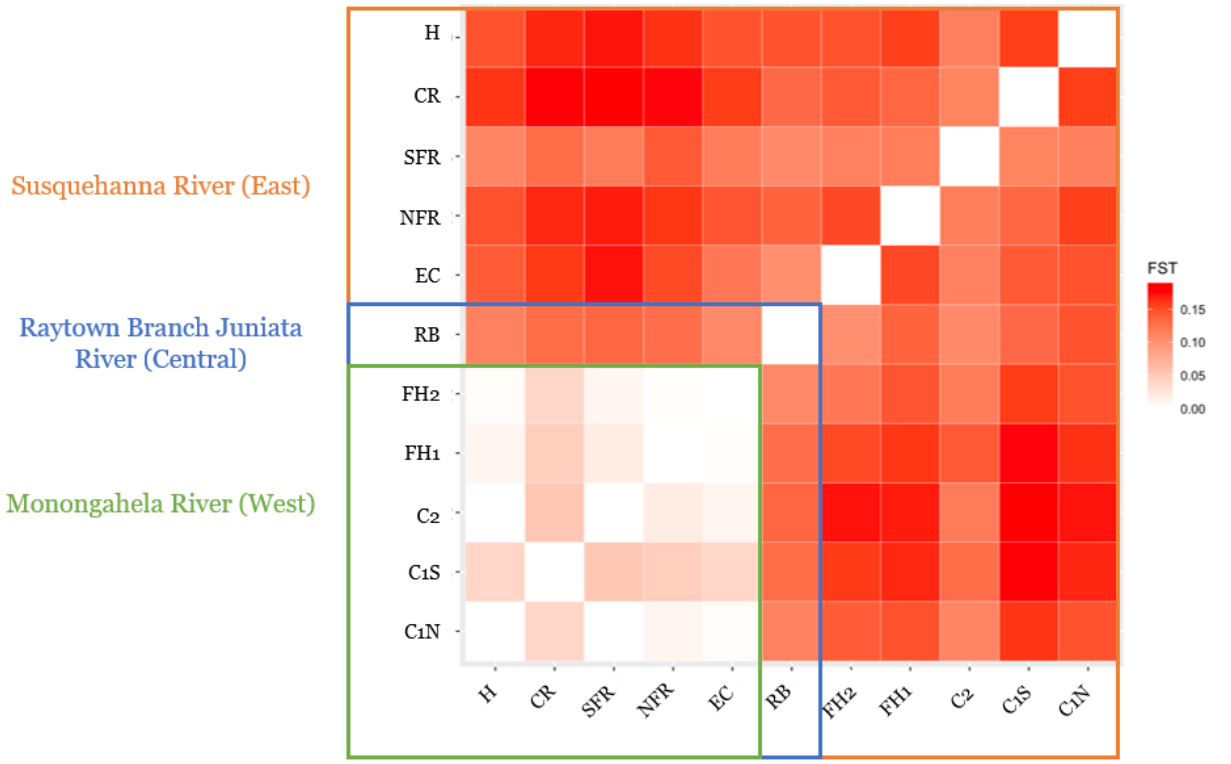

Figure 5. Heatmap of pairwise F_{ST} values. Site abbreviations correspond to Table 1. No genetic differentiation was observed within western populations, while eastern and central populations showed genetic differentiation ($F_{ST} > 0.15$).

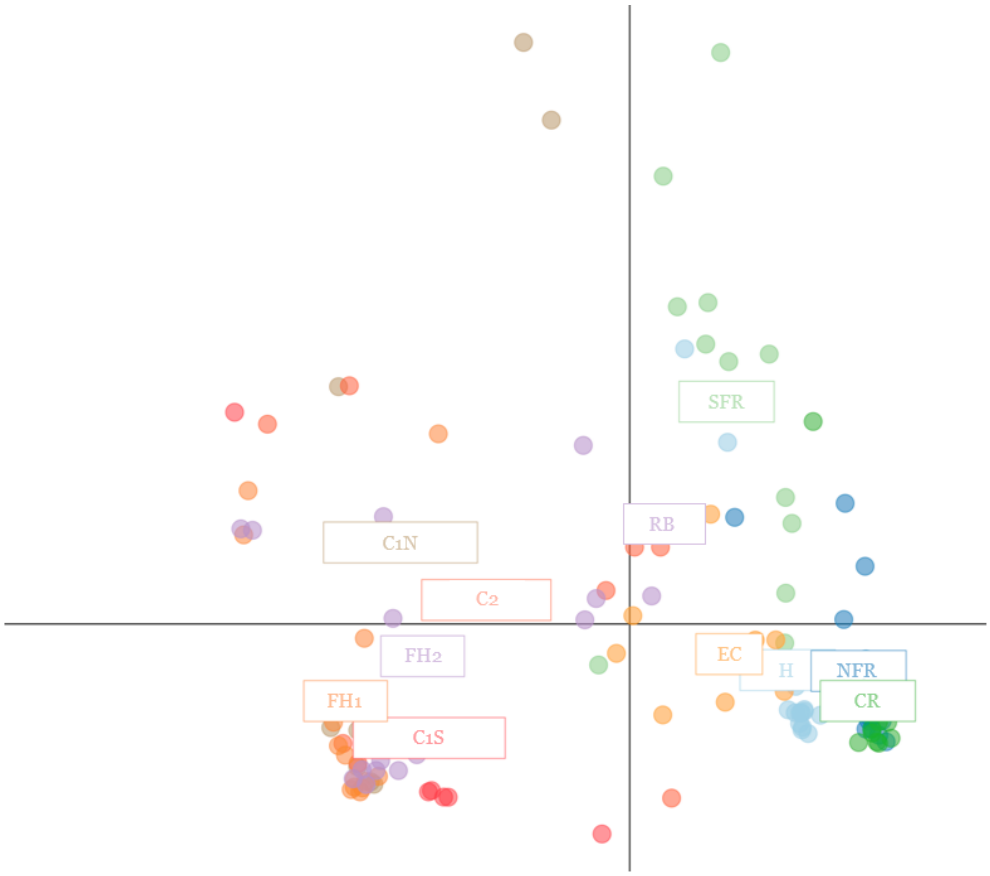

Figure 6. Principal components analysis (PCA) of SNPs from sampled *C. latifolium* shows western populations (C1N, C1S, C2, FH1, FH2) and eastern populations (H, NFR, SFR, CR, EC) clustered together, respectively. The central population (RB) is centrally located between the eastern and western clusters.

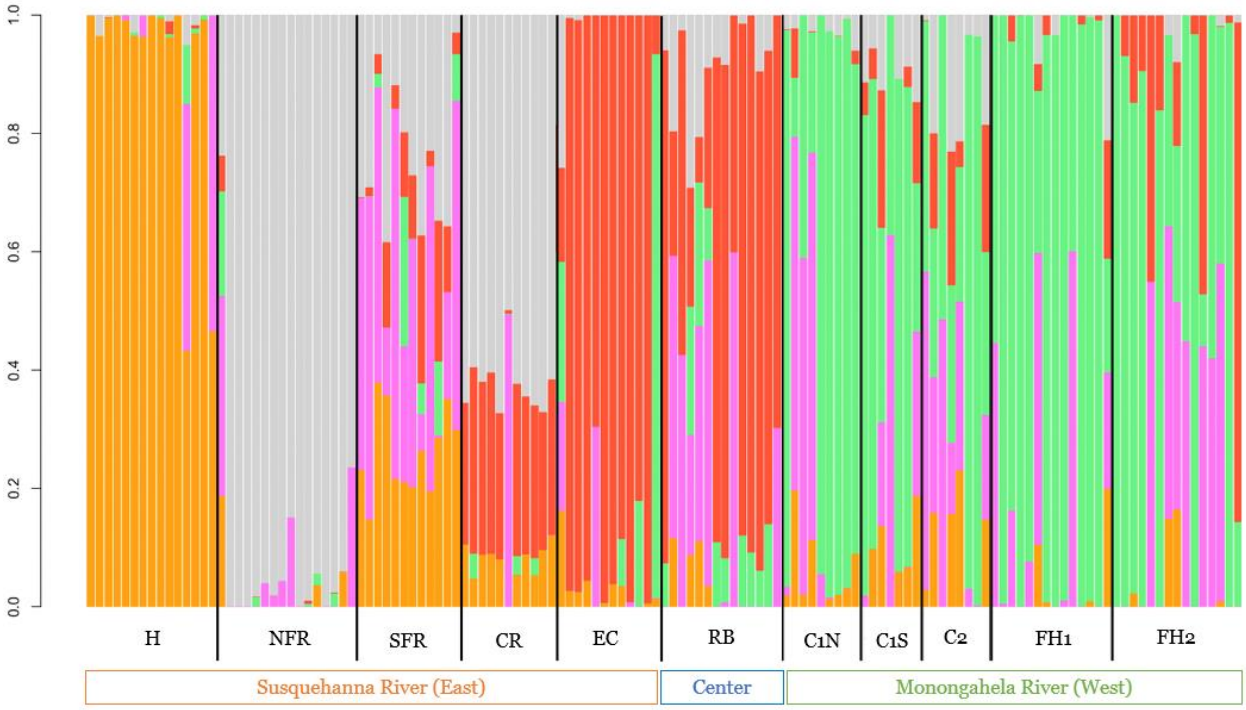

Figure 7. STRUCTURE analysis plot for K=5 genetic units. Eastern populations appear different from each other and western populations. The central population is genetically similar to EC, and western populations appear as one genetically similar unit.

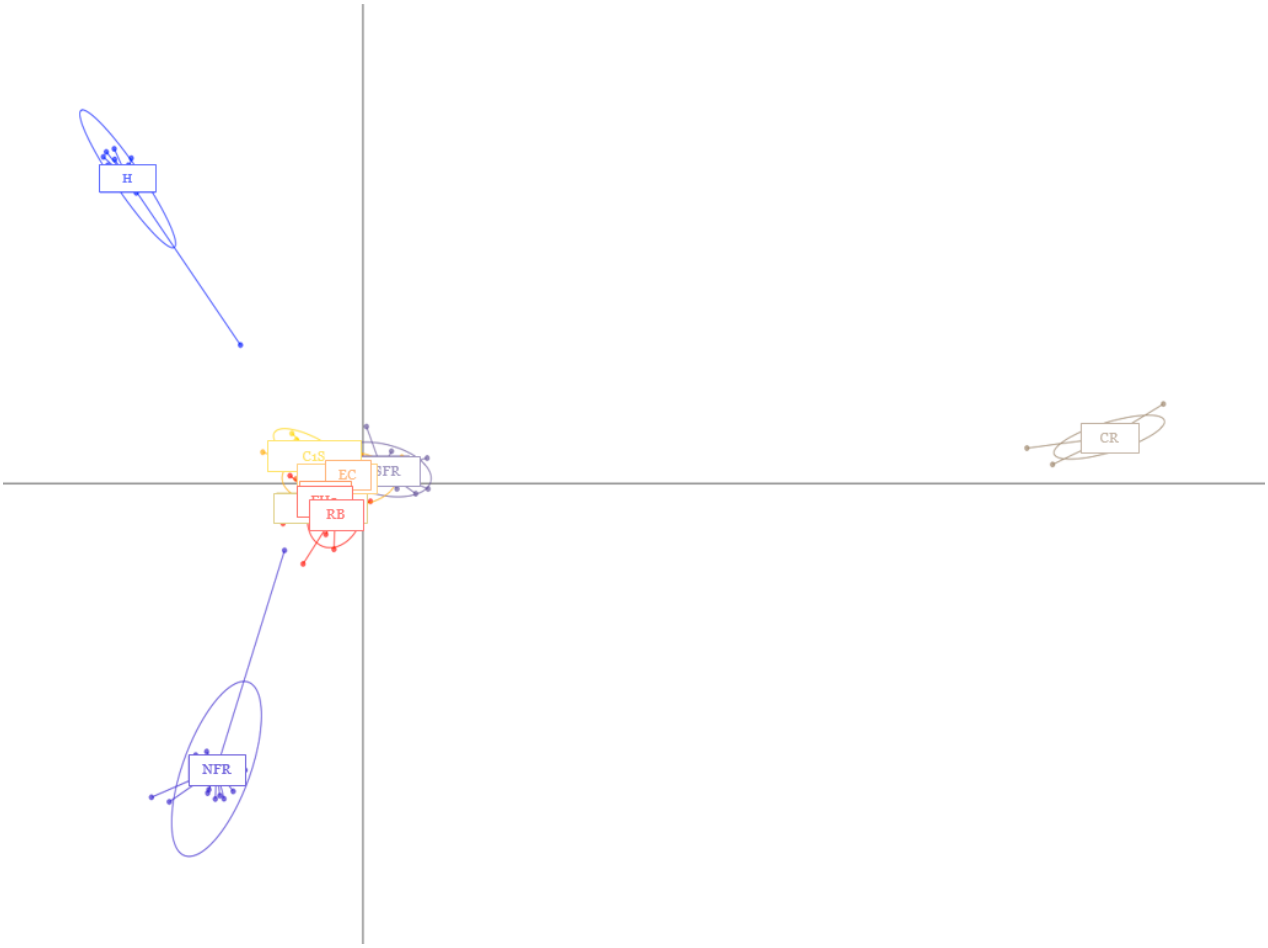

Figure 8. Discriminant analysis of principal components (DAPC) scatterplot showing the spatial relationship between populations of *C. latifolium*. All populations cluster together except for three eastern populations (H, NFR, CR).

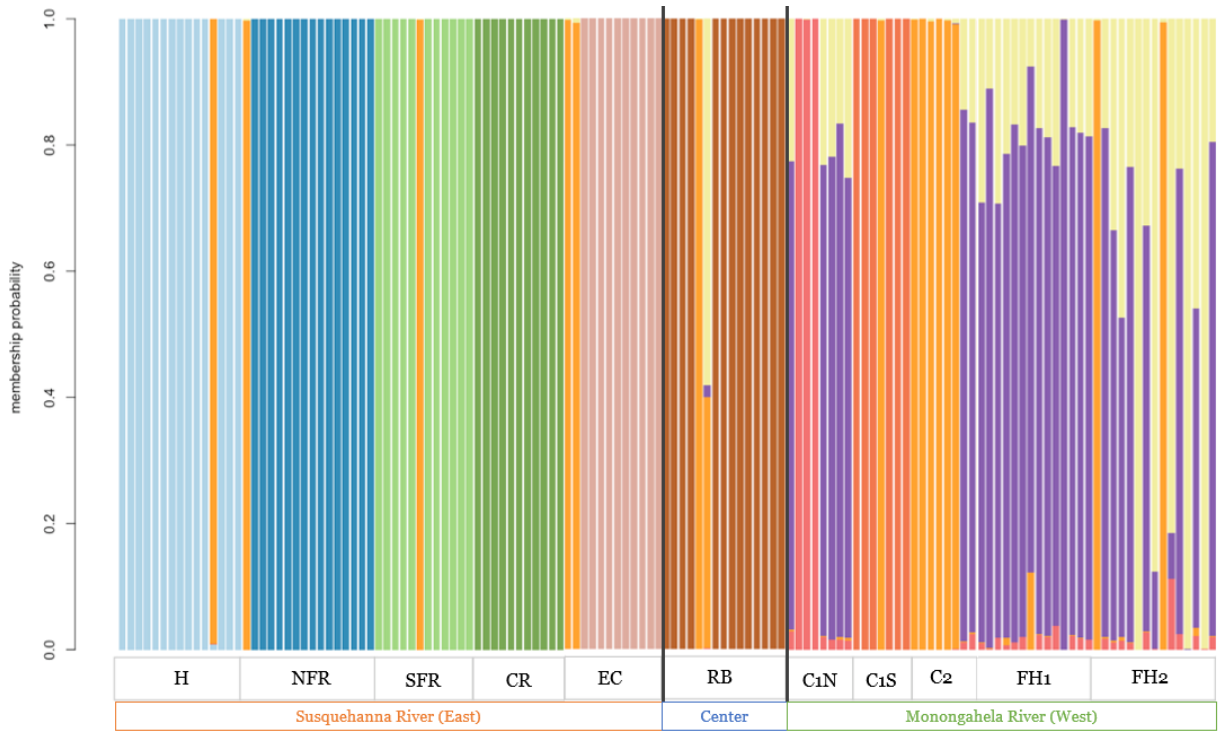

Figure 9. DAPC structure-like plot shows eastern populations are genetically distinct from each other as well as central and western populations. The central population, RB, is different from other populations, while western populations cluster together with some subdivisions within the five populations.

References

1. Abeli, T., Gentili, R., Mondoni, A., Orsenigo, S., Ross, G. (2014) Effects of marginality on plant population performance. *J Biogeogr.* 41:239–249.

2. Antonovics, J., Newman, T.J., Best, B.J. (2002). *Spatially explicit studies on the ecology and genetics of population margins*. Pages 97-117 in Silvertown, J., Antonovics, J., eds. Integrating Ecology and Evolution in a Spatial Context. Blackwell, Oxford.

3. Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A* 160, 268--282. 10.1098/rspa.1937.0109.

4. Bess, J. (2005). Conservation Assessment for Bell's Roadside Skipper (*Amblyscirtes belli Freeman*). USDA Forest Service, Eastern Region.

5. Conservation Voters of Pennsylvania. (2020). Pennsylvania's Water Sources. Retrieved from https://www.conservationpa.org/issues/water/water-sources#:~:text=Pennsylvania's%20water%20comes%20from%20several,Susquehanna%2C%20and%20Ohio%20River%20Basins.&text=Rivers%20and%20streams%20that%20eventually,the%20region's%2017%20million%20residents.

6. Darwin, C.R. (1859). *On the Origin of Species By Means of Natural Selection, or the Preservation of Favoured Races in the Struggle For Life*. London, John Murray. pp 362.

7. Davis, L. (2001). Plant Guide: Inland Sea Oats *Chasmanthium latifolium (Michx.) Yates*. *USDA NRCS*. Retrieved from https://plants.usda.gov/plantguide/pdf/pg_chla5.pdf

8. Dolgin, E.S., Charlesworth, B., Baird, S.E., & Cutter, A.D. (2007). Inbreeding and Outbreeding Depression in Caenorhabditis Nematodes. *Evolution*. 61(6):1339-1352. doi: 10.1111/j.1558-5646.2007.00118.x.

9. Eckert, C.G., Samis, K.E., Lougheed, S.C. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*. 17, 1170–1188. doi: 10.1111/j.1365-294X.2007.03659.x.

10. Fayette County Conservation District. (2016). Watersheds. Retrieved from http://www.fayettecd.org/wp-content/uploads/Pennsylvania-Watershed.pdf.

11. Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biol. Conserv.* 143:1919–1927.

12. Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* 6:925-929.

13. Friedman, J., Barrett, S.C.H. (2009). Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Ann. Bot.* 103(9):1515-1527. doi:10.1093/aob/mcp035

14. Gaston, K.J. (2003). *The Structure and Dynamics of Geographic Ranges*. Oxford Univ. Press. pp. 280.

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

16. Hohenlohe, P.A., Funk, W.C., Rajora, O.P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*. 30:62-82.

17. Holsinger, K., Weir, B. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nat Rev Genet* 10, 639–650.. doi: 10.1038/nrg2611

18. Honnay, O., Jacquemyn, H., Nackaerts, K., Breyne, P., & Van Looy, K. (2010). Patterns of population genetic diversity in riparian and aquatic plant species along rivers. *Journal of Biogeography* 37:1730–1739.

19. Hunter, M.E., Hoban, S.M., Bruford, M.W., Segelbacher, G., & Bernatchez, L. (2018). Next-generation conservation genetics and biodiversity monitoring. *Evolutionary Applications* 11(7):1029-1034. doi: 10.1111/eva.12661

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

21. Jombart, T., Collins, C. (2015). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 11:94. doi: 10.1186/1471-2156-11-94

22. Jump, A.S., Woodward, F.I. (2003). Seed production and population density decline approaching the range-edge of Cirsium species. *New Phytol* 160:349–358.

23. Keck, C., Snyder, S., Gotcher, M., Schroeder, J., & Schnelle, M., Moss, J. (2014). Drought-Tolerant Plant Selections for Oklahoma. *Oklahoma Cooperative Extension Service.* Retrieved from https://osufacts.okstate.edu/.

24. Knaus, B.J. & Grünwald, N.J. (2017). VcfR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1):44–53. doi: 10.1111/1755-0998.12549.

25. Li, Y.-S., Shih, K.-M., Chang, C.-T., Chung, J.-D. and Hwang, S.-Y. (2019). Testing the Effect of Mountain Ranges as a Physical Barrier to Current Gene Flow and Environmentally Dependent Adaptive Divergence in *Cunninghamia konishii* (Cupressaceae). *Front. Genet*. 10:742. doi: 10.3389/fgene.2019.00742

26. Lotts, K., Naberhaus, T., & coordinators. (2020). Butterflies and Moths of North America. Retrieved from http://www.butterfliesandmoths.org/ (Version March 2021).

27. Lovas-Kiss, A., Vincze, O., Kleyheeg, E., Sramkó, G., Laczkó, L., Fekete, R., Molnár V, A., & Green, A.J. (2020). Seed mass, hardness, and phylogeny explain the potential for endozoochory by granivorous waterbirds. *Ecology and evolution* 10(3):1413-1424. doi:10.1002/ece3.5997.

28. Love, H.M., Maggs, C.A., Murray, T.E., Provan, J. (2013). Genetic evidence for predominantly hydrochoric gene flow in the invasive riparian plant *Impatiens glandulifera* (Himalayan balsam). *Annals of Botany* 112(9): 1743-1750. doi: 10.1093/aob/mct227

29. Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., Buckler, E.S., & Costich, D.E. (2013). Switchgrass genomic diversity, ploidy and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* 9(1):e1003215.

30. Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., & Elshire, R. (2012). TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation.

31. MacArthur, R.H. (1972). *Geographical Ecology: Patterns in the Dispersion of Species*. New York, Harper and Row. pp 288.

32. 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. doi: https://doi.org/10.1086/713917.

33. Moore, C., (2020). *Baptisia australis* var. *australis* in Pennsylvania: A Survey of Granivory and Population Genomics in a Threatened Species. Master's Theses. 235. https://digitalcommons.bucknell.edu/masters_theses/235.

34. NatureServe. (2021). NatureServe Explorer: *Chasmanthium latifolium*. NatureServe, Arlington, Virginia. Retreived from https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.130133/Chasmanthium_latifolium.

35. National Climate Data Center (NCDC). (2009). Climate of Pennsylvania. Retreived from https://web.archive.org/web/20090225124128/http:/climate.met.psu.edu/data/ncdc_pa.pdf

36. Neill, R.H. (n.d.) Plant Fact Sheet: Inland Sea Oats *Chasmanthium latifolium (Michx.) Yates*. USDA NRCS. Retrieved from https://plants.sc.egov.usda.gov/factsheet/pdf/fs_chla5.pdf.

37. Osborne, J.L., Free, J.B. (2003). *Pollination*. Page 315-325 in Thomas B. ed. Flowering and Reproduction. Encyclopedia of Applied Plant Sciences, Elsevier. doi: 10.1016/B0-12-227050-9/00237-4.

38. Pennsylvania Department of Conservation and Natural Resources (PA DCNR). (2017). Proposed rulemaking: [17 PA. code ch. 45] conservation of Pennsylvania native wild plants. Retrieved from https://www.pabulletin.com/secure/data/vol47/47-47/1943.html.

39. Pennsylvania Department of Conservation and Natural Resources (PA DCNR). (2021). Landforms. Retreived from https://www.dcnr.pa.gov/Geology/GeologyOfPA/Landforms/Pages/default.aspx.

40. Pennsylvania Natural Heritage Program (PNHP). (2019a). Species Factsheets: *Chasmanthium latifolium*. *Pennsylvania Department of Conservation and Natural Resources*. Retrieved from https://www.naturalheritage.state.pa.us/factsheets/15537.pdf.

41. Pennsylvania Natural Heritage Program (PNHP). (2019b). Rank and Status Definitions. Retrieved from http://www.naturalheritage.state.pa.us/rank.aspx.

42. Pennsylvania State Data Center (PASDC). (2019). Harvested Cropland and Vegetables in the Commonwealth. Retrieved from https://pasdc.hbg.psu.edu/sdc/pasdc_files/researchbriefs/Oct_2019.pdf.

43. Poland, J.A., Brown, P.J., Sorrells, M.E., & Jannink, J.L., (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE* 7(2): e32253. https://doi.org/10.1371/journal.pone.0032253.

44. Pollux, J.A., Jong, M.D., Steegh, A., Ouborg, N.J., Van Groenendael, M., & Klaassen, M. (2006). The effect of seed morphology on the potential dispersal of aquatic macrophytes by the common carp (*Cyprinus carpio*). *Freshwater Biology* 51(11):2063-2071. doi: 10.1111/j.1365-2427.2006.01637.x

45. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

46. Rehm, E.M., Olivas, P. Stroud, J., Feeley, K.J. (2015). Losing your edge: climate change and the conservation value of range-edge populations. *Ecol Evol* 5:4315–4326.

47. Rhoads, A.F., Block, T.A. (2007). *The Plants of Pennsylvania: An Illustrated Manual*, 2nd Edition. University of Pennsylvania Press.

48. Rognli O.A., 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.

49. Sexton, J.P., McIntyre, P.J., Angert, A.L., & Rice, K.J. (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*. 40:415–36. doi: 10.1146/annurev.ecolsys.110308.120317.

50. Sexton, J.P., Strauss, S.Y., & Rice, K.J. (2011). Gene flow increases fitness at the warm edge of a species' range. *Proceedings of the Natural Academy of Science*. 108:11704-11709.

51. Unamba, C.I.N., Nag, A., & Sharma, R.K. (2015). Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Frontiers in Plant Science* 6:1074. doi: 10.3389/fpls.2015.01074

52. Widmann, Richard H. (2016). Forest of Pennsylvania, 2015. Resource Update FS-92. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 4 p.

53. Wright S. (1978). *Evolution and the genetics of populations, variability within and among natural populations*, vol IV. University of Chicago Press.

54. Yan, J., Zhu, M., Liu, W., Xu, Q., Zhu, C., Li, J., & Sang, T. (2016). Genetic variation and bidirectional gene flow in the riparian plant Miscanthus lutarioriparius, across its endemic range: implications for adaptive potential. *GCB Bioenergy* 8, 764–776. doi: 10.1111/gcbb.12278

55. Yates, H.O. (1966). Revision of Grasses Traditionally Referred to *Uniola*, II. *Chasmanthium*. *The Southwestern Naturalist* 11(4):415-455