



**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2017

NEUTRAL AND ADAPTIVE GENETIC VARIATION IN NORTH AMERICAN HARDWOOD TREE SPECIES

Sudhir Khodwekar

Michigan Technological University, sdkhodwe@mtu.edu

Copyright 2017 Sudhir Khodwekar

Recommended Citation

Khodwekar, Sudhir, "NEUTRAL AND ADAPTIVE GENETIC VARIATION IN NORTH AMERICAN HARDWOOD TREE SPECIES", Open Access Dissertation, Michigan Technological University, 2017.

<https://digitalcommons.mtu.edu/etdr/440>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etdr>



Part of the [Biodiversity Commons](#), [Biotechnology Commons](#), [Evolution Commons](#), [Molecular Genetics Commons](#), [Other Genetics and Genomics Commons](#), [Plant Breeding and Genetics Commons](#), and the [Population Biology Commons](#)

**NEUTRAL AND ADAPTIVE GENETIC VARIATION IN NORTH AMERICAN HARDWOOD TREE
SPECIES**

By

Sudhir D. Khodwekar

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Forest Molecular Genetics and Biotechnology

MICHIGAN TECHNOLOGICAL UNIVERSITY

2017

©2017 Sudhir D. Khodwekar

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Forest Molecular Genetics and Biotechnology.

School of Forest Resources and Environmental Science

Dissertation Advisor: *Oliver Gailing*

Committee Member: *Andrew Storer*

Committee Member: *Victor Busov*

Committee Member: *Thomas Werner*

School Dean: *Terry Sharik*

Dedication

I dedicate this dissertation to my parents, Bhimabai and Dadarao Khodwekar, sister Sukumari and my two brothers Sushil, Vishnu. I could not ask for better, thank you!

Table of Contents

Preface.....	9
Acknowledgements.....	10
Abstract.....	12
Chapter 1: Introduction	14
Background	14
Study species (Important hardwood species in North America).....	17
Biotic and abiotic threats to important hardwood species in North America	19
Hybridization in oaks	21
Development of genomic resources	22
Hardwood genomics project.....	24
Objectives.....	25
References	26
Chapter 2: Nuclear microsatellite markers for population genetic studies in sugar maple (<i>Acer saccharum</i> Marsh.)	30
Abstract	31
Introduction	32
Material and Methods	35
Species description	35
Marker development	35
PCR amplification.....	36

Data analysis	37
Results	40
Discussion	43
Conclusion.....	47
Acknowledgements	47
References	48
Table 2.1.....	51
Table 2.2.....	52
Table 2.3.....	53
Table 2.4.....	53
Figure 2.1	54
Figure 2.2	55
Figure 2.3	56
Supplementary figure 2.1	57
Appendix 1: Open-Access License	58
Chapter 3: Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations	59
Premise of the study.....	60
Methods.....	60
Key results.....	60
Conclusions	60
Key words	61
Introduction.....	62

Material and Methods	67
Plant Material	67
DNA isolation	68
Genetic markers.....	69
PCR amplification.....	69
Genetic diversity analyses	70
Genetic assignment and introgression of outlier alleles.....	70
Outlier screening	72
Gene flow	73
Results	74
Genetic diversity	74
Genetic structure assignment	74
Outlier screening	75
Introgression of outlier alleles.....	76
Ancestry and introgression of outlier alleles is related to soil properties	76
Gene flow analysis	78
Discussion	79
Evidence for symmetric interspecific gene flow.....	79
Species are stratified along soil water/nutrient gradient	80
Different levels of introgression of outlier alleles in parapatric and sympatric stands suggest different selection regimes	81
The frequency of outlier allele introgression is related to soil types	84
Conclusions	85

Acknowledgements	86
Literature cited	87
Table 3.1.....	91
Table 3.2.....	92
Table 3.3.....	93
Table 3.4.....	94
Figure 3.1	95
Figure 3.2	96
Figure 3.3	97
Figure 3.4	98
Figure 3.5	99
Supplementary table 3.1	100
Supplementary table 3.2.....	101
Supplementary table 3.3.....	102
Supplementary table 3.4.....	103
Supplementary table 3.5.....	104
Supplementary table 3.6.....	105
Supplementary table 3.7.....	106
Supplementary table 3.8.....	107
Supplementary table 3.9.....	108
Supplementary figure 3.1	109
Supplementary figure 3.2	110
Supplementary figure 3.3.....	111

Supplementary figure 3.4	112
Supplementary figure 3.5	113
Supplementary figure 3.6	114
Supplementary figure 3.7	115
Supplementary figure 3.8	116
Supplementary figure 3.9	117
Supplementary figure 3.10	118
Supplementary figure 3.11	119
Supplementary figure 3.12	120
Supplementary figure 3.13	121

Preface

This dissertation includes three chapters, of which first chapter is introduction and concept development for the thesis. The second chapter of my PhD dissertation, “Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.),” has been published in the peer-reviewed international journal ‘Annals of Forest Research.’ I am the first author on this paper; I performed most of the lab work and data analysis and wrote the paper. My co-author, Oliver Gailing, collected samples and helped in writing and editing the paper. Other co-authors also gave valuable input in editing the paper.

The third chapter, “Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations,” has been submitted to the peer-reviewed international journal ‘American Journal of Botany’. As first author, I performed all of the lab work and data analyses. My co-author, Oliver Gailing, contributed by collecting samples as well as writing and editing the paper.

As my PhD advisor, Oliver Gailing helped conceive and design the studies for this dissertation. The second chapter of my work forms part of the National Science Foundation (NSF)-funded project, *Development of genomic resources for North American hardwoods*.

Acknowledgements

This PhD dissertation is the product of four years invested in learning with valuable contributions from my peers, friends and family members. First, I would like to thank my advisor, Dr. Oliver Gailing, for accepting me as his PhD student and always encouraging me with his intellectual support throughout the four years. I am grateful to the Government of India who funded me under a “National Overseas Scholarship” for my PhD degree. Thanks also to my committee members, Dr. Andrew Storer, Dr. Victor Busov and Dr. Thomas Werner, who challenged me to think beyond my capabilities and imparted knowledge when needed.

Special thanks to Dr. Andrew Storer, who was easily approachable and ready to answer graduate school-related queries, and to Dr. Victor Busov and his post-doctoral candidate, Dr. Yordan Yordanov, for their valuable input in my PCR- and primer development-related questions.

I am grateful to Sandra Owusu and Jennifer Lind-Riehl for teaching me all of the lab procedures and software tools required for this work, I am also thankful for their wonderful friendship, which kept me going in this lonely town. Jonathan Riehl was good company in addition to Sandra and Jennifer for travelling and good food hunting. I am thankful for Mehmet Acet, Monica Harmon for their friendship and humor. Thanks to Priyanka Kadav for her great company and good pizza times, thanks to Munkaila Musah, and to Priyanka for their great company outside the lab. I would like to thank Mustafa Kagan and Sara Alian for ArcGIS help, and special thanks to Robert Richards for

valuable contributions in creating and modifying maps. Thanks to our amazing lab manager, Dr. Jennifer Sanders, for valuable input in all kinds of writings. Thanks to all other masters and undergraduate students in my lab who helped me in doing lab work. I am grateful to Bhaskar, his wife Priyanka, and their sweet little daughter who are a family to me in Houghton.

Apart from the Government of India, I would like to thank the Life Science and Technology Institute (LSTI), the School of Forest Resources and Environmental Science, and the Graduate School at Michigan Tech for providing me with stipend and tuition support in the form of finishing fellowships. I would also like to thank the Ecosystem Science Center (ESC) and Graduate School Government (GSG) for their numerous research and travel grant awards.

I was lucky to be associated with the Hardwood Genomics Project team, part of the larger NSF Plant Genome Research Program (NSF PGRP #1025974).

Thanks to my course instructors and the other great people in the School of Forestry such as David Flashpohler, Terry Sharik, Ruth Ojala, Marjorie Lindley, Andrea Longhini, Mary Jurgensen, and Laura Mattila. Thanks to all my friends in the school Alex, Igus, Akwasi, Will, Ram and my roomie Nikhil. Thanks for great support of my friends from India who were there always in my hard time Avinash, Rohit, Prashant, Anil and Yatin.

Most of all, this journey would not have been possible without the support of my amazing family: my mom and dad, my two brothers Sushil and Vishnu, my sister Sukumari, and my nephew and nieces. I missed you all.

Abstract

‘Northern Hardwoods,’ are a characteristic composition of multiple tree species and a part of the rich and diverse northeastern forests. Hardwoods e.g. oaks (*Quercus*: Fagaceae and Sugar maple), serve as foundation species and offer ecological and economical gains to wildlife and humans. Most of the forest trees go through many biotic and abiotic stresses, for example climate change in their long life span. To overcome these threats and to adapt to changing conditions, tree species need to maintain variation especially in adaptive genes. Therefore, it is necessary to analyze genetic diversity within forest tree populations and take appropriate forest management practice decisions. Genetic tools, such as microsatellite markers developed using next generation DNA sequencing technologies, can be used to address these challenges. Here we report the development of nuclear microsatellite markers in one of the most important hardwood tree species, sugar maple (*Acer saccharum* Marsh.) (Chapter II). We also use genic microsatellite markers to study the introgression patterns of adaptive alleles in two interfertile red oak species, *Quercus rubra* and *Quercus ellipsoidalis* (Chapter III).

Using the paired-end low coverage Illumina sequencing technology, we developed a set of seven nuclear microsatellite markers (nSSRs) in sugar maple (*Acer saccharum* Marsh.). Initially, we screened 96 markers in a panel of six unrelated individuals, out of which seven markers produced polymorphic PCR products. These markers were utilized to analyze genetic variation and gene flow in one sugar maple population in an urban setting consisting of 48 individuals. Additionally, 96 seeds from one open pollinated adult tree were used for the gene flow analysis. In addition to these seven markers, we

also tested six previously published microsatellites. Paternity analyses displayed effective dispersal of pollen in the sugar maple population with 76% of gene flow from outside the stand. There was no fine-scale genetic structure observed in this population, which also suggested effective dispersal of both seeds and pollen. To check the transferability of these markers to other *Acer* species, four individuals from each species of *Acer rubrum* L., *Acer saccharinum* L., *Acer platanoides* L., and *Acer ginnala* Maxim. were tested at these markers. All markers amplified PCR products in these four species. However, only two markers, AS47 and SM37, were locus-specific and polymorphic in one species (*A. ginnala*). These markers can serve as an important tool to study genetic variation patterns in sugar maple populations in the face of climate change.

Quercus rubra L. and *Q. ellipsoidalis* E. J. Hill are two interfertile red oak species with different adaptations to drought. I have studied the introgression of adaptive microsatellite alleles at a CONSTANS-Like gene (COL) as a response to micro-environment between these two hybridizing species in two sympatric populations. My results indicated that divergent selection in contrasting environments resulted in high interspecific differentiation at COL. However, interspecific differentiation was lower in sympatric stands, where both species occur in similar environments. The introgression of allele 138 from *Q. ellipsoidalis* into *Q. rubra* was higher compared to introgression of allele 141 from *Q. rubra* into *Q. ellipsoidalis* in contrast to earlier studies in parapatric populations. My results suggest that the introgression of adaptive genes between two red oak species is strongly affected by environmental selection.

Chapter 1:

Introduction

Background

Forests cover a substantial amount of Earth's land area and foster a rich and incredible biodiversity. The northeastern forest region is the most forested part of the United States, it is a principal biome, encompassing mixed and deciduous forests, which offer plentiful diversity (USDA 2005). This forest region has remnants of the 300 million hectares of forestlands that existed well before the earliest European settlements (USDA 2005).

Today hardwood and coniferous tree species dominate this area (Rustad et al. 2012). The northern hardwoods comprise a group of forest types that vary considerably in species composition, and each species differs in its characteristics according to its geographic distribution. The following species constitute a major proportion of these forests, sugar maple (*Acer saccharum* Marsh.), yellow birch (*Betula alleghaniensis* Britton), American beech (*Fagus grandifolia* Ehrh.), American basswood (*Tilia americana* L.) and hemlock (*Tsuga canadensis* (L.) Carr.). Red oaks (*Quercus* species), balsam fir (*Abies balsamia* (L.) Mill.), black ash (*Fraxinus nigra* Marsh.), and eastern white pine (*Pinus strobus* L.) are also present (Tubbs et al. 1983) and add to the species diversity.

Several climate envelope modeling studies have predicted that , as a response to climate change, primarily temperature, increase of atmospheric CO₂ concentration and changes in precipitation pattern have potential to alter the distribution ranges of forest tree species and affect species interactions (Iverson et al. 2008). Recent studies have shown that the

climate of the northeastern United States and Canada has experienced warmer and wetter periods in the past several years as compared to the previous 100 years (Rustad et al. 2012). By influencing the recruitment of particular tree species at different phases of regeneration such as seed germination, climate controls the species distribution by means of temperature and water supply (Walck et al. 2011). Hence, predicted changes in climate along with other alterations in the environment may affect the ecological dynamics and conservation of the particular species (Fitch et al. 2007, Walck et al. 2011). With a history of glacial advances and retreats, countless fires, storms, droughts and floods, the North American forests have developed great biodiversity with strong ecological associations (USDA 2005). Projections based on computer models have shown a decline in some species (e.g. spruce-fir) and expansion of the habitats for other species (e.g. oaks) (Rustad et al. 2012).

It is necessary to acquire basic knowledge of forest species populations to manage forest resources and conserve genetic diversity to adapt to future environmental changes. Forest trees species are among the most genetically diverse organisms on the planet (Schaberg et al. 2008) and are able to adapt to changing environments over time (Neale and Ingvarsson 2008). This diverse nature of forest tree species provides an excellent opportunity to understand the variation occurring in plants because of phenotypic and genotypic interplay (Neale and Ingvarsson 2008). However, it is predicted that many forest trees will not be able to adapt to rapid environmental changes because of limited genetic variation in adaptive genes (Hampe and Petit 2005, Levin 2012, Aitken and Bemmels 2016). Mutation are rare events and are unlikely to generate new adaptive

variation to cope with environmental change in the face of climate change (Barrett and Schluter 2008, Orr and Unckless 2014, Hamilton and Miller 2016). However, introgression of adaptive genes between species in hybridizing species may increase adaptive variation within species at a much higher rate than mutations alone (Aitken et al. 2008).

As research efforts in forest ecology and management have widened, there has been an increase in the number of resources (e.g., DNA markers) available for population genetic studies. The use of genetic resources for conservation is crucial for the long-term maintenance of forest tree populations. “Comparative genomic analyses have been performed in plants and animal species to fully understand the evolutionary and basis of gene function and developmental processes” (Neale and Ingvarsson 2008). Genomic databases have created new opportunities to explore forest trees at the genetic level to understand forest health, forest management and molecular evolution. Here, we generate new genetic markers resources for sugar maple and analyze the effect of interspecific hybridization in oaks as an adaptive allele introgression.

Study species (Important hardwood tree species in North America)

Sugar maple (*Acer saccharum* Marsh.) and red oaks (e.g. *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill) are important hardwood tree species in North America from both ecological (e.g., they provide food and shelter to animals) and economical (e.g., timber and timber products) perspectives (USDA 2005).

Sugar maple, which is also known as hard maple or rock maple is one of the most important hardwood species in northeastern forests of North America (Godman et al. 1990). It is well recognized as a keystone forest species and is very important to the lumber and syrup industries (Bal et al. 2013). It grows on approximately 12.5 million hectares, forms uneven-aged stands, and offers a large saw timber volume in the United States, with the highest volume present in Michigan, New York, Maine, Wisconsin and Pennsylvania (Godman et al. 1990). Sugar maple is widespread and one of the most abundant tree species in northeastern North America (Graignic et al. 2014). It is continuously distributed in the northern ranges up to the boreal mixed wood and temperate deciduous forests (Graignic et al. 2014).

Oaks also constitute a large proportion of the present northern hardwood forest species. Despite of their current widespread distribution in the temperate zone forests, in many forests oak species are failing to regenerate (Abrams 1992, Lorimer 1993, Aldrich et al. 2005). Since the last century, oaks have been widely used and studied as a model genus for understanding evolutionary processes and speciation in nature (Ayala and Fitch 1997, Neophytou et al. 2010). *Q. rubra* (northern red oak), also known as common red oak, is a

native dominant oak species, which shows a wide geographic distribution in eastern North America. Among red oaks, it is the most important timber (wood) species. Native Americans utilized northern red oak as medicine for numerous ailments. Northern red oak grows on a variety of soils and under divergent topographic conditions although it usually prefers moist, well-drained soils and uplands (Sander 1990, Nesom 2003). Northern red oak tends to form pure stands (Sander 1990).

Quercus ellipsoidalis (northern pin oak or Hill's oak) is an upland species and usually grows on dry acidic and sandy sites and maintains successful growth even on nutrient-poor soils and during times of drought. This species often prefers sandy regions and occasionally grows on moderately mesic slopes or uplands in varying mixtures with other oak species (e.g. northern red oak) (Nesom 2003). Northern pin oak is distributed primarily in the Great Lakes region and has a scattered distribution on drier sites in the northern range of eastern North America. It often shows a shrubby growth at the edge of its range (Stein et al. 2003) and tends to hybridize with northern red oak in the contact zones (Abrams 1990, Nesom 2003, Hipp 2010).

Biotic and abiotic threats to important hardwood species in North America

Forest health is influenced by many biotic and abiotic factors such as disease, fragmentation, anthropological activities, and environment-related stresses. Forests can also be threatened by interactions of biotic and abiotic stressors (Böttcher and Breisgau 2007, Moore and Allard 2011). Disturbances are intrinsic elements of forest ecosystems. Some of the key elements, which exert pressure on the ecosystem and cause disturbances are fire, drought, introduction of new invasive species, insect infestations and pathogen outbreaks, natural calamities such as hurricanes, windstorms, and ice storms (Fischer et al. 2013). Humans also exert a strong influence on ecosystems (Fischer et al. 2013). The major biotic factors affecting sugar maple are, a variety of rots, cankers, wilts, defoliators, borers, sucking insects, scale insects, bud miners, and diseases (Godman et al. 1990). The most important abiotic agents affecting sugar maple are deficiency or excess of soil moisture, atmospheric deposition, highway de-icing salts, and extreme weather events, including late spring frosts and midwinter freeze-thaw cycles (Horsley et al. 2002, Bailey et al. 2004).

Recent surveys in the northeastern United States and eastern Canada have shown that sugar maple is generally healthy throughout its range; however, over the past four decades, sporadic decline has occurred in Ontario and Quebec, Wisconsin, Vermont, New York and Pennsylvania (Horsley et al. 2002). In the boreal forests of Siberia, Canada and Alaska, sugar maple shows a modeled prediction of death, dieback, outbreaks of insect infestation, and wild fires (Horsley et al. 2002, Iverson et al. 2008, Gaignic et al. 2014). Decline in sugar maple is characterized by a slow loss of crown

vigor, dieback of fine twigs, and reduced radial increment over a period of years, frequently ending in death. High levels of sugar maple dieback have also been reported in upper Michigan because of the sapstreak disease caused by the fungus *Ceratocystis virescens* (Davidson) Moreau (Bal et al. 2013).

In the case of oaks, the overall scenario is not much different, many cases of bacterial leaf scorch (BLS) have been reported to affect species of oaks throughout the eastern United States (Gould and Lashomb 2005). Oak-dominated forests in northwestern Arkansas have recently experienced oak mortality due to an unprecedented outbreak of the native red oak borer (*Enaphalodes rufulus* (Haldeman)) (Haavik et al. 2008). Imported gypsy moth (*Lymantria dispar dispar*) has caused a tremendous destruction by attacking northern red oak; it repeatedly defoliates oak trees. Gypsy moth infection has killed oaks, including northern red oak, which covers a broad area in the northeastern United States (Sander 1990). Other diseases or disease-causing agents that affect oaks include cankers caused by *Strumella* and *Nectria* species, which damage the bole of northern red oak. Foliage diseases that attack northern red oak, but rarely results in serious damage, include: anthracnose (*Gnomonia quercina*), leaf blister (*Taphrina* spp.), powdery mildews (*Phyllactinia corylea* and *Microsphaera alni*), and eastern gall rust (*Cronartium quercuum*) (Sander 1990). Climate change is supposed to affect the migration northward in the eastern United States, leading to competition in forest trees, especially oaks, which are drought-tolerant and tend to hybridize in overlapping zones (Woodall et al. 2009, Lind and Gailing 2013).

Hybridization in oaks

Hybridization between closely related species can potentially lead to the establishment of hybrid zones in the contact regions while the identity of individual species is maintained (Genovart 2008). Extensive studies have been performed on some of the European white oak species (*Quercus petraea*, *Q. robur*) to provide evidence for hybridization and introgression using both the morphological and molecular markers (Petit et al. 2003, Curtu et al. 2007). Recent studies also provide evidence for the maintenance of species identity despite interspecific gene flow, and inclination for hybridization and introgression (Curtu et al. 2007, Sullivan et al. 2013). There is a strong evidence that divergent selection likely maintains adaptive species differences despite interspecific gene flow (Lind-Riehl et al. 2014). However, in intermediate environments hybridization may result in the transfer of adaptations between species. In my study, *Q. rubra* and *Q. ellipsoidalis*, two interfertile North American red oak species were analyzed. These two species occur on different soil types and show different adaptations to drought such as tissue elasticity (Abrams 1992). They are the only red oak species in their northern distribution range (Little 1971) and provide a model to study interspecific gene flow and transfer of adaptive traits and alleles between the species.

Development of genomic resources

Due to the growing influence of biotic and abiotic stressors on forests and major forest tree species, there is an immediate need to develop tools that help in forest management and broaden strategies to improve the health of forests. To understand the interplay between genotypes, phenotypes and surrounding environment genetic and genomic resources are valuable tools (Staton et al. 2015). The development and characterization of highly polymorphic microsatellite markers (both nuclear and genic) can greatly enhance ecological genetic studies in forest trees, including maple species (Pandey et al. 2004, Pandey et al. 2012, Khodwekar et al. 2015). Development of genomic resources for northern hardwood tree species is advantageous in many aspects. For example, markers can increase the efficiency and speed of genetic improvement programs and help build new strategies for the majority of North American hardwood tree species, which lack tree-breeding programs (Staton et al. 2015). In addition, these markers are useful in studying gene flow, mating system, population structure and for creating genetic linkage maps.

Microsatellites, or simple sequence repeats (SSRs), are generally located in non-coding regions of the genome. They are tandem repeats of simple sequences, 2-9 base pairs (bp) in length, and are highly polymorphic for repeat numbers (Conner and Hartl 2004).

Microsatellites are of co-dominant nature, multi-allelic, easily scored, and abundant throughout the genome (Selkoe and Toonen 2006). Because of their highly polymorphic nature and ability to discriminate between closely related individuals, they are widely used in population genetic analyses (Brondani et al. 2002, Selkoe and Toonen 2006,

White et al. 2009). The application of recent molecular techniques in the analysis of the genetic structure of extinction-threatened tree species has made genetic analysis more popular for conservation management (Hedrick 2001). Development of microsatellites is an expensive venture because microsatellites are largely species-specific, making their development difficult in species with little or no economic importance (Ellis and Burke 2007). Recent techniques for developing microsatellite markers using next generation sequencing, such as Illumina sequencing, have the potential to reduce the cost and time for SSR marker development through multiplexing of species and the use of high-throughput technologies (Jennings et al. 2011).

Expressed sequence tag libraries are another source of microsatellite markers. Using the BLASTX algorithm, these EST libraries are screened to develop primers for EST-SSRs (Expressed sequence tags simple sequence repeats). EST-SSR markers occur in gene-rich regions within the genome and are thus more conserved among species. They have an added advantage over genomic SSRs because they can be obtained quickly by electronic sorting, and they are present in expressed regions of the genome (Varshney et al. 2005). However, EST databases are available only for a small number of species. Together, both genic and genomic microsatellite markers can be used to elucidate the patterns of genetic variation and gene flow, which are of fundamental importance to both evolutionary ecology and conservation biology.

Hardwood Genomics Project

“The Comparative Genomics of Environmental Stress Responses in North American Hardwoods Project”, commonly referred to as the “Hardwood Genomics Project”, was funded by the National Science Foundation Plant Genome Research Program (NSF-PGRP) as a genetics and genomics resource development grant” (Carlson 2015). One of the main objectives of this Hardwood Genomics Project (<http://www.hardwoodgenomics.org/>) was to develop a suite of microsatellite marker resources for ten ecologically and economically important native hardwood tree species. Prior to this project, there were few or no sequences available for these ten hardwood species, which represent a phylogenetically wide range of families. Many of the ten hardwood species have high economic value, importance in forest ecosystems and/or are under pressure from invasive pests or pathogens (Staton et al. 2015). One project aim was to use low coverage whole genome sequencing with paired-end, multiplexed Illumina reads to economically profile the genomic content of these ten species. The advantage of low coverage sequencing technology is that it can be used to mine the genome for molecular markers, which are common throughout plant genomes (Staton et al. 2015). In addition to the development of genomic markers, stress-induced EST libraries were developed by applying artificial biotic and abiotic stress treatments to seedlings of each species. These EST-SSR resources were also used to develop gene-based markers such as EST-SSR and single nucleotide polymorphism (SNP) markers. This project also aimed to create framework-genetic linkage maps (e.g. densely populated maps for northern red

oak) based on microsatellite and SNP markers for each tree species (e.g. Northern red oak) (Sullivan et al. 2013).

Objectives

The two main objectives of my PhD dissertation are:

1. Develop genomic microsatellite markers for sugar maple using paired-end low coverage illumina sequencing.
2. Analyses of gene flow and adaptive gene introgression between two oak species with different adaptation to drought.

References

- Abrams, M. D. 1990. Adaptations and responses to drought in *Quercus* species of North America. *Tree Physiol* **7**:227-238.
- Abrams, M. D. 1992. Fire and the development of oak forests. *Bioscience* **42**:346-353.
- Aitken, S. N., and J. B. Bemmels. 2016. Time to get moving: assisted gene flow of forest trees. *Evol Appl* **9**:271-290.
- Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang, and S. Curtis-McLane. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol Appl* **1**:95-111.
- Aldrich, P. R., J. C. Glaubitz, G. R. Parker, O. E. Rhodes, Jr., and C. H. Michler. 2005. Genetic structure inside a declining red oak community in old-growth forest. *J Hered* **96**:627-634.
- Ayala, F. J., and W. M. Fitch. 1997. Genetics and the origin of species: An introduction. *Proceedings of the National Academy of Sciences of the United States of America* **94**:7691-7697.
- Bailey, S. W., S. B. Horsley, R. P. Long, and R. A. Hallett. 2004. Influence of Edaphic Factors on Sugar Maple Nutrition and Health on the Allegheny Plateau. *Soil Sci. Soc. Am. J.* **68**:243-252.
- Bal, T. L., D. L. Richter, A. J. Storer, and M. F. Jurgensen. 2013. The relationship of the sapstreak fungus, *Ceratocystis virescens*, to sugar Maple dieback and decay in northern Michigan. *American Journal of Plant Sciences* **04**:436-443.
- Barrett, R. D., and D. Schluter. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol* **23**:38-44.
- Böttcher, H., and F. Breisgau. 2007. Forest Management for climate change mitigation. Albert-Ludwigs-Universität, Freiburg im Breisgau.
- Brondani, R. P., C. Brondani, and D. Grattapaglia. 2002. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Mol Genet Genomics* **267**:338-347.
- Carlson, J. E. 2015. "TRPGR: Comparative genomics of environmental stress responses in North American hardwoods". Final Report NSF Project.
- Conner, J. K., and D. L. Hartl. 2004. DNA Markers In: A Primer of Ecological Genetics. Sinauer Associates.inc, Sunderland, MA, USA.
- Ellis, J. R., and J. M. Burke. 2007. EST-SSRs as a resource for population genetic analyses. *Heredity (Edinb)* **99**:125-132.
- Fischer, A., P. Marshall, and A. Camp. 2013. Disturbances in deciduous temperate forest ecosystems of the northern hemisphere: their effects on both recent and future forest development. *Biodivers. Conserv.* **22**:1863-1893.
- Fitch, E. A., J. I. Walck, and S. N. Hidayati. 2007. Agrosystem management for rare species of *Paysonia* (Brassicaceae): integrating their seed ecology and life cycle with cropping regimens in a changing climate. *Am. J. Bot.* **94**:102-110.
- Godman, R. M., H. W. Yawney, and C. H. Tubbs. 1990. *Acer saccharum* Marsh. Sugar maple. In: Burns R. M., Honkala B. H. (eds.). *Silvics of North America, Agricultural Handbook* **2**:78-91.

- Gould, A. B., and J. H. Lashomb. 2005. Bacterial leaf scorch of shade trees. Online (www.apsnet.org). APSnet Feature, The American Phytopathological Society, St. Paul, MN. 18p:1611-1618.
- Graignic, N., F. Tremblay, Y. Bergeron, and J. Williams. 2014. Geographical variation in reproductive capacity of sugar maple (*Acer saccharum* Marshall) northern peripheral populations. *J. Biogeogr.* **41**:145-157.
- Haavik, L. J., F. M. Stephen, M. K. Fierke, V. B. Salisbury, S. W. Leavitt, and S. A. Billings. 2008. Dendrochronological parameters of northern red oak (*Quercus rubra* L. (Fagaceae)) infested with red oak borer (*Enaphalodes rufulus* (Haldeman) (Coleoptera: *Cerambycidae*)). *For. Ecol. Manage.* **255**:1501-1509.
- Hamilton, J. A., and J. M. Miller. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv Biol* **30**:33-41.
- Hampe, A., and R. J. Petit. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecology letters* **8**:461-467.
- Hedrick, P. W. 2001. Conservation genetics: Where are we now? *Trends Ecol Evol* **16**:620-630.
- Hipp, A. L. 2010. Hill's oak: The taxonomy and dynamics of a western Great Lakes endemic. *Arnoldia* **67**:1-13.
- Horsley, S. B., R. P. Long, S. W. Bailey, R. A. Hallett, and P. M. Wagro. 2002. Health of eastern North American sugar maple forests and factors affecting decline. *Northern J. Appl. Forestry* **19**:34-44.
- Iverson, L. R., A. M. Prasad, S. N. Matthews, and M. Peters. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management* **254**:390-406.
- Jennings, T. N., B. J. Knaus, T. D. Mullins, S. M. Haig, and R. C. Cronn. 2011. Multiplexed microsatellite recovery using massively parallel sequencing. *Mol. Ecol. Resour.* **11**:1060-1067.
- Khodwekar, S., M. Staton, M. V. Coggeshall, J. E. Carlson, and O. Gailing. 2015. Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). *Annals of Forest Research* **58**:193-204.
- Levin, D. A. 2012. Mating system shifts on the trailing edge. *Ann Bot* **109**:613-620.
- Lind-Riehl, J. F., A. R. Sullivan, and O. Gailing. 2014. Evidence for selection on a CONSTANS-like gene between two red oak species. *Ann Bot* **113**:967-975.
- Lind, J. F., and O. Gailing. 2013. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* **9**:707-722.
- Little, E. L. J. 1971. Atlas of United States trees, volume 1, conifers and important hardwoods. U.S. Department of Agriculture Miscellaneous Publication 1146.
- Lorimer, C. G. 1993. Causes of the oak regeneration problem. USDA, Forest Service.
- Moore, B. A., and G. Allard. 2011. Abiotic disturbances and their influence on forest health: A review. 1-51.
- Neale, D. B., and P. K. Ingvarsson. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Curr. Opin. Plant Biol.* **11**:149-155.

- Neophytou, C., F. A. Aravanopoulos, S. Fink, and A. Dounavi. 2010. Detecting interspecific and geographic differentiation patterns in two interfertile oak species (*Quercus petraea* (Matt.) Liebl. and *Q. robur* L.) using small sets of microsatellite markers. *Forest Ecology and Management* **259**:2026-2035.
- Nesom, G. 2003. Northern pin oak In: USDA NRCS National Plant Data Center & Bioata of North America Program.1-3.
- Orr, H. A., and R. L. Unckless. 2014. The population genetics of evolutionary rescue. *PLoS Genet* **10**:e1004551.
- Pandey, M., O. Gailing, D. Fischer, H. H. Hattemer, and R. Finkeldey. 2004. Characterization of microsatellite markers in sycamore (*Acer pseudoplatanus* L.). *Molecular Ecology Notes* **4**:253-255.
- Pandey, M., O. Gailing, H. H. Hattemer, and R. Finkeldey. 2012. Fine-scale spatial genetic structure of sycamore maple (*Acer pseudoplatanus* L.). *European Journal of Forest Research* **131**:739-746.
- Rustad, L., J. Campbell, J. S. Dukes, T. Huntington, K. T. Lambert, J. Mohan, and N. Redenhouse. 2012. Changing climate, changing forests: The impacts of climate change on forests of the Northeastern United States and Eastern Canada. USDA, Forest Service.
- Sander, I., L.,. 1990. Hardwoods In: Burns Russell M. *Silvics of North America, Agriculture Handbook*. **2**:727-733.
- Schaberg, P. G., D. H. DeHayes, G. J. Hawley, and S. E. Nijensohn. 2008. Anthropogenic alterations of genetic diversity within tree populations: Implications for forest ecosystem resilience. *Forest Ecology and Management* **256**:855-862.
- Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters* **9**:615-629.
- Staton, M., T. Best, S. Khodwekar, S. Owusu, T. Xu, Y. Xu, T. Jennings, R. Cronn, A. K. Arumuganathan, M. Coggeshall, O. Gailing, H. Liang, J. Romero-Severson, S. Schlarbaum, and J. E. Carlson. 2015. Preliminary genomic characterization of ten hardwood tree species from multiplexed low coverage whole genome sequencing. *PLoS One* **10**:e0145031.
- Stein, J., D. Binion, and R. Acciavatti. 2003. Field guide to native oak species of eastern North America. USDA, Forest Service.
- Sullivan, A. R., J. F. Lind, T. S. McCleary, J. Romero-Severson, and O. Gailing. 2013. Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. *Plant Molecular Biology Reporter* **31**:231-239.
- Tubbs, C. H., R. D. Jacobs, and D. Cutler. 1983. Northern Hardwoods In: Burns Russell M. . *Silvicultural systems for the major forests types of the United States, Agriculture Handbook*:121-127. USDA.
- USDA, F. S. 2005. A snapshot of the northeastern forests. USDA, Forest Service.
- Varshney, R. K., A. Graner, and M. E. Sorrells. 2005. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* **23**:48-55.

- Walck, J. L., S. N. Hidayati, K. W. Dixon, K. E. N. Thompson, and P. Poschlod. 2011. Climate change and plant regeneration from seed. *Global Change Biology* **17**:2145-2161.
- White, T. L., T. W. Adams, and D. B. Neale. 2009. Genetic markers-Morphological, biochemical and molecular markers In: *Forest Genetics*. CABI, Oxfordshire, UK.
- Woodall, C. W., C. M. Oswalt, J. A. Westfall, C. H. Perry, M. D. Nelson, and A. O. Finley. 2009. An indicator of tree migration in forests of the eastern United States. *Forest Ecology and Management* **257**:1434-1444.

Chapter 2:

Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.)¹

Sudhir Khodwekar¹, Margaret Staton², Mark V. Coggeshall³, John E. Carlson⁴, Oliver Gailing^{1*}

¹School of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931

²Department of Entomology and Plant Pathology, The University of Tennessee, Knoxville, TN 37996

³Department of Forestry, University of Missouri, Columbia, MO 65274

⁴Department of Ecosystem Science and Management and Department of Plant Science, The Pennsylvania State University, University Park, PA 16802

*Corresponding author: ogailing@mtu.edu

¹The material in this chapter was previously published material and has been reprinted with kind permission from Executive Editor Annals of Forest Research from Khodwekar et al. (2015) Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). Annals of Forest Research **58**:193-204. Please refer to the appendix 1: Open access licence on page no. 58.

Abstract. A set of seven new nuclear microsatellite markers (nSSRs) was developed for sugar maple (*Acer saccharum* Marsh.) using paired-end Illumina sequencing. Out of 96 primers screened in a panel of six unrelated individuals, seven markers amplified polymorphic products. The utility of these markers, in addition to six already published microsatellites, for genetic variation and gene flow studies was assessed. Out of the seven newly developed markers, three amplified multiple fragments and were interpreted as dominant (absence/presence) markers, while four markers amplified a maximum of two amplification products per sample. The six published microsatellites and three of the four newly developed markers showed regular segregation in an open-pollinated single tree progeny. Observed heterozygosity (H_o) and expected heterozygosity (H_e) in 48 individuals from one population ranged from 0.436 to 0.917 and from 0.726 to 0.894, respectively. Dominant markers revealed 64 variable positions and moderate genetic variation within the population ($H_e = 0.102$, Shannon's $I = 0.193$). Paternity analyses in the program CERVUS at co-dominant markers showed effective dispersal of pollen in the sugar maple population both at 95% and 80% confidence levels. Dependent on the confidence level, the mean pollen dispersal distance within the population ranged from 33 m to 39 m and gene flow from outside the stand from 76% to 82%. The absence of fine-scale Spatial Genetic Structure (SGS) suggested effective dispersal of both seeds and pollen.

Keywords *Acer saccharum*, nuclear microsatellites, gene flow, spatial genetic structure, inheritance analysis

Introduction

Sugar maple (*Acer saccharum* Marsh.) is an important North American hardwood forest tree that is widespread and abundant in Northeastern North America (Godman et al. 1990). Despite its ecological and economic importance, genetic and genomic resources for this species are rather limited (Graganic et al. 2013). Here, we describe the development and application of Next-Generation Sequencing derived microsatellite markers for population genetic analyses in sugar maple.

The level and distribution of genetic variation is essential for the long-term survival and ability of organisms to adapt to changes in their environment (Krustovsky & Neale 2005). Recent research has predicted that climate change has the potential to affect species distributions, alter ecosystem functions and species interactions (Iverson et al. 2008, Whitham et al. 2008). Other important predicted consequences of global temperature increase are range shifts and northward migration of forest trees (Gunter et al. 2000, Graganic et al. 2014). Range edge populations are especially prone to geographical isolation and environmental changes. For example, sugar maple populations at the species' southern distribution range are threatened by fragmentation as the result of climate change, and the preservation of genetic variation in these populations is a major conservation concern (Iverson et al. 2008, Graganic et al. 2013). Also, silvicultural treatments including the regulation of stand structure through harvesting operations will impact natural regeneration patterns and be expected to have considerable effects on subsequent genetic variation patterns by affecting the reproduction system of trees

(Finkeldey & Ziehe 2004). These changes can be reflected in changes of fine scale genetic structure of the progeny generation within populations.

Genetic variation is ubiquitous in natural populations and is important for sugar maple to adapt to changing environments and cope with biotic and abiotic stresses such as increased drought, disruption of nutrient cycles, insect outbreaks and fungal diseases (Gibson & Wheelwright 1995, Whitham et al. 2008, Whitham et al. 2012). For example, sapstreak disease caused by the fungus *Ceratocystis virescens* (Davidson) Moreau leads to sugar maple dieback in Upper Michigan (Bal et al. 2013) which will likely be reflected in the level and distribution of genetic variation within species. The possibility to observe variation patterns at particular gene loci is of great significance in this context.

The development of highly polymorphic microsatellites can greatly enhance ecological genetic studies in sugar maple (see also Pandey et al. 2004, 2012), and genetic variation can be monitored in response to climate change and management. Despite its ecological and economic importance as a dominant overstory species in Northern hardwoods, there are few studies of genetic variation in sugar maple using highly variable genetic markers such as microsatellites (Graganic et al. 2013). The usefulness of microsatellites is due to the variability in the number of sequence repeats at a given gene locus within species and populations. Nuclear microsatellites are co-dominant, easily scored and abundant throughout the genome (Selkoe & Toonen 2006). Because of their high polymorphism they are suitable for investigations of the gene flow and mating system (Selkoe & Toonen 2006).

The main objective of the present study is the development and application of microsatellite markers to assess family structures and gene flow within a natural sugar maple population. These microsatellite markers can also be applied to monitor genetic variation, especially in peripheral populations, and used as predictive tools to characterize the long term adaptability of these populations to predicted impacts of climate change.

Materials and Methods

Species description

Sugar maple is a dominant overstory species in northern hardwoods and abundant in the Northeastern part of North America (Godman et al. 1990). It is a major constituent of several forest cover types and is classified as a shade tolerant hardwood tree species and considered a “climax species” throughout its extensive native range (Godman et al. 1990). It is an important commercial source of income due to its syrup and timber quality. Sugar maple is characterized by hard and strong wood, and was used as sweetener by Native Americans and early European settlers (Nesom 2006). The flowering period of sugar maple extends from late March to mid May depending upon the geographical location and seasonal variation. Sugar maple flowers can be either perfect or staminate (pistil abortive) and are generally wind pollinated, but pollination by bees was also reported (Gabriel & Garrett 1984). The fruit is a double samara and is dispersed by wind with the potential of long distance dispersal (Guries & Nordheim 1984, Godman et al. 1990). It ripens within about 16 weeks; the large wings typically range from 20 to 27 mm in length and 7 to 11 mm in width. These samara wings allow fruits to be carried away by wind at least 100 m (Godman et al. 1990).

Marker development

Sugar maple leaf samples were collected from a location near Houghton, MI, USA (47°07'07.46''N, 88°35'16.41''W) in a suburban setting (Figure 2.1) and dried in silica gel in August 2012. A total of 48 adult trees were sampled. Additionally, seeds were

collected from one adult tree. GPS locations for all 48 trees were recorded. Total genomic DNA was extracted from dried leaf tissues of adult trees and from 96 embryos of seed from one seed parent with the DNeasy Plant Mini Kit following the standard protocol (Qiagen, Hilden, Germany). Approximately 1 cm² dried leaf tissue was used for DNA extraction. For the inheritance analyses, embryos were carefully separated from the seed coat prior to DNA extraction. The fruits (samaras) contained only a single seed. The embryos ranged from 5 to 8 mm in size.

We applied low coverage whole genome paired-end Illumina sequencing in ten hardwood tree species (Owusu et al. 2013, Staton et al. 2015) including sugar maple. A total of 10,152 di-, tri and tetranucleotide microsatellite motifs were identified using an SSR finder script (Staton et al. 2015). After removal of redundant sequences (identity \geq 95%) using the program cap3 (Huang 1999), primer3 (Rozen & Skaletsky 2000) was used to identify primers with default settings. Primer pairs were designed for 8,012 di-, 1,665 tri- and 475 tetranucleotide read motifs. The genomic SSR data are deposited at NCBI (<http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA200681>, accession PRJNA200681). In total, we selected 96 primer pairs for dinucleotide repeat microsatellites for PCR amplification.

PCR amplification

For marker development, amplification and polymorphism was assessed for 96 dinucleotide repeat SSRs in a panel of six unrelated individuals on the QIAxcel Fast Analysis System with the DNA high resolution kit (Qiagen). PCR was performed in a 15

µl reaction mix and contained 3 µl of 5× HOT FIREPol® Blend Master Mix Ready to Load (contains 10 mM MgCl₂, 0.6 units of HOT FIREPol® Taq polymerase, and 2 mM dNTPs) (Solis BioDyne, Estonia), 2 µl each of 5 µM forward and reverse primers (Sigma Aldrich, St. Louis, MO and Applied Biosystems, Foster City, CA), 6 µl double deionized water (DNase and RNase free) and 2 µl DNA (~1.8 ng/µl). Amplification was done in a Peltier Thermal Cycler (GeneAmp® PCR system 2720, Applied Biosystems). The PCR profile consisted of 15 minutes denaturation at 95°C, 35 cycles denaturation at 94°C for 45s, annealing at T_a for 45s (Table 2.1), 45s elongation at 72°C, and an extension step at 72°C for 20 minutes. The amplified PCR products were separated on agarose gels and on the QIAxcel Fast Analysis System. Polymorphic loci from this screening panel and six additional SSR loci (Graignic et al. 2013) were amplified in 48 samples from a single population and in 96 single tree progenies from one seed parent using fluorescent labeled forward primers (6-FAM, PET, NED and VIC). PCR products were separated on an ABI Prism® Genetic Analyzer 3730 with GeneScan™ LIZ-500 as internal size standard (Applied Biosystems) and scored with GeneMarker® V2.6.3 (SoftGenetics). All markers were tested for marker transferability in four other maple species (*A. ginnala*, *A. platanoides*, *A. rubrum*, *A. saccharinum*). Three individual samples were collected from each of these four species surrounding the Michigan Tech campus and all markers were tested for amplification and polymorphism.

Data analysis

Genetic variation assessment was conducted for individual microsatellite markers for the 48 samples using GenAlEx 6.501 (Peakall & Smouse 2006, Peakall & Smouse 2012).

Observed and expected heterozygosities (H_o , H_e , respectively) (Nei 1973) and number of alleles (N_a) were calculated. HWE exact tests and pairwise linkage disequilibrium (LD) were calculated with GENEPOP 4.2 (Raymond & Rousset 1995, Rousset 2008). For the three markers that revealed a multibanding pattern (Table 1.1) Shannon's information index I (Brown & Weir 1983) and Nei's gene diversity H_e (Nei 1973) were calculated in GenAlEx. The frequency of the null allele at each locus was estimated by the square root method (Nei 1987).

Paternity exclusion probabilities for candidate pollen parents were calculated at co-dominant markers using the computer program CERVUS 3.0.7 (Marshall et al. 1998). Paternity analysis was carried out in CERVUS using log-likelihood ratios (LOD score) for the paternity assignment on the basis of observed multilocus genotypes at a strict (95%) and relaxed (80%) confidence level. Simulations of paternity analysis were completed to determine confidence levels of LOD scores (Pakkad et al. 2008, Pluess et al. 2009). The parameters for the simulation analyses were: number of cycles: 100,000; number of candidate parents: 200; proportion of candidate fathers sampled: 25%; proportion of loci typed: 99%; loci mistyped: 1%; error rate in likelihood calculations: 0.010 and the minimum number of typed loci: 4.

Pairwise distances between trees (m) were calculated in GenAlEx based on GPS coordinates.

Spatial genetic structure (SGS) was characterized in the population by spatial autocorrelation analysis using SPAGeDi 1.4c (Hardy & Vekemens 2002). Ten distance

classes were selected ensuring an equal number of pairwise comparisons in each distance class (113 comparisons) at an approximate increment of 15 m per distance class. To assess the association between genetic similarity and geographic distance, regression analyses of pairwise kinship coefficients (F_{ij}) (Loiselle et al. 1995) on geographical distances were performed and a jackknife method was used to estimate standard errors. Lower and upper 95% confidence intervals were established by 10,000 permutations of individuals among distance classes. Correlograms of pairwise kinship coefficients (F_{ij}) were plotted against distance and 95% confidence intervals are shown (Figure 1.3).

The Sp statistic was derived from the kinship coefficient F_{ij} between individuals of the first distance class and its regression slope b_f as $Sp = -b_f/1 - F_{ij}$.

To calculate the census density (CD) as individuals per hectare according to Stefenon et al. (2008), a polygon was drawn to represent the population area using Google Earth Pro (Source: ‘‘Houghton’’, MI, USA, 47°07’07.46’’N, 88°35’16.41’’W. Google Earth Pro. May 9, 2014. November 26, 2014). The total area was 4393.81 m² (0.493 hectare) and the census density was 109.25 individuals per hectare.

An inheritance analysis was performed for all markers using single tree progeny following the approach by Gillet and Hattemer (1989). The following assumptions were tested. For a heterozygous seed parent A_iA_j each offspring has either allele A_i or A_j . Additionally, under regular segregation and absence of selection and other evolutionary factors, the following absolute genotype frequencies are expected in the progeny: $N(A_iA_j) = N(A_iA_i) + N(A_jA_j)$, and $N(A_iA_k) = N(A_jA_k)$ with $k \neq i, j$.

Results

Initially 96 primer pairs were tested on a set of DNA samples from six individuals in the source population. Out of these 96 primer pairs, 72 primer pairs successfully amplified products in the expected size range, but only seven produced a clear discernible banding pattern on agarose gels and on the QIAxcel Fast Analysis System using the high resolution kit for microsatellite analysis. These markers were selected as potential polymorphic microsatellites for further testing on an ABI Prism® Genetic Analyzer 3730. Three markers amplified more than one gene locus and were interpreted as dominant (absence/presence) markers, while four markers amplified a maximum of two products per individual and were interpreted as co-dominant markers (see suppl. Figure 2.1). Also, all six published markers (Graignic et al. 2013) showed pattern consistent with the amplification of a single gene locus. Using the diversity panel of 48 individuals from one population, all ten co-dominant microsatellite markers showed relatively high levels of polymorphism with number of alleles per locus ranging between 9 and 16. Observed heterozygosity (H_o) ranged from 0.436 to 0.917 and expected heterozygosity (H_e) ranged from 0.726 to 0.894. The inbreeding coefficient, F_{is} , was estimated for each locus in the population using GENEPOP, and the F_{is} values ranged from -0.0581 to 0.4992 with three loci showing values significantly different from zero ($\alpha=0.05$) after Bonferroni corrections (Table 2.2). Out of forty five marker pairs, five marker pairs showed significant linkage disequilibrium at the 5% level and no marker pair was significant at the 1% level.

Shannon's information index I and gene diversity for dominant markers were 0.193 and 0.102, respectively, with 64 variable positions in 48 adults.

Results of the inheritance analysis for the ten putative single-locus markers using single tree progenies of one heterozygous seed parent are shown in Tables 2.3 and 2.4. Table 2.3 includes the comparisons for absolute genotype frequencies of the heterozygous genotype $N(A_iA_j)$ and of both homozygous genotypes $N(A_iA_i)$ and $N(A_jA_j)$. Chi-square tests of goodness-of-fit for observed and expected frequencies revealed significant deviations between observations and expectations for AS05 ($p < 0.0001$), SM21A ($p = 0.012$) and AS47 ($p = 0.029$). Table 2.4 includes absolute frequencies of heterozygous genotypes $N(A_iA_k)$ and $N(A_jA_k)$ ($k \neq i, j$) and results of the Chi-square test of goodness-of-fit for observed and expected frequencies. Only AS05 showed significant deviation from the expected 1:1 ratio in this test. The observed frequencies indicated regular segregation patterns for the other nine markers. Consequently, AS05 was excluded from the gene flow and SGS analyses.

Gene flow analyses using the computer program CERVUS showed effective dispersal of pollen in the sugar maple population both at the 95% and 80% confidence levels (Figure 2.2 A, B). The nine highly polymorphic single-locus microsatellite markers showed high exclusion probabilities for paternity analysis (Table 2.2); the total paternity exclusion probability for all loci was 0.999. At the 95% strict threshold 22% and at the 80% relaxed threshold 18% of the seeds were assigned to their putative fathers within the stand. Thus, the minimum estimates for gene flow from outside the plot are 76% and 82% under the strict and relaxed thresholds, respectively. Some seeds also resulted from the self-

pollination (2%). Figure 2.2 shows the average gene flow distance in the population stand ranging from 0 m (self-pollination) to 81 m, with a mean pollen dispersal distance ranging from 33 m to 39 m at 95% and 80% confidence intervals, respectively.

Spatial genetic structure analysis using SPAGeDi detected non-significant SGS, corroborating effective dispersal by both seeds and pollen (Figure 2.3). Consequently, the estimate for the S_p statistic was very low (0.00576). Gene dispersal estimates did not reach convergence.

Marker transferability tests showed that all thirteen markers were transferable to four species of *Acer* (*A. rubra*, *A. saccharinum*, *A. platanoides*, *A. ginnala*). Most of the markers showed a multi-banding pattern. However, AS47 and SM37 amplified a single polymorphic locus in *A. ginnala*, and markers Am116 and SM14 showed a single-locus pattern in *A. platanoides*. Finally, SM21A was heterozygous for the same alleles in *A. saccharinum*.

Discussion

Low coverage whole genome sequencing proved to be useful to generate microsatellite resources for ten North American hardwood species including sugar maple (Owusu et al. 2013, Staton et al. 2015). To date, locus-specific markers have been developed for two of the species, sugar maple (present study) and honeylocust (*Gleditsia triacanthos* L.) (Owusu et al. 2013). While most primers produced fragments in the expected size range, the percentage of polymorphic and single-locus markers was low as compared to microsatellites developed from microsatellite-enriched genomic libraries for related species using Sanger sequencing (Pandey et al. 2004). Thus, out of 96 markers, only seven were polymorphic and three of them showed regular segregation in the inheritance analysis (present study). In honey locust, 14 locus-specific markers were developed after testing 144 primer pairs using the same procedure as in sugar maple (Owusu et al. 2013). Additional studies are necessary to evaluate whether the relatively low number of polymorphic and locus-specific markers are specific for the species studied or are a peculiarity of the sequencing technique. Thus, AT-rich, relatively short motifs were preferentially recovered from Illumina sequencing for all ten species (Owusu et al. 2013, Staton et al. 2015). The relatively low number of polymorphic loci that were discovered in sugar maple may be partly related to the bias toward short and AT-rich motifs (Staton et al. 2015). In the present study, primers were developed based on low coverage genome sequencing (without enrichment step) for dinucleotide repeat motifs with repeat numbers ranging from 8 to 25 (mean of 16.8). As a comparison, pyrosequencing of an enriched microsatellite library in sugar maple resulted in the development of a moderate number of

single-locus polymorphic markers (18 new next-generation sequencing derived markers out of 90 markers tested) when motifs with the highest number of repeats were targeted (10 - 25 repeats) (Graignic et al. 2013). The relatively low read coverage for reconstructed fragments with SSR motifs in the present study (~ 1X coverage, Staton et al. 2015) may have resulted in wrong base calls in the primer binding sites and in unspecific PCR amplification. Also, amplification and sequencing errors might have resulted in an overestimation of repeat numbers for microsatellite regions. Thus, a 20 fold sequencing coverage was suggested to compensate for sequencing errors (Dohm et al. 2008).

In summary, low coverage genome sequencing was very cost-effective with regard to the generation of sequenced reads with microsatellite motifs (Staton et al. 2015), but a comparatively large number of markers needed to be tested for the development of locus-specific polymorphic markers for population genetic analyses.

In the present study, we assessed the suitability of six published (Graignic et al. 2013) and of seven newly developed nuclear microsatellites for population genetic analyses. Specifically, we tested for Hardy-Weinberg proportions and regular segregation using open pollinated single tree progeny (Gillet & Hattemer 1989). Based on these results we selected nine polymorphic and easy-to interpret markers to analyze fine scale genetic structure and gene flow in one sugar maple population. In addition to other published microsatellites for sugar maple (Graignic et al. 2013), these markers can serve as tools for future population genetic analyses. Also, all markers are transferable to four other species of *Acer* and some of them amplify single polymorphic loci opening up opportunities for

population genetic studies across different maple sections.

Gene flow analyses based on paternity assignment in CERVUS showed very low gene flow between trees within the sampled stand and high gene flow from outside the stand (76%, 82%) both at the strict and relaxed confidence levels, suggesting a prevalence of long-distance pollen dispersal. Within the sampled stand, contributions of individual pollen parents to the progeny were $\leq 10\%$ at the strict confidence level (95%) and $\leq 4\%$ at the relaxed level, and most neighboring trees did not contribute to the progeny. The sexual system (e.g. dioecy, monoecy) and differences in flowering time are important determinants for the patterns of gene flow within a population and may result in uneven contributions of pollen donors to the progeny and in the development of spatial genetic structure (SGS) at different spatial scales (Young & Merriam 1994).

Since sugar maple trees can be monoecious (male and female flowers are produced on the same tree) or functionally dioecious (separate male and female trees) (Stephens 1973, Geburek 1993, Nesom 2006), self-pollination is likely limited. Even in monoecious trees selfing rate is expected to be comparatively low since most female and male flowers come into bloom at different times (Gabriel 1968). Accordingly, only 2% of the seeds were found to be originating from self-pollination in the present study.

In summary, gene flow patterns as inferred by paternity analysis suggested effective means of pollen dispersal in sugar maple and no preferential mating between neighboring trees as would have been expected for solely insect-pollinated species. These results suggest that wind pollination is predominant and sufficient for pollination (Gabriel &

Garrett 1984). However, a spatial clustering of related individuals can also result from spatially restricted seed dispersal. In the present study, spatial autocorrelation analyses showed no significant SGS in any distance class, suggesting random and long-distance dispersal of both pollen and seeds. Since the stand is affected by sugar maple dieback, mortality could also have resulted in a secondary disintegration of family structures. In accordance with our results, most allozyme loci showed non-significant SGS at small distance classes. Thus, spatial autocorrelation analyses in six natural sugar maple populations showed significant autocorrelation coefficients (Morans I) in three out of nine allozymes in only two populations (Geburek 1993). Significant spatial aggregation of alleles was found for seven alleles for a total of 30 tests in three natural populations (Perry & Knowels 1991). In seedling cohorts of fragmented and continuous sugar maple forests, no association between genetic and geographic distance and no SGS at smaller distance classes were found at ten allozyme loci (Young & Merrium 1994). However, spatial aggregation of genotypes at different larger distance classes suggested that SGS might be influenced by the spatial patterns of flowering times within a population (Gabriel 1968, Young & Merrium 1994). In the present study, we found significant SGS between 60 m and 140 m for only a single marker (SM14). Since the ability to detect SGS is positively correlated with the number of polymorphic markers analyzed (Cavers et al. 2005), inclusion of a much larger number of SSR loci might detect weak SGS at some distance classes. However, significant SGS at a similar or smaller number of polymorphic SSRs was found in other outcrossing and wind-pollinated species (Stefenon et al. 2008) including *Acer pseudoplatanus* (Pandey et al. 2012).

Conclusion

In addition to other recently developed microsatellite markers, the newly developed and tested markers proved to be useful for the assessment of genetic variation and structure, and for gene flow analyses. They will be an important tool to monitor the level of genetic variation especially at the southern range edge of the species for which a decrease in abundance is predicted as the result of warming climates (Iverson et al. 2008, Gaignic et al. 2013).

Acknowledgements

This research was supported by grant TRPGR IOS-1025974 from the NSF Plant Genome Research program and by the Michigan Technological University Ecosystem Science and Biotech Research Centers. Sudhir Khodwekar is supported by the Government of India 'National Overseas Scholarship'.

References

- Bal, T. L., D. L. Richter, A. J. Storer, and M. F. Jurgensen. 2013. The relationship of the sapstreak fungus, *Ceratocystis virescens*, to sugar maple dieback and decay in Northern Michigan. *American Journal of Plant Sciences* 4: 436-443.
- Brown, A. H. D., and B. S. Weir. 1983. Measuring genetic variability in plant populations. In: Tanksley SD, Orton TJ (eds). *Isozymes in Plant Genetics and Breeding, Part A*. Elsevier, Amsterdam, pp 219-239.
- Cavers, S., B. Degen, H. Caron, M. R. Lemes, R. Margis, F. Salgueiro, and A. J. Lowe. 2005. Optimal sampling strategy for estimation of spatial genetic structure in tree populations. *Heredity* 95: 281-289.
- Dohm, J. C., C. Lottaz, T. Borodina, and H. Himmelbauer. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* 36: E1005.
- Finkeldey, R., and M. Ziehe. 2004. Genetic implications of silvicultural regimes. *Forest Ecology and Management* 197: 231-244.
- Gabriel, W. J. 1968. Dichogamy in *Acer saccharum*. *Botanical Gazette* 129: 334-338.
- Gabriel, W. J., and P. W. Garrett. 1984. Pollen vectors in sugar maple (*Acer saccharum*). *Canadian Journal of Botany* 62: 2889-2890.
- Geburek, T. 1993. Are genes randomly distributed over space in mature populations of sugar maple (*Acer saccharum* Marsh.)? *Annals of Botany* 71: 217-222.
- Gibson, P. J., and N. T. Wheelwright. 1995. Genetic structure in a population of a tropical tree *Ocotea tenera* (Lauraceae): influence of avian seed dispersal. *Oecologia* 103: 49-54.
- Gillet, E., and H. H. Hattemer. 1989. Genetic analysis of isozyme phenotypes using single tree progenies. *Heredity* 63: 135-143.
- Godman, R. M., H. W. Yawney, and C. H. Tubbs. 1990. *Acer saccharum* Marsh. Sugar maple. In: Burns, RM, Honkala BH (eds.) *Silvics of North America, Agricultural handbook 654 volume 2. Hardwoods-* USDA Forest Service 2: 78-91.
- Graignic, N., F. Tremblay, and Y. Bergeron. 2013. Development of polymorphic nuclear microsatellite markers in sugar maple (*Acer saccharum* Marsh.) using cross-species transfer and SSR-enriched shotgun pyrosequencing. *Conservation Genetic Resources* 5: 845-848.
- Graignic, N., F. Tremblay, Y. Bergeron, and J. Williams. 2014. Geographical variation in reproductive capacity of sugar maple (*Acer saccharum* Marshall) northern peripheral populations. *Journal of Biogeography* 41: 145-157.
- Gunter, L., E., G. Tuskan, A., C. Gunderson, A., and R. J. Norby. 2000. Genetic variation and spatial structure in sugar maple (*Acer saccharum* Marsh.) and implications for predicted global-scale environmental change. *Global Change Biology* 6: 335-344.

- Guries, R. P., and E. V. Nordheim. 1984. Flight characteristics and dispersal potential of maple samaras. *Forest Science* 30: 434-440.
- Hardy, O. J., and X. Vekemans. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at individual or population levels. *Molecular Ecology Notes* 2: 618-620.
- Huang, X. 1999. CAP3: A DNA sequence assembly program. *Genome Research* 9: 868-877.
- Iverson, L. R., A. M. Prasad, S. N. Matthews, and M. Peters. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management* 254: 390-406.
- Krustovsky, K. V., and D. B. Neale. 2005. Forest genomics and new molecular approaches to measuring and conserving adaptive genetic diversity in forest trees, In: *Conservation and Management of Forest Genetic Resources in Europe*, T. Geburek and J. Turok (eds.), Arbora Publishers, Zvolen, pp 369-390.
- Loiselle, B. A., V. L. Sork, G. Nason, and C. Graham. 1995. Spatial genetic structure of a tropical understory shrub. *American Journal of Botany* 82: 1420-1425.
- Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA* 70: 3321-3323.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nesom, G. 2006. Plant Guide, sugar maple. United States Department of Agriculture Natural Resources Conservation Service (http://plants.usda.gov/plantguide/pdf/pg_acsa3.pdf).
- Owusu, S. A., M. Staton, T. N. Jennings, S. Schlarbaum, M. V. Coggeshall, J. Romero-Severson, J. E. Carlson, and O. Gailing. 2013. Development of genomic microsatellites in *Gleditsia triacanthos* (Fabaceae) using Illumina sequencing. *Applications in Plant Sciences* 1 (12): 1300050.
- Pakkad, G., S. Ueno, and H. Yoshimaru. 2008. Gene flow pattern and mating system in a small population of *Quercus semiserrata* Roxb. (Fagaceae). *Forest Ecology and Management* 255: 3819-3826.
- Pandey, M., O. Gailing, D. Fischer, H. H. Hattemer, and R. Finkeldey. 2004. Characterization of microsatellite markers in sycamore (*Acer pseudoplatanus* L.). *Molecular Ecology Notes* 4: 253-255.
- Pandey, M., O. Gailing, H. H. Hattemer, and R. Finkeldey. 2012. Fine-scale spatial genetic structure of sycamore maple (*Acer pseudoplatanus* L.). *European Journal of Forest Research* 131: 739-746.

- Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537-2539.
- Peakall, R. O. D., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.
- Perry, D. J., and P. Knowels. 1991. Spatial genetic structure within three sugar maple (*Acer saccharum* Marsh) stands. *Heredity* 66: 137-142.
- Pluess, A. R., V. L. Sork, B. Dolan, F. W. Davis, D. Grivet, K. Merg, J. Papp, and P. E. Smouse. 2009. Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *Forest Ecology and Management* 258: 735-744.
- Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Eucemenicism. *Journal of Heredity* 86: 248 - 249.
- Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103-106.
- Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and biologist programmers. *Methods in Molecular Biology* 132: 365-386.
- Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9:615-629.
- Stefenon, V. M., O. Gailing, and R. Finkeldey. 2008. The role of gene flow in shaping genetic structures of the subtropical conifer species *Araucaria angustifolia*. *Plant Biology* 10:356-364.
- Stephens, H. A. 1973. Woody plants of the North Central plains. Lawrence, KS: The University Press of Kansas. 530 p.
- Whitham, T. G., S. P. Difazio, J. A. Schweitzer, S. M. Shuster, G. J. Allan, J. K. Bailey, and S. A. Woolbright. 2008. Extending genomics to natural communities and ecosystems. *Science* 320: 492-495.
- Whitham, T. G., C. A. Gehring, L. J. Lamit, T. Wojtowicz, L. M. Evans, A. R. Keith, and D. S. Smith. 2012. Community specificity: life and afterlife effects of genes. *Trends in Plant Science* 17: 271-281.
- Young, A. G., and H. G. Merriam. 1994. Effects of forest fragmentation on the spatial genetic structure in *Acer saccharum* Marsh. (sugar maple) populations. *Heredity* 72: 201-208.

Table 2.1 Primer sequences and characteristics of thirteen genomic microsatellite markers in one population of *Acer saccharum*

Locus name	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Repeat motif	5' dye	Amplicon size (bp)	Size range (bp)	T _a (°C)	Accession number
AS05	GGAGGTGATGAGAGAAATCTGG	GAATGTGTGGCTGTGATTTGG	(AT) ₉	VIC	124	100-147	58	PRJNA200681
AS06 ^a	GTCCACTGAAAGGAGATCGG	ACCCAGAAATCCAGAAACCC	(GA) ₁₁	6-FAM	114	98-156	58	PRJNA200681
AS09 ^a	CAACGTCTCATGCAAAATATCG	TGTGCCACAACAATGCAGGG	(CA) ₁₄	6-FAM	149	128-155	58	PRJNA200681
AS11	AACATGTGGGTCTTCTCC	AGAAGGTTCATCGACGTGC	(CT) ₉	VIC	154	89-152	58	PRJNA200681
AS47	CCATGGCCATCTATCTCTCC	TGAATCTCCATTGCTGATCG	(CT) ₁₂	6-FAM	128	106-139	56	PRJNA200681
AS48	CCAAATGACACCAAGATCACC	GTTGAATCCCTGTCCCTTCC	(GA) ₁₉	NED	115	90-113	56	PRJNA200681
AS127 ^{a,b}	TCACCGGAAAGGACAAGTGC	CCAAATCGAACCTAAACGGCG	(TC) ₁₇	6-FAM	107	100-120	56	PRJNA200681
Aml16 ^a	AACGCTACCGACTTCGCCAACT	TGGAGGTCAAGTGTGGAAACAA	(AG) ₆	PET		230-273	56	AB303350
SM14 ^a	TTTTATGTAGAGCAACTCAACCCA	TATCTGCTGTTGACATGAACTT	(GA) ₁₅	NED		72-112	56	KC751436
SM21A ^a	TAGTTGTGCACCAACCATGC	TCCATCAAAAACGCTGCTATG	(GAT) ₁₄	VIC		179-243	56	KC751437
SM22 ^a	CCAGAGCTTGAAGAAAATGTACG	GGTAAGGGGTTGTTATGCAAT	(GA) ₁₃	PET		295-323	56	KC751438
SM27 ^a	TTTCTACTTTAGAGATGGAACCGG	CCCTAAATCCCAAAATCAGTGAA	(TC) ₁₀	PET		242-260	56	KC751440
SM37 ^a	TGGGAGAAAATTAACATAGGATTT	AGACTTGACTCTCCTAATCTTGGTG	(AG) ₁₄	NED		174-198	56	KC751444

T_a – annealing temperature

* amplifies multiple loci (multi banding pattern)

^a SSRs published in Graignic et al. (2013)

^b Reverse primer sequence (5'-3') labeled fluorescently

Table 2.2 Genetic variation parameters and exclusion probabilities at codominant markers in *Acer saccharum*

Locus name	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F_{is}</i>	Exclusion probability
^a AS05	38	13	0.436	0.854	0.4992*	0.7251
AS11	48	15	0.521	0.726	0.2922*	0.5562
AS47	48	14	0.771	0.864	0.1179	0.7309
AS48	48	15	0.792	0.852	0.0815*	0.7321
Am116	47	14	0.729	0.781	0.0774	0.5908
SM14	48	16	0.750	0.880	0.1576	0.7738
SM21A	48	11	0.917	0.858	-0.0581	0.7173
SM22	48	13	0.813	0.894	0.1019	0.7940
SM27	48	9	0.708	0.757	0.0750	0.5615
SM37	48	9	0.646	0.750	0.1487	0.5553

Note: *N*- Number of samples; *N_a*- Number of alleles; *H_o*- Observed heterozygosity; *H_e*- Expected heterozygosity; *: Significantly different from Hardy-Weinberg proportions ($\alpha = 0.05$) after Bonferroni corrections. ^a: this marker showed irregular segregation and likely amplified more than one gene locus (see Table 1.3).

Table 2.3 Segregation within the progenies of the putatively heterozygous seed tree

Marker	Maternal genotype (bp)	Frequency of offspring genotypes			χ^2	<i>p</i>
		A_iA_i	A_jA_j	A_iA_j		
AS05	124/138	0	5	42	32.44*	<0.0001
AS11	120/140	16	12	35	0.77	0.38
AS47	124/126	1	3	13	4.76*	0.03
AS48	93/111	6	1	2	2.78	0.10
Am116	231/241	13	19	23	1.47	0.22
SM14	72/86	5	1	2	2.00	0.16
SM21A	201/207	2	2	15	6.37*	0.01
SM22	311/315	4	5	8	0.06	0.77
SM27	248/256	6	11	24	1.20	0.27
SM37	178/180	21	11	31	0.02	0.89

A_iA_j : genotype of the seed tree. A_iA_i and A_jA_j : frequency of putatively homozygous progenies. A_iA_j : frequency of putatively heterozygous progenies. Chi square test to analyze statistical significance between expected and observed segregation frequencies (χ^2 -test; df =1) and exact probabilities *p*.

Table 2.4 Segregation within the progenies of the putatively heterozygous seed tree


Marker	Maternal genotype (bp)	Frequency of offspring genotypes			χ^2	<i>p</i>
		A_iA_k	A_jA_k	A_iA_j		
AS05	124/138	1	37	34.11*	0.0001	
AS11	120/140	15	11	0.61	0.43	
AS47	124/126	37	42	0.32	0.57	
AS48	93/111	47	39	0.74	0.39	
Am116	231/241	22	19	0.22	0.64	
SM14	72/86	37	50	1.94	0.16	
SM21A	201/207	40	37	0.12	0.68	
SM22	311/315	34	44	1.28	0.26	
SM27	248/256	24	31	0.89	0.35	
SM37	178/180	19	11	2.13	0.14	

A_iA_j : genotype of the seed tree. A_iA_k and A_jA_k : frequency of putatively heterozygous progenies. Chi square test to analyze statistical significance between expected and observed segregation frequencies (χ^2 -test; df =1) and exact probabilities *p*.



Figure 2.1 Sugar maple natural population location in a suburban setting, Houghton, Michigan, USA (total area = 4393.81 m²).

Note: This image has been reproduced from “Khodwekar, S., M. Staton, M. V. Coggeshall, J. E. Carlson, and O. Gailing. 2015. Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). *Annals of Forest Research* **58**:193-204” with a kind permission from the executive editors of the journal.

 **Marius Teodosiu** 5:32 AM (3 minutes ago) ☆ ↶ ↷

to me, Oliver ▾

Dear Sudhir, dear Oliver,

I confirm thereby that the authors which publish in *Annals of Forest research* has permission to reproduce or publish other material e.g. images, texts, graphical elements or similar in their PhD dissertation.

Good luck with your Ph.d. thesis!

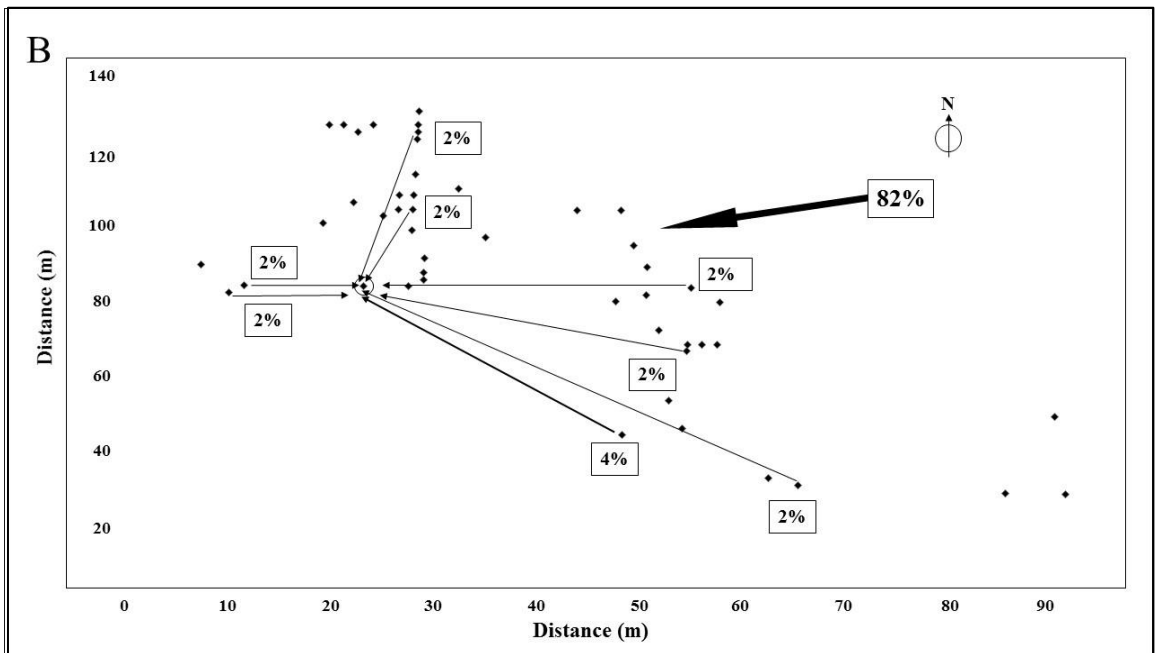
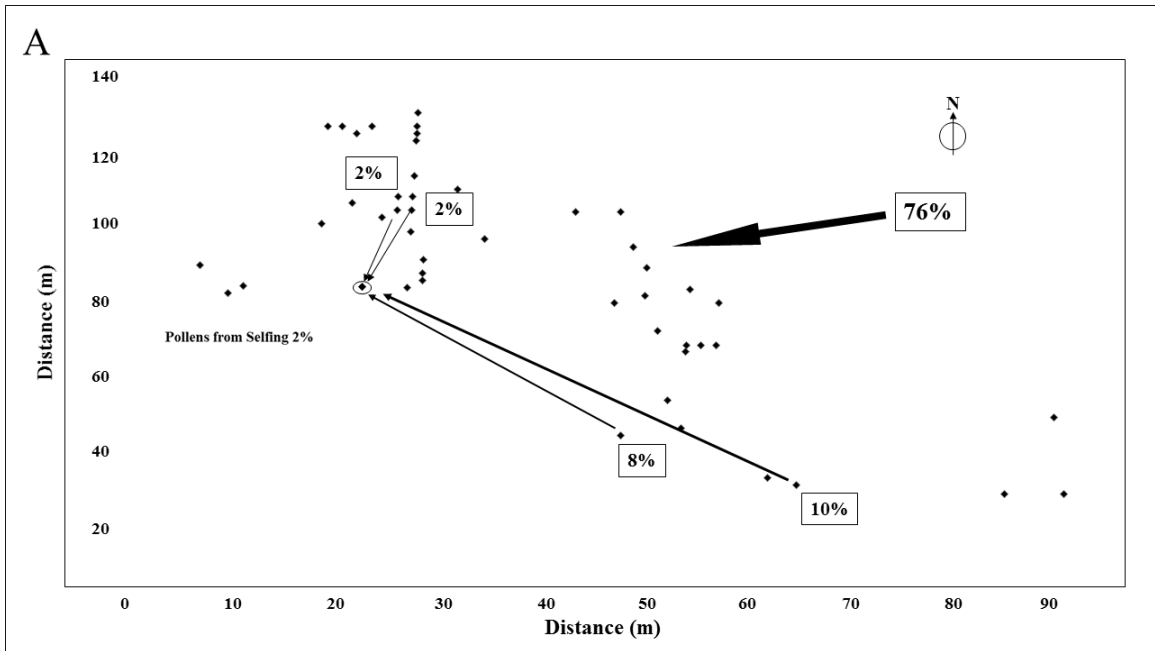


Figure 2.2 Gene flow at the 95% (A) and 80% (B) confidence level in natural population of sugar maple.

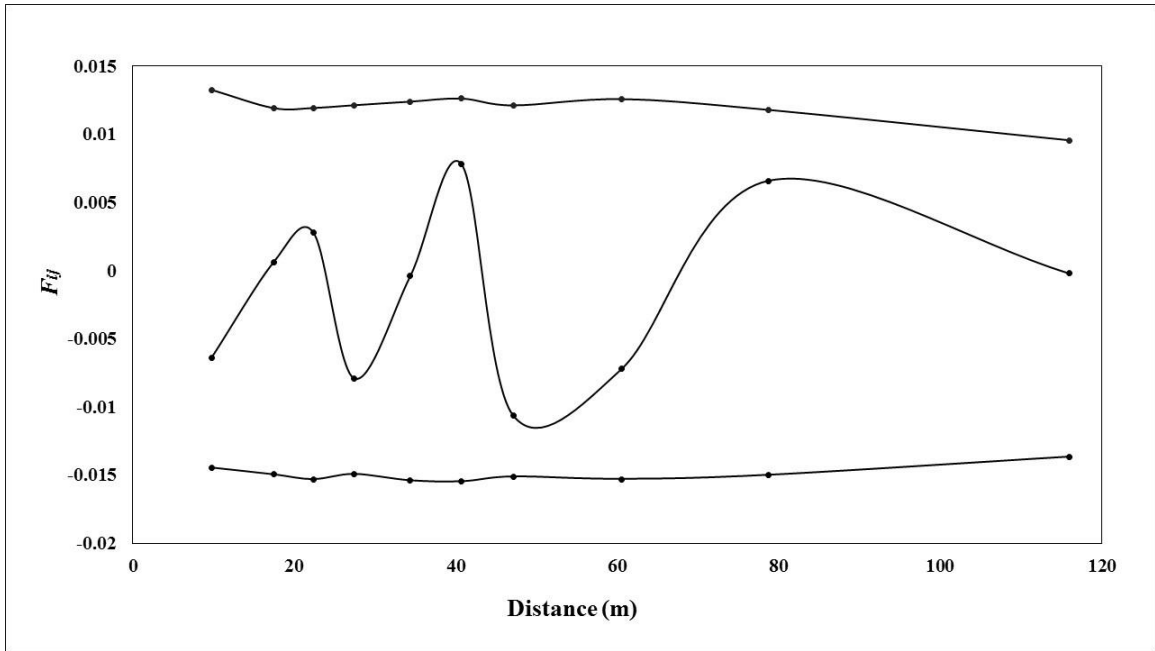
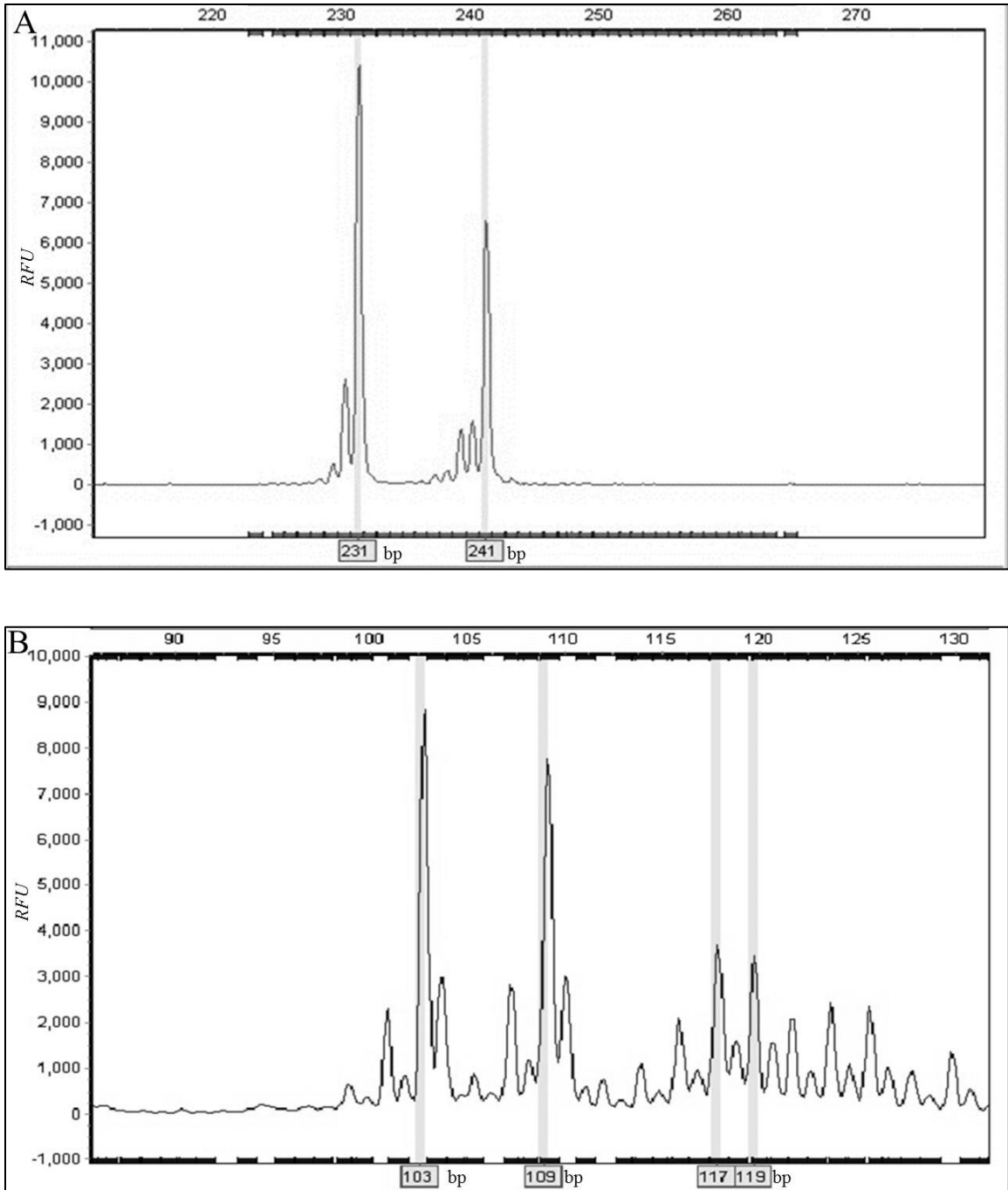


Figure 2.3 Correlograms of kinship coefficient measures (F_{ij}) plotted against distance (m).



Supplementary figure 2.1 Electropherograms showing marker amplification pattern (RFU: Relative frequency units, bp: base pairs), **(A)** Marker Am116 with co-dominant amplification pattern, **(B)** marker AS127 with multi-banding pattern.

Appendix 1: Open-Access Licence

Permission to print research paper in the PhD thesis Inbox x

Sudhir Khodwekar <sdkhodwe@mtu.edu> Feb 16 ☆

to marius.teodosiu ▾

Respected Editors,
 My name is Sudhir Khodwekar, a PhD student in Michigan Technological University, Houghton, Michigan, USA.
 In 2015 I have published a research paper 'Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.)' in your journal Annals of Forest Research as a first author. Dr. Oliver Gailing is corresponding author for the paper (Online First: April 15, 2015 | DOI: [10.15287/afr.2015.360](https://doi.org/10.15287/afr.2015.360)). I would like to reprint it in my PhD thesis, so here I would like to request you for the permission. I will submit my thesis to my committee members in first week of March and defend in third week of March. Awaiting for your kind reply.

Sincerely,
 ...

Marius Teodosiu <marius.teodosiu@gmail.com> Feb 16 ☆

to me ▾

Dear Sudhir Khodwekar,
 Please feel free to use it as you wish (more details here about our [licence](#)).
 Good luck on defending your Ph.d. and thank you for choosing AFR to disseminate your research.

Kind regards,
 Marius Teodosiu
 --
 Executive Editor Ann. For. Res.

Open-Access Licence

http://afr/pages/view/OAllicence

Jawaharlal Nehru Bo... Georg-August-Univer... Web Soil Survey Conservation and use Lect 4. Heterozygosity Measuring genetic di... Careers - ArborGen... High Rates of Gene F...

ANNALS OF FOREST RESEARCH

Forestry & Environmental Sciences
 Peer-Reviewed Open Access Journal

Search

Advanced Search

Articles | Browse | For Authors | About Us | Contact | Announcements

Journal Information | Indexing | Journal Metrics

0.741 (IF)
 0.700 (SNIP)
 0.338 (SJR)

Open-Access Licence

All the papers published in Annals of Forest Research are available under an open access policy (Gratis Gold Open Access Licence), which guaranty the free (of taxes) and unlimited access, for anyone, to entire content of the all published articles. The users are free to "read, copy, distribute, print, search or refers to the full text of these articles", as long they mention the source. Also, "users can use, reuse and build upon the material published in the journal but only for non-commercial purposes".

The other materials (texts, images, graphical elements presented on the Website) are protected by copyright.

The journal exerts a permanent quality check, based on an established protocol for publishing the manuscripts. The potential article to be published are evaluated (peer-review) by members of the Editorial Board or other collaborators with competences on the paper topics. The publishing of manuscript is free of charge, all the costs being supported by Forest Research and Management Institute.

More details about Open Access:
 Wikipedia: http://en.wikipedia.org/wiki/Open_access
 DOAJ: <http://www.doaj.org/oainfo>

Chapter 3:

Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations

Sudhir Khodwekar and Oliver Gailing*

School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI, 49931-1295, USA

*: Correspondence: ogailing@mtu.edu; Tel.: +1-906-487-1615

¹The modified and shortened version of the material in this chapter is submitted to a peer-reviewed international journal 'American Journal of Botany' and is under revision.

PREMISE OF THE STUDY: Introgression of adaptive alleles between hybridizing species is likely an important mechanism to generate new genetic variation for the adaptation to rapidly changing environmental conditions. Oaks provide a model for the study of adaptive gene introgression, since environmental selection maintains high interspecific differentiation at a few outlier loci and species-specific adaptations despite recurrent interspecific gene flow. Previously we have identified a CONSTANS-like gene under strong divergent selection between the drought tolerant *Quercus ellipsoidalis* and the drought averse *Quercus rubra*. Both species were nearly fixed on alternative alleles in neighboring (parapatric) populations, *Q. ellipsoidalis* on allele 138 and *Q. rubra* on allele 141.

METHODS: In this study we analyzed introgression of these outlier alleles in sympatric and in parapatric *Q. rubra/Q. ellipsoidalis* populations.

KEY RESULTS: We found evidence that environmental selection affected the level and direction of outlier allele introgression. Thus, in the face of symmetric interspecific gene flow, outlier allele introgression was asymmetric and introgression of allele 138 into *Q. rubra* was consistently higher than introgression of allele 141 into *Q. ellipsoidalis* in sympatric populations, but the opposite pattern was found in parapatric populations. Furthermore, the rate of introgression of outlier alleles between species in sympatric stands was related to soil quality (i.e. soil water holding capacity, nutrient availability).

CONCLUSIONS: This pattern suggests that introgression of adaptive genes between these two red oak species is strongly affected by environmental selection, and is an

important mechanism for species' adaptation to changing environmental conditions in a changing climate.

KEY WORDS: Outlier genes, *Quercus rubra*, *Quercus ellipsoidalis*, CONSTANS-like, EST-SSRs

Introduction

Global warming is expected to result in changes in tree species distribution and in an increase of biotic (e.g. insect pests) and abiotic (e.g. drought) stressors, especially in trailing (rear) edges of the species' range (Hampe and Petit, 2005; Levin, 2012; Aitken and Bemmels, 2016). Adaptation of long-lived forest trees to these stressful conditions is largely dependent on standing genetic variation within and among populations (Krutovsky and Neale, 2005; Barrett and Schluter, 2008). Since new advantageous mutations are rare events, they are unlikely to generate enough new genetic variation for local adaptation (adaptive genetic variation) to cope with rapidly changing climatic conditions (Barrett and Schluter, 2008; Orr and Unckless, 2014; Hamilton and Miller, 2016). Introgression of adaptive alleles and traits as result of interspecific hybridization has been largely neglected as a mechanism to increase adaptive genetic variation within forest tree species for the adaptation to climate change (Aitken et al., 2008).

On the other hand, natural hybridization has been common in plant species and played an important role in the evolution of higher plants and in adaptation processes (Ellstrand, Whitkus, and Rieseberg, 1996; Barton, 2001; Rieseberg et al., 2003; Aitken et al., 2008; Abbott et al., 2013). Furthermore, a survey of natural hybridization in vascular plants based on the Flora of the British Isles (Stace, 1975, 1997) revealed that at least 25% of species hybridize with other species (Mallet, 2005). Interspecific gene flow has been described as a mechanism for the transfer of adaptive alleles between interfertile species providing a source of adaptive genetic variation (Arnold, 2004, 2006; Abbott et al., 2013; Arnold, 2016a), and for the invasion of late-successional species into the range of pioneer

species (Petit et al., 2003). The frequency of interspecific hybridization implicates that introgression produces more new genetic variants as raw material for selection than mutations (Anderson, 1949; Abbott et al., 2013). Furthermore the high frequency of hybridization in island floras suggests that hybridization and adaptive introgression maintains adaptability in geographically isolated populations (Carlquist, 1974).

Climate change will likely increase the frequency of interspecific hybridization and the potential of adaptive trait transfer as result of differential shifts of species distribution ranges (Levin, 2012; Gailing and Curtu, 2014), and the generation of new habitats for natural hybrids (Anderson, 1948; Grant, 1971; Arnold, 2016b; Hamilton and Miller, 2016; Miller and Hamilton, 2016). Thus, for the conservation of forests in the face of climate change it is mandatory to gain a better understanding of the frequency of interspecific hybridization and its importance for the transfer of adaptive genes (alleles) and traits.

Due to recurrent interspecific gene flow, oaks provide a model to analyze species coherence in the face of interspecific gene flow and to identify adaptive genes under strong divergent selection (outlier loci) that resist the homogenizing effect of gene flow (Scotti-Saintagne et al., 2004; Goicoechea, Petit, and Kremer, 2012; Lind-Riehl, Sullivan, and Gailing, 2014; Goicoechea et al., 2015). For example, genome scans between the interfertile but ecologically divergent white oak species, *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl., revealed genome regions with high interspecific differentiation (outlier regions) surrounded by regions with low interspecific differentiation (Scotti-

Saintagne et al., 2004), a pattern predicted by models of ecological speciation in the face of interspecific gene flow and strong divergent selection (Via, 2009, 2012).

The North American interfertile red oak species, *Quercus rubra* L. and *Quercus ellipsoidalis* E.J. Hill, are constrained to different soil types where they co-occur and reveal different adaptations to drought such as differences in tissue elasticity, root depth and leaf conductance (Abrams, 1992). *Quercus rubra* is a mesophytic species while *Q. ellipsoidalis* can grow on dry outwash plains (Abrams, 1990, 1992). They are the only red oak species in their northern distribution range (Little, 1971). Despite recurrent and symmetric interspecific gene flow in contact zones between both species (Lind and Gailing, 2013; Owusu et al., 2015; Lind-Riehl and Gailing, 2017), they remain morphologically and ecologically distinct, and hybrids and introgressive forms in adult trees are relatively rare in neighboring (parapatric) populations (Lind and Gailing, 2013; Collins, Sullivan, and Gailing, 2015). Interspecific gene flow was reflected in low interspecific genetic differentiation at most nuclear and genic microsatellites and sharing of regional chloroplast haplotypes between species in sympatric and parapatric populations (Lind and Gailing, 2013; Zhang, Hipp, and Gailing, 2015).

An outlier screen using 36 genic microsatellites (Expressed Sequence Tag- Simple Sequence Repeats, EST-SSRs) and eight non-genic nuclear SSRs (nSSRs) in four *Q. rubra*/*Q. ellipsoidalis* population pairs from three geographic regions identified one outlier EST-SSR, FIR013, under strong divergent selection in all population pairs and regions (Lind-Riehl, Sullivan, and Gailing, 2014). The microsatellite was nearly fixed on two different alleles in both species, allele 138 in *Q. ellipsoidalis* and allele 141 in *Q.*

rubra. This outlier locus with very high interspecific differentiation can be used as diagnostic marker to follow introgression of outlier alleles between species (Arnold, 2006; Lind-Riehl and Gailing, 2017). The EST, in which FIR013 is located, was annotated as CONSTANS-like (*COL*) (Lind-Riehl, Sullivan, and Gailing, 2014), a circadian clock gene with putative function in flowering time and growth (Putterill et al., 1995; Yano et al., 2000; Holefors et al., 2009; Amasino and Michaels, 2010; Lind-Riehl, Sullivan, and Gailing, 2014). Sequencing of both alleles in *Q. rubra* and in the drought tolerant clade, *Q. ellipsoidalis*, *Quercus velutina* Lam. and *Quercus coccinea* Münchh., revealed that the trinucleotide SSR encodes a poly (Q) repeat (poly-glutamine repeat) and that *Q. ellipsoidalis* and the other more drought tolerant species, *Q. velutina* and *Q. coccinea*, are missing one repeat unit (Lind-Riehl and Gailing, 2017). In the European white oak species *Q. petraea*, a single nucleotide polymorphism (SNP) located in the same *COL* gene was associated with vegetative bud burst (Alberto et al., 2013), a trait directly correlated with flowering time in oaks (Chesnoiu et al., 2009; Collins, Sullivan, and Gailing, 2015). Also, in *Populus tremula* L., variation in a poly (E) repeat in *COL2* was related to growth cessation (Ma et al., 2010) which is associated with leaf senescence (Weih, 2009). Variation in poly (Q) repeat number in other circadian clock genes was found to be under selection in several animal and plant species (Faux et al., 2005; Haerty and Golding, 2010). Signatures of strong divergent selection for poly (Q) repeat number in *COL* between *Q. rubra* and *Q. ellipsoidalis* make this gene a prime candidate for reproductive isolation and adaptive divergence between both species.

Characterization of the outlier locus *COL* in neighboring (parapatric) *Q. rubra*/*Q. ellipsoidalis* populations from three regions (Lind-Riehl, Sullivan, and Gailing, 2014; Lind-Riehl and Gailing, 2017) showed signatures of asymmetric introgression of outlier alleles; introgression of the 141 allele (*Q. rubra*-allele) into *Q. ellipsoidalis* was consistently higher than introgression of the 138 allele into *Q. rubra* (Lind-Riehl and Gailing 2017). In the only sympatric *Q. rubra*/*Q. ellipsoidalis*/*Q. velutina* population an opposite pattern was observed, introgression of the 138 allele into *Q. rubra* was higher than introgression of the 141 allele into *Q. ellipsoidalis*. Since we found evidence for symmetric gene flow between species (Lind and Gailing, 2013; Lind-Riehl and Gailing, 2017), different directions of asymmetric introgression of outlier alleles in sympatric and parapatric populations can be interpreted as result of different selection regimes.

The objective of this study is to characterize introgression of outlier alleles at FIR013 in two additional sympatric *Q. rubra*/*Q. ellipsoidalis* stands, outside the distribution range of other hybridizing red oak species, in comparison to parapatric stands.

Our overarching hypothesis is that environmental selection affects the level of introgression of outlier alleles between species. Specifically, we hypothesize that (1) the level of introgression of outlier alleles differs in sympatric and parapatric populations as result of different selection regimes potentially exerted by the interspecific competition in sympatric stands and that (2) the frequency of outlier allele introgression is related to soil water holding capacity and fertility in sympatric stands. To separate effects of interspecific gene flow and selection on outlier allele introgression, we estimate the level and direction of interspecific gene flow using genetic assignment and parentage analyses.

MATERIALS AND METHODS

Plant material

Sampling of adult trees (>7 cm diameter at breast height) of northern red oak (*Quercus rubra*) and northern pin oak (*Quercus ellipsoidalis*) from two sympatric natural oak populations in the contact zones between species was performed along a soil moisture gradient (Figure 2.2, 2.3). The populations are located in the western Upper Peninsula of Michigan, Baraga Plains, FC (Ford center, Baraga county) (46°67'67.46''N, 88°53'49.41''W) and Escanaba region, ES (Delta county) (45°83'07.46''N, 86°35'16.41''W) (Table 2.1). The average annual precipitation in the Baraga Plains region is 938 mm including 3472 mm of average winter snowfall, with mean average summer (April to October) and winter (November to March) temperatures of 12.27°C and -7.34° C, respectively (Soil Survey of Baraga County, MI, 1988). The mean annual precipitation in the Escanaba region is 800 mm including 1628 mm of winter snowfall, with mean average summer (April to October) and winter (November to March) temperatures of 12.8 °C and -3.7 °C, respectively (Soil Survey of Delta County, MI, 1977). From each sympatric population leaf samples were collected from 96 adult trees. Additionally, a total of 466 acorns were collected from 15 seed producing adult trees from both regions (Table 2.2). Initial species assignment in the field was based on leaf morphological traits and stem morphology, e.g. *Q. ellipsoidalis* often has low hanging dead branches. Genetic assignment analysis was used for the final identification of species, hybrids and introgressive forms (see below). GPS locations for all trees were recorded using a Garmin (eTrex H) GPS device, and diameter at breast height (DBH) was

measured using a diameter tape. One to two healthy leaves from individual adult trees were sampled and stored in paper bags filled with silica gel for fast moisture absorption to keep DNA intact in the leaf tissues. Data on water holding capacity (amount of water content, awc), site index, soil composition (% organic matter, % sand) and pH were retrieved from the USDA soil surveys of the Baraga and Delta county areas, Michigan (<http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx>). In both sympatric stands soils were ranked from 1 (rich) to 6 based on these soil properties. As reference, we included two populations of each species from the Baraga Plains region and one population from each species from the Chequamegon National Forest and the Nicolet National Forest (parapatric populations) (Table 2.1) that were characterized by Lind-Riehl and Gailing (2017).

DNA isolation

Total genomic DNA (~20 ng) was extracted from dried leaf tissues of adult trees and from the embryos of acorns collected from the individual adult trees. Approximately 1 cm² dried leaf tissue and 1/8th of the size of the mature embryo was used for DNA extraction. DNA from dried leaf tissues was isolated with the DNeasy96 Plant Kit and DNA from acorns was isolated with the DNeasy Plant Mini Kit following the standard protocols (Qiagen, Hilden, Germany) and stored at -20 °C. DNA quality was checked on 0.8% agarose gels stained with GelRed Nucleic Acid stain (Phenix Research Products) relative to the DNAL-1000-100 DNA ladder (Life Technologies).

Genetic markers

A total of six nSSRs (Aldrich et al., 2002; Sullivan et al., 2013) and 24 EST-SSRs originally developed for *Q. robur* (Durand et al., 2010) were selected and tested initially for amplification and polymorphism on eight DNA samples from each population. Finally, 24 markers (5 nSSRs and 19 EST-SSRs) were selected for amplification in all the samples based on clear amplification patterns and amplification of a single polymorphic locus. The selected nSSRs and EST-SSRs cover 10 of the 12 linkage groups of oak (n=12) (Supplementary Table 3.1).

PCR amplification

A PCR protocol modified from (Owusu et al., 2015) was used for PCR amplification. We grouped markers for pre-PCR multiplexes based on fragment size and color of the fluorescence dye attached (Supplementary Table 3.2). The 10 µl reaction contained 2 µl of 5× HOT FIREPol® Blend Master Mix Ready to Load (contains 10 mM MgCl₂, 0.6 units of HOT FIREPol® Taq polymerase, and 2 mM dNTPs) (Solis BioDyne, Estonia), 1.34 µl of each 5 µM forward and reverse primers (Sigma Aldrich, St. Louis, MO and Applied Biosystems, Foster City, CA), 4 µl double deionized water (DNase and RNase free) and 1.34 µl DNA (~1.8 ng/µl). Forward primers were labelled with fluorescent dyes (6-FAM, VIC, NED, PET) (Supplementary Table 3.1). Amplification was done in a Peltier Thermal Cycler (GeneAmp® PCR system 2720, Applied Biosystems). The PCR profile consisted of 15 min denaturation at 95°C, 35 cycles denaturation at 94°C for 45s, annealing at T_a for 45s (Supplementary Table 2.1), 45s elongation at 72°C, and an

extension step at 72°C for 20 min. Fragment sizes were determined on a capillary electrophoresis system (Applied Biosystems Genetic Analyzer ABI 3730 at Biotechnology Resource Center, Cornell University) with the internal size standard Gene-Scan™ LIZ-500 (Applied Biosystems). Alleles were scored with GeneMarker® V2.6.3 (SoftGenetics) after careful visual inspection and editing.

Genetic diversity analyses

Genetic variation within populations was characterized in GenAlEx 6.502 (Peakall and Smouse, 2012) as number of different alleles per locus (N_a), observed heterozygosity (H_o) and expected heterozygosity (H_e) per sympatric population and separated by genetically assigned species. The inbreeding coefficient (F) was calculated for each species in all populations. Significant deviation from Hardy-Weinberg proportions was tested in GenePop 4.1 using probability tests for each marker and population separated by species (Raymond and Rousset, 1995). Pairwise genetic differentiation between populations (F_{ST}) within and between species with corresponding significances was calculated in GenePop 4.1 with the exact G test using default settings (Raymond and Rousset, 1995).

Genetic assignment analysis and introgression of outlier alleles

Multi-locus genotype data were used to investigate population structure and correspondence with morphological species assignment in STRUCTURE 2.3.4 (Pritchard, Stephens, and Donnelly, 2000). STRUCTURE implements a model-based genetic admixture analysis using Bayesian methods to assign individuals to species and detect putative hybrids. This method requires the markers to be in linkage equilibrium

with one another within populations. Thus, linkage disequilibrium for all pairs of loci was tested using Fisher's exact test with the program GenePop for genetically assigned species (Raymond and Rousset, 1995). For comparability we used the same parameters as described in Lind and Gailing (2013) and Collins et al. (2015) for the genetic assignment analysis. Specifically, to determine the number of clusters K to best fit our data, we performed five independent runs with a burn-in period of 30,000 iterations followed by 10^6 iterations for each value of K (K=1-4) under the admixture model without *a priori* information regarding species identity. Pure species were defined as having a proportion of ancestry of ≥ 0.9 in one cluster. Hybrids were defined as having a proportion of ancestry between 0.4 and 0.6 and introgressive forms were defined as having a proportion of ancestry of 0.61 to 0.89 in one of the clusters (Lind and Gailing, 2013).

Introgression of species-specific (outlier) alleles 138 and 141 between species was estimated in each population as the relative frequency of the *Q. ellipsoidalis*-specific allele 138 in *Q. rubra* samples and as the frequency of the *Q. rubra*-specific allele 141 in *Q. ellipsoidalis* samples. Also, the relative frequency of homozygotes for the introgressed alleles (i.e. 138/138 in *Q. rubra* and 141/141 in *Q. ellipsoidalis*) was calculated for each population and species. Since introgression of outlier alleles is the result of interspecific hybridization and of repeated backcrosses of hybrids with one species (introgression), the frequency of *Q. rubra*-like and *Q. ellipsoidalis*-like introgressive forms was calculated as an estimate of recurrent interspecific gene flow. To adjust for the effect of different levels of interspecific gene flow, preferential introgression of outlier alleles was estimated as the ratio of the frequency of outlier alleles in one species (e.g. as the relative frequency of

allele 138 in *Q. rubra*) and the frequency of introgressive forms (e.g. the frequency of *Q. rubra*-like introgressive forms with *Q. ellipsoidalis*). Data on the frequency of hybrids/introgressive forms and introgression of outlier alleles in parapatric populations were taken from Lind-Riehl and Gailing (2017).

The number of copies of introgressed allele 138 in *Q. rubra* individuals (e.g., 0 (141/141), 1 (138/141), 2 (138, 138)) and of introgressed allele 141 in *Q. ellipsoidalis* was plotted against soil quality (sorted from rich to poor) using linear regression analysis in WINSTAT (Fitch, 2006). Likewise, the *Q. rubra* ancestry coefficient as determined by genetic assignment analysis was plotted against soil quality.

Outlier screening

Outlier screens between species were performed using the LOSITAN-Selection Workbench (Antao et al., 2008) in the two sympatric populations. LOSITAN uses the F_{dist} approach to calculate genetic differentiation between populations (F_{ST}) weighted by the heterozygosity between populations in an island model of migration (Antao et al., 2008). For the calculation of the neutral envelope 100,000 simulations were run at the 95% and 99% confidence levels and a false discount rate of 0.05 and 0.01 under the stepwise mutation model using both the ‘neutral’ mean F_{ST} and the “force” mean F_{ST} option. The ‘neutral’ mean F_{ST} option is used in a first simulation run to remove loci under selection for the computation of the neutral mean F_{ST} . The “force” mean F_{ST} option is used to obtain a precise F_{ST} estimate by running a bisection algorithm over repeated simulations.

Gene flow

Paternity analysis was carried out using the computer program CERVUS 3.0.7 (Marshall et al., 1998). Log-likelihood ratios (LOD scores) were calculated at a strict (95%) and relaxed (80%) confidence level for the paternity assignment on the basis of observed multilocus genotypes. The simulation parameters used for paternity analysis were as follows: error rate used in likelihood calculations: 0.010; minimum number of typed loci: 12; mistyped loci: 1%; proportions of loci typed: 99%; proportion of candidate fathers sampled: 25%; number of candidate parents: 400; number of cycles: 100,000.

RESULTS

Genetic diversity

Mean genetic variation (H_e , H_o and N_a) was slightly higher in *Q. rubra* than in *Q. ellipsoidalis* for both geographic locations, but differences were not statistically significant (paired t-test, $p = 0.396$) (Table 3.3). For each species and region mean genetic variation was higher at genomic SSRs than at EST-SSRs (Supplementary Tables 2.3, 2.4, 2.5, 2.6, 2.7).

Genetic structure assignment

In accordance with the observation of two morphological species, genetic assignment analysis performed in STRUCTURE 2.3.4 designated $K=2$ as the most likely number of genetic clusters according to the log likelihood of the multilocus genotypes $\Pr(X/K)$. Initial morphological species assignment and genetic species assignment did not correspond for 10% of the individuals in the Baraga Plains and for 7% in the Escanaba region (Figure 2.4). Genetic differentiation between species (ancestry coefficient $Q \geq 0.9$) was higher (0.067-0.071) between populations of different species than between populations from different regions of the same species (0.012-0.027, Supplementary table 2.8). Similar frequencies of *Q. rubra*-like (8.92-12.00%) and *Q. ellipsoidalis*-like (8.10-16.66%) introgressive forms and a low frequency of F_1 hybrids (2.10-4.16%) were observed in sympatric populations of both regions indicating moderate and symmetric gene flow between species (Table 3.4). Also, the frequency of introgressive forms and hybrids suggests symmetric gene flow between parapatric *Q. rubra/Q. ellipsoidalis*

populations (Table 3.4, data from Lind-Riehl et al. 2014, Lind-Riehl and Gailing 2017). As compared to the sympatric population (Baraga plains, FC) the frequency of introgressive forms was slightly higher in both parapatric *Q. ellipsoidalis* and in one of the two parapatric *Q. rubra* populations from the Baraga Plains regions. No F₁ hybrids were reported in the parapatric *Q. rubra* populations (FC-A, FC-B) and low frequencies (1.00-2.00%) were found in the parapatric *Q. ellipsoidalis* populations (FC-C, FC-C). Relatively high numbers of introgressive forms and F₁ hybrids were found in the other parapatric populations of NN and CN (Table 3.4). The highest percentage of F₁ hybrids was found in CN-QR and CN-QE (10.00% each).

Outlier screening

The genic microsatellite FIR013 had two major alleles (138, 141). In both geographic regions the 138 allele was predominant in *Q. ellipsoidalis* (75% in the Baraga Plains region and 80% in Escanaba) and the 141 allele was predominant in *Q. rubra* (70% in the Baraga Plains region and 78% in Escanaba). These differences in allele frequencies between species are reflected in much higher interspecific differentiation at FIR013 than at the remaining neutral markers (Baraga Plains: $F_{ST} = 0.38$ as compared to a mean neutral $F_{ST} = 0.05$; Escanaba: $F_{ST} = 0.51$ as compared to a mean neutral $F_{ST} = 0.08$). As a result, outlier screening between species identified FIR013 as outlier under divergent selection above the 99% confidence interval in the Baraga Plains region at a false discovery rate of 0.01 and above the 95% confidence interval in the Escanaba region at a false discovery rate of 0.05 (Suppl. Figure 3.1).

Introgression of outlier alleles

Even though the frequency of hybrids and introgressive forms indicated symmetric interspecific gene flow, introgression of outlier allele 138 into *Q. rubra* (FC: 27.00%, ES: 20.53%) was more frequent than introgression of outlier allele 141 into *Q. ellipsoidalis* (FC: 13.09%, ES: 14.86%) in sympatric populations (Table 3.4). In contrast, in parapatric populations the opposite pattern was observed, introgression of the 141 allele into *Q. ellipsoidalis* was much higher than introgression of the 138 allele into *Q. rubra* in each of the three geographic regions (Table 3.4, Lind-Riehl and Gailing 2017). In the Baraga Plains and Escanaba regions homozygotes for the 138 allele were generally rare with the exception of population FC (16.66%) and mainly found in sympatric populations. Homozygotes for the 141 allele were largely absent (2.70% in ES) in these regions, but frequent in NN-QE (15.38%) and in CN-QE (20.00%). Introgression of outlier alleles between species was highest in the CN and NN regions.

Ancestry and introgression of outlier alleles is related to soil properties

Quercus rubra is predominant on soils with comparatively high water holding capacity, while *Q. ellipsoidalis* is predominant on nutrient poor and excessively drained soils with very low water holding capacity (Grayling sand) (Table 3.1). In both sympatric stands, species occur next to each other on Rubicon sand while *Q. rubra* is dominant on adjacent richer soils (e.g. Keweenaw Kalkaska complex, Croswell sand) and *Q. ellipsoidalis* is dominant on adjacent poorer soil (Grayling sand) (Figure 3.2). Also, the distribution of hybrids and introgressive forms is strongly constrained by soil types in both regions. In

the Baraga Plains region F₁ hybrids and introgressive forms are largely restricted to Rubicon sand and *Quercus ellipsoidalis*-like introgressive forms and F₁ hybrids are only found on Rubicon sand. In the Escanaba region, hybrids and introgressive forms are largely restricted to Grayling sand and Rubicon sand. Only two *Q. rubra*-like introgressive forms are found on Au Gres sand, a richer soil with higher water holding capacity (awc values in Table 3.1). *Quercus ellipsoidalis*-like introgressive forms are only found on Grayling sand and both F₁ hybrids are found in the contact zone between both Grayling and Rubicon sand. Overall, ancestry in *Q. rubra* as determined by genetic assignment analysis significantly increased from poor to rich soil in the Baraga Plains region ($R^2 = 0.146$, $P = 0.00012$) and in the Escanaba region ($R^2 = 0.409$, $P < 0.00001$) (suppl. Figure 3.3, 3.4).

Also, introgression of outlier alleles between species seems to be restricted by water availability (water holding capacity of different soil types) in the Baraga Plains region. Thus, the frequency of allele 138 in *Q. rubra* increased from rich to poor soil (0%, 15%, 24.19%) and the frequency of allele 141 increased from poor to rich soil in *Q. ellipsoidalis* (12.16%, 50%, 50%). Consequently, the copy number of introgressed allele 138 per individual tree was significantly associated with the occurrence of *Q. rubra* on poorer soils ($R^2 = 0.091$, $P = 0.037$) and allele 141 was significantly associated with the occurrence of *Q. ellipsoidalis* on richer soils ($R^2 = 0.134$, $P = 0.0199$) (suppl. Figure 3.5, 3.6). No such association was found in the Escanaba region (suppl. Figure 3.7, 3.8).

Gene flow analysis

A total of 466 seeds from 15 seed parents and all 192 adult trees in both regions were characterized at the SSR markers for parentage analyses. Genetic assignment analysis identified 14 seed parents as *Q. ellipsoidalis* or *Q. ellipsoidalis*-like introgressive forms, and one seed parent as F₁ hybrid (Table 3.2). The absence of *Q. rubra* as seed parent was the result of differences between initial morphological and genetic species assignment and of irregular fruiting. Out of the 466 seeds, 339 seeds (73%) were sired by external unidentified pollen parents, for the remaining 127 seeds (27%) the pollen parent was identified within the study plots. Out of 127 seeds that were collected from *Q. ellipsoidalis* or *Q. ellipsoidalis*-like introgressive forms in both regions and for which pollen parents were identified, only five seeds from three seed parents of the Escanaba region were derived from interspecific crosses (6.7% in the Escanaba region, 3.9% for all seeds). Interspecific gene flow for *Q. ellipsoidalis* and *Q. ellipsoidalis*-like introgressive forms ranged from 0% to 25%.

DISCUSSION

Evidence for symmetric interspecific gene flow

We hypothesize that introgression of outlier alleles at *COL* is affected by environmental selection. Outlier screens in neighboring (parapatric) *Q. rubra*/*Q. ellipsoidalis* populations on contrasting soil types suggested that strong environmental selection maintains species identity and very high frequency differences at outlier EST-SSR FIR013, which encodes for a poly (Q) repeat in a *COL* gene (Lind-Riehl, Sullivan, and Gailing, 2014). If high interspecific differentiation at outlier alleles is the result of environmental selection we also expect that introgression of outlier alleles between species is related to environmental conditions such as soil water availability in the face of interspecific gene flow.

Low interspecific differentiation at other genic and nuclear microsatellites (Lind-Riehl and Gailing, 2014) and sharing of geographically restricted chloroplast haplotypes between species (Zhang, Hipp, and Gailing, 2015) were indicative of recurrent interspecific gene flow. We recognize that low interspecific differentiation could also reflect shared ancestral variation between species (Muir and Schlötterer, 2005). However, sharing of geographically restricted chloroplast haplotypes between species and the location of genetically intermediate individuals (putative F₁ hybrids) in contact zones between species gives strong evidence for recurrent interspecific gene flow (Collins, Sullivan, and Gailing, 2015; Owusu et al., 2015). The present study provides direct evidence for recurrent interspecific gene flow. Thus, parentage analysis of 466 seeds

collected from 15 genetically assigned seed parents in sympatric stands revealed that 3.9% of the seedlings (0% to 25 % for individual seed parents) were derived from interspecific matings (Table 3.2). Similar frequencies of *Q. rubra*-like and *Q. ellipsoidalis*-like introgressive forms in each region in both parapatric (Lind and Gailing, 2013; Collins, Sullivan, and Gailing, 2015) and sympatric populations (this study) indicated symmetric gene flow between species. Similarly, symmetric gene flow between *Q. rubra* and *Q. ellipsoidalis* was observed across the Great Lakes region based on genetic assignment analysis of 103 *Q. ellipsoidalis* individuals from nine populations and 72 *Q. rubra* individuals from eight populations (Owusu et al., 2015). If there was no selection on outlier alleles, we would expect that introgression of outlier allele 138 into *Q. rubra* is as frequent as introgression of outlier allele 141 into *Q. ellipsoidalis* in both sympatric and parapatric populations and proportional to interspecific gene flow.

Species are stratified along soil water/nutrient gradients

The spatial distribution of related hybridizing oak species in sympatric stands is often related to different soil/environmental preferences (Curtu, Gailing, and Finkeldey, 2007; de Heredia, Valbuena-Carabaña, and Gil, 2009; Owusu et al., 2015). When we ranked soil types according to nutrient richness and water holding capacity, we found a significant association between soil quality and the occurrence of both species. Thus ancestry in *Q. rubra* as identified by genetic assignment analysis was significantly associated with richer soils in both geographic regions. This observation provides strong empiric evidence for environmental selection maintaining species differences in the face of interspecific gene flow.

Different levels of introgression of outlier alleles in parapatric and sympatric stands suggest different selection regimes

Introgression of outlier alleles between species was asymmetric in all regions both in sympatric and parapatric populations. Since interspecific gene flow was mostly symmetric between species based on genetic assignment analysis, asymmetric introgression of outlier alleles is likely the result of environmental selection. Even if we scale introgression of outlier alleles by interspecific gene flow estimated as the frequency of introgressive forms in each species, preferential introgression of outlier alleles is observed in each region for sympatric and parapatric populations. Also, the different direction of outlier allele introgression in sympatric and parapatric populations is strong evidence for different selection regimes on outlier alleles. In both sympatric populations (FC, ES) introgression of allele 138 into *Q. rubra* (27.00%, 20.53 %) is considerably higher than introgression of allele 141 into *Q. ellipsoidalis* (13.09%, 14.86%). In another sympatric population (Lind-Riehl and Gailing, 2017) the same pattern was observed, the introgression of allele 138 into *Q. rubra* (19%) was higher than introgression of allele 141 into *Q. ellipsoidalis* (5%). In contrast, in parapatric populations from each region the introgression of allele 141 into *Q. ellipsoidalis* was higher than introgression of allele 138 into *Q. rubra*. When we adjusted introgression of outlier alleles by intraspecific gene flow (ratio of outlier allele frequency and frequency of introgressive forms), preferential introgression of allele 138 into *Q. rubra* in sympatric populations (ratio of 3 and 2.30) as compared to parapatric populations (ratio of 0.25 to 1.25) was evident. In contrast, the ratio of 141 allele introgression into *Q. ellipsoidalis* relative to the frequency of *Q.*

ellipsoidalis-like introgressive forms showed no clear-cut difference between sympatric (0.79, 2.17) and parapatric populations (0.73 to 1.46). This pattern suggests that introgression of allele 138 into *Q. rubra* is favored by natural selection in sympatric populations and selected against in parapatric populations. Since individuals of both species occur next to each other in sympatric stands in intermediate environments with regard to water availability competition between species likely affects the strength of selection on outlier alleles at *COL*. Selection in nutrient-poor and dry environments as in the contact zones between both species in sympatric stands is expected to be on traits that reduce water and nutrient losses, while selection in more nutrient-rich environments (*Q. rubra* environments) is expected to be on traits that maximize growth rate to compete for light (Aerts, 1999). Thus, introgression of *COL* outlier allele 138 or of physically linked genes into *Q. rubra* may confer higher water use efficiency and nutrient retention which is reflected in a slow growth rate. Slow growth is likely an adaptation to xeric and nutrient-poor environments (Aerts, 1999; Ouédraogo et al., 2013). *Quercus ellipsoidalis* seedlings showed a much slower growth and later bud burst and leaf fall than *Q. rubra* in a common garden seedling trial (Gailing, 2013).

CONSTANS-like genes play an important role in shoot growth and the photoperiodic control of flowering and growth cessation (Holefors et al., 2009). Association genetic studies in an altitudinal common garden in *Quercus petraea* identified a Single Nucleotide Polymorphism (SNP) associated with vegetative bud burst in the same *COL* gene (Alberto et al., 2013) which was identified as outlier locus between *Q. rubra* and *Q. ellipsoidalis* (Lind-Riehl, Sullivan, and Gailing, 2014). A Blast-search against the NCBI

Arabidopsis nucleotide database (Zhang et al., 2000) revealed the highest similarity to the *Arabidopsis thaliana* mRNA for the CONSTANS gene (Putterill et al., 1995) ($E = 1E-95$, 67% identity, accession X94937.1), a circadian clock gene of the flowering pathway whose expression accelerates flowering in response to long days by activation of the FLOWERING LOCUS T (Suarez-Lopez et al., 2001; Amasino and Michaels, 2010). A search against the NCBI nucleotide plant database (Zhang et al., 2000) revealed the highest similarity of *COL* to CONSTANS-like2 (*COL2*) of *Juglans regia* L. and *Prunus mume* Siebold & Zucc. and to the CONSTANS-like protein *CO2* in *Populus trichocarpa* Torr. & A.Gray. Overexpression of *CO1* and *CO2* in *Populus* spp. resulted in smaller trees, but had no effect on phenology (Hsu et al., 2012). In *Populus tremula*, variation in the poly (E) repeat microsatellite of *COL2B* was associated with growth cessation (Ma et al., 2010). Also, in *Picea abies* (L.) H. Karst. a decrease of transcript levels of *COL1* and *COL2* prior to growth cessation suggested a role of these genes in the photoperiodic control of shoot elongation (Holefors et al., 2009). Single Nucleotide Polymorphisms in another CONSTANS-like gene in *Medicago sativa* L. were associated with flowering date and stem height and explained up to 4.2% of the total phenotypic variation in an association mapping population (Herrmann et al., 2010). Thus, it is possible that outlier alleles at *COL* (or of physically linked genes) are involved in adaptive divergence and reproductive isolation between *Q. rubra* and *Q. ellipsoidalis* by affecting growth and phenology.

The frequency of outlier allele introgression is related to soil types

The direction and frequency of introgression in sympatric and parapatric stands suggested adaptive introgression of allele 138 into *Q. rubra* in sympatric stands. Indeed, the introgression of allele 138 into *Q. rubra* is much higher in sympatric stands on very well drained soils than in pure *Q. rubra* stands of the same region which grow on moderately well drained soils. The higher level of introgression of allele 138 in sympatric stands is not the result of higher effective interspecific gene flow in sympatric stands since the frequency of *Q. rubra*-like introgressive forms is similar or higher in parapatric than in sympatric stands. Also, *Q. rubra* individuals homozygous for the 138 allele are found in higher frequencies (16.00%) only on very well drained soils in the sympatric stand FC, but are absent (FC-A) or rare (2.5%, FC-B) in pure *Q. rubra* populations growing on moderately drained soils. In Escanaba, *Q. ellipsoidalis* is largely restricted to excessively drained soils (Grayling sand), the only three *Q. rubra* individuals on this soil are either homozygous or heterozygous for the 138 allele. The predominant occurrence of *Q. rubra* trees with introgressed 138 allele on poorer soils (Rubicon sand and transition zone between Rubicon sand and Keweenaw-Kalkaska complex) and the predominant occurrence of *Q. ellipsoidalis* trees with introgressed 141 allele on richer soils in the Baraga Plains sympatric stand gives further evidence for environment-dependent introgression of outlier alleles (see significant results of the regression analyses). Also the restriction of pure stands of both species to contrasting soil types and of hybrids and introgressive forms to the contact zones between both species on intermediate soils (e.g. Rubicon sand) suggests that water and nutrient availability are important selective forces

for the maintenance of species integrity and the introgression of potentially adaptive alleles between species.

CONCLUSIONS

Adaptive introgression between species can be an important mechanism to rapidly adapt to changing environments in the face of climate change. Asymmetric introgression of outlier alleles at *COL* and higher introgression of allele 138 into *Q. rubra* in sympatric stands than in parapatric stands, despite similar levels of interspecific gene flow, provided evidence of adaptive introgression. Introgression of outlier alleles from the drought tolerant *Q. ellipsoidalis* into the drought averse *Q. rubra* and vice versa occurs preferentially in intermediate environments in the contact zone between both species and was associated with soil quality. Assessment of growth and of drought related traits such as water use efficiency and nutrient retention in each species for different genotypes of *COL* (138/138, 138/141, 141/141) in a common garden setting could reveal whether introgression of outlier alleles is associated with drought tolerance and growth. To better evaluate the potential function of *COL* in adaptive divergence and reproductive isolation between oak species, transcript levels of *COL* should be analyzed by quantitative PCR for each genotype and species prior to leaf flush and growth cessation.

ACKNOWLEDGEMENTS

Funding was provided by the McIntire Stennis fund (USDA-NIFA #1001471).

Additional funding was provided by the Life Science and Technology Institute (LSTI) and Ecosystem Science Center (ESC) at Michigan Technological University. Sudhir Khodwekar has been supported by the Government of India “National Overseas Scholarship” and by finishing fellowships of the LSTI and the Graduate School of Michigan Technological University.

LITERATURE CITED

- ABBOTT, R., D. ALBACH, S. ANSELL, J. W. ARNTZEN, S. J. BAIRD, N. BIERNE, J. BOUGHMAN, et al. 2013. Hybridization and speciation. *J Evol Biol* 26: 229-246.
- ABRAMS, M. D. 1990. Adaptations and responses to drought in *Quercus* species of North America. *Tree Physiology* 7: 227-238.
- ABRAMS, M. D. 1992. Fire and the development of oak forests. *Bioscience* 42: 346-353.
- AERTS, R. 1999. Interspecific competition in natural plant communities: mechanisms, trade-offs and plant-soil feedbacks. *Journal of Experimental Botany* 50: 29-37.
- AITKEN, S. N., AND J. B. BEMMELS. 2016. Time to get moving: assisted gene flow of forest trees. *Evolutionary Applications* 9: 271-290.
- AITKEN, S. N., S. YEAMAN, J. A. HOLLIDAY, T. WANG, AND S. CURTIS-MCLANE. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95-111.
- ALBERTO, F. J., J. DERORY, C. BOURY, J. M. FRIGERIO, N. E. ZIMMERMANN, AND A. KREMER. 2013. Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195: 495-512.
- ALDRICH, P. R., C. H. MICHLER, W. SUN, AND J. ROMERO-SEVERSON. 2002. Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). *Molecular Ecology Notes* 2: 472-474.
- AMASINO, R. M., AND S. D. MICHAELS. 2010. The timing of flowering. *Plant Physiology* 154: 516-520.
- ANDERSON, E. 1948. Hybridization of the habitat. *Evolution* 1: 1-9.
- ANDERSON, E. 1949. Introgressive hybridization. John Wiley and Sons, New York.
- ANTAO, T., A. LOPES, R. J. LOPES, A. BEJA-PEREIRA, AND G. LUIKART. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9: 323.
- ARNOLD, M. L. 2004. Transfer and origin of adaptations through natural hybridization: Were Anderson and Stebbins right? *Plant Cell* 16: 562-570.
- ARNOLD, M. L. 2006. Evolution through genetic exchange. Oxford University Press. Oxford.
- ARNOLD, M. L. 2016a. Divergence with genetic exchange. Oxford University Press. Oxford.
- ARNOLD, M. L. 2016b. Anderson's and Stebbins' Prophecy Comes True: Genetic Exchange in Fluctuating Environments. *Systematic Botany* 41: 4-16.
- BARRETT, R. D., AND D. SCHLUTER. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol* 23: 38-44.
- BARTON, N. H. 2001. The role of hybridization in evolution. *Molecular Ecology* 10: 551-568.
- CARLQUIST, S. J. 1974. Island Biology, 512-513. Columbia University Press, New York.
- CHESNOIU, E. N., N. SOFLETEA, A. L. CURTU, A. TOADER, R. RADU, AND M. ENESCU. 2009. Bud burst and flowering phenology in a mixed oak forest from Eastern Romania. *Annals of Forest Research* 52: 199-206.

- COLLINS, E., A. R. SULLIVAN, AND O. GAILING. 2015. Limited effective gene flow between two interfertile red oak species. *Trees-Structure and Function* 29: 1135-1148.
- CURTU, A. L., O. GAILING, AND R. FINKELDEY. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evol Biol* 7: 218.
- DE HEREDIA, U. L., M. C. VALBUENA-CARABAÑA, M., AND L. GIL. 2009. Variation components in leaf morphology of recruits of two hybridising oaks [*Q. petraea* (Matt.) Liebl. and *Q. pyrenaica* Willd.] at small spatial scale. *European Journal of Forest Research* 128: 543-554.
- DURAND, J., C. BODÉNÈS, E. CHANCEREL, J.-M. FRIGERIO, G. VENDRAMIN, F. SEBASTIANI, A. BUONAMICI, et al. 2010. A fast and cost effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* 11: 570.
- ELLSTRAND, N., E., R. WHITKUS, AND L. RIESEBERG, H. 1996. Distribution of spontaneous hybrids. *Proceedings of the National Academy of Sciences of the United States of America* 93: 5090-5093.
- FAUX, N. G., S. P. BOTTOMLEY, A. M. LESK, J. A. IRVING, J. R. MORRISON, M. BANDA, GARCIA, DE LA., AND J. C. WHISSTOCK. 2005. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Research* 15: 537-551.
- FITCH, R. 2006. WinSTAT. The statistics add-in for Microsoft Excel. R. Fitch Software.
- GAILING, O. 2013. Differences in growth, survival and phenology in *Quercus rubra* and *Q. ellipsoidalis* seedlings. *Dendrobiology* 70: 73-81.
- GAILING, O., AND A. L. CURTU. 2014. Interspecific gene flow and maintenance of species integrity in oaks. *Annals of Forest Research* 57: 5-18.
- GOICOECHEA, P. G., R. J. PETIT, AND A. KREMER. 2012. Detecting the footprints of divergent selection in oaks with linked markers. *Heredity (Edinb)* 109: 361-371.
- GOICOECHEA, P. G., A. HERRAN, J. DURAND, C. BODENES, C. PLOMION, AND A. KREMER. 2015. A linkage disequilibrium perspective on the genetic mosaic of speciation in two hybridizing Mediterranean white oaks. *Heredity (Edinb)* 114: 373-386.
- GRANT, V. 1971. Plant speciation. 1 ed. Columbia University Press, New York.
- HAERTY, W., AND G. B. GOLDING. 2010. Low-complexity sequences and single amino acid repeats: not just "junk" peptide sequences. *Genome* 53: 753-762.
- HAMILTON, J. A., AND J. M. MILLER. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* 30: 33-41.
- HAMPE, A., AND R. J. PETIT. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* 8: 461-467.
- HERRMANN, D., P. BARRE, S. SANTONI, AND B. JULIER. 2010. Association of a CONSTANS-LIKE gene to flowering and height in autotetraploid alfalfa. *Theor Appl Genet* 121: 865-876.
- HOLEFORS, A., L. OPSETH, A. K. REE ROSNES, L. RIPEL, L. SNIPEN, C. G. FOSSDAL, AND J. E. OLSEN. 2009. Identification of PaCOL1 and PaCOL2, two CONSTANS-like genes showing decreased transcript levels preceding short day induced growth cessation in Norway spruce. *Plant Physiol Biochem* 47: 105-115.

- HSU, C. Y., J. P. ADAMS, K. NO, H. LIANG, R. MEILAN, O. PECHANOVA, A. BARAKAT, et al. 2012. Overexpression of CONSTANS homologs CO1 and CO2 fails to alter normal reproductive onset and fall bud set in woody perennial poplar. *PLoS One* 7: e45448.
- KRUTOVSKY, K. V., AND D. B. NEALE. 2005. Forest genomics and new molecular genetic approaches to measuring and conserving adaptive genetic diversity in forest trees 369-390. Arbora Publishers. Zvolen.
- LEVIN, D. A. 2012. Mating system shifts on the trailing edge. *Ann. Bot.* 109: 613-620.
- LIND-RIEHL, J., AND O. GAILING. 2014. Fine-scale spatial genetic structure of two red oak species, *Quercus rubra* and *Quercus ellipsoidalis*. *Plant Systematics and Evolution* 301: 1601-1612.
- LIND-RIEHL, J., AND O. GAILING. 2017. Adaptive variation and introgression of a CONSTANS-like gene in North American red oaks. *Forests* 8: 3.
- LIND-RIEHL, J. F., A. R. SULLIVAN, AND O. GAILING. 2014. Evidence for selection on a CONSTANS-like gene between two red oak species. *Ann. Bot.* 113: 967-975.
- LIND, J. F., AND O. GAILING. 2013. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* 9: 707-722.
- LITTLE, E. L. J. 1971. Atlas of United States trees, volume 1, conifers and important hardwoods. U.S. Department of Agriculture Miscellaneous Publication 1146.
- MA, X. F., D. HALL, K. R. ONGE, S. JANSSON, AND P. K. INGVARSSON. 2010. Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* 186: 1033-1044.
- MALLET, J. 2005. Hybridization as an invasion of the genome. *Trends Ecol Evol* 20: 229-237.
- MARSHALL, T. C., J. SLATE, L. E. B. KRUK, AND J. M. PEMBERTON. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.
- MILLER, J. M., AND J. A. HAMILTON. 2016. Interspecies hybridization in the conservation toolbox: response to Kovach et al. (2016). *Conservation Biology* 30: 431-433.
- MUIR, G., AND C. SCHLÖTTERER. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* 14: 549-561.
- ORR, H. A., AND R. L. UNCKLESS. 2014. The population genetics of evolutionary rescue. *PLoS Genet* 10: e1004551.
- OUÉDRAOGO, D.-Y., F. MORTIER, S. GOURLET-FLEURY, V. FREYCON, N. PICARD, AND M. TURNBULL. 2013. Slow-growing species cope best with drought: evidence from long-term measurements in a tropical semi-deciduous moist forest of Central Africa. *Journal of Ecology* 101: 1459-1470.
- OWUSU, S. A., A. R. SULLIVAN, J. A. WEBER, A. L. HIPPI, AND O. GAILING. 2015. Taxonomic relationships and gene flow in four North American *Quercus* species (section *Lobatae*). *Systematic Botany* 40: 510-521.
- PEAKALL, R., AND P. E. SMOUSE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537-2539.

- PETIT, R. J., C. BODÉNÈS, A. DUCOUSSO, G. ROUSSEL, AND A. KREMER. 2003. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 161: 151-164.
- PRITCHARD, J. K., M. STEPHENS, AND P. DONNELLY. 2000. Inference of population structure using multilocus genotypic data. *Genetics* 156: 945-969.
- PUTTERILL, J. R., F. LEE, K. R. SIMON, AND G. COUPLAND. 1995. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc-finger transcription factors. *Cell* 80: 847-857.
- RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *The Journal of Heredity* 86: 248-249.
- RIESEBERG, L. H., O. RAYMOND, D. M. ROSENTHAL, Z. LAI, K. LIVINGSTONE, T. NAKAZATO, J. L. DURPHY, et al. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211-1216.
- SCOTTI-SAINTAGNE, C., S. MARIETTE, I. PORTH, P. G. GOICOECHEA, T. BARRENECHE, C. BODENES, K. BURG, AND A. KREMER. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* 168: 1615-1626.
- STACE, C. A. 1975. Hybridization and the flora of British isles. Academic Press. London.
- STACE, C. A. 1997. The new flora of the British isles. Cambridge University Press. Cambridge.
- SUAREZ-LOPEZ, P., K. WHEATLEY, F. ROBSON, H. ONOUCHI, F. VALVERDE, AND G. COUPLAND. 2001. CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410: 1116-1120.
- SULLIVAN, A. R., J. F. LIND, T. S. MCCLEARY, J. ROMERO-SEVERSON, AND O. GAILING. 2013. Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. *Plant Molecular Biology Reporter* 31: 231-239.
- VIA, S. 2009. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America* 106: 9939-9946.
- VIA, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci* 367: 451-460.
- WEIH, M. 2009. Genetic and environmental variation in spring and autumn phenology of biomass willows (*Salix* spp.): effects on shoot growth and nitrogen economy. *Tree Physiology* 29: 1479-1490.
- YANO, M., Y. KATAYOSE, M. ASHIKARI, Y. UTAKO, L. MONNA, T. FUSE, T. BABA, et al. 2000. Hd1, a major photoperiod sensitivity quantitative trait locus in Rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* 12: 2473-2483.
- ZHANG, R., A. L. HIPPI, AND O. GAILING. 2015. Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations. *Botany* 93: 691-700.
- ZHANG, Z., S. SCHWARTZ, L. WAGNER, AND W. MILLER. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7: 203-214.

TABLE 3.1. Sampled *Quercus rubra* and *Q. ellipsoidalis* populations

Abbreviation	Region	Population	Species	Sample size (n)	Soil characteristics	Site Index	Latitude	Longitude	Mean altitude (m)
FC-QR/QE	FC/Baraga Plains: MI	Stand F	<i>Q. rubra</i>	96	1:Keweenaw-Kalkaska Complex; 1-8% slopes; 0.11awc	61 ^a	46°40'27"N	88°32'06"W	393
ES-QR/QE	Escanaba Plains: MI	Escanaba	<i>Q. ellipsoidalis</i>	95	5:Rubicon sand; 0-6% slopes; 0.07awc	53 ^b	45°50'01"N	88°32'04"W	393
			<i>Q. rubra</i>		4:Croswell sand; 0-4% slopes; 0.06awc	53 ^b	45°50'01"N	86°59'38"W	179
			<i>Q. ellipsoidalis</i>		3:Au Gres sand; 0-6% slopes; 0.05awc	51 ^b			
					2:Roscommon mucky sand; 0.13awc	74 ^c			
					5:Rubicon sand; 0-6% slopes; 0.07awc	53 ^b	45°50'26"N	86°05'58.37"W	186
					6:Grayling sand; 0-6% slopes; 0.08awc	48 ^b			
FC-A*	FC/Baraga Plains: MI	Stand A	<i>Q. rubra</i>	40	Keweenaw-Kalkaska Complex; 1-12% slopes	61 ^a	46°39'09"N	88°36'06"W	297
FC-B*	FC/Baraga Plains: MI	Stand B	<i>Q. rubra</i>	40	Munising-Yalmer loamy sand; 1-6% slopes; 0.05awc	63 ^a	46°40'27"N	88°31'27"W	423
FC-C*	FC/Baraga Plains: MI	Stand C	<i>Q. ellipsoidalis</i>	40	Grayling sand; 0-6% slopes; 0.07awc	48 ^b	46°39'14"N	88°35'25"W	394
FC-E*	FC/Baraga Plains: MI	Stand E	<i>Q. ellipsoidalis</i>	40	Grayling sand; 0-6% slopes; 0.07awc	48 ^b	46°39'55"N	88°33'19"W	398
NNF-QR*	Nicolet National Forest: WI	Nicolet	<i>Q. rubra</i>	40	Kemman fine sandy loam; 6-15% slopes	na	45°20'53"N	88°23'17"W	354
NNF-QE*	Nicolet National Forest: WI	Nicolet	<i>Q. ellipsoidalis</i>	39	Rousseau fine sand; 1-6% slopes; 0.08awc	na	45°19'19"N	88°19'53"W	301
CNF-QR*	Chequamegon National Forest: WI	Chequamegon	<i>Q. rubra</i>	40	Portwing-Herbstler Complex; 0-6% slopes	na	46°42'54"N	91°02'08"W	323
CNF-QE*	Chequamegon National Forest: WI	Chequamegon	<i>Q. ellipsoidalis</i>	40	Vilas loamy sand; 0-6% slopes; 0.04awc	na	46°44'53"N	91°04'20"W	384

Official soil survey descriptions are available online at <http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx> (Information downloaded and plotted on maps as a layer on September 5, 2016) (wss_SSA_MI013_soildb_MI_2003_[2015-09-21]_Baraga county and wss_SSA_MI013_soildb_MI_2003_[2015-09-21]_Delta county). *: Populations were genetically characterized by Lind-Riehl et al. (2014). Soils in the two sympatric stands are ranked from rich (1) to poor (6) (wss_SSA_MI061_soildb_MI_2003_[2016-09-22]_Houghton county). ^a: Sugar maple; ^b: Jack pine; ^c: Quaking aspen; na: not available.

TABLE 3.2. Gene flow analysis in sympatric populations of *Quercus rubra* and *Q. ellipsoidalis*.

Seed parent at FIR013	Genotype	Species	Seeds	External gene flow	Pollen parent										Interspecific gene flow (%)			
					<i>Q. ellipsoidalis</i> /genotype at FIR013					<i>Q. rubra</i> /genotype at FIR013								
FE50	138/144	QE	46	33	13	10	2	0	1	0	0	0	0	0	0	0	0	0
FE55	138/138	QE-I	42	34	8	7	0	0	1	0	0	0	0	0	0	0	0	0
EE59	138/138	QE	50	38	11	6	5	0	0	1	0	0	0	0	1	0	8.3	0
EE69	138/138	QE	29	21	6	3	3	0	0	2	0	1	1	1	1	0	25.0	0
ER81	138/141	QE	8	6	2	1	1	0	0	0	0	0	0	0	0	0	0	0
EE82	138/138	QE-I	30	9	19	17	2	0	0	2	0	0	0	0	2	0	9.5	0
EE84	138/138	QE	47	15	32	25	5	0	2	0	0	0	0	0	0	0	0	0
FR501	138/138	QE	24	14	10	2	1	0	0	7	0	0	0	0	0	0	0	0
FR502	138/144	QE	24	17	7	7	0	0	0	0	0	0	0	0	0	0	0	0
FR503	138/144	QE	32	23	9	4	0	0	5	0	0	0	0	0	0	0	0	0
FR505	138/138	QE	32	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FR508	138/138	QE-I	30	29	1	1	0	0	0	0	0	0	0	0	0	0	0	0
FR509	138/138	F ₁	24	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FR510	138/144	QE	24	22	2	1	0	0	1	0	0	0	0	0	0	0	0	0
FR512	138/138	QE	24	22	2	2	0	0	0	0	0	0	0	0	0	0	0	0

Notes: Paternity assignment at the strict confidence level (95%). QE: *Q. ellipsoidalis*; QR: *Q. rubra*, QE-I: *Q. ellipsoidalis*-like introgressive form, QR-I: *Q. rubra*-like introgressive form, F₁: first generation hybrid. Interspecific gene flow is shown for seeds with pollen parents identified within the stand as percentage of seeds derived from interspecific crosses. Gene flow between introgressive forms and the other species (e.g. between a *Q. ellipsoidalis*-like seed parent and *Q. rubra*) is considered as interspecific gene flow.

TABLE 3.3. Genetic variation parameters for each population at 23 SSR loci.

Population	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>
FC-QR/QE	96	9	0.53	0.63	0.16
ES-QR/QE	95	9	0.48	0.64	0.25
FC-QR	43	8	0.50	0.60	0.16
FC-QE	37	6	0.53	0.56	0.06
ES-QR	51	8	0.51	0.63	0.17
ES-QE	34	6	0.43	0.56	0.21
FC-Acorns	302	14	0.58	0.66	0.15
ES-Acorns	164	11	0.55	0.58	0.05

Notes: *N_a*: Number of alleles; *H_o*: Observed heterozygosity; *H_e*: Expected heterozygosity; *F*: Inbreeding coefficient calculated as $F = 1 - H_o/H_e$.

TABLE 3.4. Frequency of hybrids and introgressive forms and introgression of outlier alleles at *COL* between species.

Pops	Introgressive forms		<i>F₁</i> hybrids (%)		Introgression of outlier alleles (138, 141) between species			
	Species	QR-like (%)	QE-like (%)	QR-like (%)	Frequency of 138/138 (%)	Frequency of 141/141 (%)	QR Frequency of 138 (%)	QE Frequency of 141 (%)
FC	QR/QE	12.00	16.66	4.16	16.00	0	27.00 ⁺	13.09 ⁺
ES	QR/QE	8.92	8.10	2.10	5.36	2.70	20.53	14.86
FC-A*	QR	10.00	0	0	0	-	2.50	-
FC-B*	QR	4.00	2.00	0	2.50	-	5.00	-
FC-C*	QE	4.00	12.00	2.00	-	0	-	17.50
FC-E*	QE	1.00	12.00	1.00	-	0	-	8.80
NN-QR*	QR	20.00	0	0	0	-	2.50	-
NN-QE*	QE	0	18.00	5.12	-	15.38	-	25.64 ⁺
CN-QR*	QR	23.00	0	10.00	17.50	0	21.25	-
CN-QE*	QE	0	20.00	10.00	-	20.00	-	27.50

Notes: Relative frequency of individuals identified as *Q. rubra* and *Q. ellipsoidalis*, *F₁* hybrids, and introgressive forms using genetic assignment analysis in STRUCTURE without *a priori* species information (Pritchard et al. 2000). Pure species were defined as having a proportion of ancestry of ≥ 0.9 in one cluster. *F₁* hybrids were defined as having a proportion of ancestry of 0.4 to 0.6 in one cluster and introgressive forms were defined as having a proportion of ancestry of 0.61 to 0.89 in one of the clusters (Lind and Gailing, 2013). QR: *Quercus rubra*; QE: *Quercus ellipsoidalis*. *: Populations have been characterized by Lind-Riehl and Gailing (2017).



FIGURE 3.1 Northern red oak (*Quercus rubra* L.) young leaves (above), Northern pin oak (*Quercus ellipsoidalis* E. J. Hill) leaves (below).

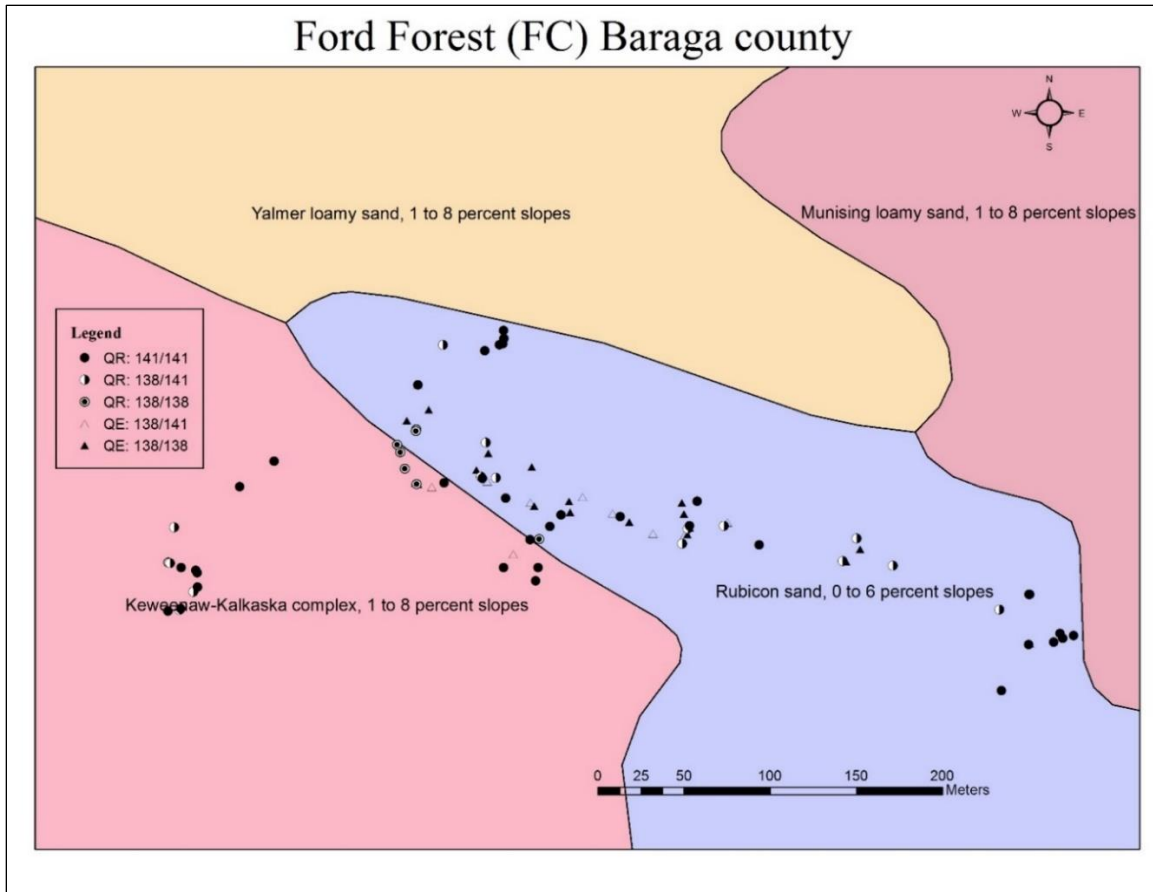


FIGURE 3.2 Location of genetically assigned species with different genotypes at FIR013 (138/138, 138/141, 141/141). A high level of introgression of alleles 138 and 141 between species is observed on Rubicon sand and at the transition zone between Keweenaw-Kalaska complex and Rubicon sand.

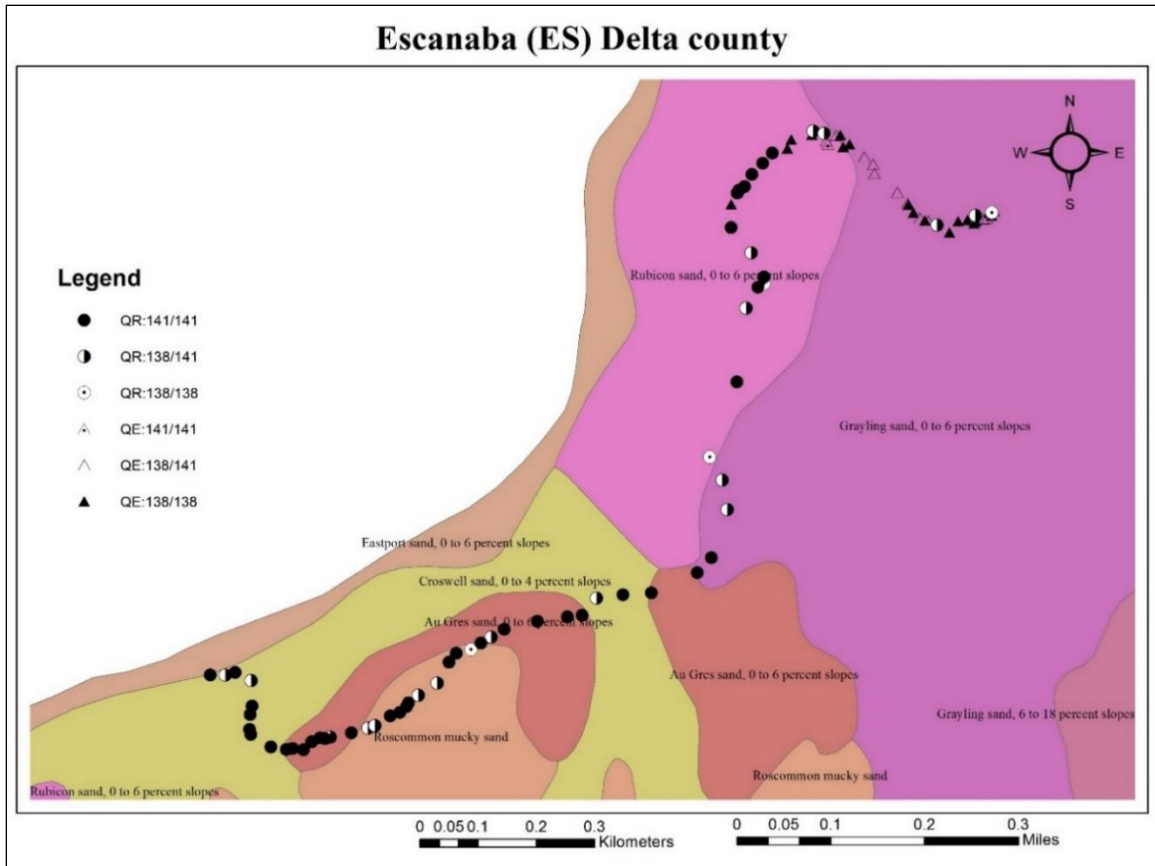


FIGURE 3.3 Location of genetically assigned species with different genotypes at FIR013 (138/138, 138/141, 141/141). A high level of introgression of alleles 138 and 141 between species is observed in the transition zone of Rubicon sand and Grayling sand.

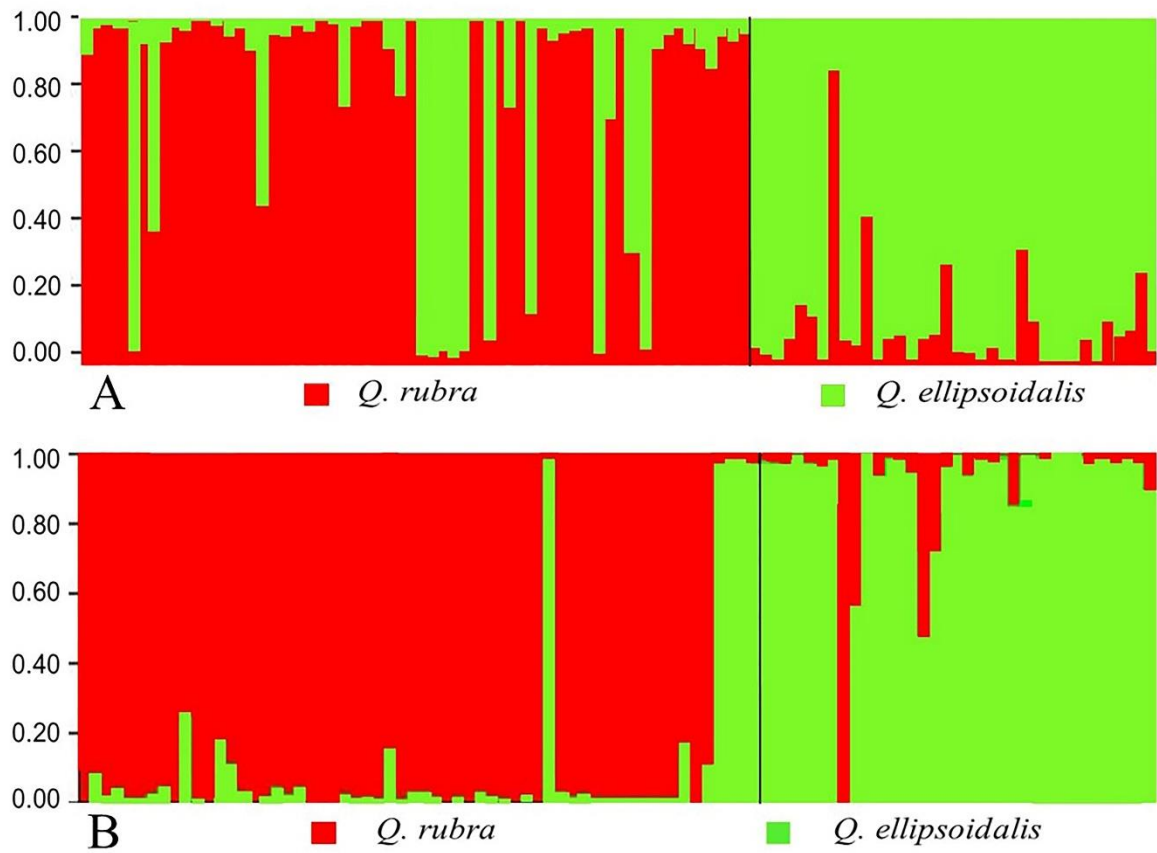


FIGURE 3.4 Genetic assignment of individuals in the Baraga Plains (FC) (A) and in the Escanaba (ES) sympatric stand (B) using STRUCTURE 2.3.4. Individuals are grouped by morphological species. The proportion of ancestry in one of the two genetic groups is shown for each sample

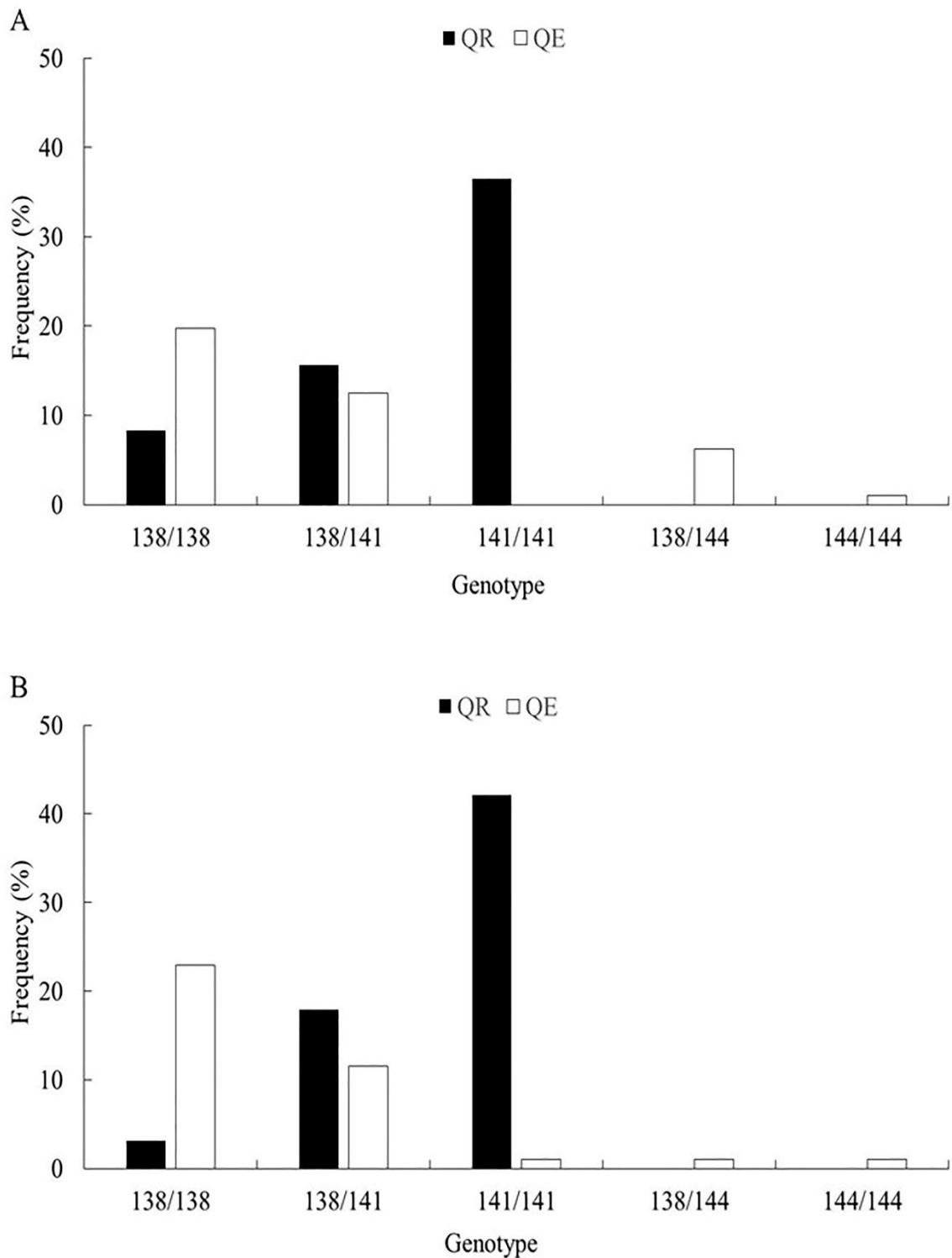


FIGURE 3.5 Genotype frequency distribution for each species at EST-SSR FIR013 (CONSTANS-like) in the FC/Baraga Plains population (above) and in the Escanaba population (below). QR: *Quercus rubra*. QE: *Quercus ellipsoidalis*.

SUPPLEMENTARY TABLE 3.1 Microsatellite marker information.

Marker	Repeat	Forward sequence (5'-3') Reverse sequence (3'-5')	T _m (°C)	Dye	Size range (bp)	Marker type	LG	Annotated function
FIR013	CAG ₅	CGGGAGGTTGATGATTT AACTGTCACCCCAIAGC	56	6-FAM	138-144	EST-SSR	2	CONSTANS-like protein COL
FIR030	AG ₇	GGACATATTCTAGGAGCGAGGT ATGCCATAGCACAGAGA	56	6-FAM	161-179	EST-SSR	7	u6 snrna-associated sm-like protein lsm3
FIR039	CT ₇	GAGCCTTTTCATCGCTCAC TCACACCCCAAAACTCCAT	56	PET	109-125	EST-SSR	1	histone deacetylase
FIR048	CT ₅	TGCACAAAATTTGGAGATG TTGATGCAAGGTGCGATTC	56	PET	187-219	EST-SSR	2	cell division protein ftsH, putative
FIR053	GTG ₇	AGTTTCCCACATTTGTGC TACCATGCCACCAAGCAAATC	56	NED	127-151	EST-SSR	5	glutaredoxin, grx, putative
FIR089	GA ₅	AGCGACTAACCCAACTTCCA GCGGATTCGATAGCAATTTT	56	6-FAM	165-175	EST-SSR	7	abscisic acid receptor pyd4
FIR104	GGT ₇	TFACTGGHTTGGACTCA AGCAGTGACTGACCTGTA	56	VIC	217-223	EST-SSR	12	r2L3-myb transcription factor
GOT009	TC ₇	CACCTCACTAAGCAACTGTCA TTTTGGAGGGCGAGATAATG	59	6-FAM	231-249	EST-SSR	12	unknown protein
GOT021	AT ₁₃	AGAAAGTTCAGGGAAGCA CTTCGTCCTGATGAAATGT	59	VIC	91-103	EST-SSR	3	histidine kinase 1,2,3 plant, putative
GOT040	GA ₁₁	AAAGGACTCGTGGTTTCTA ACCGATTTGAAGCTGGAGAA	56	VIC	236-260	EST-SSR	6	ribosomal protein, S9, putative
PIE040	TTC ₅	GTGAGAGAGAGAGACAAAGAGAAAA AAATTCGCCCACTTGAG	59	NED	155-173	EST-SSR	11	unknown protein
PIE099	TC ₉	GGTACCGACTACTACCCTTC CGGTGGACCCAAATATGTAAC	56	VIC	197-219	EST-SSR	8	unknown protein
PIE101	AT ₁₃	GGCAGAGTCAAAATAAGCTAC CACCCTTTTCAATCTGTG	56	PET	145-157	EST-SSR	8	unknown protein
POR003	CT ₅	CTCGCTCTCTCCCTTAATC AGCTTTGATCGAGTCCGAAA	56	PET	99-113	EST-SSR	3	bzip transcription factor bzip109
POR016	GGT ₆	AGCAAAGCAGAGCCAAAT CAGCGGCTTTGAGGTAATTC	56	NED	114-120	EST-SSR	6	heat shock protein 70 (stress)
QuoGAOC11**	GA ₁₅	ATACCCAGCTCCCATGACCA TCCCAAAATTCAGTAGTGT	56	VIC	199-219	SSR	-	-
VIT057	AACTCG ₅	TCAGCAAAATCCCACCTTGT ACACTTCGCTGTCTCTGGAT	56	VIC	133-147	EST-SSR	9	ap2_erf domain-containing transcription factor
VIT107	TA ₁₅	TGATCACAGATGGAGGTTAAACA CCCCCACTTAGAAAGAACG	56	NED	130-142	EST-SSR	3	lhca2 protein
WAG018	GT ₇	GGTTCGGATTTGTTGATGCC TACAAAACCCAAAGTCCCTG	56	6-FAM	143-151	EST-SSR	6	beta-amylase
3A05*	CA ₁₁ CT ₂	AACTGACCTCTCTCACAGC AGTGTGGAGTCTCATGG	62	PET	237-269	SSR	-	-
3D15*	CA ₁₅	GGTGGTGCAGATACACTGG GACTCAGACACCAACTTCAAG	62	NED	208-238	SSR	-	-
1P10*	CA ₁₂ GCC(TG) ₅	ATTTCGATGCGAGGGTGTGC TAGCCAAAGGACCAGAGACC	62	6-FAM	138-160	SSR	-	-
2P24*	CA ₁₄	GCAAGAGATCACACAAACTTAGC CTTTGGGTTCAACCAACAGC	62	NED	136-164	SSR	-	-

SUPPLEMENTARY TABLE 3.2 PCR protocol for pre-PCR multiplex of markers based on size (bp) and fluorescent dye color.

PCR Group	Markers	Water (µl)	HOT FIREPol® (µl)	Primer forward (µl)	Primer Reverse (µl)	DNA (µl)	Total reaction volume (µl)
OMP1	3A05, 3D15, 1P10	4	2	1.34	1.34	1.34	10
OMP2	GOT021, QuruGAOC11, PIE040	4	2	1.34	1.34	1.34	10
OMP3	GOT040, POR016	4	2	1.34	1.34	1.34	10
OMP4	GOT09, FIR053	4	2	1.34	1.34	1.34	10
OMP5	WAG018, FIR048	4	2	1.34	1.34	1.34	10
OMP6	FIR013, FIR039	4	2	1.34	1.34	1.34	10
OMP8	FIR089, PIE101	4	2	1.34	1.34	1.34	10
OMP12	VIT057, POR003	4	2	1.34	1.34	1.34	10

Notes: For the pre-PCR multiplex all forward and reverse primers (5µM) were diluted 1:10 and mixed in equal proportions and the final volume was adjusted to 1.34µl with double deionized water.

SUPPLEMENTARY TABLE 3.3 Genetic variation parameters at each locus in the sympatric FC/Baraga Plains population.

Loci	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>
FIR013	96	3	0.36	0.54	0.34*
FIR030	96	12	0.50	0.51	0.02
FIR039	96	13	0.38	0.63	0.41
FIR048	96	18	0.70	0.89	0.22
FIR053	95	4	0.52	0.54	0.05
FIR089	96	7	0.66	0.78	0.16
FIR104	96	4	0.44	0.38	-0.16
GOT009	96	12	0.66	0.79	0.17
GOT021	96	4	0.09	0.15	0.39
GOT040	96	7	0.33	0.43	0.22
PIE040	96	8	0.57	0.69	0.18
PIE099	96	14	0.72	0.87	0.18
PIE101	95	14	0.79	0.80	0.02
POR003	96	14	0.83	0.88	0.05
POR016	96	2	0.01	0.01	0.00
VIT057	96	3	0.36	0.44	0.18
VIT107	96	6	0.65	0.73	0.12
WAG018	96	2	0.20	0.24	0.18
Mean^a	96	8	0.49	0.57	0.15
QuruGAOC11	96	17	0.87	0.89	0.03
3A05	96	11	0.36	0.77	0.53
3D05	96	12	0.72	0.82	0.12
1P10	96	13	0.67	0.82	0.18
2P24	95	12	0.70	0.81	0.12
Mean^b	96	12	0.66	0.82	0.20
Overall mean	96	9	0.53	0.63	0.16

Notes: *N*: Number of samples; *N_a*: Number of alleles; *H_o*: Observed heterozygosity; *H_e*: Expected heterozygosity; *F*: Inbreeding coefficient. Significant deviation from Hardy-Weinberg proportions at the 5% level (*) after Bonferroni corrections are shown; ^a: Mean at gene-based EST-SSRs; ^b: Mean at SSRs.

SUPPLEMENTARY TABLE 3.4 Genetic variation parameters at each locus in the sympatric Escanaba population.

Loci	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>
FIR013	95	4	0.33	0.50	0.35*
FIR030	95	12	0.29	0.59	0.50
FIR039	95	8	0.43	0.73	0.40
FIR048	94	19	0.83	0.88	0.06
FIR053	94	7	0.53	0.56	0.05
FIR089	95	12	0.53	0.79	0.33
FIR104	95	8	0.53	0.54	0.02
GOT009	91	11	0.52	0.72	0.28
GOT021	95	4	0.12	0.26	0.52
GOT040	95	6	0.40	0.40	-0.005
PIE040	95	11	0.52	0.62	0.16
PIE099	92	13	0.52	0.88	0.41
PIE101	93	15	0.66	0.76	0.14
POR003	95	13	0.74	0.90	0.18
POR016	95	2	0.01	0.01	-0.10
VIT057	95	3	0.40	0.49	0.19
VIT107	93	6	0.39	0.67	0.43
WAG018	95	3	0.29	0.38	0.21
Mean^a	94	9	0.45	0.60	0.24
QuruGAOC11	95	15	0.80	0.87	0.09
3A05	95	9	0.39	0.80	0.51
3D05	95	13	0.74	0.81	0.09
1P10	94	14	0.70	0.79	0.11
2P24	95	11	0.36	0.78	0.54
Mean^b	95	12	0.60	0.81	0.27
Overall mean	94	9	0.48	0.64	0.25

Notes: *N*: Number of samples; *N_a*: Number of alleles; *H_o*: Observed heterozygosity; *H_e*: Expected heterozygosity; *F*: Inbreeding coefficient. Significant deviation from Hardy-Weinberg proportions at the 5% level (*) after Bonferroni corrections are shown; ^a: Mean at gene-based EST-SSRs; ^b: Mean at SSRs.

SUPPLEMENTARY TABLE 3.5 Genetic variation parameters at each locus in the FC/Baraga Plains population separated by genetically assigned species.

Loci	FC-QR					FC-QE				
	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>
FIR013	43	3	0.26	0.23	-0.13	37	3	0.35	0.34	-0.04
FIR030	42	9	0.43	0.45	0.05	36	8	0.53	0.55	0.04
FIR039	41	8	0.54	0.76	0.30*	37	4	0.22	0.29	0.25
FIR048	43	16	0.70	0.88	0.21	37	14	0.68	0.88	0.23
FIR053	42	4	0.48	0.62	0.23	37	3	0.46	0.41	-0.13
FIR089	42	7	0.74	0.80	0.07	37	5	0.62	0.73	0.15
FIR104	43	4	0.23	0.25	0.06	37	3	0.62	0.46	-0.36
GOT009	43	12	0.65	0.78	0.17	37	8	0.68	0.73	0.08
GOT021	42	4	0.10	0.23	0.59	37	2	0.03	0.03	-0.01
GOT040	43	6	0.44	0.58	0.24	37	4	0.19	0.22	0.15
PIE040	43	6	0.44	0.58	0.24	37	6	0.68	0.71	0.05
PIE099	43	10	0.51	0.84	0.39	37	10	0.86	0.85	-0.02
PIE101	42	10	0.76	0.81	0.06	37	11	0.86	0.75	-0.15
POR003	43	13	0.84	0.88	0.04	37	12	0.81	0.85	0.05
POR016	43	2	0.02	0.02	-0.01	37	1	0.00	0.00	0.00
VIT057	43	3	0.37	0.43	0.12	37	3	0.41	0.49	0.18
VIT107	40	6	0.57	0.70	0.19	32	5	0.72	0.75	0.04
WAG018	43	2	0.12	0.11	-0.02	36	2	0.36	0.39	0.07
Mean^a	42	7	0.45	0.55	0.15	37	6	0.50	0.52	0.03
QuruGAOC11	43	13	0.79	0.87	0.09	37	12	0.92	0.86	-0.06
3A05	43	10	0.42	0.75	0.44	37	5	0.22	0.52	0.59
3D05	43	11	0.72	0.79	0.09	37	9	0.68	0.68	0.01
1P10	43	13	0.72	0.84	0.14	37	9	0.65	0.78	0.16
2P24	43	9	0.77	0.80	0.04	36	9	0.64	0.67	0.05
Mean^b	43	11	0.68	0.81	0.16	37	9	0.62	0.70	0.15
Overall mean	43	8	0.50	0.60	0.15	37	6	0.53	0.56	0.06

Notes: *N*: Number of samples; *N_a*: Number of alleles; *H_o*: Observed heterozygosity; *H_e*: Expected heterozygosity; *F*: Inbreeding coefficient; Significant deviation from Hardy-Weinberg proportions at the 5% level (*) after Bonferroni corrections are shown; ^a: Mean at gene based EST-SSRs; ^b: Mean at SSRs.

SUPPLEMENTARY TABLE 3.6 Genetic variation parameters at each locus in the Escanaba population separated by genetically assigned species.

Loci	ES-QR					ES-QE				
	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>
FIR013	51	3	0.29	0.28	-0.05	34	2	0.32	0.27	-0.19
FIR030	50	10	0.28	0.59	0.52	34	5	0.29	0.55	0.46*
FIR039	51	6	0.45	0.76	0.41	34	7	0.38	0.60	0.36*
FIR048	51	16	0.86	0.86	-0.01	33	10	0.82	0.85	0.04
FIR053	51	7	0.69	0.67	-0.03	34	3	0.26	0.24	-0.11
FIR089	51	10	0.63	0.80	0.21	34	9	0.41	0.76	0.46
FIR104	50	8	0.56	0.54	-0.04	34	4	0.50	0.53	0.05
GOT009	48	9	0.54	0.72	0.25	33	9	0.48	0.70	0.30
GOT021	51	2	0.18	0.19	0.08	34	3	0.06	0.26	0.77
GOT040	51	4	0.55	0.50	-0.10	34	4	0.15	0.14	-0.06
PIE040	51	9	0.47	0.56	0.15	34	5	0.53	0.65	0.18
PIE099	48	10	0.44	0.85	0.48	34	10	0.68	0.83	0.19
PIE101	51	10	0.76	0.78	0.02	32	11	0.50	0.71	0.30
POR003	50	11	0.74	0.89	0.17	34	11	0.74	0.87	0.16
POR016	50	2	0.02	0.02	-0.01	34	1	0.00	0.00	0.00
VIT057	51	3	0.33	0.50	0.33	33	3	0.52	0.46	-0.11
VIT107	50	6	0.40	0.66	0.40	34	5	0.41	0.57	0.28
WAG018	51	3	0.16	0.18	0.13	33	2	0.48	0.50	0.03
Mean^a	50	7	0.46	0.57	0.16	34	6	0.42	0.53	0.17
QuruGAOC11	51	13	0.80	0.87	0.07	34	10	0.74	0.79	0.07
3A05	51	7	0.53	0.70	0.25	34	7	0.26	0.71	0.63
3D05	51	12	0.71	0.83	0.15	34	6	0.79	0.75	-0.06
1P10	51	12	0.75	0.78	0.04	33	6	0.61	0.69	0.12
2P24	51	11	0.49	0.84	0.42	34	4	0.15	0.44	0.67
Mean^b	51	11	0.65	0.80	0.19	34	7	0.51	0.68	0.29
Overall mean	51	8	0.51	0.62	0.17	34	6	0.44	0.56	0.20

Notes: *N*: Number of samples; *N_a*: Number of alleles; *H_o*: Observed heterozygosity; *H_e*: Expected heterozygosity; *F*: Inbreeding coefficient; Significant deviation from Hardy-Weinberg proportions at the 5% level (*) after Bonferroni corrections are shown; ^a: Mean at gene based EST-SSRs; ^b: Mean at SSRs.

SUPPLEMENTARY TABLE 3.7 Genetic variation parameters at each locus for the acorns collected from the FC/Baraga Plains and Escanaba populations.

Loci	FS-acorns					ES-acorns				
	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>	<i>N</i>	<i>N_a</i>	<i>H_o</i>	<i>H_e</i>	<i>F</i>
FIR013	299	5	0.43	0.39	-0.10	164	4	0.31	0.31	-0.01
FIR030	299	9	0.66	0.57	-0.14	160	6	0.50	0.49	-0.03
FIR039	287	8	0.16	0.31	0.46	162	12	0.39	0.51	0.24
FIR048	294	27	0.81	0.89	0.08	158	25	0.92	0.90	-0.03
FIR053	296	9	0.54	0.60	0.10	160	12	0.34	0.41	0.17
FIR089	300	16	0.71	0.81	0.12	161	14	0.60	0.79	0.25
FIR104	298	6	0.57	0.69	0.17	161	6	0.48	0.48	-0.01
GOT009	300	20	0.79	0.86	0.09*	162	14	0.65	0.60	-0.09
GOT021	298	4	0.01	0.01	0.50	162	4	0.04	0.04	-0.02
GOT040	301	6	0.27	0.26	-0.04	161	5	0.30	0.27	-0.09
PIE040	297	11	0.71	0.68	-0.04	161	11	0.63	0.59	-0.08
PIE099	300	17	0.90	0.88	-0.02	161	19	0.79	0.88	0.10
PIE101	296	16	0.76	0.83	0.08	159	16	0.74	0.73	-0.02*
POR003	299	23	0.85	0.86	0.00	164	13	0.88	0.86	-0.03
POR016	299	3	0.08	0.54	0.84	162	5	0.02	0.41	0.94
VIT057	301	10	0.54	0.71	0.23	162	4	0.51	0.48	-0.06
VIT107	300	15	0.71	0.85	0.16	160	7	0.76	0.63	-0.21*
WAG018	300	8	0.46	0.69	0.33	161	3	0.53	0.50	-0.06
Mean^a	298	12	0.56	0.64	0.16	161	10	0.52	0.55	0.05
QuruGAOC11	301	25	0.88	0.93	0.05	161	15	0.86	0.74	-0.15
3A05	302	9	0.39	0.62	0.36	159	11	0.72	0.71	-0.02
3D05	302	10	0.74	0.73	0.00	162	7	0.30	0.68	0.56
1P10	300	14	0.62	0.71	0.12	161	10	0.68	0.71	0.05*
2P24	300	10	0.68	0.67	-0.02	163	16	0.68	0.55	-0.23
Mean^b	299	12	0.66	0.73	0.10	161	12	0.64	0.68	0.04
Overall mean	301	14	0.58	0.66	0.15	161	10	0.55	0.58	0.05

Notes: *N*: Number of samples; *N_a*: Number of alleles; *H_o*: Observed heterozygosity; *H_e*: Expected heterozygosity; *F*: Inbreeding coefficient; Significant deviation from Hardy-Weinberg proportions at the 5% level (*) after Bonferroni corrections are shown; ^a: Mean at gene-based EST-SSRs; ^b: Mean at SSRs.

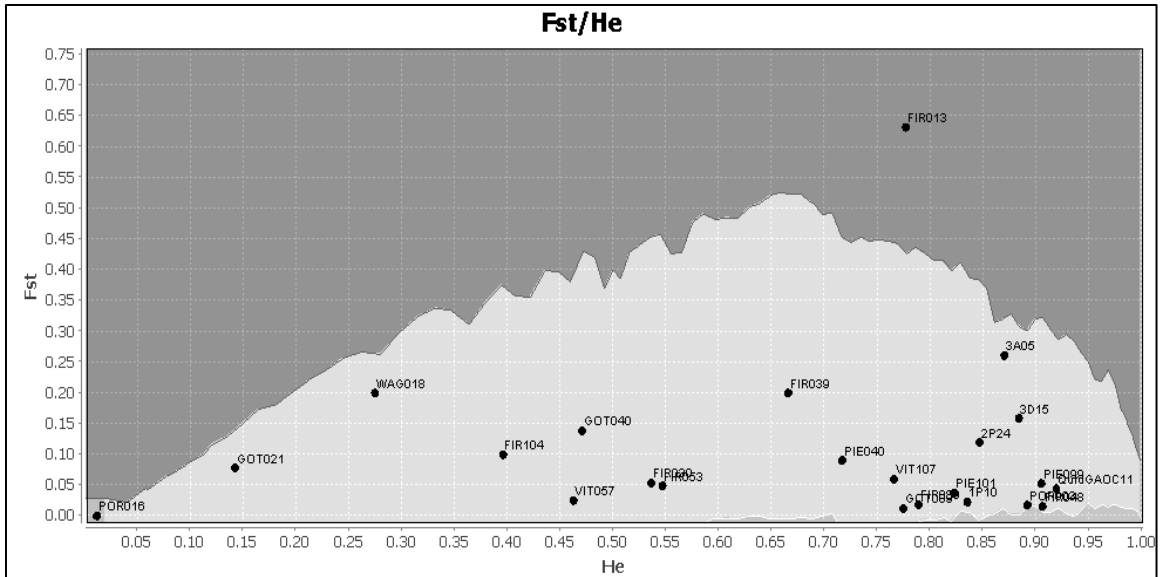
SUPPLEMENTARY TABLE 3.8 Pairwise genetic differentiation (F_{ST}) values between populations for genetically confirmed species and seeds (acorns).

Species	FC-QR	FC-QE	ES-QR	ES-QE	FC-acorns	ES-acorns
FC-QR	0.000					
FC-QE	0.067	0.000				
ES-QR	0.012	0.063	0.000			
ES-QE	0.083	0.027	0.071	0.000		
FC-acorns	0.077	0.038	0.074	0.052	0.000	
ES-acorns	0.089	0.036	0.081	0.025	0.034	0.000

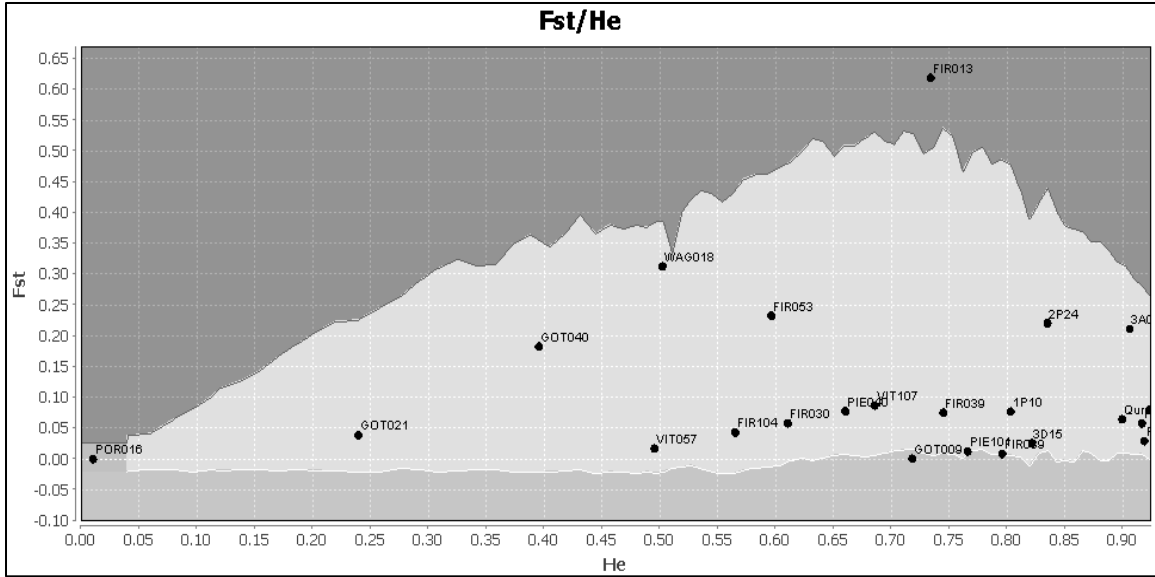
Notes: All pairwise F_{ST} values are significant ($p < 0.01$)

SUPPLEMENTARY TABLE 3.9 Scaled introgression rate of outlier alleles.

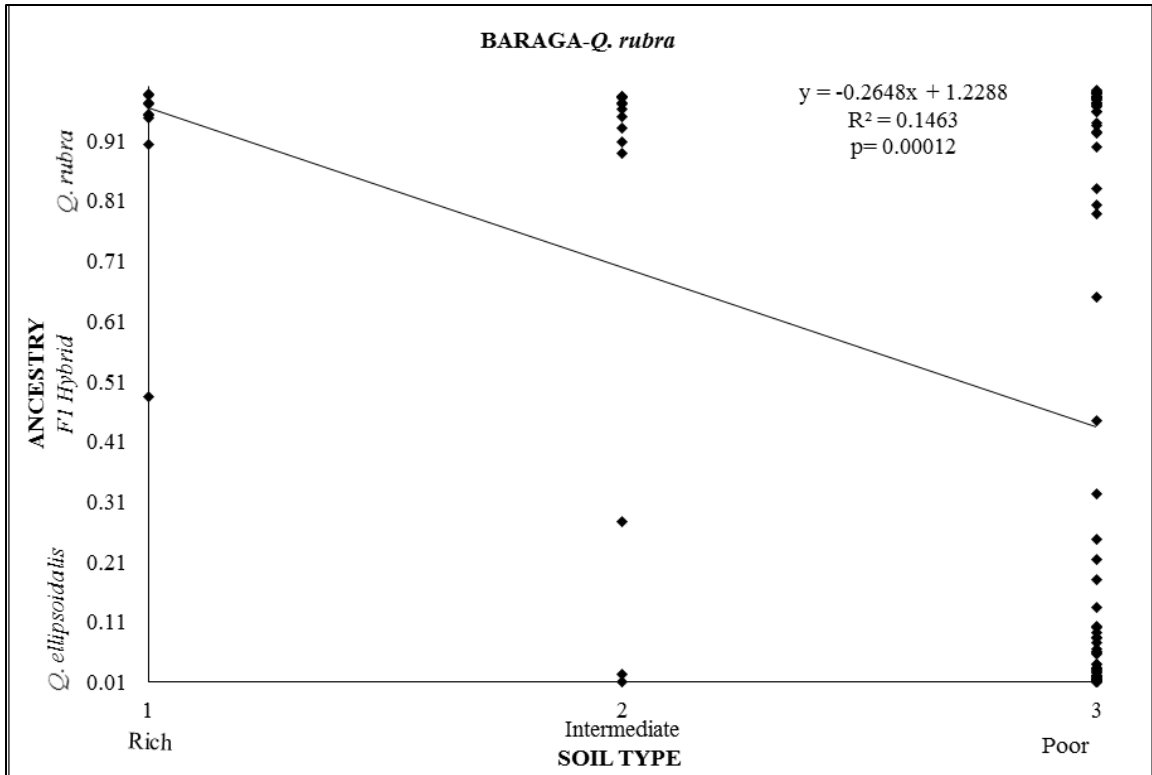
Population	Species	Introgression of allele 138 or 141	Frequency of introgressive forms	Ratio of 138 or 141 introgression/ Frequency of introgressive forms
FC	QR	27.00	12.00	2.25
FC	QE	13.09	16.66	0.79
ES	QR	20.53	8.92	2.30
ES	QE	14.86	8.1	1.84
FC-A	QR	2.5	10.00	0.25
FC-B	QR	5.00	4.00	1.25
FC-C	QE	17.5	12.00	1.46
FC-E	QE	8.8	12.00	0.73
NN-QR	QR	2.5	20.00	0.13
NN-QE	QE	25.64	18.00	1.42
CN-QR	QR	21.25	23.00	0.92
CN-QE	QE	27.5	20.00	1.38



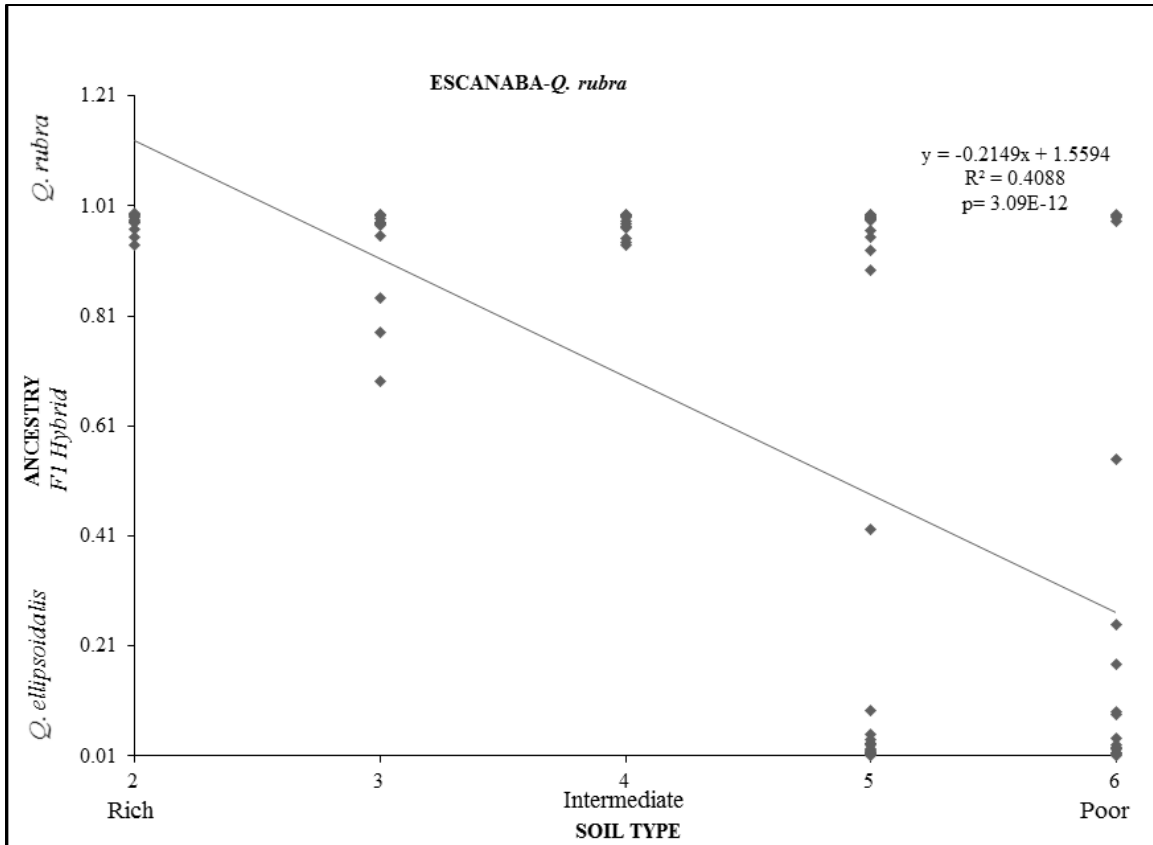
SUPPLEMENTARY FIGURE 3.1 Outlier analysis of 23 markers between species in the Baraga Plains sympatric population under the stepwise mutation model as calculated in LOSITAN. The neutral 99% confidence envelope is shown in white. Outlier locus FIR013 showed F_{ST} values above the neutral confidence envelope indicative of positive selection, the false discovery rate was 0.01.



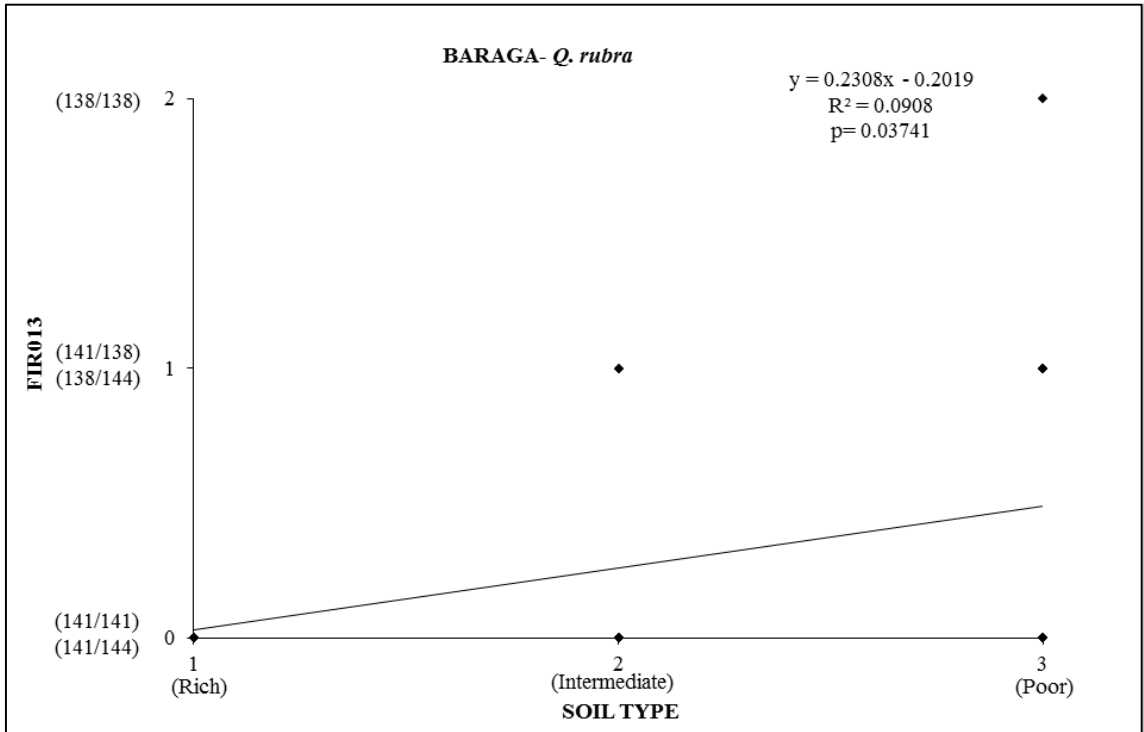
SUPPLEMENTARY FIGURE 3.2 Outlier analysis of 23 markers between species in the Escanaba sympatric population under the stepwise mutation model as calculated in LOSITAN. The neutral 95% confidence envelope is shown in white. Outlier locus FIR013 showed F_{ST} values above the neutral confidence envelope indicative of positive selection, the false discovery rate was 0.05.



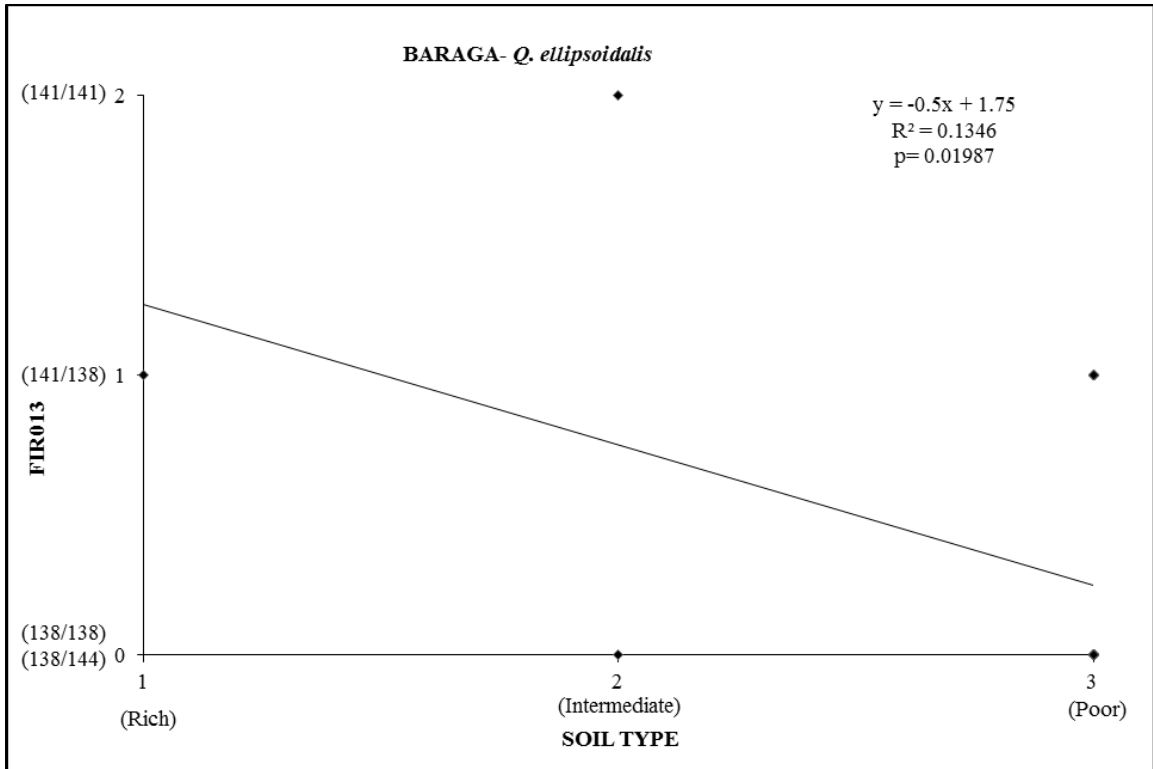
SUPPLEMENTARY FIGURE 3.3 Linear regression of *Q. rubra* ancestry according to assignment analysis in STRUCUTRE against soil type in Baraga.



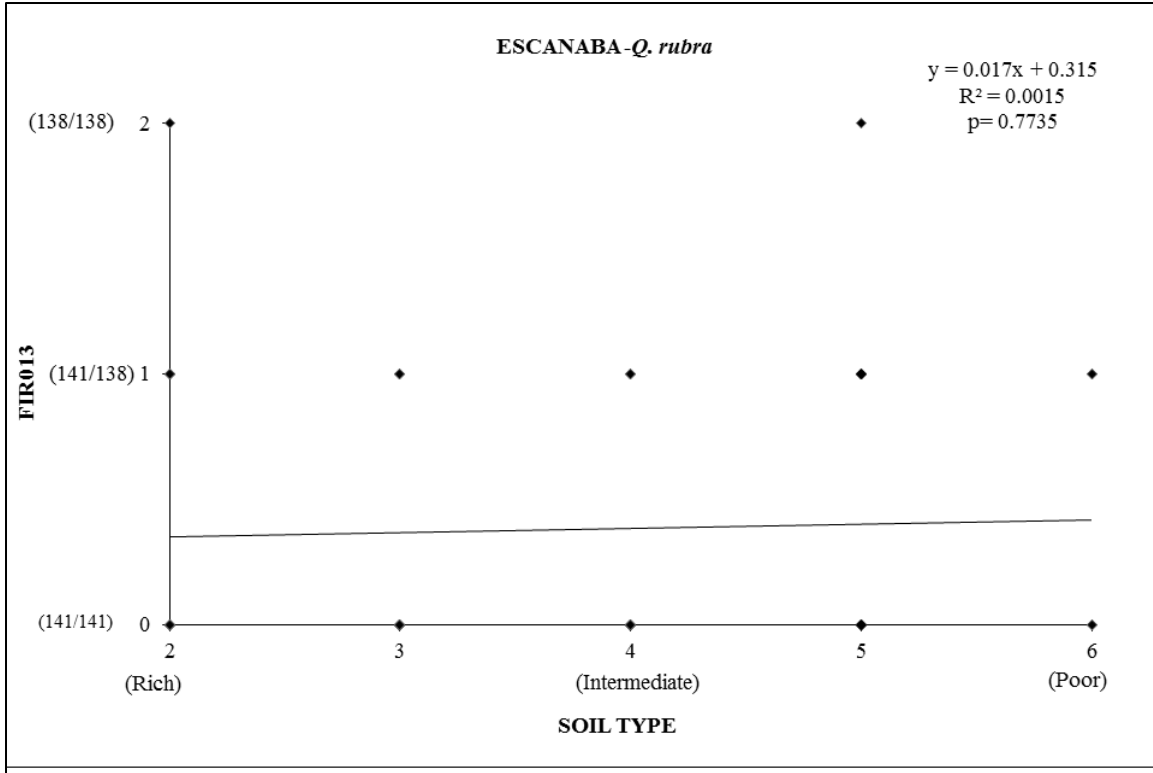
SUPPLEMENTARY FIGURE 3.4 Linear regression of *Q. rubra* ancestry according to assignment analysis in STRUCUTRE against soil type in Escanaba.



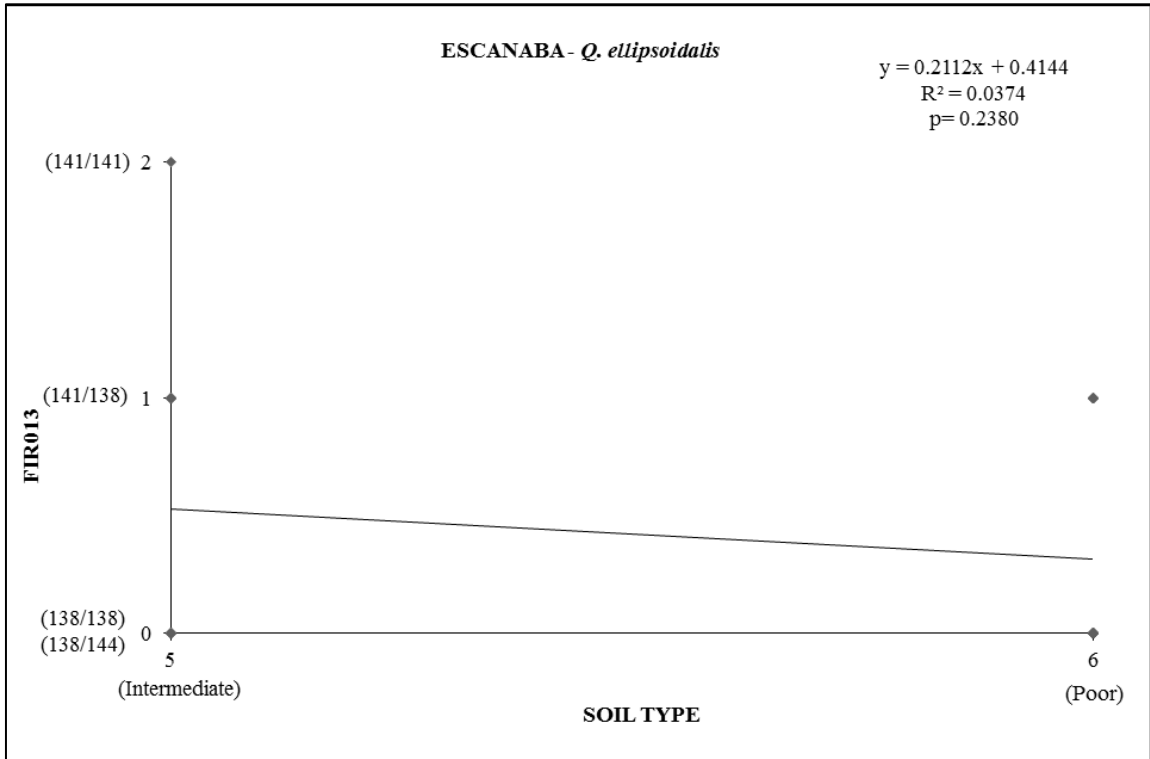
SUPPLEMENTARY FIGURE 3.5 Linear regression of the soil type against the number of *Q. ellipsoidalis* (138) alleles at FIR013 in each *Q. rubra* tree (0:138/138; 1:138/141; 2:141/141) in Baraga Plains population samples (n=48).



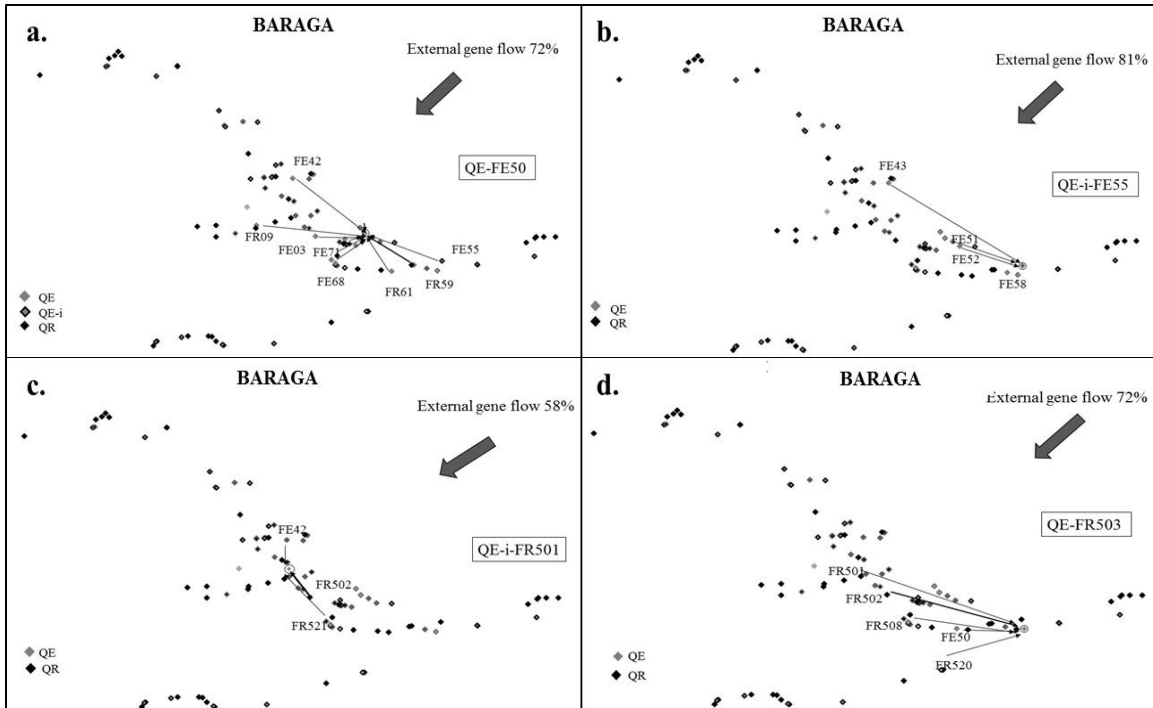
SUPPLEMENTARY FIGURE 3.6 Linear regression of soil type against the number of *Q. rubra* (141) alleles at FIR013 in each *Q. ellipsoidalis* tree (0: 141/141 and 141/144; 1: 138/141 and 138/144; 2: 138/138) in Baraga Plains population samples (n=35).



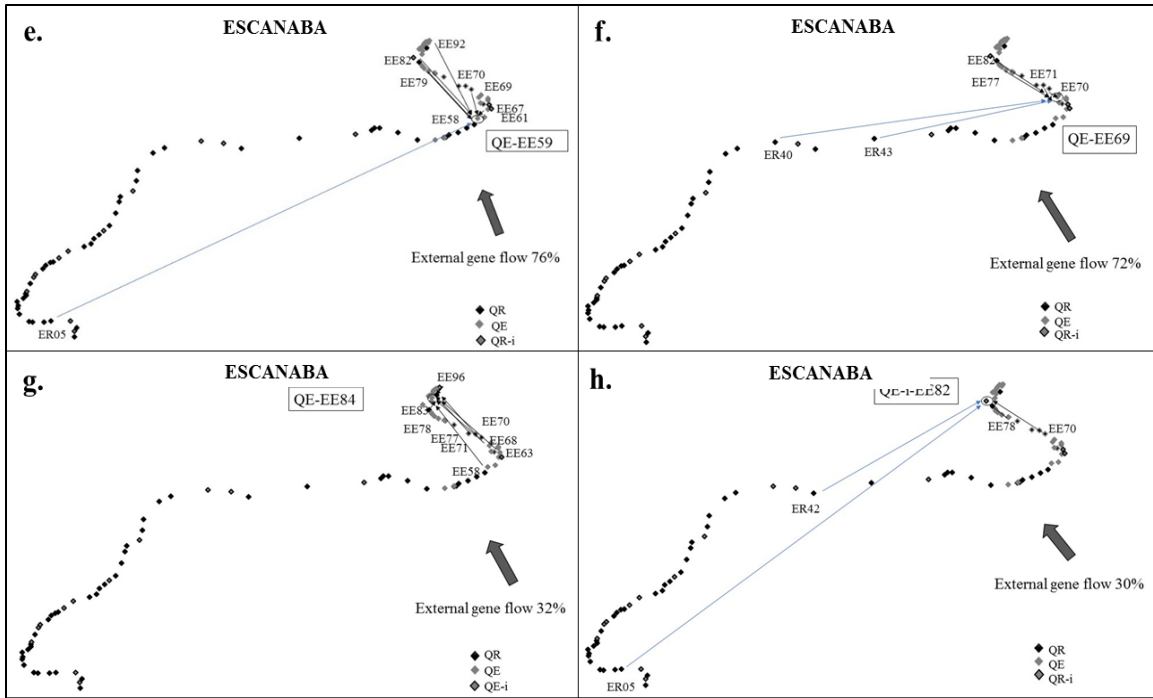
SUPPLEMENTARY FIGURE 3.7 Linear regression of the soil type against the number of *Q. ellipsoidalis* (138) alleles at FIR013 in each *Q. rubra* tree (0: 141/141; 1: 138/141; 2: 138/138) in Escanaba population samples (n=58).



SUPPLEMENTARY FIGURE 3.8 Linear regression of the soil type against the number of *Q. rubra* (141) alleles at FIR013 in each *Q. ellipsoidalis* tree (0: 138/138; 1: 138/141; 2: 141/141) in Escanaba population samples (n=39).

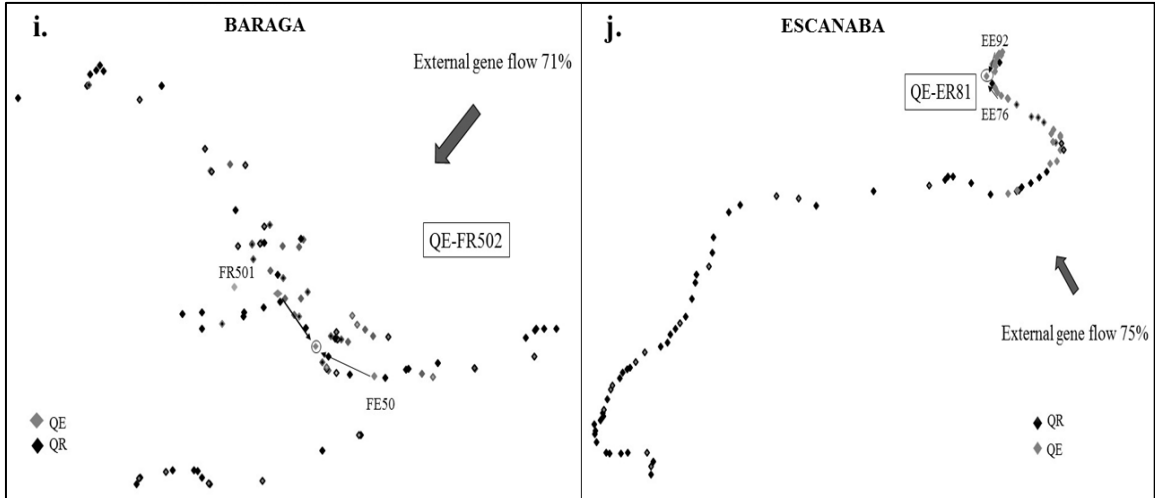


SUPPLEMENTARY FIGURE 3.9 Gene flow analysis in the sympatric Baraga Plains population. Pollen donors within the stand and pollen flow from outside the stand are shown. The line weight is proportional to the contribution of the pollen donor. QE: *Quercus ellipsoidalis*; QE-i: *Q. ellipsoidalis* introgressive; QR: *Q. rubra*.

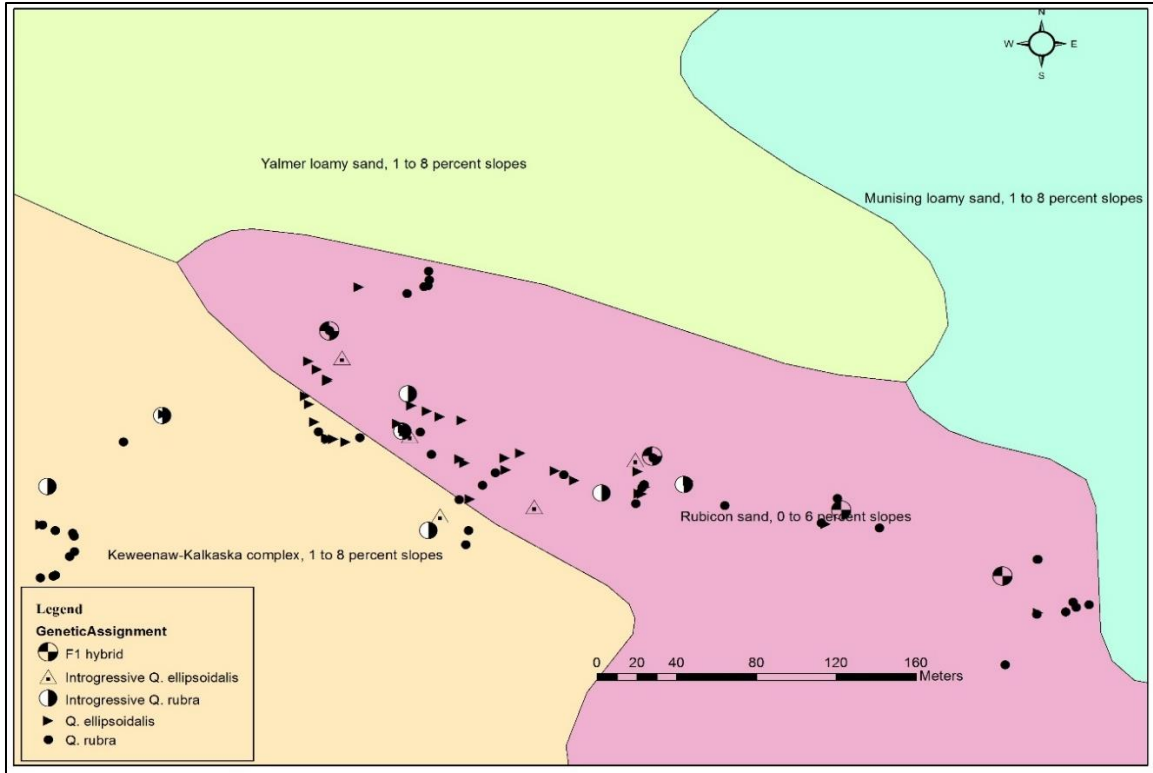


SUPPLEMENTARY FIGURE 3.10 Gene flow analysis in the sympatric Escanaba population. Pollen donor within the stand and pollen flow from outside the stand are shown. The line weight is proportional to the contribution of the pollen donor. Blue lines indicate interspecific gene flow.

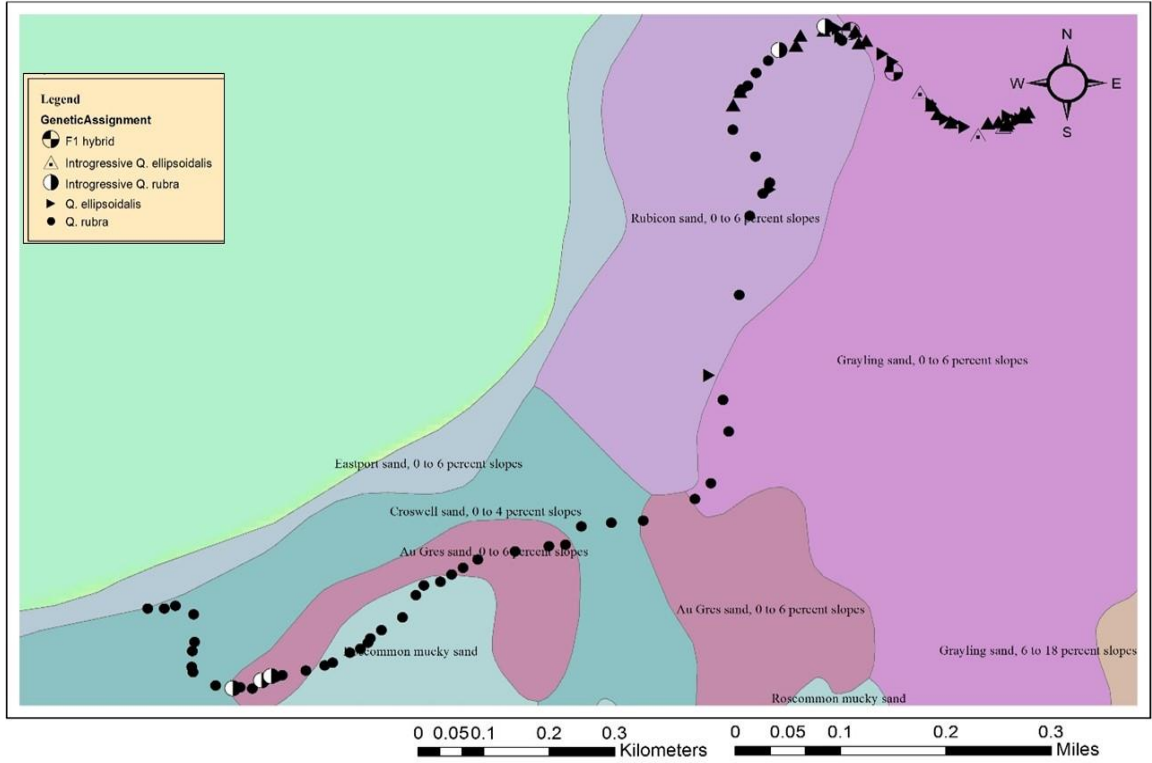
QE: *Quercus ellipsoidalis*; QE-i: *Q. ellipsoidalis* introgressive; QR: *Q. rubra*.



SUPPLEMENTARY FIGURE 3.11 Gene flow analysis in the sympatric Baraga Plains and Escanaba population. Pollen donors within the stand and pollen flow from outside the stand are shown. The line weight is proportional to the contribution of the pollen donor. QE: *Quercus ellipsoidalis*; QR: *Q. rubra*.



SUPPLEMENTARY FIGURE 3.12 Location of genetically assigned species, F₁ hybrids and introgressive forms in the Baraga Plains.



SUPPLEMENTARY FIGURE 3.13 Location of genetically assigned species, F₁ hybrids and introgressive forms in the Escanaba region.