2014

GENETIC VARIATION, LOCAL ADAPTATION AND POPULATION STRUCTURE IN NORTH AMERICAN RED OAK SPECIES, QUERCUS RUBRA L. AND Q. ELLIPSOIDALIS E. J. HILL

Jennifer F. Lind-Riehl
Michigan Technological University

GENETIC VARIATION, LOCAL ADAPTATION AND POPULATION STRUCTURE IN NORTH AMERICAN RED OAK SPECIES, *QUERCUS RUBRA* L. AND *Q. ELLIPSOIDALIS* E. J. HILL

By

Jennifer F. Lind-Riehl

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

2014

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: *Erik Lilleskov*

Committee Member: *Victor Busov*

Committee Member: *Thomas Snyder*

School Dean: *Terry Sharik*

Table of Contents

Preface

Chapters within this dissertation include three multi-authored peer reviewed journal articles of which two have been published and one has been accepted for publication. I am the primary author on all three articles. Previously published studies are documented with a footnote citation on the first page of the publication.

"Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations" has been published in Tree Genomes and Genetics. As first author, I carried out the majority of the laboratory work, performed the analyses, and wrote the paper. Oliver Gailing, my co-author, supported me with sample collection, analytical, writing, and editorial guidance.

"Evidence for selection on a *CONSTANS*-like gene between *Quercus rubra* and *Q. ellipsoidalis*" has been published in Annals of Botany. As first author, I carried out the majority of the laboratory work, performed the analyses, and wrote the paper. My co-author, Alexis Sullivan, supported me with some laboratory work, figure creation and editorial guidance. My other co-author, Oliver Gailing, supported me with analytical, writing and editorial guidance.

"Fine-scale spatial genetic structure of two red oak species, *Quercus rubra* L. *Q. ellipsoidalis*" has been published in Plant Systematics and Evolution. As first author, I carried out the majority of the laboratory work, performed the analyses, and wrote the paper. My co-author, Oliver Gailing, supported me with analytical, writing and editorial guidance.

Oliver Gailing, my advisor, helped conceive and design the study and Alexis Sullivan assisted with some of the sample collection and microsatellite characterization.

Acknowledgements

Many people were instrumental in the completion of this dissertation. I would first like to thank my advisor, Oliver Gailing, for initially encouraging me to begin this degree and providing financial and intellectual support. I especially want to thank him for giving me great freedom and understanding in completing this work despite my chronic migraines. Great thanks to my committee members, Victor Busov, Erik Lilleskov, and Thomas Snyder, all of whom continually challenged me to think more like a scientist and willingly shared their knowledge and advice even when I was less than receptive.

I must also extend my gratitude to the many fellow students, both graduate and undergraduate, who helped bring this dissertation to fruition and made this experience bearable. To Kristina Flesher, Erin Hickey, and Joanna Rogers for their invaluable help with the more mundane laboratory and greenhouse work. To all of my lab mates over the years, Yang Li, Sandra Owusu, Sudhir Khodwekar, Alexis Sullivan, Roba Bdeir, Kristina Flesher, Sirikorn Khumwan, Jessica Goodin, Cengiz Yucedag, and Bader Awama; without the support and practice that comes with working in a busy (constantly buzzing) lab, I would not have grown as much in my knowledge of procedures, trouble shooting and camaraderie. Specifically, Yang, "Jessie" Li for keeping me company that first lonely year. Alexis Sullivan graciously shared some of her population samples and data, let me bounce thoughts off of her, and has remained a collegial contact. Sandra and Sudhir always checked on me when I (frequently) went MIA holed up writing towards the end, and have become my friends as well. Anna Hess has been and is a good friend and colleague whom I've had many intellectual and not so intellectual conversations with over good food and beer. May this collaboration continue to be a part of both of our future endeavors.

I must also extend my heartfelt thanks to my close friends and family whom I would not have gotten this far without; their love, support, and laughter have both prevented an early death and created fond memories. To my sisters, Rebecca and Sarah, whom

I couldn't possibly live without; to my mother, Jill, who has always believed in me and pushed me to live up to my potential; to my girlfriends, Dianne, Lynn, Danielle, and Nicole, who have and continue to provide necessary stress relief and fabulous life experiences. Finally, I want to end this with a special thanks to my wonderful and amazing husband, Jon, who has stuck by me all these years and always supported me and my abilities even when I wasn't sure I had any. He has been the best partner in work, play, and crime that I could have ever hoped for. I love you.

Abstract

Forest trees, like oaks, rely on high levels of genetic variation to adapt to varying environmental conditions. Thus, genetic variation and its distribution are important for the long-term survival and adaptability of oak populations. Climate change is projected to lead to increased drought and fire events as well as a northward migration of tree species, including oaks. Additionally, decline in oak regeneration has become increasingly concerning since it may lead to decreased gene flow and increased inbreeding levels. This will in turn lead to lowered levels of genetic diversity, negatively affecting the growth and survival of populations. At the same time, populations at the species' distribution edge, like those in this study, could possess important stores of genetic diversity and adaptive potential, while also being vulnerable to climatic or anthropogenic changes. A survey of the level and distribution of genetic variation and identification of potentially adaptive genes is needed since adaptive genetic variation is essential for their long-term survival.

Oaks possess a remarkable characteristic in that they maintain their species identity and specific environmental adaptations despite their propensity to hybridize. Thus, in the face of interspecific gene flow, some areas of the genome remain differentiated due to selection. This characteristic allows the study of local environmental adaptation through genetic variation analyses. Furthermore, using genic markers with known putative functions makes it possible to link those differentiated markers to potential adaptive traits (e.g., flowering time, drought stress tolerance). Demographic processes like gene flow and genetic drift also play an important role in how genes (including adaptive genes) are maintained or spread. These processes are influenced by disturbances, both natural and anthropogenic. An examination of how genetic variation is geographically distributed can display how these genetic processes and geographical disturbances influence genetic variation patterns. For example, the spatial clustering of closely related trees could promote inbreeding with associated negative effects (inbreeding depression), if gene flow is limited. In turn this can have

negative consequences for a species' ability to adapt to changing environmental conditions. In contrast, interspecific hybridization may also allow the transfer of genes between species that increase their adaptive potential in a changing environment.

I have studied the ecologically divergent, interfertile red oaks, *Quercus rubra* and *Q. ellipsoidalis,* to identify genes with potential roles in adaptation to abiotic stress through traits such as drought tolerance and flowering time, and to assess the level and distribution of genetic variation. I found evidence for moderate gene flow between the two species and low interspecific genetic differences at most genetic markers (Lind and Gailing 2013). However, the screening of genic markers with potential roles in phenology and drought tolerance led to the identification of a *CONSTANS*-like (*COL*) gene, a candidate gene for flowering time and growth. This marker, located in the coding region of the gene, was highly differentiated between the two species in multiple geographical areas, despite interspecific gene flow, and may play a role in reproductive isolation and adaptive divergence between the two species (Lind-Riehl et al. 2014). Since climate change could result in a northward migration of trees species like oaks, this gene could be important in maintaining species identity despite increased contact zones between species (e.g., increased gene flow). Finally I examined differences in spatial genetic structure (SGS) and genetic variation between species and populations subjected to different management strategies and natural disturbances. Diverse management activities combined with various natural disturbances as well as species specific life history traits influenced SGS patterns and inbreeding levels (Lind-Riehl and Gailing submitted).

Introduction

Forest trees make up 80% of the world's biomass and contain more carbon in their biomass and soils than the atmosphere (Pan et al. 2014). They are integral to the maintenance of biological diversity, soil and water resource protection, and carbon sequestration in addition to their aesthetic, cultural and economic value (FAO 2011). They are largely outcrossing and hold large amounts of genetic diversity making them capable of adapting to changing environments over their long life spans (Petit and Hampe 2006). However, over the past 150 years anthropogenic impacts (e.g., deforestation, fragmentation, etc.) have devastated forests across the globe (FAO 2011). To complicate this, climate change is predicted to accelerate the pace of environmental changes such as increased droughts and floods, warmer temperatures, increased insect and pest attacks and overall more frequent and extreme natural disturbances (Aitken et al. 2008; Bréda et al. 2006; Woodall et al. 2009). The success of forest tree responses will be dependent on the genetic stores available to adapt to changing local conditions (Aitken et al. 2008). Thus, it becomes imperative to assess existing genetic diversity and understand how forest trees currently survive and adapt to local environmental conditions in order to develop management and conservation plans that will be effective in the future. Specifically, genetic information is lacking at the regional and population level for many temperate forest trees in North America. This is concerning since climate change is likely to display varying impacts dependent on location (Aitken et al. 2008; Gibson et al. 2009). In fact, a recent study analyzed geographic patterns of genetic variation in *Quercus lobata* and how this might impact the species' ability to survive the effects of climate change (Sork et al. 2010). The authors found that populations appeared to be adapted to local environmental conditions through the examination of historical colonization patterns and the correlation between multivariate nuclear genetic variation and climatic variation. Regional differences with regard to the impact of climate change will lead to variable stressors for populations of *Q. lobata*, with some populations facing a greater likelihood of extinction than others. For example, temperature increases will

result in greater spatial shifts in the northern range versus the southern range of *Q. lobata*. As forest trees, including oaks, often encompass large environmentally heterogeneous ranges, this finding is likely applicable to other tree species as well (Kremer et al. 2012; Petit and Hampe 2006). In order to accurately predict the impact of climate change on forest trees, local adaptation and the distribution and amount of genetic variation within and between species needs to be understood more clearly.

The ability of populations to adapt to changing environmental conditions depends on genetic variation in adaptive genes (Aitken et al. 2008). However, forest trees present a challenge for genetic studies like breeding and genetic experiments (e.g., linkage and Quantitative Trait Loci (QTL) mapping, map-based cloning, genome sequencing) due to their long generation times and large genomes, respectively. Fortunately, with the advent of PCR-based markers (e.g., microsatellites, single nucleotide polymorphisms or SNPs), genomic resources like cDNA libraries and next generation sequencing advancements, studying non-model organisms like forest trees is becoming more feasible (Bodénès et al. 2012; Ekblom and Galindo 2011; Ellis and Burke 2007; Neale and Kremer 2011). The most commonly used markers, microsatellites are tandemly repeated segments of DNA that are generally evenly dispersed throughout the genome and highly polymorphic making them ideal for examining genetic variation in populations and species (Weising 2005). Use of SNPs has recently begun to increase as well. SNPs represent single base substitutions at a particular location in the genome, are found frequently in genes and are often biallelic making them useful for DNA and RNA profiling and genetic variation studies (Weising 2005). In particular, the development of Expressed Tag Sequence (EST) libraries has provided a way to mine for microsatellites and SNPs directly linked to genes. Existing and expanding EST resources thus present an opportunity to relatively quickly and inexpensively develop gene-associated microsatellite markers. These EST-Simple Sequence Repeats (EST-SSRs) are generally more conserved than traditional microsatellite markers and are often transferable among species within genera and even sometimes between genera (Bodénès et al. 2012;

Ellis and Burke 2007). EST-SSRs are not only used to examine within and between population genetic diversity and structure, but can also be used to link phenotypic traits with potentially underlying genes. Furthermore, if these markers are genetically mapped, genomic regions of interest such as those under selection or involved in reproductive isolation can be identified and compared between species (Bodénès et al. 2012).

The use of EST-SSRs is of great interest for genetic studies in forest trees because they link genetic variation to potential adaptive traits. Forest trees possess high genetic diversity within species and populations as well as specific environmental adaptations, but at the same time many species exhibit large amounts of gene flow within and among populations and species (Hamrick et al. 1992). High gene flow might prevent local adaptation unless genomic areas involved in that adaptation are under strong selection (Via 2012). Markers that show higher differentiation than expected under neutrality (outlier loci) between species or populations with different environmental niches could point towards a gene involved in local adaptation. Using EST-SSRs to identify outlier loci, which potentially represent or are linked to candidate genes, provides a targeted search method for markers that have putative functions related to environmental adaptations in species that do not have a sequenced genome. For example, a study looking at populations of sunflowers (*Helianthus annuus*) with differences in adaptations to drought and salt conditions using EST-SSRs (some with putative functions in drought and salt tolerance) found that a substantial proportion of the outliers detected are linked to genes with putative abiotic stress response functions (Kane and Rieseberg 2007). Currently studies using EST-SSRs to detect outliers are limited, but growing in both animals (Shikano et al. 2010; Vasemagi et al. 2005; Vilas et al. 2010) and plant species (Kane and Rieseberg 2007; Lind-Riehl et al. 2014; Scotti-Saintagne et al. 2004b; Sullivan et al. 2013). These outlier loci are candidates for further investigation through sequencing and genetic mapping to examine the molecular basis for differentiation and confirm potential involvement in local adaptation. For example, Wood et al. (2008) followed

up a outlier scan that used Amplified Fragment Length Polymorphisms (AFLPs) in the marine gastropod, *Littorina saxatilis*, by analyzing sequence variation in population samples at and close to the outlier genomic location. Their results indicated that two outliers contained large indels and the sequence variation in these corresponded to populations in two different environments, high shore and mid shore. In oaks, Vornam et al. (2011) recently characterized a full length dehydrin gene (involved in abiotic stress response) and its promoter region and assessed the allelic variation in natural populations of *Q. petraea* discovering nucleotide diversity within the coding region. This gene was not originally an identified outlier, but *Q. petraea* is known to reside on drier soils than other interfertile European oak species such as *Q. robur* (Curtu et al. 2007a). Studies to associate observed nucleotide diversity with phenotypic variation in larger populations along environmental gradients are still needed to confirm the involvement of these candidate genes in local adaptation. For example, a candidate gene was identified in the outlier study (Lind-Riehl et al. 2014). The same gene was associated with bud burst along an altitudinal gradient in a *Q. petraea* provenance trial (Alberto et al. 2013).

Oaks as a model

Oaks play key roles in temperate North American forests by providing habitat and food for many wildlife species and comprise a large proportion of trees in the Northern Hemisphere (Aldrich and Cavender-Bares 2011; McShea et al. 2007). Additionally oaks are economically valuable for their use in wood and paper products. Oaks possess many life history characteristics associated with high levels of genetic diversity including an outcrossing mating system, wind pollination, high fecundity and generally encompassing large ranges (Aldrich and Cavender-Bares 2011; Hamrick et al. 1979). Oak regeneration rates have been declining over the past 60 years potentially due to fire suppression, elevated herbivory, and competition with invasive plant species (Huebner 2003; Lorimer 1993). However, generally studies have shown maintenance of high levels of genetic diversity despite low regeneration

rates likely due to the high outcrossing nature of oaks (Aldrich et al. 2005; Muir et al. 2004). Despite this, ageing populations are still a concern. Fragmented populations or those at the species' distribution edge are of particular concern because of the potential for increased inbreeding. This can lead to decreased genetic variation, which leading to negative effects on growth and survival (e.g., inbreeding depression) and the ability to adapt to changing environmental conditions (Muir et al. 2004).

Interspecific gene flow and selection

Oaks (*Quercus* spp.) provide an excellent model to study genetic variation patterns and local adaptation. Oaks hybridize frequently with other closely related oak species (Rushton 1993). Consequently oaks have displayed low levels of genetic differentiation between related species, and most genetic diversity is distributed within species and populations (e.g., (Moran et al. 2012; Scotti-Saintagne et al. 2004b). Despite these often high levels of interspecific gene flow they consistently maintain their species identity and different local adaptations in sympatry (Gailing and Curtu 2014; Whittemore and Schaal 1991). In both North America and Europe, oaks grow over a wide range of environmental conditions (Gailing et al., 2009). For example, the interfertile European oak species, *Q. robur* and *Q. petraea*, display preferences with regard to edaphic conditions, with *Q. petraea* growing on drier soils than *Q. robur* (Curtu et al. 2007b). Local adaptations to these environmental conditions may be responsible for this seeming paradox since gene flow is expected to prevent genetic differentiation. In fact, most regions of the genome are homogenized by gene flow and exhibit low interspecific differentiation. However, a few genomic regions display strong interspecific differentiation and resistance to gene flow (Scotti-Saintagne et al. 2004b). This could be the result of strong divergent selection on those genes that are involved in species specific local adaptations (Curtu et al. 2007a; Gailing and Curtu 2014; Scotti-Saintagne et al. 2004b).

Basic genetic variation studies and outlier screens using nSSRs and EST-SSRs can reveal areas that show high interspecific differentiation as described in the previous section. This provides a way to focus research efforts on particular genomic regions that may be of importance for local adaptation in oaks, while using neutral markers (nSSRs) as a reference. EST-SSRs are particularly suited for this endeavor for two reasons. First, plentiful EST resources already exist for white oaks (Durand et al. 2010; Bodénès et al. 2012) and are now available for red oaks (Fageaceae Genomics Project: http://www.fagaceae.org/, see Gailing and Curtu 2014). Since the markers are located in expressed genes, outliers can be directly linked to genes with known functions, and variation in these genes might be associated with specific traits. For example, genic markers could be selected with potential functions in drought tolerance or flowering time, two traits that may play a role in local adaptation and reproductive isolation (Petit et al. 2013; Gailing and Curtu 2014). EST-SSRs are highly transferable among species within the same genera (or even family), so they can be applied to discover outliers possibly under divergent selection between populations within species as well as between interfertile species pairs. Currently, high density genetic linkage maps exist in *Q. robur* to which many EST-SSRs are currently mapped. These can be used to identify the genomic location of potential outlier loci and allow comparative studies between species.

Oaks are also not exempt from the effects of climate change. Increased drought and fire events may challenge their long-term survival and productivity. Additionally a northward migration of some oak species in North America is predicted (Woodall et al. 2009) that may lead to increased species competition and potentially increased hybridization with other oak species. High genetic diversity within oaks and the potential to share genes through hybridization may help them to adapt to future conditions (Gailing and Curtu 2014), but limited seed dispersal and competition for space with other species may slow down adaptation (Savolainen et al. 2007). The ability for forest trees, including oaks, to adapt to climate change will be dependent on existing genetic variation as well as dispersal and establishment rates (Savolainen

et al. 2007). Thus it becomes important to understand how the genetic variation in populations is arranged spatially.

Level and distribution of genetic variation

Oaks often represent foundation species in their ecosystems impacting nutrient and energy fluxes, food webs and biodiversity. The loss foundation species can detrimentally impact ecosystem functioning (Ellison et al. 2005). Natural regeneration in oaks has markedly decreased over the last 60 years. Increased competition with other more shade tolerant tree species, fire suppression, and increased herbivory may be responsible for this trend (Lorimer 1993). To add to this, anthropogenic influences are also leading to increased fragmentation of populations through urbanization of the landscape, which can lead to decreases in genetic variation and gene flow (McShea et al. 2007). Most studies have found high and similar levels of genetic variation in oak populations subject to different management regimes as well as different natural and anthropogenic environmental impacts such as fragmentation, which has been attributed to the highly outcrossing nature of these species (Cottrell et al. 2003; Craft and Ashley 2007; Muir et al. 2004). However, this may not be the case in the future; as displayed by Sork et al. (2002), pollen donors in a declining *Q. lobata* population have decreased since 1944 showing that there may be limits to gene flow from other populations reaching fragmented or range edge populations as time passes. Similar observations have been made for *Q. ellipsoidalis* populations in the Baraga Plains of Upper Michigan (Lind and Gailing 2013). Other studies have noted lower numbers of individuals belonging to smaller size classes in both *Q. rubra* (Aldrich et al. 2005) and *Q. petraea* (Muir et al. 2004) stands. This means that the standing genetic diversity in the populations may not persist over time. Specifically, with fewer individuals contributing to future generations, spatial clustering of related trees could result in inbreeding with negative effects on survival and growth (inbreeding depression). Spatial genetic structure (SGS) analysis allows the characterization of genetic variation in a two dimensional space. Examining SGS

can reveal patterns of gene flow, genetic drift, and genetic variation and distribution (Epperson 1992). An understanding of how the genetic variation of populations is spatially arranged at a fine scale allows current and future conservation and management plans to prepare for potential long term impacts on standing genetic diversity.

Given the importance of oaks both ecologically and economically and the threats of climate change and continued anthropogenic impacts, is it important to understand the standing genetic diversity and its distribution at the population level as well as the role that diversity plays in local environmental adaptation and gene flow within and between species.

Red oak species in North America

There are approximately 400 species of oaks across five different continents that occupy a large proportion of the Northern Hemisphere's forests (Aldrich and Cavender-Bares 2011). While local adaptation and genetic spatial structure and variation in European white oaks (section *Quercus*) have been studied extensively (e.g., (Bacilieri et al. 1996; Brendel et al. 2008; Cottrell et al. 2003; Curtu et al. 2007a, 2009; Jimenez et al. 2004; Kremer et al. 2002; Lepais et al. 2009; Lexer et al. 2006; Mariette et al. 2002; Petit et al. 2002; Salvini et al. 2009; Scotti-Saintagne et al. 2004a; Streiff et al. 1998; Streiff et al. 1999; Valbuena-Carabana et al. 2007)), studies in North American temperate oak species including the red oaks (section *Lobatae*) are more limited (Cavender-Bares and Pahlich 2009; Craft et al. 2002; Dodd and Afzal-Rafii 2004; Moran et al. 2012; Peñaloza-Ramírez et al. 2010; Sork et al. 1993). There are approximately 195 species of red oaks which are only found in the Americas (Aldrich and Cavender-Bares 2011). They have been shown to exhibit more extensive introgression and hybridization as compared to the white oaks (Guttman and Weigt 1989). Furthermore, current genetic information on the basis of species differences is much less as compared to the well-studied white oaks, but their ecological and economic importance is just as high (Aldrich and Cavender-Bares

2011). As such, more research on the delineation and maintenance of species boundaries including information about genetic variation, gene flow, local adaptation and the distribution of these genetic resources is warranted.

Quercus rubra L. is a foundation species within North American temperate forests (Ellison et al. 2005). It has a wide distribution range that extends from the east coast of North America to the Mississippi River (Barnes 2004). In contrast, the distribution of *Q. ellipsoidalis* E. J. Hill is more restricted and fragmented within the Midwestern Great Lakes region of the United States. While *Q. rubra* resides on north facing and bottom slopes with mesic soils, *Q. ellipsoidalis* is the most drought tolerant red oak species and resides on sandy dry barrens habitats (Barnes 2004; Hipp 2010). These two species have different adaptations to drought as displayed by differences in morphological and physiological characteristics. *Q. ellipsoidalis* has thicker leaves and a deeper penetrating root system, and displays higher rates of photosynthesis at low leaf water potentials, and later stomatal closure. (Abrams 1990). Both are outcrossing wind pollinated species with primarily gravity dispersed seeds (Burns and Honkala 1990). Evidence for hybridization was previously indicated by morphological and isozyme studies (Hokanson et al. 1993; Jensen et al. 1993). However species delineation and extent of hybridization was insufficiently understood as isozymes possess far less variation as compared to microsatellite markers. Additionally, leaf morphology is often not a reliable way to identify species and their hybrids due to phenotypic plasticity (Blue and Jensen 1988; Tomlinson and Jensen 2000). In fact, morphological differentiation does not always correspond to molecular differentiation, as found in a hybrid zone of Mexican red oaks (González-Rodríguez et al. 2004). Using genetic characteristics at molecular markers has also been shown to resolve species relationships better than morphological methods (Curtu et al. 2007a; Gailing et al. 2012; Lind and Gailing 2013).

The ecological niche differences with respect to drought tolerance combined with the presence of interspecific gene flow makes this species pair ideal for studying local

adaptation and genetic variation patterns in red oaks. The use of genic markers (EST-SSRs) will allow simultaneous identification of genes under divergent selection (outlier loci), species and hybrids as well as estimate levels of interspecific gene flow. Additionally, the populations under study are at the northern range limit of red oaks and may harbor important genetic stocks (Gibson et al. 2009). This has been displayed for other peripheral populations of European oaks with specific local adaptations (Jimenez et al. 2004; Lorenzo et al. 2009). Northern range edge populations have been shown to genetically differentiate from core populations and sometimes hold greater genetic diversity that the core populations. Genetic distinctness and diversity in northern peripheral populations is likely due to limited gene flow from core populations as well as local selective pressures, thus making these populations particularly evolutionarily dynamic (Gibson et al. 2009). In addition, hybridization between these two species could increase due to the predicted climatically induced northward migration that may allow for a novel source of diversity and adaptive potential. For example, increased genetic variation in hybrid zones of the flowering plant species, *Cyclamen spp.*, has been found in northern peripheral populations (Thompson et al. 2010). Alternatively these populations may also be more susceptible to losses of genetic diversity through inbreeding, since their distribution is fragmented and there is increased competition with other species for space. Anthropogenic factors such as land use changes may exacerbate this further (Muir et al. 2004; Gibson et al. 2009). The assessment of the spatial distribution of genetic variation and potential levels of gene flow provides a way to assess the formation of family structures, level of inbreeding, and genetic diversity. This information coupled with historical knowledge of environmental disturbances, both natural and anthropogenic, can provide a way to determine the impact of those disturbances on genetic variation and distribution.

Aims and objectives

The ecologically divergent interfertile red oaks, *Q. rubra* and *Q. ellipsoidalis,* were the subject of my dissertation. My main goals were to identify genes with potential roles in adaptation to environmental stresses and to assess the level and distribution of genetic variation essential for their long-term survival. Specifically I have chosen to address these goals through three objectives.

1. To assign species and identify hybrids and introgressive forms using both genic (EST-SSRs) and neutral microsatellite markers (nSSRs), and assess within and between species genetic diversity. This is described in the first part of my thesis entitled "Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations." (Lind and Gailing 2013)
2. To identify outlier loci potentially under divergent selection using EST-SSRs. The second part, entitled "Evidence for selection on a *CONSTANS*-like gene between *Quercus rubra* and *Q. ellipsoidalis*," describes the findings of this study. (Lind-Riehl et al. 2014)
3. To assess spatial genetic structure of managed and unmanaged *Q. rubra* populations and managed *Q. ellipsoidalis* populations. This is described in the third part, entitled "Fine scale spatial genetic structure of two red oak species, *Quercus rubra* and *Q. ellipsoidalis*." (Lind-Riehl and Gailing 2014)

Material and methods

Sampling

Eleven morphologically identified *Q. rubra* and *Q. ellipsoidalis* populations in the Western Upper Peninsula of Michigan (USA) were sampled for genetic assignment and variation analyses (Lind & Gailing 2013, Table 1). Four more populations of *Q. rubra* and *Q. ellipsoidalis* were sampled from Wisconsin to provide two geographically distant population pairs to distinguish between isolation by distance and isolation by adaptation in the outlier screening (Lind-Riehl et al. 2014, Table 1).

Populations used in the spatial genetic structure analysis included nine previously collected populations: four incidentally managed *Q. ellipsoidalis* populations, four managed *Q. rubra* populations and three unmanaged (primary forest) *Q. rubra* populations. An additional unmanaged (primary forest) *Q. rubra* population was collected so we had an equal number of managed and unmanaged populations (Lind-Riehl and Gailing 2014, Table 1). Leaf samples for DNA analysis and GPS coordinates were taken for each tree. Sampling consisted of only adult trees occupying a dominant or co-dominant canopy with a focus on identifying populations growing on contrasting sites or along environmental gradients. DNA was isolated from leaf samples using the Qiagen DNeasy Plant Mini and 96 Plant Kit (Qiagen, Valencia, CA) and corresponding procedures.

Marker analyses

A set of 8 nuclear microsatellite markers (nSSRs) developed in *Q. rubra* were used to represent neutral variation (Aldrich et al. 2002; Steinkellner et al. 1997; Sullivan et al. 2013). A set of 7 genic microsatellite markers (EST-SSRs) were adapted from *Q. robur* (Durand et al. 2010) for use in *Q. rubra* and *Q. ellipsoidalis* (Lind and Gailing 2013). These 15 microsatellite markers were used to perform genetic assignment of individuals and to assess genetic variation within and among the populations as well as to perform the spatial genetic analyses. This 15 marker set covers eight of the 12 *Q. robur* linkage groups. An additional set of 21 EST-SSRs (15 from Sullivan et al. 2013) with putative functions in drought tolerance, flowering time and other functions were adapted to *Q. rubra* for use in the outlier screens (Lind-Riehl et al. 2014). The 44 marker set used in the outlier screens covers 10 of the 12 *Q. robur* linkage groups. Polymerase chain reaction and electrophoretic separation were performed according to Lind and Gailing (2013).

Statistical analyses

Various population genetic programs, including GeneAlEx 6.41 (Peakall and Smouse 2006), GenePop 4.2 (Raymond and Rousset 1995) and FSTAT (Goudet 1995), were used to assess basic genetic variation parameters including observed and expected heterozygosity, average number of alleles, the inbreeding coefficient, pairwise F_{ST}, and to perform Hardy Weinberg equilibrium tests. STRUCTURE (Pritchard et al. 2000), a model based clustering method that utilizes a Bayesian inference approach, was applied to delineate species and population groups and assign individuals to species as well as to identify hybrids and introgressive forms. Both a model based method using coalescent simulations (LOSITAN, Antao et al. 2008) and a model-independent statistic (LnRH, Schlötterer and Dieringer 2005) were applied to minimize false positive detection of outliers in the outlier screening. Spatial autocorrelation analyses (SPAGeDI, Hardy and Vekemans 2002) were performed to assess the degree of dependency of genetic variation on geographic space.

Summary of results

"Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers" by Lind, JF and Gailing O, *Tree Genetics and Genomes* 9:707-722

We found evidence for low to moderate gene flow between the two species and low interspecific genetic differences at most genetic markers. Despite overall low interspecific genetic differentiation, individuals were largely grouped into two distinct clusters corresponding to the two species. A comparison between seedling populations and adult populations showed similar levels of hybrids and introgressive forms indicating that gene flow is likely low to moderate. Initial outlier testing using Arlequin (Excoffier and Lischer 2010) indicated GOT021 as an outlier with a putative function as a histidine kinase 4 between *Q. rubra* and *Q. ellipsoidalis*

populations in the Ford Center Baraga Plains region. It was also identified as an outlier by a second outlier screening method, LOSITAN (Antao et al. 2008), between the same populations (Sullivan et al. 2013). Histidine kinases have been shown to be involved in abitoic stress response in *Arabidopsis* (Desikan et al. 2008; Tran et al. 2007).

"Evidence for selection on a *CONSTANS*-like gene between two red oak species" by Lind-Riehl JF, Sullivan AR, and Gailing O, *Annals of Botany* 113: 967-975

An additional outlier screening with increased numbers of genic markers (36 EST-SSRs and 8 nSSRs) including those with putative roles in phenology and drought tolerance led to the identification of a *CONSTANS*-like (*COL*) gene. This gene was highly differentiated between the two species despite interspecific gene flow in all population pairs from three different geographical regions. Specifically, this *COL* gene has putative functions in flowering time and growth. A seedling common garden trial also showed differences in flowering time between the two species (Gailing 2013). Thus, differences in flowering time may also be involved in the low to moderate effective gene flow observed in the genetic assignment analyses. This suggests that the *Quercus COL* gene may play a role in reproductive isolation and adaptive divergence between the two species.

"Fine-scale spatial genetic structure of two red oak species, *Quercus rubra* and *Q. ellipsoidalis*" by Lind-Riehl JF and Gailing O *Plant Systematics and Evolution* DOI: 10.1007/s00606-014-1173-y

While no consistent differences in spatial genetic structure (SGS) were noted between unmanaged and managed *Q. rubra* populations, diverse management activities did show differences in SGS patterns. For example, even-aged management exhibited no significant SGS, while uneven-aged management showed significant SGS up to 36 m. This is likely a result of mating between overlapping generations in the uneven-aged stands. Between species comparisons showed that *Q. ellipsoidalis* displayed much stronger SGS than *Q. rubra*. The *Q. ellipsoidalis*

populations in this study are largely isolated, and one of them (FC-E) showed evidence for high levels of inbreeding. Reduced genetic diversity could impact the future ability of this species to adapt to changing environmental conditions.

Additional related studies

These two studies were not a part of my dissertation, but I played a role in their completion and was listed as an author on the publication of the results.

"Development and characterization of genomic and gene-based microsatellite markers in North American red oak species" by Sullivan AR, Lind JF, McCleary TS, Romero-Severson J, Gailing O in *Plant Molecular Biology Reporter* 30: 231-239

This study validated four new nSSRs in *Q. rubra* as well as adapted 21 EST-SSRs from *Q. robur* for use in *Q. rubra*. These markers were subsequently used in all three parts of my dissertation. Initial outlier screens were also performed that included one population pair of *Q. rubra* and *Q. ellipsoidalis* from the Ford Research Forest and Baraga Plains region. GOT021 was identified as an outlier between that population pair and has a putative function as a histisdine kinase-4.

"Leaf morphological and genetic differentiation between *Q. rubra* L. and *Q. ellipsoidalis* E. J. Hill populations in contrasting environments" by Gailing O, Lind JF, Lilleskov E in *Plant Systematics and Evolution* 275: 1533-1545

This was a sister study to the first part of my dissertation that assessed morphological characters in the same populations. The goal was to test whether the large continuous leaf morphological variation observed in the field was a result of phenotypic plasticity within species or between two interfertile species and compare the morphological assignment to the genetic assignment (Gailing et al. 2012). This study found that the morphological variation was largely due to phenotypic plasticity within *Q. rubra* and that although *Q. ellipsoidalis* was differentiated from *Q. rubra*

at the 17 morphological characters analyzed, the genetic assignment analysis delineated between the species more clearly.

General discussion

Understanding genetic variation patterns provides a way to effectively conserve and manage oak populations for long term survival and productivity. Since climate change is predicted to impact migration patterns, water availability, and temperatures differently across the landscape of North American forests (Aitken et al. 2008; Sork et al. 2010), it is important to develop regionally specific sources of data on the growth and survival of important trees species like oaks. Specifically, range edge population may harbor important genetic resources that may aid in future adaptation of these species to climatic changes (Gibson et al. 2009). These studies have provided a solid base of information on genetic variation patterns and the impacts of various management practices for populations of *Q. rubra* and *Q. ellipsoidalis* at the northern edge of their distribution range. Additionally, a potential candidate gene (*COL*) has been identified that may be involved in adaptive species differences. However, more work is certainly needed and further studies including direct analysis of gene flow and sequence variation analysis in the *COL* gene for the two species are planned. Whole genome scans using recently developed genomic resources for *Q. robur* (a closely related white oak species) and *Q. rubra* (Gailing and Curtu 2014) as well as associations of potential outliers with phenotypic variation will also provide a better understanding of how these genetic differences translate in the environment (Sork et al. 2010).

Genetic variation and distribution

A set of genic and nuclear microsatellite markers were successfully applied to delineate species boundaries between the interfertile *Q. ellipsoidalis* and *Q. rubra* and to identify hybrids and introgressive forms. Generally high levels of genetic diversity were found in populations of both species whether managed or unmanaged.

This is consistent with the wind pollinated, highly outcrossing nature of oaks (Hamrick et al. 1992) and in correspondence with other studies within Fagaceae (Buiteveld et al. 2007; Craft and Ashley 2007; Muir et al. 2004). However, potential inbreeding was noted in at least one *Q. ellipsoidalis* population. Stronger spatial genetic structure (SGS) was also noted in the *Q. ellipsoidalis* populations indicating more extensive family structures. The geographic isolation of the *Q. ellipsoidalis* populations combined with the harsher environmental conditions (dry sandy pine barrens) may be responsible for the higher level of inbreeding and lower fecundity observed as well as slightly, but significantly less genetic diversity and stronger SGS when compared to adjacent *Q. rubra* populations. In another similar interfertile oak species pair, *Q. robur* and *Q. petraea*, more pronounced SGS was found in *Q. petraea* populations. The more drought tolerant *Q. petraea*'s range is more limited than *Q. robur*'s (Streiff et al. 1998). Another study on a related species with similar life history traits, *Fagus sylvatica*, found that fragmentation of populations resulted in a decrease in rare alleles (Paffetti et al. 2012). In contrast, Muir et al. (2004) did not find reduced genetic variation in fragmented populations of *Q. petraea* in Ireland, but noted low size class diversity in the stands and that naturally regenerated seedlings were also susceptible to animal herbivory. Likewise, Aldrich et al. (2005) found high levels of genetic diversity within an old growth *Q. rubra* forest, but observed that the smaller diameter size classes were conspicuously missing from the stands. Since trees can hold high levels of genetic variation, the impacts of fragmentation and human management regimes may not be apparent for several generations. Case in point, Jump and Peñuela (2006) studied the long-term (>600 years) impact of fragmentation in *Fagus sylvatica* populations and found that genetic diversity was reduced. The *Q. rubra* and *Q. ellipsoidalis* populations still show high levels of genetic diversity, but the isolated *Q. ellipsoidalis* populations show some signs of inbreeding and stronger SGS. The long term effects of fragmentation coupled with apparent decreases in natural regeneration rates due to increased competition with other species and herbivory (both of which will be exacerbated by

climate change) may leave these oak populations less equipped to deal with future environmental changes.

Gene flow and outlier scans

Gene flow was noted to be relatively low between adjacent *Q. rubra* and *Q. ellipsoidalis* populations in the Ford Research Forest and Baraga Plains area, with similar numbers of statistically significant hybrids and introgressive forms in both seedling and adult populations. Our outlier screening discovered a putative *CONSTANS*-like (*COL*) gene with functions in flowering time and growth containing a microsatellite (FIR013) that was highly differentiated between the species in all three separate geographical locations. Additionally, this marker was not differentiated within species, even from different geographical regions, providing strong evidence for this marker being under divergent selection or closely linked to a gene under divergent selection. In greenhouse experiments, *Q. ellipsoidalis* seedlings were found to have significantly later budburst (a proxy for flowering time, see Chesnoiu et al. (2009) than *Q. rubra* seedlings (Gailing 2013) and low effective gene flow may be due to these differences in flowering time. Additionally, the *Q. rubra* populations were located in more mesic soils mixed with maple (*Acer spp*.) and/or white pine (*Pinus strobus*), while the *Q. ellipsoidalis* populations were growing on very dry sandy plains with jack pine (*Pinus bansksiana*). Slower growth characteristics have been shown to be characteristic of species residing on dry sandy outwash plains like those inhabited by *Q. ellipsoidalis* populations (Motzkin et al. 2002). Furthermore other studies have shown a putative role for this gene in flowering time and growth (Herrmann et al. 2010; Hsu et al. 2012), making *COL* a candidate gene that may be involved in divergent selection and adaptive speciation between *Q. ellipsoidalis* and *Q. rubra*. This *COL* gene may be a "magic gene" that simultaneously acts as a reproductive barrier (i.e., differences in flowering time) and is under divergent selection, and may lead to the evolution of adaptive species differences (Servedio et al. 2011). The ecological species concept suggests that

ecological specialization is the mechanism by which species like oaks maintain their species despite recurrent gene flow (Van Valen 1976), and divergent speciation through local adaptation is now viewed as a legitimate evolutionary path since genetic evidence for this theory has been building (see (Schluter 2009)). Further examination of this candidate gene will need to be explored to ascertain its role in adaptive divergent speciation of course, such as examining sequence variation in the gene and its promoter region in contrasting populations and association studies between contrasting populations to connect phenotypic and genetic variation.

Outlook

These three studies have provided genetic variation data and its distribution for the species pair *Q. rubra* and *Q. ellipsoidalis* that has provided a better understanding of the effective reproductive isolation between these interfertile species and a model to identify genes under divergent selection that may be involved in adaptive species differences. However, continued research is needed. Given the differences in drought tolerance between the two species and the identification of a potential outlier involved in abiotic stress response (GOT021) in the Ford Research Forest and Baraga Plains region, examining ecophysiological traits such as water use efficiency will begin to shed light on whether this genetic difference in drought tolerance is significant physiologically. Preliminary analyses of water use efficiency and soil moisture differences between *Q. rubra* and *Q. ellipsoidalis* populations in that region are in progress to better understand the role drought tolerance has in the formation of ecological niches for these two species. Additionally, the gene flow estimates obtained in this study could be affected by two factors. First, the populations were not sympatric but rather adjacent within the range of gene flow, meaning that we do not have a complete understanding of how these two species interact when growing sympatrically. Second, indirect measures of gene flow, like those carried out in this thesis, have well documented shortcomings since the assumptions of the models used often do not reflect natural conditions in populations (e.g., equilibrium between

genetic drift and migration, equal contribution of migrants from all populations) (Burczyk et al. 2004). Since the inception of this study, a few populations of truly sympatric *Q. rubra* and *Q. ellipsoidalis* have been identified and sampled. Species assignment and hybrid identification along with parentage analysis are under way for those newly identified populations (PhD study, Sudhir Khodwekar). Parentage analysis can provide a direct, non-model based estimate of gene flow without making assumptions about equal contribution of migrants from all populations and equilibrium between genetic drift and migration. In addition to providing better estimates of contemporary interspecific gene flow, parentage analysis could provide estimates of genetic differences in adaptive traits. Since the species are nearly fixed on alternative alleles, a reciprocal transplant experiment with *Q. rubra*, *Q. ellipsoidalis* and hybrid seedlings with different genotypes for *COL* could be conducted. If this gene is involved in adaptive species divergence with gene flow, an increase of species specific alleles from the seedling to adult generation along with higher growth and survival of *Q. rubra* and *Q. ellipsoidalis* seedlings with species specific alleles in the parental environments would be expected.

Our outlier screening identified a marker (FIR013) located in the coding region of a *COL* gene. Further work will need to be conducted to elucidate the importance of this gene in adaptive species divergence. For example, the whole gene and possibly its promoter region could be sequenced to characterize sequence variation in the gene and/or its promoter region among contrasting populations. Alternatively, since variation for adaptive traits is likely to be spread across many loci with varying effects (Savolainen et al. 2007), restriction site-associated DNA (RAD) sequencing (Baird et al. 2008) or genotyping by sequencing (GBS) (Elshire et al. 2011) could be used to conduct genome-wide scans to identify other outlier regions that may also play a role in the local adaptation or reproductive isolation. Finally high density genetic linkage maps of *Q. rubra* are being constructed currently and could be used to see if QTL for local adaptations of species co-locate with identified outlier regions (Gailing and Curtu 2014).

My SGS study revealed that both anthropogenic and natural disturbances impact the genetic structure of *Q. rubra* populations. Since the impacts appear to be variable from one population to the next, a landscape genetic study could prove useful to correlate these spatial genetic patterns as well as genetic variation at outlier loci with environmental variables. Manel et al. (2003) defines landscape genetics as the detection and correlation of genetic discontinuities to environmental features. This type of study can help inform conservation efforts on how to account for regional and local environmental characteristics. For example, Ackerman et al. (2013) examined the spatial genetic structure of two salmon species using both neutral and outlier SNPs and showed that local adaptation can play a role in shaping genetic structure. A SNP, identified as an outlier and located in a gene associated with adaptive differences in other salmon ecotypes, differentiated between populations with different spawning behaviors and locations related to temperature and water flow gradients. This could impact how civic water projects or fishery restocking may be conducted in order to preserve these distinct ecotypes. In Sork et al. (2010), mentioned in the introduction, the authors found that regional climatic gradients were associated with genetic variation in *Quercus lobata***.** The authors are currently working on identifying SNP variation in candidate genes linked to traits like drought tolerance and flowering time. SNPs can be associated with climatic variables allowing adaptive variation to be mapped to the landscape. This will help provide regionally relevant predictions about how climatic changes may impact growth and survival.

References

Abrams MD (1990) Adaptations and responses to drought in *Quercus* species of North America. Tree Physiol 7:227-238

Ackerman M, Templin W, Seeb J, Seeb L (2013) Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. Conservation Genetics 14:483-498.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1:95-111

Alberto FJ, Derory J, Boury C, Frigerio J-M, Zimmermann NE, Kremer A (2013) Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. Genetics 195:495-512

Aldrich PR, Cavender-Bares J (2011) *Quercus*: Wild Crop Relatives: Genomic and Breeding Resources. In: Kole C (ed). Springer Berlin Heidelberg, pp 89-129

Aldrich PR, Glaubitz JC, Parker GR, Rhodes OE, Michler CH (2005) Genetic structure inside a declining red oak community in old-growth forest. J Hered 96:627-634

Aldrich PR, Michler CH, Sun WL, Romero-Severson J (2002) Microsatellite markers for northern red oak (Fagaceae : *Quercus rubra*). Mol Ecol Notes 2:472-474

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323

Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution 50:900-908

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376

Barnes BVW, Warren Herbert Jr. (2004) Michigan Trees, Revised and Updated: A Guide to the Trees of the Great Lakes Region. University of Michigan Press Ann Arbor

Blue MP, Jensen RJ (1988) Positional and seasonal variation in oak (*Quercus*; Fagaceae) leaf morphology. Am J Bot 75:939-947

Bodénès C, Chancerel E, Gailing O, Vendramin G, Bagnoli F, Durand J, Goicoechea P, Soliani C, Villani F, Mattioni C, Koelewijn H, Murat F, Salse J, Roussel G, Boury C, Alberto F, Kremer A, Plomion C (2012) Comparative mapping in the Fagaceae and beyond with EST-SSRs. BMC Plant Biol 12:153

Bréda N, Huc R, Granier A, Dreyer E (2006) Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. Ann For Sci 63:625-644

Brendel O, Thiec D, Scotti-Saintagne C, Bodénès C, Kremer A, Guehl J-M (2008) Quantitative trait loci controlling water use efficiency and related traits in *Quercus robur* L. Tree Genet Genomes 4:263-278

Buiteveld J, Vendramin GG, Leonardi S, Kamer K, Geburek T (2007) Genetic diversity and differentiation in European beech (*Fagus sylvatica* L.) stands varying in management history. Forest Ecol Manag 247:98-106

Burczyk J, DiFazio S, Adams W (2004) Gene flow in forest trees: how far do genes really travel? Forest Genet 11:179

Burns RM, Honkala BH (1990) Silvics of North America Volume 2: Hardwoods. vol 2. Washington D.C.

Cavender-Bares J, Pahlich A (2009) Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). Am J Bot 96:1690-1702

Chesnoiu E, Sofletea N, Curtu AL, Toader A, Radu R, Enescu M (2009) Bud burst and flowering phenology in a mixed oak forest from Eastern Romania. Ann For Res 52:199-206

Cottrell JE, Munro RC, Tabbener HE, Milner AD, Forrest GI, Lowe AJ (2003) Comparison of fine-scale genetic structure using nuclear microsatellites within two British oakwoods differing in population history. Forest Ecol Manag 176:287-303

Craft KJ, Ashley MV (2007) Landscape genetic structure of bur oak (*Quercus macrocarpa*) savannas in Illinois. Forest Ecol Manag 239:13-20

Craft KJ, Ashley MV, Koenig WD (2002) Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. Am J Bot 89:1792-1798

Curtu A, Gailing O, Finkeldey R (2007a) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. BMC Evol Biol 7:218

Curtu A, Gailing O, Finkeldey R (2009) Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. BMC Evol Biol 9:284

Curtu AL, Gailing O, Leinemann L, Finkeldey R (2007b) Genetic variation and differentiation within a natural community of five oak species (*Quercus* spp.). Plant Biol 9:116-126

Desikan R, Horák J, Chaban C, Mira-Rodado V, Witthöft J, Elgass K, Grefen C, Cheung M-K, Meixner AJ, Hooley R, Neill SJ, Hancock JT, Harter K (2008) The histidine kinase AHK5 integrates endogenous and environmental signals in *Arabidopsis* guard cells. PLoS One 3:e2491

Dodd RS, Afzal-Rafii Z (2004) Selection and dispersal in a multispecies oak hybrid zone. Evolution 58:261-269

Durand J, Bodenes C, Chancerel E, Frigerio J-M, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn H-P, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herran A, Ikaran Z, Cabane C, Ueno S, Alberto F, Dumoulin P-Y, Guichoux E, de Daruvar A, Kremer A, Plomion C (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. J Heredity 107:1-15

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. J Heredity 99:125-132

Ellison, AM, Bank, MS, Clinton, BD, Colburn EA, Elliott K, Ford CR, Foster DR,Kloeppel BD, Knoepp JD, Lovett GM, Mohan J, Orwig DA, Rodenhouse NL, SobczakWV, Stinson KA, Stone JK, Swam CM, Thompson J, Von Holle B, Webster JR (2005) Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Front Ecol Environ 3:479-486

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for high diversity species. PLoS ONE 6:e19379

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. In: Adams WT, Strauss S, Copes D, Griffin AR (eds) Population Genetics of Forest Trees, vol 42. Forestry Sciences. Springer Netherlands, pp 257-278

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567

FAO (2011) Global Forest Resources Assessment 2005-2010, Food and Agricultural Organization of the United Nations

Gailing O (2013) Difference in growth, survival and phenology in *Q. rubra* and *Q. ellipsoidalis* seedlings. Dendrobiology 70:71-79

Gailing O, Curtu AL (2014) Interspecific gene flow and maintenance of species integrity in oaks. Ann For Res 57:5-18

Gailing O, Lind J, Lilleskov E (2012) Leaf morphological and genetic differentiation between Quercus rubra L. and Q. ellipsoidalis E.J. Hill populations in contrasting environments. Plant Syst Evol 298:1533-1545

Gibson SY, Van Der Marel RC, Starzomski BM (2009) Climate change and conservation of leading-edge peripheral populations. Conserv Biol 23:1369-1373

González-Rodríguez A, Arias DM, Valencia S, Oyama K (2004) Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. Am J Bot 91:401-409

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. J Hered 86:485-486

Guttman SI, Weigt LA (1989) Electrophoretic evidence of relationships among *Quercus* (oaks) of eastern North America. Can J Bot 67:339-351

Hamrick JL, Godt M, Sherman-Broyles S (1992) Factors influencing levels of genetic diversity in woody plant species. In: Adams WT, Strauss S, Copes D, Griffin AR (eds) Population Genetics of Forest Trees, vol 42. Forestry Sciences. Springer Netherlands, pp 95-124

Hamrick JL, Linhart YB, Mitton JB (1979) Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. Annu Rev Ecol Syst 10:173-200

Hardy OJ, Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618-620

Herrmann D, Barre P, Santoni S, Julier B (2010) Association of a *CONSTANS-LIKE* gene to flowering and height in autotetraploid alfalfa. Theor Appl Genet 121:865-876

Hipp AL (2010) Hill's oak: The taxonomy and dynamics of a Western Great Lake endemic. Arnoldia 67

Hokanson SC, Isebrands JG, Jensen RJ, Hancock JF (1993) Isozyme variation in oaks of the Apostle Islands in Wisconsin: genetic structure and levels of inbreeding in *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). Am J Bot 80:1349-1357

Hsu C-Y, Adams JP, No K, Liang H, Meilan R, Pechanova O, Barakat A, Carlson JE, Page GP, Yuceer C (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

Huebner CD (2003) Vulnerability of oak-dominated forests in West Virginia to invasive exotic plants: temporal and spatial patterns of nine exotic species using herbarium records and land classification data. Castanea 68:1-14

Jensen RJ, Hokanson SC, Isebrands JG, Hancock JF (1993) Morphometric variation in oaks of the Apostle Islands in Wisconsin: evidence of hybridization between *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). Am J Bot 80:1358-1366

Jimenez P, Lopez de Heredia U, Collada C, Lorenzo Z, Gil L (2004) High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history. J Heredity 93:510-515

Jump AS, Peñuelas J (2006) Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. PNAS 103: 8096-8100

Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. Genetics 175:1823-1834

Kremer A, Abbott A, Carlson J, Manos P, Plomion C, Sisco P, Staton M, Ueno S, Vendramin G (2012) Genomics of Fagaceae. Tree Genet Genomes 8:583-610

Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Dam BV, Ducousso A, Forrest I, Heredia ULd, Lowe AJ, Tutkova M, Munro RC, Steinhoff S, Badeau V (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. Ann For Sci 59:777-787

Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S (2009) Species relative abundance and direction of introgression in oaks. Mol Ecol 18:2228-2242

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. Mol Ecol 15:2007-2012

Lind-Riehl JF, Gailing O (in press) Fine-scale spatial genetic structure of two red oak species, *Quercus rubra* and *Q. ellipsoidalis* Plant Sys Evol

Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a *CONSTANS*-like gene between two red oak species. Ann Bot London 113:967-975

Lind J, Gailing O (2013) Genetic structure of *Quercus rubra* L and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genet Genomes 9:707-722

Lorenzo Z, Burgarella C, de Heredia UL, Lumaret R, Petit RJ, Soto Á, Gil L (2009) Relevance of genetics for conservation policies: the case of Minorcan cork oaks. Ann Bot-London 104:1069-1076

Lorimer CG (1993) Causes of oak regeneration problem. United Staes Department of Agriculture, Washington D.C.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189-197

Mariette S, Cottrell J, Csaikl UM, Goikoechea P, Konig A, Lowe AJ, Van Dam BC, Barreneche T, Bodenes C, Streiff R, Burg K, Groppe K, Munro RC, Tabbener H, Kremer A (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. Silvae Genet 51:72-79

McShea WJ, Healy WM, Devers P, Fearer T, Koch FH, Stauffer D, Waldon J (2007) Forestry matters: Decline of oaks will impact wildlife in hardwood forests. J Wildlife Manage 71:1717-1728

Moran EV, Willis J, Clark JS (2012) Genetic evidence for hybridization in red oaks (*Quercus* sect. Lobatae, Fagaceae). Am J Bot 99:92-100

Motzkin G, Ciccarello SC, Foster DR (2002) Frost pockets on a level sand plain: does variation in microclimate help maintain persistent vegetation patterns? Journal Torrey Bot Society 129:154-163

Muir G, Lowe AJ, Fleming CC, Vogl C (2004) High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of *Quercus petraea* at the margin of its range in Ireland. Ann Bot London 93:691-697

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111-122

Paffetti D, Travaglini D, Buonamici A, Nocentini S, Vendramin GG, Giannini R, Vettori C (2012) The influence of forest management on beech (*Fagus sylvatica* L.) stand structure and genetic diversity. Forest Ecol Manag 28434-44

Pan Y, Birdsey RA, Phillips OL, Jackson RB (2014) The structure, distribution, and biomass of the world's forests. Annu Rev Ecol Evol Syst 44: 593-622

Peakall ROD, Smouse PE (2006) GeneAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288-295

Peñaloza-Ramírez JM, González-Rodríguez A, Mendoza-Cuenca L, Caron H, Kremer A, Oyama K (2010) Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. Ann Bot London 105:389-399

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu J-L, Kremer A (2002) Identification of refugia and post-glacial

colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecol Manage 156:49-74

Petit RJ, Carlson J, Curtu AL, Loustau M-L, Plomion C, González-Rodríguez A, Sork VL, Ducousso A (2013) Fagaceae trees as models to integrate ecology, evolution and genomics. New Phytol 197:369-371

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol S 37:187-214

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

Raymond M, Rousset F (1995) GenePop (Version-1.2) - Population genetics software for exact tests and ecumenicism. J Hered 86:248-249

Rushton B (1993) Natural hybridization within the genus *Quercus* L. Ann For Sci 50:73s-90s

Salvini D, Bruschi P, Fineschi S, Grossoni P, Kjær ED, Vendramin GG (2009) Natural hybridisation between *Quercus petraea* (Matt.) Liebl. and *Quercus pubescens* Willd. within an Italian stand as revealed by microsatellite fingerprinting. Plant Biol 11:758-765

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene flow and local adaptation in trees. Annu Rev Ecol Evol S 38:595-619

Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: Nurminsky D (ed) Selective Sweep. Landes Bioscience, Georgetown, TX

Schluter D (2009) Evidence for ecological speciation and its alternative. Science 323:737-741

Scotti-Saintagne C, Bodénès C, Barreneche T, Bertocchi E, Plomion C, Kremer A (2004a) Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. Theor Appl Genet 109:1648-1659

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénès C, Burg K, Kremer A (2004b) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. Genetics 168:1615-1626

Servedio MR, Doorn GSV, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: magic but not rare? Trends Ecol Evol 26:389-397

Shikano T, Ramadevi J, Merilä J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). Mol Biol Evol 27:2775-2789

Sork VL, Davis FW, Smouse PE, Apsit VJ, Dyer RJ, Fernandez-M J, Kuhn B (2002) Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? Mol Ecol 11:1657-1668

Sork VL, Davis FW, Westfall R, Flint A, Ikegami M, Wang H, Grivet D (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. Mol Ecol 19:3806-3823

Sork VL, Stowe KA, Hochwender C (1993) Evidence for local adaptation in closely adjacent subpopulations of northern red oak (*Quercus rubra* L.) expressed as resistance to leaf herbivores. Am Nat 142:928-936

Steinkellner H, Lexer C, Turetschek E, Glössl J (1997) Conservation of (GA)n microsatellite loci between *Quercus* species. Mol Ecol 6:1189-1194

Streiff, Ducousso, Lexer, Steinkellner, Gloessl, Kremer (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. Mol Ecol 8:831-841

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. Mol Ecol 7:317-328

Sullivan AR, Lind JF, McCleary TS, Romero-Severson J, Gailing O (2013) Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. Plant Mol Biol Rep 31:231-239

Thompson JD, Gaudeul M, Debussche M (2010) Conservation value of sites of hybridization in peripheral populations of rare plant species. Conserv Biol 24:236-245

Tomlinson PT, Jensen RJ (2000) Do whole tree silvic characters indicate hybridization in red oak (*Quercus* section Lobatae)? Am Midl Nat 143

Tran L-SP, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K, Yamaguchi-Shinozaki K (2007) Functional analysis of *AHK1/ATHK1* and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in *Arabidopsis*. P Natl Acad Sci 104:20623-20628

Valbuena-Carabana M, Gonzalez-MartÍnez SC, Hardy OJ, Gil L (2007) Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. Mol Ecol 16:1207-1219

Van Valen L (1976) Ecological species, multispecies, and oaks. Taxon 25:233-239

Vasemagi A, Nilsson J, Primmer C (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting

signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol Biol Evol 22:1067-1076

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos T R Soc B 367:451-460

Vilas R, Bouza C, Vera M, Millan A, Martinez P (2010) Variation in anonymous and EST-microsatellites suggests adaptive population divergence in turbot. Mar Ecol-Prog Ser 420:231-239

Vornam B, Gailing O, Derory J, Plomion C, Kremer A, Finkeldey R (2011) Characterisation and natural variation of a dehydrin gene in *Quercus petraea* (Matt.) Liebl. Plant Biol 13:881-887

Weising K, Nybom, Hilde, Wolff, Kirsten, Kahl, Gunther (2005) DNA Fingerprinting in Plants: Principles, Methods, and Applications. Second Edition edn. CRC Press, Taylor & Francis Group, Boca Raton

Whittaker R (1975) Communities and Ecosystems 2nd ed. Macmillian, London

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. P Natl Acad Sci 88:2540-2544

Wood H, Grahame J, Humphray S, Rogers J, Butlin R (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. Mol Ecol 17:3123-3135

Woodall CW, Oswalt CM, Westfall JA, Perry CH, Nelson MD, Finley AO (2009) An indicator of tree migration in forests of the Eastern United States. Forest Ecol Manag 257:1434-1444

Table 1 Sampled populations and associated studies

Abbreviation	Population	Region	Species	Publication
BR-1	Brockway 1	Brockway Mountain	*Q. rubra*	Lind & Gailing 2013

CNF-QE	Chequamegon QE	Chequamegon National Forest	*Q. ellipsoidalis*	Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
CNF-QR	Chequamegon QR	Chequamegon National Forest	*Q. rubra*	Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
FC-A	Stand A	Ford Research Forest and Baraga Plains	*Q. rubra*	Lind & Gailing 2013, Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
FC-B	Stand B	Ford Research Forest and Baraga Plains	*Q. rubra*	Lind & Gailing 2013, Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
FC-C	Stand C	Ford Research Forest and Baraga Plains	*Q. ellipsoidalis*	Lind & Gailing 2013, Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
FC-E	Stand E	Ford Research Forest and Baraga Plains	*Q. ellipsoidalis*	Lind & Gailing 2013, Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
HMR-IH	Ives Hill	Huron Mountain Reserve	*Q. rubra*	Lind & Gailing 2013, Lind-Riehl & Gailing 2014
HMR-LI	Lake Ives	Huron Mountain Reserve	*Q. rubra*	Lind & Gailing 2013
HMR-LP	Lily Pond	Huron Mountain Reserve	*Q. rubra*	Lind & Gailing 2013, Lind-Riehl & Gailing 2014
HMR-MI	Mount Ives	Huron Mountain Reserve	*Q. rubra*	Lind & Gailing 2013, Lind-Riehl & Gailing 2014
HMR-PL	Pine Lake	Huron Mountain Reserve	*Q. rubra*	Lind & Gailing 2013
MTU-1	Michigan Tech Trails 1	Michigan Tech Trails	*Q. rubra*	Lind & Gailing 2013
NNF-QE	Nicolet National Forest	Nicolet National Forest	*Q. ellipsoidalis*	Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
NNF-QR	Nicolet National QE	Nicolet National Forest	*Q. rubra*	Lind-Riehl et al. 2014, Lind-Riehl & Gailing 2014
PM-QR	Porcupine Mountains QR	Porcupine Mountains Wilderness State Park	*Q. rubra*	Lind-Riehl & Gailing 2014

Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers [1]

Jennifer Lind[†], Oliver Gailing[†‡]

[†]Michigan Technological University, 1400 Townsend Drive, 49931 Houghton, MI, United States, ogailing@mtu.edu, [‡] corresponding author

[1] The material in this chapter was previously published material and has been reprinted with kind permission from Springer Science and Business Media from Lind JF and Gailing O (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 and Genomes 9: 707-722.

Abstract

Sympatric hybridizing oak species provide a model system for studying local adaptation. Disjunct populations of *Quercus rubra* L. and *Q. ellipsoidalis* E. J. Hill at the northern edge of their distribution may harbor important reservoirs of adaptive genetic variation. Genic (Expressed Sequence Tag- Simple Sequence Repeat = EST-SSR) and non-genic nuclear microsatellite (nuclear SSR = nSSR) markers were used to estimate neutral and potentially adaptive genetic variation in these two supposedly interfertile oak species showing different adaptations to drought. Eleven populations of putative *Q. rubra* and *Q. ellipsoidalis* located in the Western Upper Peninsula of Michigan were characterized using seven EST-SSRs and eight nSSRs. Bayesian cluster analysis revealed two distinct groups corresponding to each species with evidence of low levels of potential introgression. A comparison of the genetic structure of adult trees and seedlings revealed no evidence for selection against hybrids. Overall, similar levels of genetic variation and differentiation between populations and species were found at both EST-SSRs and SSRs indicating that most EST-SSRs chosen reflect neutral variation. Two loci, 3A05 (nSSR) and GOT021 (EST-SSR, putative histidine kinase 4-like), were identified as putative outlier loci between species showing largely reduced variation in *Q. ellipsoidalis*. Future analyses of an increased number of EST-SSRs located in functional genes will allow the identification of genes involved in the reproductive isolation between both species.

Keywords EST-SSRs, nuclear SSRs, oaks, genetic assignment

Introduction

The evolutionary role of hybridization has been a controversial topic. The prevailing view is that hybridization plays an important role in adaptive evolution and speciation (e.g., by formation of new species, transfer of adaptations and invasion of one species into the range of another species, Arnold and Martin 2010). In particular, oaks (*Quercus* spp.) have become model taxa to investigate the potential role of hybridization and introgression in adaptive evolution (Aldrich and Cavender-Bares 2011). Hybridization between related oak species is a common phenomenon and has challenged the biological species concept even resulting in the development of a species concept that relies on ecological criteria (Van Valen 1976). Due to their propensity to hybridize (e.g., Curtu et al. 2009; Mir et al. 2006) and high within-species genetic and phenotypic variation, species boundaries within the genus *Quercus* are not clear-cut (Burger 1975; Rushton 1993). Genetic differentiation between hybridizing species is generally low, as the result of porous species boundaries (Moran et al. 2012; Scotti-Saintagne et al. 2004) and/or a short divergence time between closely related species (Muir et al. 2001). Despite frequent interspecific gene flow between closely related oak species (Bacilieri et al. 1996; Curtu et al. 2009, 2007a; Dodd and Afzal-Rafii 2004; Howard et al. 1997; Moran et al. 2012; Valbuena-Carabaña et al. 2005), species identity is often maintained possibly as the result of directional selection (Whittemore and Schaal 1991). For example, in a central Romanian stand of four hybridizing oak species (*Q. robur*, *Q. petraea*, *Q. pubescens* and *Q. frainetto*), their spatial distribution was associated with different environmental conditions (e.g., soil conditions, elevation) and adult hybrids and introgressive forms were located at the sympatric edges of the species' ecological niches, suggesting that selection maintains different species' adaptations (Curtu et al. 2009, 2007a). Hybrid seedlings of *Q. petraea* and *Q. pyrenaica* have also been shown to be located in micro-sites of maximum contact between recruits of both parental species (López de Heredia et al. 2009). Furthermore, even moderate levels of gene flow are expected to lead to a highly homogenized genome as

observed in the closely related white oak species *Q. robur* and *Q. petraea* (Scotti-Saintagne et al. 2004). However, some genic regions exhibited interspecific differentiation that exceeded the neutral expectation. These regions were designated as outlier loci and may play a role in the control of adaptive traits involved in pre- or post-zygotic isolation between species (e.g., differences in flowering time or drought tolerance).

In addition to being a model for studying the evolutionary effects of hybridization, oaks are important forest tree species in the Northern Hemisphere, both economically, through wood and paper production, and ecologically, as a source of food and shelter for countless wildlife (McShea et al. 2007). However, a northward migration of temperate forest trees in North America is predicted along with changes in precipitation patterns that will lead to increased drought and fire events as a result of climate change (Pautasso 2009; Woodall et al. 2009). These changes will affect distribution ranges and competition between forest tree species (Aldrich and Cavender-Bares 2011; Doak and Morris 2010). Current disjunct populations of *Q. rubra* L. and *Q. ellipsoidalis* E. J. Hill at the northern range limit of oaks, such as those analyzed in the present study, may harbor important reservoirs of genetic variation as has been observed in other peripheral or marginal populations of oaks with particular local adaptations (Jimenez et al. 1999; Lorenzo et al. 2009). Additionally, hybridization between these closely related oak species with different drought adaptations might result in the transfer of adaptations or new adaptations due to novel gene combinations. Assessment of this species pair in the study area may provide information relevant for management and conservation of these species in the face of climate change.

Q. rubra and *Q. ellipsoidalis* were chosen as model taxa, since they overlap in their distribution ranges, while also showing different local adaptations to drought. *Q. rubra* extends from southern Ontario in the north to central Georgia in the south and from the Atlantic coast to the Mississippi River in the west (Sork et al. 1993). It

prefers mesic to well drained uplands and is one of the most genetically diverse oak species in North America according to a review of nuclear and organelle gene diversity studies of 33 species of oak (Kremer and Petit 1993). The range of *Q. ellipsoidalis* is more scattered, including Michigan, Wisconsin, most of Iowa and Minnesota and the northern parts of Indiana and Illinois (Hipp 2010). In contrast to *Q. rubra*, *Q. ellipsoidalis* prefers very dry sandy sites and is the most drought tolerant red oak species in North America showing several morphological and physiological adaptations related to drought tolerance (e.g., tissue elasticity, leaf conductance, xylem anatomy, root depth) (Hipp 2010; Abrams 1990). The two species also show differences in leaf and acorn morphology. *Q. ellipsoidalis* has smaller ellipsoid acorns and deeply lobed glabrous leaves, while *Q. rubra* has larger ovoid acorns with a short cap and leaves with shallower lobes (Barnes 2004). While hybridization between the species was suggested based on the observation of continuous leaf morphological variation (Jensen et al. 1993) and low genetic differentiation between species at isozyme markers (Hokanson et al. 1993), these markers did not allow an assignment of individual samples to species.

In the present study, a total of 582 individuals from 11 locations in the Upper Peninsula, where *Q. rubra* and *Q. ellipsoidalis* are the only species present, were characterized at both potentially selectively neutral non-genic microsatellites (nuclear Simple Sequence Repeats or nSSRs) and at genic microsatellites (Expressed Sequence Tag or (EST)-SSRs) (Aldrich et al. 2002; Durand et al. 2010; Sullivan et al. 2013). Most populations were tentatively characterized as *Q. rubra* and *Q. ellipsoidalis* based on multivariate analyses of leaf characters (Gailing et al. 2012). In one location, adjacent *Q. rubra* and *Q. ellipsoidalis* populations were identified on contrasting soil types with *Q. ellipsoidalis* growing on very dry outwash sands (Baraga Plains region).

The specific objectives of the present study were to (1) to develop a set of discriminating microsatellite markers that allow the assignment of individuals to

species and to identify hybrids and introgressive forms, (2) test whether there is evidence for selection against hybrids by comparing the number of hybrids and introgressive forms in the adult and seedling generation, (3) compare genetic variation within and among populations and species at potentially selectively neutral nSSRs and at genic EST-SSRs, (4) identify outlier loci putatively under divergent selection between species.

We hypothesize that (1) hybrids and introgressive forms can be identified in the contact zones between species, but their number is low as result of pre- and/or post-zygotic isolation, and (2) selection against hybrids is reflected in a lower number of hybrids in the adult tree generation as compared to the seedling generation, (3) genetic differentiation between species is low as a result of gene flow or shared common characters, but higher than among populations within species, (4) the level of interspecific differentiation is higher at potentially adaptive EST-SSRs than at selectively neutral nSSRs, (5) outlier loci putatively under divergent selection between species can be identified.

Materials and methods

Plant material

A total of 582 trees were sampled from nine *Q. rubra* and two *Q. ellipsoidalis* populations in four regions of the western Upper Peninsula of Michigan and their locations mapped using a GPS unit (Table 2, Fig. 1). In the Baraga Plains (near the Ford Center, FC) two *Q. ellipsoidalis* populations (FC-C, FC-E) on dry sites and two *Q. rubra* populations (FC-A, FC-B) on neighboring mesic sites were sampled. In the Huron Mountain Reserve (HMR) sampling was conducted along an altitudinal gradient with soil moisture decreasing as altitude increased. The HMR and Brockway Mountain (BR) populations included individuals collected from granite rock outcrops showing a shrubby growth habit, while the MTU-1 samples were collected from a single old-growth *Q. rubra* stand on loamy fine sand. In addition,

leaves from 131 and 34 naturally regenerated seedlings were collected from FC-A and FC-C, and from 86 one and a half year old seedlings grown from acorns collected from FC-E. An additional set of samples were included that had previously been genetically assigned to species using AFLPs (Amplified Fragment Length Polymorphisms) (Hipp and Weber 2008). These included 12 and 27 individuals that were assigned to *Q. rubra* and *Q. ellipsoidalis*, respectively. The samples were collected from eight populations (*Q. rubra*) in the Lower Peninsula of Michigan and three populations (*Q. ellipsoidalis*) in Indiana (41° 31' N, 87° 26' W), the Lower Peninsula of Michigan (44° 26' N, 84° 12' W), and Wisconsin (44° 40' N, 88° 5' W). Another 30 trees that were morphologically identified as *Q. rubra* from Henderson County in Tennessee (TN-HC, 18 individuals from one population) (35° 43'N, 88° 17'W) and Madison County in Alabama (AL-MC, 12 individuals from one population) (34° 45'N, 86° 30'W) outside of the *Q. ellipsoidalis* distribution range were also included in the analyses (Jeanne Romero-Severson, personal communication). The additional *Q. ellipsoidalis* samples are referred to as QE and the additional *Q. rubra* samples as QR throughout the remainder of the manuscript. All leaf samples were stored in a -20°C freezer prior to DNA isolation.

DNA extraction and microsatellite genotyping

Total genomic DNA (~20 ng) was isolated from leaf material using the DNeasy96 Plant Kit (Qiagen, Valencia, CA) following the manufacturer's instructions. A total of sixteen microsatellite markers were amplified in the populations covering eight of the 12 linkage groups in *Q. robur* (Table 3). The eight EST-SSRs were originally developed in *Q. robur* (Durand et al. 2010) and adapted for use in *Q. rubra* (Sullivan et al. 2013). Seven of the eight nSSRs were developed in *Q. rubra* (Aldrich et al. 2002; Sullivan et al. 2013) and one was developed in *Q. robur* (Steinkellner et al. 1997) and adapted for use in *Q. rubra* (Table 3). EST-SSRs were assigned functional annotations using the Blast2GO software (Conesa et al. 2005) employing the BLASTx algorithm (Altschul et al. 1997) by comparing the reassembled ESTs

(Durand et al. 2010) to homologous sequences in the non-redundant NCBI database (Table 3) following the method by Luro et al. (2008).

The samples were amplified in a Peltier Thermal Cycler (GeneAmp® PCR system 2700, Applied Biosystems, Foster City, CA), using the following program: (1) initial denaturation at 95°C for 15 min (2) 35 cycles at 94°C for 45 s, annealing at 54-62°C for 45 s (see Table 3), and elongation at 72°C for 45 s and (3) a final extension step at 72°C for 20 min. The 10 µL PCR reaction was composed of 2 µL 5x HOTFIREPol Blended Master Mix from Solis BioDyne, Estonia, containing 10mM $MgCl_2$, 0.4 U of HOTFIREPol Taq polymerase, 2mM dNTPs, 1.3 µL of each 5µM fluorescently labeled forward primer from Applied Biosystems, Foster City, CA, (see Table 3) and reverse primer from Sigma Aldrich, St. Louis, MO, and 1.5 µL of template DNA (~2 ng). The PCR products were separated on an ABI Prism® Genetic Analyzer 3730 and scored using GeneMapper® v.4.0 (Applied Biosystems, Foster City, CA). One marker, FIR110, amplified two loci. In the *Q. rubra* populations, the size range was 168bp to 188bp for FIR110L1 and 180bp to 206bp for FIRl10L2, with 29 individuals (5.7%) showing fragments in the overlapping size range. In *Q. ellipsoidalis* populations, the size range was 166bp to 194bp for FIR110L1 and 180bp to 236bp for FIR110L2, with 54 individuals (30.3%) showing fragments in the overlapping size range. As a result, this marker was excluded from further analyses.

Population structure analysis

Null alleles might slightly decrease the power to correctly assign individuals in STRUCTURE (Carlsson 2008). Since most markers did not show high null allele frequencies using MICROCHECKER (van Oosterhout et al. 2004) all markers were used for the assessment of population structure using the software STRUCTURE 2.2 (Pritchard et al. 2000). While the selected microsatellite markers were appropriate to assign samples to species in one geographic region (see below), the number of markers might be too scarce to classify samples from a range-wide sampling. Thus,

Q. ellipsoidalis and *Q. rubra* reference samples consisting of few individuals from different geographic locations showed considerable population substructure, resulting in incorrectly assigned mixed shared ancestry for these individuals (data not shown). In the final STRUCTURE analyses the latter samples were excluded and only populations from the western Upper Peninsula of Michigan were included, where no other oak species besides *Q. rubra* and *Q. ellipsoidalis* are present. The program was set to run five independent runs of a burn-in period of 30000 iterations followed by 10^6 iterations for each value of K (K= 1 through 11) using the admixture model with correlated allele frequencies without *a priori* information regarding species identification. We used the admixture model because it allows for mixed ancestry of individuals and the correlated allele frequencies model because it can improve clustering of closely related populations or species. The most likely number of groups (K) was chosen by comparing logarithmized probabilities of data [Pr(X|K)] according to Pritchard et al. (2000) and by the calculation of an ad hoc statistic, ΔK, according to Evanno et al. (2005). 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 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 according to Curtu et al. (2007a). In addition, we used *a priori* species information derived from the morphological classification of samples (Gailing et al. 2012) to calculate posterior probabilities for pure species and putative hybrids using the USEPOPINFO option, which assumes that most individuals classified as belonging to one group have pure ancestry from that group, while a small percentage of individuals may share some proportion of ancestry from another group (species in our case) (Falush et al. 2007). The GENSBACK option was set to three meaning that the probability of ancestry from the other species is expressed as proportions attributable to their parent, grandparent and great-grandparent. The MIGRPRIOR option was run at three different values of immigration rate ν (0.01, 0.05, 0.10) to

assess the robustness of the assignment as suggested by Pritchard et al. (2000) and as used in other similar studies in oaks (Craft et al. 2002; Curtu et al. 2007a).

Genetic diversity and differentiation

For all populations with a sample size of ≥49, the following summary statistics were calculated in GeneAlEx v.6.41 (Peakall and Smouse 2006): number of different alleles per locus (N_a), observed heterozygosity (H_o), and Nei's unbiased gene diversity (H_e) (Nei 1973). Wright's inbreeding coefficient (F) (Wright 1922), F_{ST} and pairwise F_{ST} with corresponding significances, and linkage disequilibrium (LD) for all pairs of loci were calculated in GenePop 4.1 (Raymond and Rousset 1995). These calculations were done using all individuals and after exclusion of potentially introgressed and hybrid individuals (data not shown) resulting in only minor differences. Sequential Bonferroni adjustments of the significance level (α) were made to correct for multiple comparisons (Rice 1989).

A genetic distance matrix (F_{ST}, Latter 1973) was calculated and an unrooted tree was created using the un-weighted pair-group method with arithmetic means (UPGMA) algorithm (Sneath and Sokal 1973) with 1000 bootstrap replicates in Populations 2.0 (Langella 1999) for populations with a sample size of ≥49. TreeViewX (Page 1996) was used to visualize the dendrogram. *Q. rubra* and *Q. ellipsoidalis* reference samples (QE, QR) were included in these analyses to corroborate the species identity of the *Q. ellipsoidalis* populations (FC-C, FC-E) from the northern distribution edge of the species. Arlequin 3.5 was used to analyze population differentiation through hierarchical analysis of molecular variance (AMOVA) using default settings (Excoffier and Lischer 2010). A Mantel test (Mantel 1967) was performed in GeneAlEx v.6.41 to compare pairwise F_{ST} values of nSSRs and EST-SSRs.

Multilocus scans

Wright's fixation index, F_{ST}, (Weir and Cockerham 1984) was calculated for each locus weighted by its heterozygosity between populations to identify outlier loci

under the assumption that gene loci under selection will show higher differentiation between species than expected under neutral evolutionary conditions (Beaumont 2005). The two outlier screening programs LOSITAN (Antao et al. 2008) and Arlequin 3.5 (Excoffier and Lischer 2010), both based on the Fdist method of Beaumont and Nichols (1996), were used for the calculations. For both LOSITAN and Arlequin, populations were pooled by species under the finite island model, and additionally a hierarchical island model of population structure was used in Arlequin 3.5 (Excoffier and Lischer 2010) to account for population structure within species. In addition, calculations were performed only in the four populations of the Baraga Plains region where both species occur within gene flow distance (no isolation by distance). Only populations with a sample size of ≥49 were included in the analyses. For LOSITAN, fifty thousand simulations were run at a 95% confidence interval and a false discount rate of 0.1 under the stepwise mutation model using both the "neutral" mean F_{ST} option to exclude loci potentially under selection for the computation of the initial mean F_{ST} and the "Force" mean F_{ST} option to increase the precision of the simulated mean F_{ST}. In Arlequin, the finite island and the hierarchical island model with default settings were used and included 50000 simulations at a 95% confidence interval under the stepwise mutation model for microsatellite data.

Results

Population structure

As the model used in STRUCTURE 2.2 assumes that all loci are unlinked, we performed an LD analysis of all pairs (136 total pairwise comparisons) of loci and only two pairs of loci exhibited significant LD at $p < 0.001$ (data not shown). Given that the majority of markers used were unlinked and the program can handle weakly linked loci, we proceeded with the analyses. The initial analysis was performed without any *a priori* species information. Figure 2 shows the values of the log likelihood of the multilocus genotypes, Pr(X|K), as a function of the number of

clusters, K, given the observed genotypes, X. The values of ln Pr(X|K) "plateaued" at K= 2, indicating that two clusters, corresponding to the two species *Q. rubra* and *Q. ellipsoidalis*, best fit the multi-locus genotypic data. Additionally, using the method developed by Evanno et al. (2005), the values of ΔK peaked at K= 2, further confirming that two clusters best fit the multi-locus genotypic data (Fig. 2).

Without *a priori* species information using the thresholds defined in the material and methods, low numbers of introgressive forms and hybrids were found in both species, although in slightly higher numbers in the *Q. ellipsoidalis* populations (Suppl. 1). The two populations, FC-C and FC-E, were comprised mostly of individuals that were assigned to *Q. ellipsoidalis,* while for all other populations the vast majority of individuals were assigned to *Q. rubra* (Fig. 3). Statistical support for introgressive forms and hybrids at three different values of "immigration" rate (ν = 0.01, 0.05, 0.1) using *a priori* species assignment information was detected for an even lower number of samples (Suppl. 2). Thus at ν = 0.01 no introgressive forms and only two putative hybrids were found, one in the *Q. rubra* population BR-1 and one in the *Q. ellipsoidalis* seedling population of FC-C. For higher values of ν (0.05, 0.1) the number of introgressive forms (0% to 4%) and hybrids (0% to 3%) was still very low, again with slightly higher values in *Q. ellipsoidalis* adult and seedling populations. *Q. ellipsoidalis*-like introgressive forms were found in low numbers only in the *Q. ellipsoidalis* populations (FC-C, FC-E) and in one neighboring *Q. rubra* population (FC-B) in the Baraga Plains.

Introgressive forms and hybrids found in the adjacent *Q. rubra* (FC-A, FC-B) and *Q. ellipsoidalis* (FC-C, FC-E) populations were visualized spatially in ArcMap (ESRI 2009) but revealed no spatial clustering. An examination of the population structure of naturally regenerated seedlings from these populations showed similar numbers of statistically significant introgressive forms for *Q. ellipsoidalis* (1%) and *Q. rubra* seedlings (1%) than in the adult generation and hybrids were identified only in the *Q. ellipsoidalis* seedlings (3%) (Suppl. 2). Likewise, among the *Q. ellipsoidalis*

greenhouse seedlings, the number of introgressive forms was low (1%) and no hybrid individuals were detected, but there was a high mortality rate (~40%) for these seedlings before leaf material was collected.

Genetic variation within populations and species

Eight out of the total 15 markers showed no deviation from Hardy Weinberg Equilibrium (HWE) in all or most populations (Suppl. 3). However, both FIR004 and quru-GA-0E09 displayed significant excess of homozygotes relative to HWE in all eight populations, while several other markers showed significant deviations in at least six populations (GOT004, PIE099, quru-GA-1F07) which may be the result of the presence of null alleles or population substructure. Nuclear SSR 3A05 showed significant HWE deviations (homozygote excess) in both the *Q. ellipsoidalis* populations, FC-C and FC-E, but not in the vast majority of the other *Q. rubra* populations. No population showed deviations from HWE across all loci (see Suppl. 3 and 4). However, one *Q. ellipsoidalis* population FC-E showed high and positive F values for several loci and the mean F value was at least 30% higher than any other population's mean F value. EST-SSRs showed lower genetic variation (H_e= 0.73) than nSSRs (e.g., H_e= 0.86) (Table 5). However, this difference was shown to be non-significant (p= 0.1552) by a two-sided t-test. For both EST-SSRs and SSRs, *Q. ellipsoidalis* and *Q. rubra* populations showed similar levels of genetic variation (Table 4). However, GOT021 showed largely reduced genetic variation in *Q. ellipsoidalis* populations (H_o= 0.05, H_e= 0.07), as compared to *Q. rubra* populations (H_o= 0.32, H_e= 0.28). A reduced genetic variation at GOT021 was also observed in the *Q. ellipsoidalis* reference samples, albeit less pronounced (H_o= 0.18, H_e= 0.17). Likewise, 3A05 showed a low H_o (0.36) and a high mean F value (F= 0.45, H_e= 0.66) in *Q. ellipsoidalis* populations that was not observed in *Q. rubra* populations (H_o= 0.67, H_e= 0.78) (Table 4).

Genetic variation between populations and species

All fifteen loci showed significant F_{ST} values that ranged from 0.02 to 0.08 across all samples with 3A05 (F_{ST}= 0.08) showing the highest differentiation (Table 4). All loci except GOT021, GOT009, FIR048, quru-GA-1F07, and 2P24, showed significant differentiation between *Q. ellipsoidalis* populations, while all loci showed significant differentiation among *Q. rubra* populations. A hierarchical AMOVA (Table 6) showed that most of the variation was present within populations (93.6%), while differentiation between species and among populations within species accounted for 4.95% and 1.41% of the total variation, respectively, with all components being highly significant. Similar results were obtained from the hierarchical AMOVA examining the four Baraga Plains populations alone (data not shown). Between the four *Q. ellipsoidalis* and *Q. rubra* populations in the Baraga Plains, mean differentiation across all loci was 5.8%, while the highest interspecific differentiation was accounted for by three loci, GOT021 (F_{ST}= 0.167), 3A05 (F_{ST}= 0.136) and 3D15 (F_{ST}= 0.109).

Considering all populations the largest interspecific differentiation was observed at 3A05 (pairwise F_{ST}= 0.147), while other loci showed moderate genetic differentiation, but all pairwise F_{ST} values were significant (Table 4). Pairwise F_{ST} values between populations revealed a strong differentiation between *Q. ellipsoidalis* populations FC-C and FC-E and *Q. rubra* populations (Suppl. 5). Moreover, statistical support for the *Quercus ellipsoidalis* cluster was high (96%), while it was low for the *Quercus rubra* cluster (15%) (Fig. 4). Specifically, *Q. ellipsoidalis* populations (FC-C, FC-E) were clearly separated from adjacent *Q. rubra* populations (FC-A, FC-B). Notably, the QR and QE samples grouped with *Q. rubra* and *Q. ellipsoidalis* populations, respectively.

A Mantel test between pairwise F_{ST} matrices of the two marker types displayed a significant correlation ($y= 0.7382x + 0.0024$, $R^2= 0.861$, $p= 0.002$). The mean differentiation between species for EST-SSRs and nSSRs was 4.7% and 6.1%,

respectively. Interspecific differentiation for the four *Q. rubra* and *Q. ellipsoidalis* populations in the Baraga Plains was 6.0% for EST-SSRs and 6.5% for nSSRs.

Under the hierarchical island model, Arlequin identified FIR004 (F_{ST}= 0.026, p= 0.033) as an outlier under balancing selection and 3A05 (F_{ST}= 0.15, p= 0.018) under divergent selection. Locus quru-GA-1F07 (F_{ST}= 0.022, p= 0.038) was identified as an outlier under balancing selection, and GOT021 (F_{ST}= 0.19, p= 0.037) and 3A05 (F_{ST}= 0.14, p= 0.049) as outliers under divergent selection when only the four populations in the Baraga Plains were included in the analysis. When populations were pooled by species, again FIR004 (F_{ST}= 0.014, p= 0.019) and 3A05 (F_{ST}= 0.15, p= 0.016) were detected as outlier loci in the whole sample set, and only locus GOT021 (F_{ST}= 0.17, p= 0.047) was identified as under divergent selection in the four Baraga Plains populations (Fig. 5). LOSITAN did not identify outliers between species.

Discussion

Genetic species assignment

Previous variation studies of the species pair *Q. rubra* and *Q. ellipsoidalis* were mainly restricted to morphological characters and isozyme markers (Hokanson et al. 1993; Jensen et al. 1993). Potential introgressive forms and hybrids could not be identified with these methods due the plasticity inherent in leaf morphological characters (Aldrich and Cavender-Bares 2011) and the limited amount of genetic variation within and among species at isozyme markers (Hokanson et al. 1993). Due to the availability of numerous nSSRs and EST-SSRs with high transferability among species in *Quercus* and their highly polymorphic nature (Aldrich et al. 2002; Dow et al. 1995; Durand et al. 2010; Kampfer et al. 1998; Steinkellner et al. 1997) they are the markers of choice to identify potential hybrids and introgressive forms.

A clear separation between *Q. ellipsoidalis* populations growing on dry sandy sites versus *Q. rubra* populations was shown in the present study at both nSSR and EST-

SSR markers. The high correspondence between morphological and genetic assignment for "pure" species (e.g., 87.7 % for *Q. ellipsoidalis* and 97.5% for *Q. rubra*) and the assignment of *Q. ellipsoidalis*-like and *Q. rubra*-like introgressive forms to respective morphological species suggest that leaf morphological differences reflect genetic differences between species in most cases (Gailing et al. 2012). It has been shown repeatedly in different plant species that molecular and morphological assignments do not always correspond and phenotypically intermediate individuals are not always of hybrid origin due to wide intraspecific variation or convergent evolution (Curtu et al. 2007a; Craft et al. 2002; Moran et al. 2012; Rieseberg et al. 1993). Thus most studies in oaks, including the present study, have shown generally high, but imperfect correspondence between morphological and molecular assignment which has been attributed to intraspecific variation, hybridization and maternal effects (*Q. virginiana* and *Q. geminata*, Cavender-Bares and Pahlich 2009; *Q. robur*, *Q. petraea*, *Q. pubescens* and *Q. frainetto,* Curtu et al. 2007a*; Q. grisea* and *Q. gambelii*, Howard et al. 1997, *Q. robur* and *Q. petraea*, Kremer et al. 2002).

Gene flow and hybridization

While introgressive forms and hybrids were identified in the genetic assignment analysis in the adult and seedling generations, their numbers were very low (hypothesis 1). Thus, there was no evidence for selection against hybrids in the seedling generation as it displayed similar numbers of introgressive forms and hybrids as the adult generation (rejection of hypothesis 2). However, selection against hybrids in early seedling stages cannot be excluded completely since the genotypic structure was assessed in young seedlings, not in seeds. In contrast to our study, previous studies have shown indirect evidence for selection against hybrids as one potential post-zygotic isolation mechanism involved in the maintenance of species identity in oaks, including North American red oaks (Curtu et al. 2009; Emms and Arnold 1997; Moran et al. 2012; Nagy and Rice 1997). The low number

of hybrids and introgressive forms in the present study and pronounced differences in vegetative bud burst (unpublished results), as a proxy for flowering time in oaks (e.g., Chesnoiu et al. 2009), for *Q. rubra* and *Q. ellipsoidalis* seedlings suggest pre-zygotic isolation between both species. Likewise, low levels of introgression mainly due to differences in flowering time were identified in the species pair *Q. virginiana* and *Q. geminata* (Cavender-Bares and Pahlich 2009).

While there is no evidence for high levels of gene flow between species, the number of introgressive forms was slightly higher in *Q. ellipsoidalis* than in neighboring *Q. rubra* populations indicating low levels of asymmetrical gene flow as a result of the much higher abundance of *Q. rubra* in the region. In contrast, previous studies of different oak species revealed that interspecific gene flow was relatively common but often asymmetric (Boavida et al. 2001; Jimenez et al. 2004), and intermediate forms and hybrids were mostly located at the environmental edges of the species respective ecological niches (Curtu et al. 2007a; Dodd and Afzal-Rafii 2004; Howard et al. 1997; Lepais et al. 2009; Peñaloza-Ramírez et al. 2010).

Genetic variation and differentiation at non-genic and genic markers

Studies comparing genic and non-genic markers have found that generally EST-SSRs show lower levels of polymorphism which is often attributed to their location in more conserved regions of the genome (reviewed in Ellis and Burke 2007; Varshney et al. 2005). While EST-SSRs may show lower levels of polymorphism relative to putatively neutral nSSRs, they still exhibit a high amount of polymorphism compared to less variable markers such as isozyme markers (Ellis and Burke 2007). Likewise, in the present study slightly lower levels of genetic variation were observed at EST-SSRs than at nSSRs (see Tables 3 and 4).

On average, levels of genetic variation were comparable within *Q. rubra* and *Q. ellipsoidalis* populations with one notable exception. One EST-SSR, GOT021, showed strongly reduced genetic variation (H_o= 0.05, H_e= 0.07) in the *Q.*

ellipsoidalis but not in *Q. rubra* populations (H_o= 0.32, H_e= 0.28). As a result the genetic differentiation between the four *Q. rubra* and *Q. ellipsoidalis* populations in the Baraga Plains was comparatively high at this marker (F_{ST}= 0.167, p≤ 0.0001) and GOT021 was identified as an outlier locus potentially under divergent selection by Arlequin (see below). The near fixation of one allele in the more drought adapted *Q. ellipsoidalis* populations at this gene locus might be the result of directional selection. GOT021 has a putative function as a histidine kinase-4 like protein (Table 3). Histidine kinases have been shown to be involved in the control of water use efficiency in *Arabidopsis* as an important part of a signaling cascade to effect stomatal closure in response to environmental and endogenous stimuli (Desikan et al. 2008). Also GOT021 is associated with leaf shape variation in a *Q. robur* full-sib family (Gailing et al. submitted) and triple mutants of histidine kinase genes in *Arabidopsis* showed altered leaf shape and vasculature (Nishimura et al. 2004).

Genetic differentiation between populations and species was overall low, but higher between species (F_{ST}= 0.05-0.07) than among populations within species (F_{ST}= 0.01-0.03) (Suppl. 5) (hypothesis 3). Similar results were obtained in other sympatric hybridizing oak species with low overall genetic differentiation between species, and this pattern has been attributed to either interspecific gene flow or shared ancestral variation (Cavender-Bares and Pahlich 2009; Curtu et al. 2007b; Peñaloza-Ramírez et al. 2010; Salvini et al. 2009). However, introgression has been shown to be the more parsimonious explanation in most cases with selection maintaining species integrity despite interspecific gene flow (Lexer et al. 2006). In the present study, genetic differentiation between species at nSSRs (F_{ST}= 0.061) was slightly higher than at EST-SSRs (F_{ST}= 0.047) mostly due to the high interspecific differentiation at locus 3A05 (F_{ST}= 0.147, p≤ 0.001). After the exclusion of the two loci with the highest interspecific differentiation (nSSR 3A05 and EST-SSR GOT021) differentiation between species was similar for both EST-SSRs (F_{ST}= 0.042) and nSSRs (F_{ST}= 0.049). Examination of species pairs in four *Q. rubra* and *Q. ellipsoidalis* populations in the Baraga Plains region showed differentiation at nSSRs

(F_{ST}= 0.065) to be slightly higher than at EST-SSRs (F_{ST}= 0.060). Thus, based on the limited number of markers analyzed there was no evidence that EST-SSR markers show a higher level of interspecific genetic differentiation than nSSR markers (rejection of hypothesis 4). More markers with known genomic locations need to be screened to find genomic regions under divergent selection between species.

Of the two outlier programs (LOSITAN and Arlequin) utilized, only Arlequin identified potential outliers, both under the finite and hierarchical island model. This discrepancy has been observed in numerous studies using more than one outlier screening method (see Coyer et al. 2011; Goicoechea et al. 2012) and is likely due to differences in the algorithms used by each program. For example, LOSITAN computes heterozygosities between populations directly, while Arlequin computes average heterozygosity within populations and the average F_{ST} from which heterozygosities between populations are deduced (Excoffier et al. 2009). Additionally, LOSITAN has options to exclude loci potentially under selection when calculating the mean F_{ST} and to increase the precision of the simulated F_{ST}. A study examining the performance of different programs in identifying outliers found that LOSITAN had lower type I and type II errors than two other outlier programs including Arlequin 3.5, but false positives were described for both programs especially for loci supposedly under balancing selection (Narum and Hess 2011). While we cannot discard the possibility of false positives, it is worth noting that one of the outliers detected between species among the four Baraga Plains populations, GOTO21, has a known putative function in drought tolerance and leaf shape variation consistent with species differences between *Q. rubra* and *Q. ellipsoidalis* (see above). Additionally, nSSR 3A05 was identified as an outlier under both migration models in Arlequin and may be a candidate for further investigation as it could be in linkage disequilibrium with a genomic area under selection (hypothesis 5). Since these loci were only detected by one outlier program, we argue that the application of different outlier detection programs (e.g., LOSITAN and Arlequin) and the inclusion of additional evidence (e.g., derived from functional or QTL

analyses) can provide a comprehensive approach to detect loci potentially under selection as candidates for further investigation.

Increasing the number of markers analyzed, particularly genic markers with putative functions that may be involved in different local adaptations of the two species may enhance the discriminative power and allow for the identification of outlier loci that are involved in pre- or post-zygotic isolation between the species. For example, the two loci (3A05, GOT021) that revealed relatively high levels of genetic variation between species and were identified as putative outlier loci among the four *Q. rubra* and *Q. ellipsoidalis* populations in the Baraga Plains region could be further characterized by sequencing. Gene flow analyses based on paternity exclusion at microsatellites can reveal the actual level of contemporary gene flow between species. Also, the analysis of additional populations and markers in the contact zones between both species will be necessary to characterize levels of hybridization and introgression and to better understand the isolation mechanisms between *Q. rubra* and *Q. ellipsoidalis*.

Acknowledgements

We would like to thank James Schmierer for his help in sample collection in the Baraga Plains and Dr. Kerry Woods for his help in the identification of populations in the Huron Mountain Wildlife Reserve. *Q. rubra* and *Q. ellipsoidalis* reference samples were provided by Jeanne Romero-Severson and Andrew Hipp. Additionally we would like to thank Jonathan Riehl for invaluable advice and guidance on the construction of many of the graphics. We also thank two anonymous reviewers for their very helpful comments on earlier drafts. Funding for the study came from Michigan Technological University start-up funds to Oliver Gailing, the Michigan Technological University Research Excellence fund, the USDA McIntire Stennis fund, the Huron Mountain Wildlife Foundation, the Hanes Trust and the NSF Plant Genome Research program (NSF 1025974).

References

Abrams MD (1990) Adaptations and responses to drought in *Quercus* species of North America. Tree Physiol 7:227-238

Aldrich PR, Cavender-Bares J (2011) *Quercus* Wild Crop Relatives: Genomic and Breeding Resources. In: Kole C (ed). Springer Berlin Heidelberg, pp 89-129

Aldrich PR, Michler CH, Sun WL, Romero-Severson J (2002) Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). Mol Ecol Notes 2:472-474

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. BMC Bioinformatics 9:323

Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. Trends Ecol Evol 25:530-536

Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution 50:900-908

Barnes BVW, Warren Herbert Jr. (2004) Michigan Trees, Revised and Updated: A Guide to the Trees of the Great Lakes Region. University of Michigan Press Ann Arbor

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? Trends Ecol Evol 20:435-440

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. P Roy Soc B-Biol Sci 263: 1619-1626

Boavida LC, Silva JP, Feijo JA (2001) Sexual reproduction in the cork oak (*Quercus suber* L).II. crossing intra- and interspecific barriers. Sex Plant Reprod 14:143-152

Burger WC (1975) The species concept in *Quercus*. Taxon 24:45-50

Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. Heredity 99: 616-623

Cavender-Bares J, Pahlich A (2009) Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). Am J Bot 96:1690-1702

Chesnoiu EN, Sofletea N, Curtu AL, Toader A, Radu R, Enescu M (2009) Bud burst and flowering phenology in a mixed oak forest from Eastern Romania. Ann For Res 52: 199-206

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676

Coyer JA, Hoarau G, Pearson G, Mota C, Juterbock A, Alpermann T, John U, Olsen JL (2011) Genomic scans detect signatures of selection along a salinity gradient in populations of the intertidal seaweed *Fucus serratus* on a 12km scale. Mar Genomics 4:41-49

Craft KJ, Ashley MV, Koenig WD (2002) Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. Am J Bot 89 (11):1792-1798

Curtu A, Gailing O, Finkeldey R (2007a) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. BMC Evol Biol 7:218

Curtu A, Gailing O, Leinemann L, Finkeldey R (2007b) Genetic variation and differentiation within a natural community of five oak species (*Quercus* spp.). Plant Biol 9:116-126

Curtu A, Gailing O, Finkeldey R (2009) Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. BMC Evol Biol 9:284

Desikan R, Horák J, Chaban C, Mira-Rodado V, Witthöft J, Elgass K, Grefen C, Cheung M-K, Meixner AJ, Hooley R, Neill SJ, Hancock JT, Harter K (2008) The histidine kinase AHK5 integrates endogenous and environmental signals in *Arabidopsis* guard cells. PLoS One 3:e2491

Doak DF, Morris WF (2010) Demographic compensation and tipping points in climate-induced range shifts. Nature 467:959-962

Dodd RS, Afzal-Rafii Z (2004) Selection and dispersal in a multispecies oak hybrid zone. Evolution 58:261-269

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (GA/CT) microsatellites in the bur oak, *Quercus macrocarpa*. Theor Appl Genet 91:137-141

Durand J, Bodenes C, Chancerel E, Frigerio J-M, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn H-P, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herran A, Ikaran Z, Cabane C, Ueno S, Alberto F, Dumoulin P-Y, Guichoux E, de Daruvar A, Kremer A, Plomion C (2010) A

fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11:570

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132

Emms SK, Arnold ML (1997) The effect of habitat on parental and hybrid fitness: transplant experiments with Louisiana irises. Evolution 51:1112-1119

ESRI (Environmental Systems Resource Institute) (2009) ArcMap 9.2: ArcGIS for desktop. Redlands, California (http://www.esri.com/software/arcgis/arcgis-for-desktop)

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611-2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchical structured population. Heredity 103: 285-298

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7: 574-578

Gailing O, Lind J, Lilleskov E (2012) Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* E. J. Hill populations in contrasting environments. Plant Syst Evol 298: 1533-1545

Goicoechea PG, Petit RJ, Kremer A (2012) Detecting the footprints of divergent selection in oaks with linked markers. Heredity DOI:10.1038/hdy.2012.51

Hipp AL (2010) Hill's oak: The taxonomy and dynamics of a Western Great Lake endemic. Arnoldia 67

Hipp AL, Weber JA (2008) Taxonomy of Hill's oak (*Quercus ellipsoidalis*: Fagaceae): evidence from AFLP data. Syst Bot 33:148-158

Hokanson SC, Isebrands JG, Jensen RJ, Hancock JF (1993) Isozyme variation in oaks of the Apostle Islands in Wisconsin: genetic structure and levels of inbreeding in *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). Am J Bot 80:1349-1357

Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ (1997) How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. Evolution 51:747-755

Jensen RJ, Hokanson SC, Isebrands JG, Hancock JF (1993) Morphometric variation in oaks of the Apostle Islands in Wisconsin: evidence of hybridization between *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). Am J Bot 80:1358-1366

Jimenez P, Lopez de Heredia U, Collada C, Lorenzo Z, Gil L (2004) High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history. Heredity 93: 510-515

Jimenez P, Agundez D, Alia R, Gil L (1999) Genetic variation in central and marginal populations of *Quercus suber* L. Silvae Genet 48: 278-284

Kampfer S, Lexer C, Glössl J, Steinkellner H (1998) Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. Hereditas 129:183-186

Kremer A, Petit R (1993) Gene diversity in natural populations of oak species. Ann For Sci 50:186s-202s

Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Van Dam B, Ducousso A, Forrest I, Lopez de Heredia U, Lowe AJ, Tutkova M, Munro RC, Steinhoff S, Badeau V (2002) Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. Ann For Sci 59:777-787

Langella O (1999) Populations 1.2.30: a population genetic software CNRS UPR9034

Latter BDH (1973) The island model of population differentiation: a general solution. Genetics 73:147-157

Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S (2009) Species relative abundance and direction of introgression in oaks. Mol Ecol 18:2228-2242

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. Mol Ecol 15:2007-2012

Lopez de Heredia U, Valbuena-Carabaña M, Cordoba M, Gil L. (2009) Variation components in leaf morphology of recruits of two hybridising oaks [*Q. petraea* (Matt.) Liebl. and *Q. pyrenaica* Willd.] at a small spatial scale. Eur J Forest Res 128: 543-554

Lorenzo Z, Burgarella C, Lopez de Heredia U, Lumaret R, Petit RJ, Soto A, Gil L (2009) Relevance of genetics for conservation policies: the case of Minorcan cork oaks. Ann Bot-London 104:1069-1076

Luro F, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R (2008) Transferability of the EST-SSRs developed on *Nules clementine* (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. BMC Genomics 9:287

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

McShea WJ, Healy WM, Devers P, Fearer T, Koch FH, Stauffer D, Waldon J (2007) Forestry matters: Decline of oaks will impact wildlife in hardwood forests. J Wildlife Manage 71:1717-1728

Mir C, Toumi L, Jarne P, Sarda V, Di Guisto F, Lumaret R (2006) Endemic North African *Quercus afares* Pomel originates from hybridisation between two genetically very distant oak species (*Q. suber* L. and *Q. canariensis* Willd.): evidence from nuclear and cytoplasmic markers. Heredity 96:175-184

Moran EV, Willis J, Clark JS (2012) Genetic evidence for hybridization in red oaks (*Quercus* sect. Lobatae, Fagaceae). Am J Bot 99:92-100

Muir G, Fleming CC, Schlötterer C (2001) Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Mol Biol Evol 18:112-119

Nagy ES, Rice KJ (1997) Local adaptation in two subspecies of an annual plant: implications for migration and gene flow. Evolution 51:1079-1089

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. Mol Ecol Resour 11:184-194

Nei M (1973) Analysis of gene diversity in subdivided populations. P Natl Acad Sci USA 70:3321-3323

Nishimura C, Ohashi Y, Sato S, Kato T, Tabata S, Ueguchi C (2004) Histidine kinase homologs that act as cytokinin receptors possess overlapping functions in the regulation of shoot and root growth in *Arabidopsis*. Plant Cell 16:1365-1377

Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357-358

Pautasso M (2009) Geographical genetics and the conservation of forest trees. Perspect Plant Ecol 11:157-189

Peakall ROD, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288-295

Peñaloza-Ramírez JM, González-Rodríguez A, Mendoza-Cuenca L, Caron H, Kremer A, Oyama K (2010) Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. Ann Bot 105:389-399

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

Raymond M, Rousset F (1995) GenePop (Version-1.2) - Population genetics software for exact tests and ecumenicism. J Hered 86:248-249

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223-225

Rieseberg LH, Ellstrand NC, Arnold M (1993) What can molecular and morphological markers tell us about plant hybridization? CRC Cr Rev Plant Sci 12:213-241

Rosenburg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137-138

Rushton B (1993) Natural hybridization within the genus *Quercus* L. Ann For Sci 50:73s-90s

Salvini D, Bruschi P, Fineschi S, Grossoni P, Kjær ED, Vendramin GG (2009) Natural hybridisation between *Quercus petraea* (Matt.) Liebl. and *Quercus pubescens* Willd. within an Italian stand as revealed by microsatellite fingerprinting. Plant Biol 11:758-765

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénès C, Burg K, Kremer A (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. Genetics 168:1615-1626

Sneath PHA, Sokal RR (1973) Numerical Taxonomy: The Principles and Practice of Numerical Classification. Freeman, San Fransisco

Sork V, Huang S, Wiener E (1993) Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. Ann For Sci 50:261s-270s

Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glossl J (1997) Identification and characterization of $(GA/CT)_n$-microsatellite loci from *Quercus petraea*. Plant Mol Biol 33:1093-1096

Sullivan AR, Lind JF, McCleary TS, Romero-Severson J, Gailing O (2012) Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. Plant Mol Biol Rep. (DOI: 10.1007/s11105-012-0495-6)

Valbuena-Carabaña M, Gonzáles-Martínez SC, Sork VL, Collada C, Soto A, Goicoechea PG, Gil L (2005) Gene flow and hybridisation in a mixed oak

forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. Heredity 95: 457-465

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4: 535-538

Van Valen L (1976) Ecological species, multispecies, and oaks. Taxon 25:233-239

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48-55

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. Evolution 38:1358-1370

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. P Natl Acad Sci 88:2540-2544

Woodall CW, Oswalt CM, Westfall JA, Perry CH, Nelson MD, Finley AO (2009) An indicator of tree migration in forests of the Eastern United States. Forest Ecol Manag 257:1434-1444

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330-338

Table 2 Sampled *Quercus rubra* and *Q. ellipsoidalis* populations

Abbreviation	Region	Population	Species	Sample size (n)	Soil characteristics[b]	Forest type[a]	Latitude	Longitude	Mean Altitude (m)
BR-1	Brockway Mountain	Brockway 1	*Q. rubra*	50	2 (53F: Arcadian-Michigamme-Rock outcrop)	Quercus, Betula, Pinus	47° 27' 57.478" N	87° 54' 59.209" W	407
FC-A	Baraga Plains, Ford Center	Stand A	*Q. rubra*	49	3 (25B: Munising-Yalmer loamy sand)	Quercus, Acer, Pinus	46° 39' 9.407" N	88° 30' 6.962" W	297
FC-B	Baraga Plains, Ford Center	Stand B	*Q. rubra*	53	3 (25B: Munising-Yalmer loamy sand)	Quercus, Acer, Pinus	46° 40' 27.937" N	88° 31' 27.397" W	423
FC-C	Baraga Plains, Ford Center	Stand C	*Q. ellipsoidalis*	56	1 (10B: Grayling sand)	Quercus, Pinus	46° 39' 14.454" N	88° 35' 25.616" W	394
FC-E	Baraga Plains, Ford Center	Stand E	*Q. ellipsoidalis*	69	1 (10B: Grayling sand)	Quercus, Pinus	46° 39' 55.879" N	88° 33' 19.775" W	398
HMR-IH	Huron Mountain Reserve	Ives Hill	*Q. rubra*	88	2 (56E: Peshekee-Rock outcrop)	Quercus, Acer, Pinus	46° 51' 12.884" N	87° 50' 42.824" W	257
HMR-LI	Huron Mountain Reserve	Lake Ives	*Q. rubra*	19	2 (56E: Peshekee-Rock outcrop)	Quercus, Acer, Pinus	46° 50' 39.462" N	87° 51' 17.978" W	246
HMR-LP	Huron Mountain Reserve	Lily Pond	*Q. rubra*	69	2 (56E: Peshekee-Rock outcrop)	Quercus, Acer, Pinus	46° 50’ 59.813” N	87° 49’ 48.806” W	246
HMR-MI	Huron Mountain Reserve	Mount Ives	*Q. rubra*	81	2 (56E: Peshekee-Rock outcrop)	Quercus, Acer, Pinus	46° 51’ 20.783” N	87° 51’ 24.026” W-	307
HMR-PL	Huron Mountain Reserve	Pine Lake	*Q. rubra*	12	1 (11C: Deer Park sand)	Quercus, Pinus	46° 53' 19.442" N	87° 52' 49.888" W	194
MTU-1	Houghton	Michigan TechTrails 1	*Q. rubra*	36	3 (10B: Munising loamy fine sand)	Quercus, Pinus	47° 6' 24.649" N	88° 32' 51.209" W	266

[a] Field observations

[b] Rating developed from drainage classes with 1 being excessively drained, 2 being well drained, and 3 being moderately well drained; soil type and drainage class identified according to the Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Official Soil Series Descriptions available at http://websoilsurvey.nrcs.usda.gov/app/HomePage.htm Accessed [11/29/2012]

Table 3 Microsatellite marker characteristics

Locus	Repeat motif	Forward primer sequence (5'-3') Reverse primer sequence (5'-3')	T_a (°C)	Size Range (in base pairs)	Species	Linkage group[g]	Putative function[b]
G0T021[d]	$(AT)_{13}$	VIC-AGAAAGTTCCAGGGAAAGCA CTTCGTCCCCAGTTGAATGT	59	95-103	*Q. robur*	3	Histidine kinase-4 like
PIE040[d]	$(TTC)_8$	NED- GTGAGAGAGAGAGAGACAAAGAAGAAAAA AAATTCTCCGCCACATTGAG	59	155-179	*Q. robur*	11	Basic leucine zipper transcription factor-like protein
GOT004[d]	$(TG)_{12}$	6-FAM-GGGCATATTGATCGCTTAGG TGACGATTCATACATTCCATGAT	59	209-298	*Q. robur*	7	Tonoplast intrinsic protein
FIR004[d]	$(CT)_{18}$	6-FAM-TCTCTCTCAGGGCAGCTTCT AACCAAACTCAGATCCAGATTCA	59	122-186	*Q. robur*	3	no hit
FIR110 L1[a,d]	$(AG)_{12}$	NED-ACTTGCTCGCTTCAACCTTC ATTCCTCCTCATCAGGCTCA	56	166-200	*Q. robur*	6	transport protein sec23
FIR110 L2[a,d]	$(AG)_{12}$	NED-ACTTGCTCGCTTCAACCTTC ATTCCTCCTCATCAGGCTCA	56	180-236	*Q. robur*	6	unknown
GOT009[d]	$(TC)_7$	6-FAM-CACCTCACTAAGCAACCTGTCA TTTTGGAGGCGGAGATAATG	56	221-249	*Q. robur*	12	uncharacterized protein
PIE099[d]	$(TC)_9$	VIC- GTAAAACGACGGCCAGTGTGGCTACCGACTA CTACCACTTC CGGTGGACCCAATATGTAAC	56	179-209	*Q. robur*	8	no hit

[a] primer pair amplifies two loci

[b] Putative function determined (BLAST search following method in (Luro et al. 2008))

[c] (SSRs) (Aldrich et al. 2002)

[d] (EST-SSRs) (Durand et al. 2010; Sullivan et al. 2013)

[e] (SSRs) (Sullivan et al. 2013)

[f] (SSRs) (Steinkellner et al. 1997)

[g] on *Q. robur* linkage map (Durand et al. 2010)

Table 3 continued

Locus	Repeat motif	Forward primer sequence (5'-3') Reverse primer sequence (5'-3')	T_a (°C)	Size Range (in base pairs)	Species	Linkage group[g]	Putative function[b]
FIR048[d]	$(CT)_9$	PET-TGCACCAAAATTGGAGGATG TTGATGCAAGGTGCAGTTTC	56	185-229	*Q. robur*	2	cell division protein
quru-GA-1F07[c]	$(GA)_{22}$	NED-CCGGTCAAAGAAGTTATCAGA GGGTGGATTGGGTTTCTACCTA	59	294-352	*Q. rubra*	NA	NA
quru-GA-0E09[c]	$(GA)_{16}$	6-FAM-TGCCATCCCTATACACAACCA CCTCCATCACAAAGTTGCC	59	174-250	*Q. rubra*	NA	NA
quru-GA-0C11[c]	$(GA)_{15}$	VIC-ATACCCAGCTCCCATGACCA TCCCCAAATTCAGGTAGTGT	59	201-241	*Q. rubra*	NA	NA
QpZAG15[f]	$(AG)_{23}$	HEX-CGATTTGATAATGACACTATGG CATCGACTCATTGTTAAGCAC	54	103-148	*Q. petraea*	9	NA
3D15[e]	$(CA)_{15}$	NED-GGTGGTGGCAGATACACTGG GACTCAGACAACCAACTTCAGG	62	208-238	*Q. rubra*	NA	NA
1P10[e]	$(TG)_{12}$	6-FAM-ATTTCTGATGCAGGGTGTCG TAGGCCAAGGACCAGAGACC	62	237-269	*Q. rubra*	NA	NA
2P24[e]	$(CA)_{14}$	VIC-GCAAGAGATCACACACAAACTAGC CTTTGGGTTCACCAAACAGC	62	130-176	*Q. rubra*	NA	NA
3A05[e]	$(CT)1_2$	PET-AACGTGACCTCTCTCACAGC AGTGCTGGAGTGCTCATGG	62	137-162	*Q. rubra*	NA	NA

Table 4 Genetic variation at 15 microsatellite loci in eight populations with a sample size of N ≥ 49

	All populations						***Q. ellipsoidalis* populations (N = 2)**						***Q. rubra* populations (N = 6)**						**Pairwise F_{ST} between species**
Locus	**N**	**N_a**	**H_o**	**H_e[b]**	**F**	**F_{ST}[a]**	**N**	**N_a**	**H_o**	**H_e[b]**	**F**	**F_{ST}[a]**	**N**	**N_a**	**H_o**	**H_e[b]**	**F**	**F_{ST}[a]**	**F_{ST}[a]**
G0T021[e]	64	3	0.25	0.23	-0.07	**0.06**	63	3	0.05	0.07	0.13	0.005	65	3	0.32	0.28	-0.13	**0.03**	**0.07**
PIE040[e]	64	8	0.59	0.64	0.08	**0.02**	63	8	0.67	0.73	0.06	**0.01**	65	9	0.56	0.61	0.08	**0.005**	**0.03**
GOT004[e]	64	10	0.61	0.79	0.22	**0.04**	62	10	0.41	0.84	0.51	**0.02**	64	10	0.67	0.77	0.13	**0.01**	**0.075**
FIR004[e]	64	21	0.65	0.91	0.29	**0.02**	63	21	0.83	0.85	0.30	**0.02**	65	21	0.66	0.91	0.29	**0.01**	**0.014**
GOT009[e]	64	10	0.72	0.76	0.05	**0.02**	63	9	0.75	0.80	0.05	0.0	65	10	0.71	0.75	0.05	**0.02**	**0.027**
PIE099[e]	64	12	0.62	0.86	0.29	**0.03**	63	11	0.81	0.87	0.08	**0.02**	64	12	0.55	0.86	0.36	**0.02**	**0.027**
FIR048[e]	64	14	0.79	0.81	0.03	**0.04**	63	13	0.87	0.88	0.01	0.007	65	15	0.76	0.79	0.04	**0.01**	**0.078**
EST-SSRs Mean	64	11	0.60	0.71	0.13	0.03	63	11	0.63	0.72	0.16	0.01	65	11	0.60	0.71	0.12	0.01	**0.047**

[a] significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of α= 0.05

[b]unbiased expected heterozygosity (Peakall and Smouse 2006)

[c] primer pair amplifies two loci

[d] (SSRs) (Aldrich et al. 2002)

[e] (EST-SSRs) (Durand et al. 2010; Sullivan et al. 2013)

[f] (SSRs) (Sullivan et al. 2013)

[g](SSRs) (Steinkellner et al. 1997)

Table 4 continued

	All populations						*Q. ellipsoidalis* populations (N = 2)						*Q. rubra* populations (N = 6)						Pairwise F_{ST} between species
Locus	**N**	**N_a**	**H_o**	**H_e[b]**	**F**	**F_{ST}[a]**	**N**	**N_a**	**H_o**	**H_e[b]**	**F**	**F_{ST}[a]**	**N**	**N_a**	**H_o**	**H_e[b]**	**F**	**F_{ST}[a]**	**F_{ST}**
quru-GA-1F07[d]	63	20	0.75	0.91	0.17	**0.02**	61	16	0.66	0.88	0.25	0.0	64	22	0.78	0.92	0.14	**0.01**	**0.026**
quru-GA-0E09[d]	64	21	0.67	0.91	0.27	**0.03**	63	21	0.72	0.86	0.18	**0.008**	65	21	0.65	0.93	0.30	**0.02**	**0.030**
quru-GA-0C11[d]	64	14	0.83	0.87	0.05	**0.03**	63	14	0.87	0.87	0.007	**0.02**	65	14	0.82	0.87	0.06	**0.01**	**0.037**
QpZAG15[g]	64	15	0.83	0.87	0.04	**0.04**	63	15	0.83	0.85	0.03	**0.005**	65	15	0.83	0.88	0.05	**0.02**	**0.070**
3D15[f]	64	11	0.68	0.78	0.13	**0.06**	62	8	0.67	0.67	0.003	**0.02**	65	12	0.68	0.82	0.17	**0.02**	**0.090**
1P10[f]	64	12	0.68	0.72	0.05	**0.02**	63	12	0.67	0.72	0.07	**0.03**	65	13	0.69	0.72	0.05	**0.02**	**0.012**
2P24[f]	64	10	0.69	0.81	0.15	**0.05**	63	10	0.59	0.77	0.24	0.009	65	10	0.73	0.82	0.12	**0.02**	**0.081**
3A05[f]	64	9	0.59	0.75	0.22	**0.08**	62	8	0.36	0.66	0.45	**0.01**	65	9	0.67	0.78	0.15	**0.009**	**0.147**
SSRs Mean	64	14	0.72	0.83	0.14	0.04	63	13	0.67	0.79	0.15	0.01	65	15	0.73	0.84	0.13	0.02	**0.061**
Overall Mean	64	13	0.66	0.78	0.13	**0.04**	62	12	0.64	0.76	0.16	**0.01**	65	13	0.67	0.78	0.12	**0.02**	**0.054**

Table 5 Total and mean genetic variation in eight populations with sample size N≥ 49 at all 15 microsatellite loci and at the 8 SSRs and 7 EST-SSRs separately

All Markers						
Population	**Species[b]**	**N**	**N_a**	**H_o**	**H_e[a]**	**F**
BR-1	*Q. rubra*	50	13	0.67	0.76	0.11
FC-A	*Q. rubra*	49	12	0.66	0.80	0.16
FC-B	*Q. rubra*	53	12	0.66	0.79	0.15
FC-C	*Q. ellipsoidalis*	56	12	0.68	0.77	0.10
FC-E	*Q. ellipsoidalis*	69	12	0.60	0.75	0.21
HMR-IH	*Q. rubra*	88	14	0.70	0.78	0.09
HMR-LP	*Q. rubra*	69	14	0.66	0.78	0.13
HMR-MI	*Q. rubra*	81	13	0.69	0.77	0.09
Mean Variation		64	13	0.66	0.78	0.13
Total		-	20	-	0.80	-
SSRs						
Population	**Species[b]**	**N**	**N_a**	**H_o**	**H_e[a]**	**F**
BR-1	*Q. rubra*	49	15	0.76	0.83	0.07
FC-A	*Q. rubra*	49	13	0.71	0.86	0.17
FC-B	*Q. rubra*	53	14	0.74	0.85	0.12
FC-C	*Q. ellipsoidalis*	56	13	0.72	0.81	0.10
FC-E	*Q. ellipsoidalis*	69	13	0.62	0.77	0.19
HMR-IH	*Q. rubra*	88	15	0.73	0.83	0.12
HMR-LP	*Q. rubra*	69	15	0.70	0.85	0.16
HMR-MI	*Q. rubra*	81	15	0.75	0.84	0.10
Mean Variation		64	14	0.72	0.83	0.13
Total		-	22	-	0.86	-
EST-SSRs						
Population	**Species[b]**	**N**	**N_a**	**H_o**	**H_e[a]**	**F**
BR-1	*Q. rubra*	50	11	0.56	0.69	0.15
FC-A	*Q. rubra*	49	10	0.62	0.74	0.13
FC-B	*Q. rubra*	53	10	0.57	0.71	0.16
FC-C	*Q. ellipsoidalis*	56	10	0.64	0.72	0.08
FC-E	*Q. ellipsoidalis*	69	11	0.57	0.74	0.23
HMR-IH	*Q. rubra*	88	12	0.66	0.72	0.05
HMR-LP	*Q. rubra*	69	12	0.61	0.70	0.08
HMR-MI	*Q. rubra*	80	12	0.61	0.70	0.08
Mean Variation		64	11	0.61	0.70	0.12
Total		-	18	-	0.73	-

[a] unbiased expected heterozygosity (Peakall and Smouse 2006)

[b]genetic species assignment made using STRUCTURE (Pritchard et al. 2000)

Table 6 Hierarchical Analysis of Molecular Variance (AMOVA) calculated in Arlequin 3.5 (Excoffier and Lischer 2010)

Source of variation	*df*	Sum of squares	Variance components	Percentage variation	p-value
Among species	1	133	0.304	4.95	0.0387
Among populations within species	6	106	0.087	1.41	<0.001
Within populations	1022	5872	5.751	93.64	<0.001
Total	1029	6110	6.14	-	-

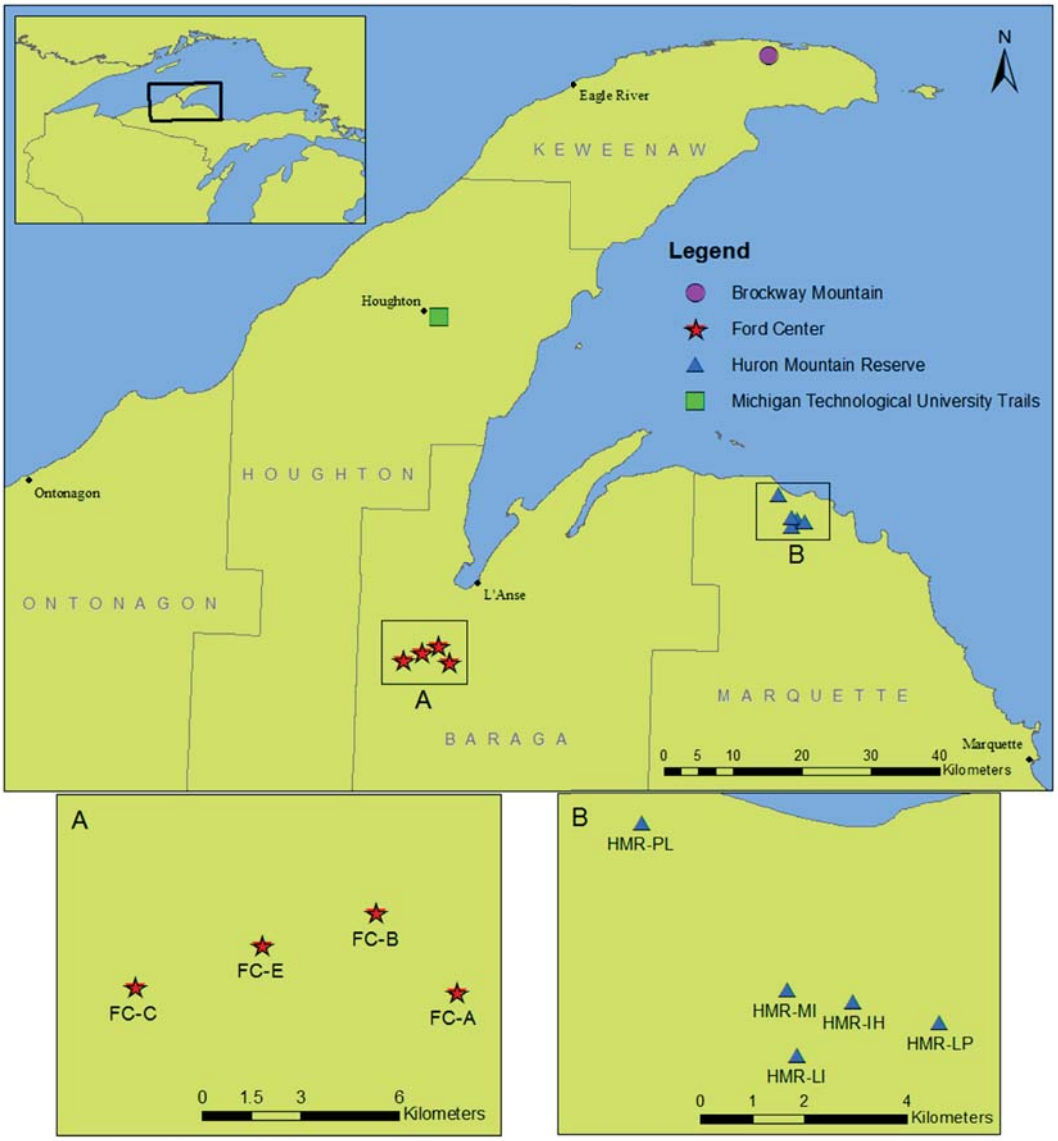


Figure 1 Sampling locations

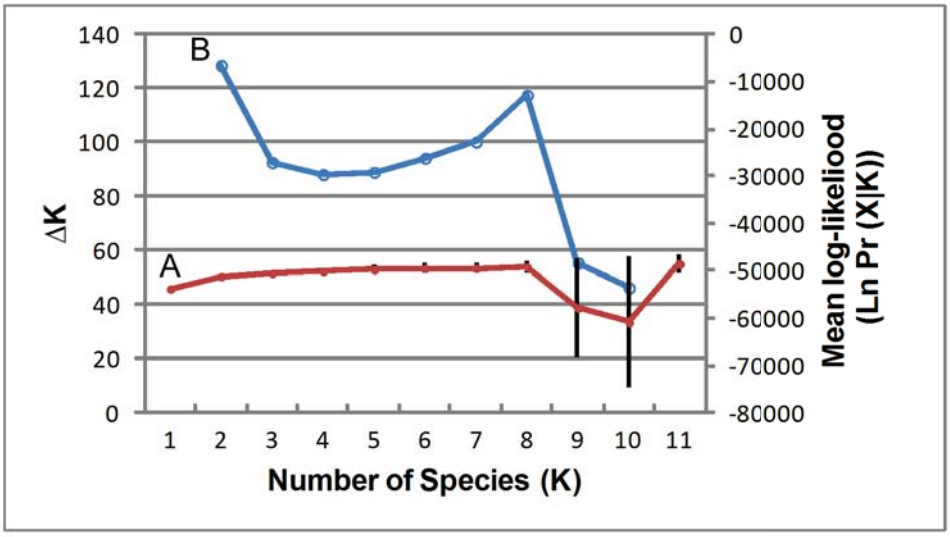


Figure 2 The number of populations best fitting the microsatellite data by A) mean log-likelihood and associated variance across five independent runs for each value of K ranging from 1 to 11, and B) value of ΔK as a function of K where the modal value points to the most likely number of partitions (Evanno et al 2005)

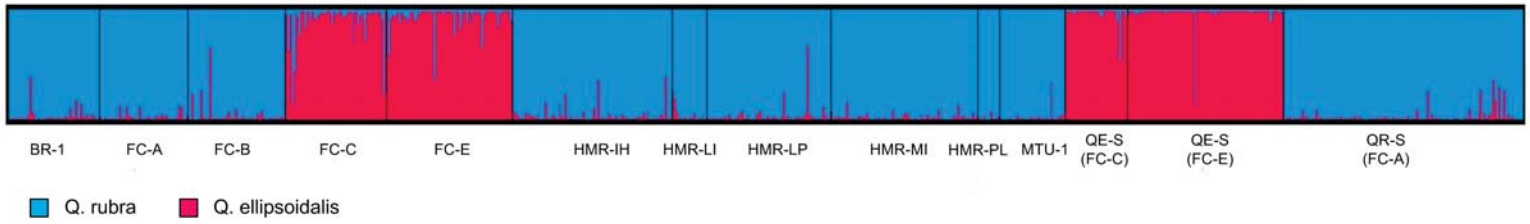


Figure 3 Genetic assignment of individuals and populations according to the Bayesian method implemented in the program STRUCTURE 2.2 (Pritchard, 2000) displayed using DISTRUCT 1.1 (Rosenburg 2004). One sample of 5 iterated runs considered to best explain the data is shown. Each thin vertical line represents an individual and the proportion of each color is the proportion of ancestry derived from each of the two main genetic groups (K= 2) inferred. Populations are separated by black vertical lines

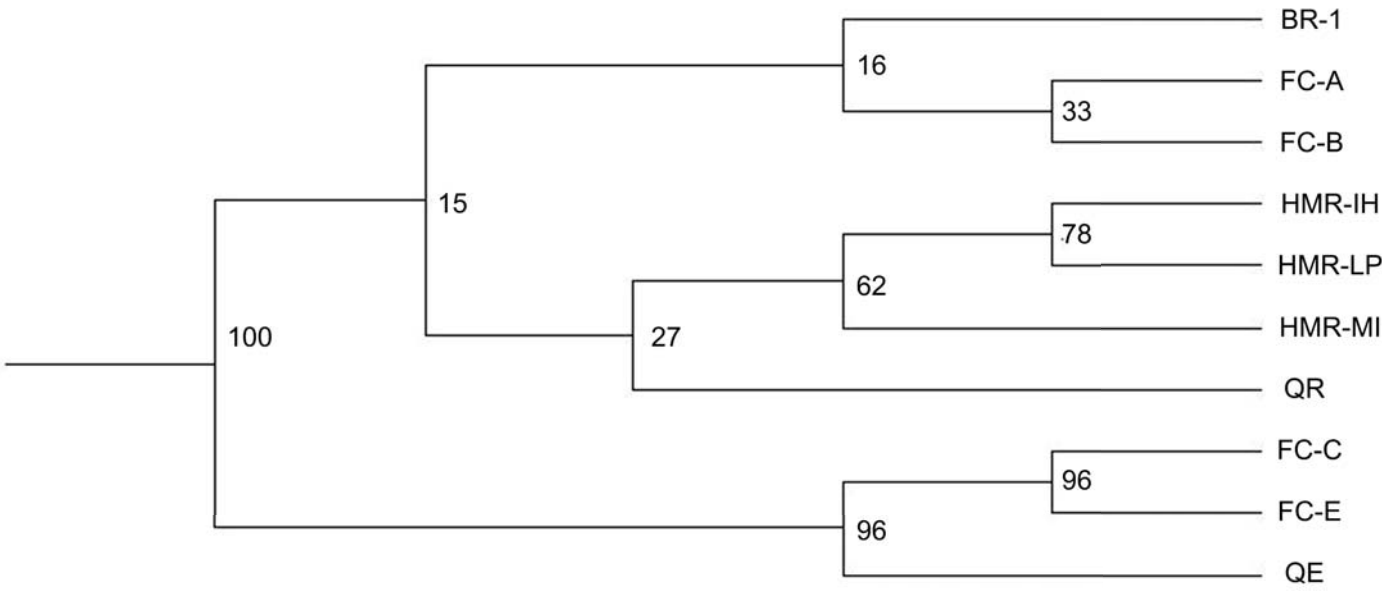


Figure 4 Dendrogram (UPGMA method) based on genetic distance (F_{ST}) at EST-SSRs and SSRs; numbers at nodes are percentages over 1000 bootstrap replicates using Populations 2.0 (Langella 1999)

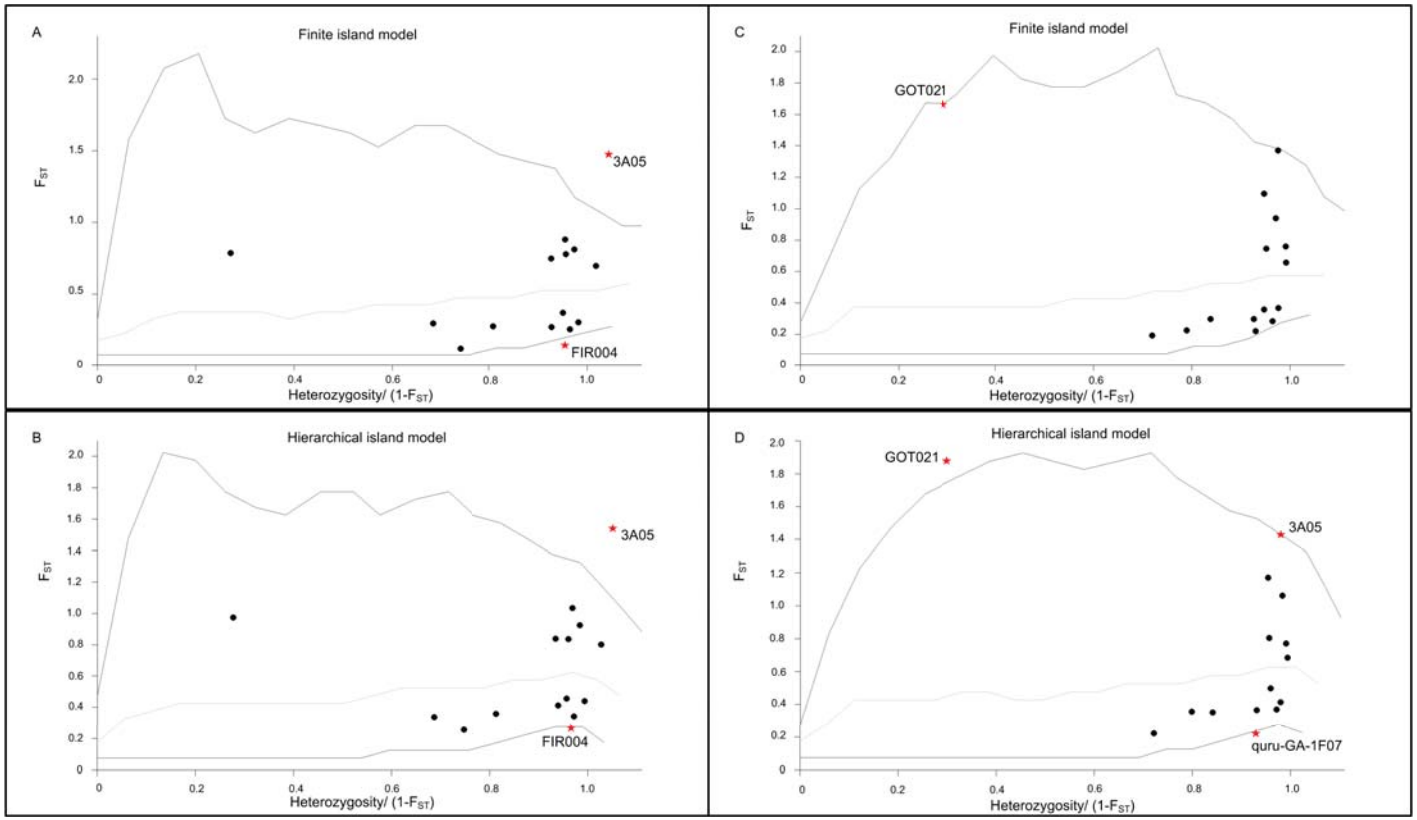


Figure 5 Outlier loci analysis of fifteen markers (A) between species among all populations (N≥ 49) (finite island model), (B) among all populations (N≥ 49) (hierarchical island model), (C) between species among the four Baraga Plains populations (finite island model), (D) among the four Baraga Plains populations (hierarchical island model), as calculated by Arlequin 3.5 (Excoffier and Lischer 2010). Outliers are indicated by a filled star shape while all others are filled circles and dark grey lines represent the 5% quantiles with markers outside the lower 5% quantile under balancing selection and the markers outside the upper 5% quantile under divergent selection while the light gray line represents the median (50%)

Supplement 1 Relative frequency of individuals identified as *Q. rubra* and *Q. ellipsoidalis*, hybrids, and introgressive forms using genetic assignment in STRUCTURE without *a priori* species information (Pritchard, 2000)

Population	*Q. rubra*	*Q. ellipsoidalis*	Introgressive forms (*Q. rubra*)	Introgressive forms (*Q. ellipsoidalis*)	Hybrids
BR-1	0.92	0	0.06	0	0.02
FC-A	0.90	0	0.10	0	0
FC-B	0.94	0	0.04	0.02	0
FC-C	0	0.82	0.04	0.12	0.02
FC-E	0	0.86	0.01	0.12	0.01
HMR-IH	0.93	0	0.06	0	0.01
HMR-LI	0.89	0	0.11	0	0
HMR-LP	0.96	0	0.03	0.01	0
HMR-MI	0.98	0	0.02	0	0
HMR-PL	1.00	0	0.00	0	0
MTU-1	0.97	0	0.03	0	0
QE-S (FC-C)	0	0.97	0	0	0.03
QE-S (FC-E)	0	0.98	0.01	0.01	0
QR-S (FC-A)	0.95	0	0.05	0	0

Supplement 2 Relative frequency of statistically significant hybrids and introgressive forms using genetic assignment in STRUCTURE with *a priori* morphological species information at two different immigration rates (Pritchard, 2000)

Population	Introgressive forms (*Q. rubra*) ν= 0.1	Introgressive forms (*Q. rubra*) ν= 0.05	Introgressive forms (*Q. ellipsoidalis*) ν= 0.1	Introgressive forms (*Q. ellipsoidalis*) ν= 0.05	Hybrids ν= 0.1	Hybrids ν= 0.05
BR-1	0	0	0	0	0.02	0.02
FC-A	0.02	0	0	0	0	0
FC-B	0	0	0.02	0	0	0
FC-C	0.04	0.04	0.02	0	0.02	0.02
FC-E	0	0	0.04	0.01	0.01	0.01
HMR-IH	0.03	0.02	0	0	0	0
HMR-LI	0	0	0	0	0	0
HMR-LP	0	0	0	0	0	0
HMR-MI	0.01	0	0	0	0	0
HMR-PL	0	0	0	0	0	0
MTU-1	0.03	0.03	0	0	0	0
QE-S (FC-C)	0	0	0	0	0.03	0.03
QE-S (FC-E)	0.01	0.01	0	0	0	0
QR-S (FC-A)	0.01	0	0	0	0	0

Supplement 3 HWE Exact Tests (Weir and Cockerham, 1984) by locus in each population. Significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of α= 0.05

Population	F	p-value	Standard Error
GOT021			
BR-1	-0.1136	1	0
FC-A	-0.2462	0.1651	0.0006
FC-B	-0.0209	1	0
FC-C	-0.0046	1	0
FC-E	0.2641	0.0257	0.0002
HMR-IH	-0.1428	0.4149	0.0008
HMR-LP	-0.088	1	0
HMR-MI	-0.194	0.1111	0.0002
Mean	-0.068	-	-
PIE040			
BR-1	0.1507	0.0721	0.0014
FC-A	0.1445	0.144	0.0027
FC-B	0.1695	0.2019	0.0023
FC-C	-0.0357	0.4851	0.0014
FC-E	0.1651	0.3806	0.0022
HMR-IH	-0.0482	0.9485	0.0005
HMR-LP	0.0624	0.1132	0.0023
HMR-MI	0.0051	0.2325	0.0033
Mean	0.077	-	-
GOT004			
BR-1	**0.1469**	0.0015	0.0001
FC-A	**0.2101**	0.0023	0.0002
FC-B	**0.2206**	0.0023	0.0001
FC-C	**0.3066**	0	0
FC-E	**0.7111**	0	0
HMR-IH	-0.0275	0.658	0.0035
HMR-LP	**0.0839**	0.008	0.0006
HMR-MI	**0.1224**	0.0016	0.0002
Mean	0.222	-	-
FIR004			
BR-1	**0.2292**	0.0001	0.0001
FC-A	**0.4209**	0	0
FC-B	**0.3566**	0	0
FC-C	**0.271**	0	0
FC-E	**0.3293**	0	0
HMR-IH	**0.2559**	0	0
HMR-LP	**0.1614**	0.0081	0.0009
HMR-MI	**0.327**	0	0
Mean	0.294	-	-

S3 continued

Population	F	p-value	Standard Error
GOT009			
BR-1	**0.3289**	0	0
FC-A	-0.0243	0.6203	0.0021
FC-B	-0.0575	0.6608	0.0022
FC-C	0.1164	0.3085	0.0027
FC-E	-0.007	0.8919	0.0008
HMR-IH	-0.0745	0.9921	0.0003
HMR-LP	0.0657	0.2378	0.0024
HMR-MI	0.0452	0.5333	0.0032
Mean	0.049	-	-
PIE099			
BR-1	**0.2827**	0.0003	0.0001
FC-A	**0.4298**	0	0
FC-B	**0.4254**	0	0
FC-C	0.0204	0.3534	0.0026
FC-E	0.1308	0.1113	0.0012
HMR-IH	**0.3554**	0	0
HMR-LP	**0.3939**	0	0
HMR-MI	**0.2709**	0	0
Mean	0.289	-	-
FIR048			
BR-1	0.0689	0.3486	0.0051
FC-A	0.0161	0.5016	0.0062
FC-B	0.1186	0.3314	0.0063
FC-C	-0.0233	0.4786	0.0024
FC-E	0.0407	0.2735	0.0029
HMR-IH	0.058	0.1241	0.0034
HMR-LP	-0.0469	0.8577	0.0025
HMR-MI	0.0021	0.9577	0.0019
Mean	0.029	-	-
quru-GA-1F07			
BR-1	0.0233	0.6984	0.0051
FC-A	**0.2233**	0.0001	0.0001
FC-B	**0.1876**	0.0041	0.0007
FC-C	**0.2045**	0.0029	0.0005
FC-E	**0.2919**	0	0
HMR-IH	**0.2072**	0	0
HMR-LP	**0.1689**	0.001	0.0002
HMR-MI	0.0817	0.0187	0.0015
Mean	0.174	-	-
quru-GA-0E09			
BR-1	**0.3109**	0	0
FC-A	**0.2158**	0	0
FC-B	**0.2974**	0	0
FC-C	**0.2455**	0	0
FC-E	**0.1192**	0.006	0.0011
HMR-IH	**0.2612**	0	0
HMR-LP	**0.4429**	0	0
HMR-MI	**0.2673**	0	0
Mean	0.270	-	-

S3 continued

quru-GA-0C11			
Population	**F**	**p-value**	**Standard Error**
BR-1	0.0915	0.0375	0.0014
FC-A	0.0663	0.0544	0.0018
FC-B	0.0341	0.1325	0.0024
FC-C	-0.0723	0.276	0.0032
FC-E	**0.0872**	0.0055	0.0004
HMR-IH	0.0552	0.1594	0.0024
HMR-LP	-0.0194	0.687	0.0034
HMR-MI	0.1345	0.0231	0.0009
Mean	0.047	-	-
QpZAG15			
BR-1	-0.0096	0.4706	0.0038
FC-A	0.1315	0.3704	0.0028
FC-B	0.0577	0.5806	0.0035
FC-C	0.0308	0.0207	0.0013
FC-E	0.021	0.4339	0.0035
HMR-IH	-0.0082	0.7338	0.0038
HMR-LP	0.0971	0.6682	0.0035
HMR-MI	0.0336	0.4159	0.0044
Mean	0.044	-	-
3D15			
BR-1	**0.1791**	0.0051	0.0004
FC-A	0.1585	0.1284	0.0018
FC-B	0.1501	0.0858	0.002
FC-C	-0.1517	0.2314	0.0013
FC-E	0.1582	0.185	0.0027
HMR-IH	0.1356	0.0787	0.0021
HMR-LP	**0.2404**	0.0004	0.0001
HMR-MI	0.1492	0.0703	0.0022
Mean	0.127	-	-
1P10			
BR-1	-0.0271	0.8097	0.0045
FC-A	0.0305	0.7353	0.0022
FC-B	-0.0562	0.5458	0.0052
FC-C	0.0678	0.2056	0.003
FC-E	0.0662	0.2869	0.0044
HMR-IH	0.0944	0.23	0.0057
HMR-LP	0.0406	0.793	0.0038
HMR-MI	**0.2082**	0.0054	0.0006
Mean	0.053	-	-
2P24			
BR-1	0.0041	0.7169	0.0023
FC-A	**0.2679**	0.0007	0.0001
FC-B	**0.1686**	0.0012	0.0001
FC-C	**0.2547**	0.0006	0.0001
FC-E	0.2223	0.0139	0.0004
HMR-IH	0.1057	0.06	0.001
HMR-LP	**0.1395**	0.0036	0.0003
HMR-MI	0.024	0.502	0.0024
Mean	0.148	-	-

S3 continued

3A05			
Population	**F**	**p-value**	**Standard Error**
BR-1	0.0475	0.423	0.0027
FC-A	**0.3415**	0.0005	0.0001
FC-B	0.1963	0.1187	0.001
FC-C	**0.3208**	0.0058	0.0002
FC-E	**0.5787**	0	0
HMR-IH	0.1398	0.1168	0.002
HMR-LP	**0.2177**	0.0076	0.0003
HMR-MI	-0.0635	0.6884	0.002
Mean	0.222	-	-
Mean	0.044	-	-

Supplement 4 HWE Exact Tests (Weir and Cockerham, 1984) by population at each locus . Significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of α= 0.05

BR-1			
Locus	**F**	**p-value**	**Standard Error**
GOT021	-0.1136	1	0
PIE040	0.1507	0.0721	0.0014
GOT004	**0.1469**	0.0015	0.0001
FIR004	**0.2292**	0.0001	0.0001
GOT009	**0.3289**	0	0
PIE099	**0.2827**	0.0003	0.0001
FIR048	0.0689	0.3486	0.0051
quru-GA-1F07	0.0233	0.6984	0.0051
quru-GA-0E09	**0.3109**	0	0
quru-GA-0C11	0.0915	0.0375	0.0014
QpZAG15	-0.0096	0.4706	0.0038
3D15	0.1791	0.0051	0.0004
1P10	-0.0271	0.8097	0.0045
2P24	0.0041	0.7169	0.0023
3A05	0.0475	0.423	0.0027
Mean	0.1142	-	-
FC-A			
GOT021	-0.2462	0.1651	0.0006
PIE040	0.1445	0.144	0.0027
GOT004	**0.2101**	0.0023	0.0002
FIR004	**0.4209**	0	0
GOT009	-0.0243	0.6203	0.0021
PIE099	**0.4298**	0	0
FIR048	0.0161	0.5016	0.0062
quru-GA-1F07	**0.2233**	0.0001	0.0001
quru-GA-0E09	**0.2158**	0	0
quru-GA-0C11	0.0663	0.0544	0.0018
QpZAG15	0.1315	0.3704	0.0028
3D15	0.1585	0.1284	0.0018
1P10	0.0305	0.7353	0.0022
2P24	**0.2679**	0.0007	0.0001
3A05	**0.3415**	0.0005	0.0001
Mean	0.1591	-	-

S4 continued

Locus	F	p-value	Standard Error
FC-B			
GOT021	-0.0209	1	0
PIE040	0.1695	0.2019	0.0023
GOT004	**0.2206**	0.0023	0.0001
FIR004	**0.3566**	0	0
GOT009	-0.0575	0.6608	0.0022
PIE099	**0.4254**	0	0
FIR048	0.1186	0.3314	0.0063
quru-GA-1F07	**0.1876**	0.0041	0.0007
quru-GA-0E09	**0.2974**	0	0
quru-GA-0C11	0.0341	0.1325	0.0024
QpZAG15	0.0577	0.5806	0.0035
3D15	0.1501	0.0858	0.002
1P10	-0.0562	0.5458	0.0052
2P24	**0.1686**	0.0012	0.0001
3A05	0.1963	0.1187	0.001
Mean	0.1499	-	-
FC-C			
GOT021	-0.0046	1	0
PIE040	-0.0357	0.4851	0.0014
GOT004	**0.3066**	0	0
FIR004	**0.271**	0	0
GOT009	0.1164	0.3085	0.0027
PIE099	0.0204	0.3534	0.0026
FIR048	-0.0233	0.4786	0.0024
quru-GA-1F07	**0.2045**	0.0029	0.0005
quru-GA-0E09	**0.2455**	0	0
quru-GA-0C11	-0.0723	0.276	0.0032
QpZAG15	0.0308	0.0207	0.0013
3D15	-0.1517	0.2314	0.0013
1P10	0.0678	0.2056	0.003
2P24	**0.2547**	0.0006	0.0001
3A05	0.3208	0.0058	0.0002
Mean	0.1034	-	-
FC-E			
GOT021	0.2641	0.0257	0.0002
PIE040	0.1651	0.3806	0.0022
GOT004	**0.7111**	0	0
FIR004	**0.3293**	0	0
GOT009	-0.007	0.8919	0.0008
PIE099	0.1308	0.1113	0.0012
FIR048	0.0407	0.2735	0.0029
quru-GA-1F07	**0.2919**	0	0
quru-GA-0E09	0.1192	0.006	0.0011
quru-GA-0C11	0.0872	0.0055	0.0004
QpZAG15	0.021	0.4339	0.0035
3D15	0.1582	0.185	0.0027
1P10	0.0662	0.2869	0.0044
2P24	0.2223	0.0139	0.0004
3A05	**0.5787**	0	0
Mean	0.2119	-	-

S4 continued

Locus	F	p-value	Standard Error
HMR-IH			
GOT021	-0.1428	0.4149	0.0008
PIE040	-0.0482	0.9485	0.0005
GOT004	-0.0275	0.658	0.0035
FIR004	**0.2559**	0	0
GOT009	-0.0745	0.9921	0.0003
PIE099	**0.3554**	0	0
FIR048	0.058	0.1241	0.0034
quru-GA-1F07	**0.2072**	0	0
quru-GA-0E09	**0.2612**	0	0
quru-GA-0C11	0.0552	0.1594	0.0024
QpZAG15	-0.0082	0.7338	0.0038
3D15	0.1356	0.0787	0.0021
1P10	0.0944	0.23	0.0057
2P24	0.1057	0.06	0.001
3A05	0.1398	0.1168	0.002
Mean	0.0911	-	-
HMR-LP			
GOT021	-0.088	1	0
PIE040	0.0624	0.1132	0.0023
GOT004	0.0839	0.008	0.0006
FIR004	0.1614	0.0081	0.0009
GOT009	0.0657	0.2378	0.0024
PIE099	**0.3939**	0	0
FIR048	-0.0469	0.8577	0.0025
quru-GA-1F07	**0.1689**	0.001	0.0002
quru-GA-0E09	**0.4429**	0	0
quru-GA-0C11	-0.0194	0.687	0.0034
QpZAG15	0.0971	0.6682	0.0035
3D15	**0.2404**	0.0004	0.0001
1P10	0.0406	0.793	0.0038
2P24	**0.1395**	0.0036	0.0003
3A05	0.2177	0.0076	0.0003
Mean	0.1307	-	-
HMR-MI			
GOT021	-0.194	0.1111	0.0002
PIE040	0.0051	0.2325	0.0033
GOT004	**0.1224**	0.0016	0.0002
FIR004	**0.327**	0	0
GOT009	0.0452	0.5333	0.0032
PIE099	**0.2709**	0	0
FIR048	0.0021	0.9577	0.0019
quru-GA-1F07	0.0817	0.0187	0.0015
quru-GA-0E09	**0.2673**	0	0
quru-GA-0C11	0.1345	0.0231	0.0009
QpZAG15	0.0336	0.4159	0.0044
3D15	0.1492	0.0703	0.0022
1P10	0.2082	0.0054	0.0006
2P24	0.024	0.502	0.0024
3A05	-0.0635	0.6884	0.002
Mean	0.0942	-	-

Supplement 5 Pairwise F_{ST} values and significances calculated in GenePop (Raymond and Rousset 1995)

	BR-1	**FC-A**	**FC-B**	**FC-C**	**FC-E**	**HMR-IH**	**HMR-LP**	**HMR-MI**	**QE**	**QR**
BR-1	-	0.1	0.14	**0.006**	**0.0004**	0.04	0.10	0.04	0.02	0.12
FC-A	0.02	-	0.19	**0.0003**	**<0.0001**	0.07	0.17	0.04	0.01	0.07
FC-B	0.01	0.01	-	**0.007**	**0.0003**	0.20	0.07	0.07	**0.003**	0.10
FC-C	0.06	**0.06**	0.06	-	0.05	**0.002**	**0.007**	**0.002**	0.12	**0.005**
FC-E	**0.07**	**0.07**	**0.07**	0.02	-	**<0.0001**	**0.0005**	**0.0001**	**0.01**	**0.006**
HMR-IH	0.01	0.02	0.01	**0.06**	**0.07**	-	0.16	0.11	0.009	0.04
HMR-LP	0.02	0.02	0.02	0.05	**0.06**	0.009	-	0.02	0.01	0.10
HMR-MI	0.02	0.02	0.02	**0.06**	**0.07**	0.01	0.02	-	**0.0003**	**0.01**
QE	0.06	0.05	0.06	0.02	0.03	0.06	0.06	**0.06**	-	0.07
QR	0.02	0.02	0.02	0.05	0.06	0.02	0.02	0.02	0.05	-

Lower triagonal: F_{ST} values, upper triagonal: p-values. Significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of $\alpha = 0.05$

License agreement

This is a License Agreement between Jennifer F Lind-Riehl ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

License Number 3357810603310
License date Mar 28, 2014
Licensed content publisher Springer
Licensed content publication Tree Genetics & Genomes
Licensed content title: Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers
Licensed content author: Jennifer F. Lind
Licensed content date: Jan 1, 2013
Volume number: 9
Issue number: 3
Type of Use: Thesis/Dissertation
Number of copies: 1
Author of this Springer article: Yes and you are the sole author of the new work
Title of your dissertation: GENETIC VARIATION, LOCAL ADAPTATION AND POPULATION STRUCTURE IN NORTH AMERICAN RED OAK SPECIES, QUERCUS RUBRA L. AND Q. ELLIPSOIDALIS E. J. HILL

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com). Rightslink Printable License: https://s100.copyright.com/App/PrintableLicenseFrame.jsp?publisherID...1

Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry. Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process. This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other

electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com). The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper. Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well). Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:

Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms

Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut'(Netherlands Institute of Arbitration).

OR: All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions: v1.3

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501263840. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006
For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Evidence for selection on a CONSTANS-like gene between two red oak species[2]

Jennifer Lind-Riehl[†], Alexis Sullivan[†‡], Oliver Gailing[†*]

[†] Michigan Technological University, 1400 Townsend Drive, 49931 Houghton, MI, United States, [‡] Current address: Umeå University, SE-901 87 Umeå, Sweden

*For correspondence: ogailing@mtu.edu

[2] The material in this chapter was previously published material and has been reprinted with kind permission from Oxford University Press from Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a *CONSTANS*-like gene between two red oak species. Annals of Botany 113: 967-975

Abstract

- Background and Aims: Hybridizing species such as oaks may provide a model to study the role of selection in speciation with gene flow. Discrete species identities and different adaptations are maintained among closely related oak species despite recurrent gene flow. This is likely due to ecologically-mediated selection at a few key genes or genomic regions. Neutrality tests can be applied to identify so-called outlier loci, which demonstrate locus-specific signatures of divergent selection and are candidate genes for further study.
- Methods: We screened 36 genic microsatellite markers, some with putative functions in flowering time and drought tolerance, and eight non-genic microsatellite markers in two population pairs (n = 160) of the interfertile species *Quercus rubra* and *Q. ellipsoidalis*, which are characterized by contrasting adaptations to drought. Putative outliers were then tested in additional population pairs from two different geographic regions (n = 159) to further support their potential role in adaptive divergence.
- Key Results: A marker located in the coding sequence of a putative *CONSTANS*-like (*COL*) gene was repeatedly identified as under strong divergent selection across all three geographically disjunct population pairs. *COL* genes are involved in the photoperiodic control of growth and development and are implicated in the regulation of flowering time.
- Conclusions: The location of the polymorphism in the *Quercus COL* gene and given the potential role of *COL* genes in adaptive divergence and reproductive isolation makes this a promising candidate speciation gene. Further investigation of the phenological characteristics of both species and flowering time pathway genes is suggested to elucidate the importance of phenology genes for the maintenance of species integrity. Next generation sequencing in multiple population pairs in combination with high-density genetic linkage maps could reveal the genome wide distribution of outlier genes and their potential role in reproductive isolation between these species.

Introduction

Natural selection has long been recognized as an important driver of speciation, particularly in the adaptive radiation of allopatric populations (Schluter 2000). More recently, divergent selection has been suggested to play an important role in the development of intrinsic barriers to gene flow (Barluenga et al. 2006; Savolainen et al. 2006) in a process termed ecological speciation (Schluter 2009; Via 2012). In ecological speciation, adaptive divergence and reproductive isolation are both products of genetic adaptation to differing ecological conditions, potentially even in the presence of recurrent gene flow (Rundle and Nosil 2005; Schluter 2009). However, both the prevalence and the genetic mechanisms underlying this mode of speciation are still largely unknown. Linkage disequilibrium may facilitate ecological speciation if alleles involved in reproductive isolation are linked to genomic regions subject to strong divergent selection (Via 2012). Alternatively, a single "magic" trait could be associated with simultaneous adaptive divergence and non-random mating such as flowering phenology in plants (Servedio et al. 2011). By extension, magic genes may potentially underlie such traits involved in both ecological adaptation and non-random mating. In either scenario, the genomes of recently diverged “ecological species” are marked by broad regions exhibiting little or no differentiation punctuated by islands of pronounced genetic differentiation. Accordingly, genome screens can be used to identify regions resisting the homogenizing effect of gene flow, which may contain genes subject to divergent selection and/or may be involved in the development of reproductive isolation (Nosil et al. 2009; Via 2012).

A variety of marker types can be employed in F_{ST}-based tests to detect loci putatively affected by divergent selection. In these tests, F_{ST} values of individual loci are compared to a simulated null distribution derived from a neutral island model of migration (Beaumont and Nichols 1996). Loci significantly deviating from neutral expectations of differentiation, or outlier loci, may be closely linked to the target of

natural selection or even directly under selection themselves. Such model-based approaches can be used concurrently with other statistics to reduce the rate of false positives due to underlying population structure, varying mutation and/or recombination rates, and population admixture (Storz 2005). In particular, the LnRH statistic provides a model-independent test of selection and compares the reduction of locus-specific gene diversity between two populations to the approximately normal distribution of genome-wide heterozygosity (Schlötterer and Dieringer 2005).

Here, we apply complementary selection-detection methods to oaks (*Quercus*: Fagaceae) to identify genes involved in adaptive divergence and reproductive isolation. Hybridization is common in oaks, yet species identities are maintained despite non-zero levels of interspecific gene flow (Curtu et al. 2009; Lepais and Gerber 2011). Consistent with ecological speciation, genetic mapping and outlier analyses using 389 genetic markers in two sympatric European oak species with different adaptations to water availability (*Q. robur* and *Q. petraea*) revealed largely undifferentiated genomes marked by a few clusters of highly differentiated loci, a pattern likely the result of divergent selection with recurrent interspecific gene flow (Scotti-Saintagne et al. 2004). Furthermore, Goicoechea et al. (2012) studied the same oak species pair and their results indicated lower recombination rates for genomic regions containing outliers as compared to a control region.
We focus on the interfertile species *Q. rubra* and *Q. ellipsoidalis* (section *Lobatae*) which maintain varied adaptations to drought despite recurrent interspecific gene flow (Lind and Gailing 2013; Sullivan in prep). Both species occur in sympatry in the Great Lakes region of eastern North America but differ in their ecological niches. While *Q. ellipsoidalis* grows on sandy outwash plains and is considered the most drought tolerant red oak species (Abrams 1988; Burns and Honkala 1990), *Q. rubra* is more common on north facing and bottom slopes with fine soils containing more organic matter (Abrams 1990; Burns and Honkala 1990). Consistent with their occurrence on sites with differing moisture availability, these two species also differ in morphological and physiological characteristics related to drought response and

water-use efficiency (e.g., tissue elasticity, leaf conductance, xylem anatomy, and root depth; Abrams 1990). Despite the ecological and morphological differences between *Q. rubra* and *Q. ellipsoidalis*, they generally show very low genetic differentiation at most genetic markers, and the presence of morphologically and genetically intermediate individuals is consistent with recurrent interspecific gene flow (Hokanson et al. 1993; Lind and Gailing 2013). In addition, genetic assignment analyses using nuclear (nSSR) and genic microsatellite (EST-SSR) markers have identified putative first generation hybrids and introgressive forms in adult trees and seedlings. However, adult hybrids were relatively infrequent, suggesting the absence of hybrid swarms and maintenance of species identity by some sort of pre- and/or post-zygotic isolation mechanism (Lind and Gailing 2013). Differences in flowering time and different adaptations to drought might contribute to the effective reproductive isolation between both species. For example, *Q. ellipsoidalis* seedlings showed a significantly later, albeit overlapping, bud burst and higher mortality than *Q. rubra* seedlings from neighboring populations in a common garden experiment (Gailing 2013).

Our main objective was to identify genes involved in the adaptive species divergence and reproductive isolation between *Q. rubra* and *Q. ellipsoidalis.* For this purpose, we selected genetically mapped genic markers, some of which had annotated functions in drought tolerance and flowering time, to identify loci under divergent selection between these two species. Outlier tests were replicated in three distinct *Q. rubra*/*Q. ellipsoidalis* population pairs. Replicated population pairs can provide greater power to outlier screens because divergence in multiple pairs of populations is less likely to be due to nonselective factors such as false positives or genetic drift (Nosil et al. 2009). Specifically, in addition to the eight nSSRs developed for *Q. rubra*, we adapted 36 gene-linked EST-SSRs to *Q. rubra* and *Q. ellipsoidalis* that were originally developed and mapped in *Q. robur* (Durand et al. 2010). These EST-SSRs provide a focused way to detect selection because they represent gene-associated polymorphic regions (Sullivan et al. 2013). We expect to see loci involved

in adaptive species differences to be under selection in population pairs across different geographic regions. Such replicated outlier loci are candidate genes potentially involved in the evolution and maintenance of species identity between these ecologically divergent oak species.

Materials and methods

Plant material

Two neighboring population pairs of *Q. rubra* and *Q. ellipsoidalis* within the Ford Center and Baraga Plains area (Baraga County, Michigan), one population pair in the Chequamegon National Forest (Bayfield County, Wisconsin) and one population pair in the Nicolet National Forest (Oconto County, Wisconsin) were sampled resulting in a total number of 319 adult trees (Fig. 6 and Table 7). Only adult trees occupying a dominant or codominant canopy were sampled. A minimum distance of approximately 30 m was kept between sampled trees to minimize family structure. Species identity in the Baraga Plains was confirmed both through morphological (Gailing et al. 2012) and genetic assignment analyses (Lind and Gailing 2013). Species identity in the Chequamegon-Nicolet National Forest was determined through whole tree silvic characters and confirmed by genetic species assignment (Suppl. 6). Total genomic DNA (~20 ng) was isolated from leaves using the DNeasy96 Plant Kit (Qiagen, Valencia, CA) following the manufacturer's instructions.

Marker selection and microsatellite genotyping

For all EST-SSRs, the repeat-containing *Quercus* unigene elements (Durand et al. 2010) were reassembled in CAP3 (Huang and Madan 1999) and functionally annotated using the BLASTx algorithm as implemented in the Blast2GO software package (Altschul et al. 1997) by comparing the reassembled EST contigs to homologous sequences in the non-redundant NCBI database. Results are reported in Supplement 7 for sequences that had an expect value of $\leq 10^{-4}$ and a sequence

similarity of >50 %. Fourteen EST-SSRs originally described for *Q. robur* (Durand et al. 2010) that had putative functions in drought tolerance, flowering time and other functions were chosen and adapted for use in *Q. rubra*. Previously, 30 markers including seven putatively neutral nSSRs were developed for *Q. rubra* (Aldrich et al. 2002; Sullivan et al. 2013), as well as one one nSSR (Steinkellner et al. 1997) and 22 EST-SSRs adapted for use in *Q. rubra* (Lind and Gailing 2013; Sullivan et al. 2013).

All 44 microsatellite markers (8 nSSRs and 36 EST-SSRs), covering 10 of the 12 linkage groups in *Q. robur* (Durand et al. 2010), were amplified in the Ford Center and Baraga Plains populations (FC). Identified outliers in the FC population pairs, along with 8 EST-SSRs with pairwise F_{ST} values close to the mean and the 8 nSSRs were amplified in the Chequamegon National Forest (CNF) and Nicolet National Forest (NNF) populations (Suppl. 7).

PCR amplification and electrophoretic separation were performed according to Lind and Gailing (2013) with one modification: the 10 µL PCR reaction mix was scaled to a 15 µL reaction mix. Marker VIT081 amplified two loci: VIT081L1 had a range of 108-112 bp and VIT081L2 had a range of 115-136 bp. Both loci, VIT081L1 and VIT081L2, were in Hardy-Weinberg equilibrium in all populations.

Genetic structure

MICROCHECKER was used to assess all markers in each population for null alleles ($\alpha = 0.05$) since they can lead to overestimation of measures of differentiation such as F_{ST} (Van Oosterhout et al. 2004). MICROCHECKER identified 13 loci as being in Hardy-Weinberg (HW) disequilibrium in at least three of the FC populations. Only two loci, FIR013 and GOT004, were identified as being in HW disequilibrium in at least three of the CNF and NNF populations, respectively. Deviations from HW equilibrium may be due to the presence of null alleles, but the high number of markers in HW disequilibrium in one *Q. rubra* and both *Q. ellipsoidalis* populations in the FC indicates potential inbreeding (Suppl. 8 and 9). All outlier screens were

conducted with and without corrections using the van Oosterhout correction algorithm, for potential null alleles resulting in only minor differences in the outliers detected (Suppl. 10-12).

STRUCTURE 2.3 (Pritchard et al. 2000) was used to determine the population structure under the admixture model with correlated allele frequencies without a priori information regarding species identity. Five independent runs of 10^6 iterations with a burn-in period of 30000 were performed for each K value (K= 1-4 for FC populations with 44 markers and K= 1-8 for FC, CNF and NNF populations with 20 markers). We did both analyses to see if the lower number of markers was sufficient to delineate the species. The most likely number of groups (K) was chosen by using the ad hoc statistic presented in Evanno et al. (2005).

Genetic variation parameters were calculated in GeneAlEx 6.41 (Peakall and Smouse 2006) including the average number of alleles (N_a), Nei's unbiased gene diversity (H_e), and Wright's inbreeding coefficient (F). F_{ST} and pairwise F_{ST} with corresponding significances were calculated in GenePop 4.1 (Raymond and Rousset 1995). To correct for multiple comparisons, sequential Bonferroni adjustments of the significance level (α) were made. An F_{ST} based UPGMA dendrogram was created in Populations 2.0 (Langella 1999).

Outlier screens

F_{ST} (Wright's fixation index) can be used to identify outlier loci between populations based on the assumption that gene loci under selection will show higher differentiation between species than expected under neutral evolutionary conditions (Beaumont and Nichols 1996). Using coalesecent simulations, the program LOSITAN (Antao et al. 2008) creates a null distribution of F_{ST} values from the multilocus data that can be used to identify outlier loci under the finite island model. The run parameters in LOSITAN included 100000 simulations at a 95 % confidence interval under the stepwise mutation model. In order to account for multiple testing and resultant potential false positives, the default false discount rate (FDR) of 10 %

was used. Both the "neutral" mean F_{ST} option to exclude loci potentially under selection for the computation of the initial mean F_{ST} and the "force" mean F_{ST} option to increase the precision of the simulated mean F_{ST} were applied.

The LnRH test statistic estimates variability between populations at individual loci instead of population divergence to identify selection and is a powerful test statistic for the detection of selective sweeps. LnRH is generally normally distributed as determined by simulation studies, and when plotted against F_{ST} values outliers can be found in the tails of the distribution (Schlötterer and Dieringer 2005). LnRH measures relative gene diversity between two populations to identify putative loci under selection through the calculation of the natural logarithm of the ratio of gene diversity (H) for each locus between a pair of populations as follows:

$$Ln\ RH = \ ln\left[\frac{\left({}^{1}/_{1 - H_{pop1}}\right)^{2} - 1}{\left({}^{1}/_{1 - H_{pop2}}\right)^{2} - 1}\right]$$

The Kolmogorov-Smirnov test for normal distribution was performed on the observed data for each comparison. Significance was determined by standardizing the observed values and applying a 95 % confidence interval, where outliers were identified as being more than ±1.96 standard deviations away from the mean.

Outlier tests were performed for all between and within species pairwise comparisons as replicated population pairs provide increased power and reduce the chance of observed divergence being due to nonselective factors (Nosil et al. 2009). They were also performed for all populations within a region pooled by species.

Isolation by distance vs. isolation by adaptation

To further confirm potential loci under divergent selection, identified outliers were amplified in two additional interspecific population pairs in two different regions, CNF and NNF. Each identified outlier was paired with another marker with an F_{ST}

value close to the mean identified in the FC populations and amplified in the same populations to prevent upward biasing of F_{ST}. The outlier methods described above were used to screen this marker set in these additional populations. Outliers that are due to isolation by adaptation between species (IBA) should be identified as outliers in both allopatric and sympatric geographic scenarios. In contrast, loci identified as potentially under divergent selection due to isolation by distance (IBD) should only be identified as distance increases between populations, so that geographically proximal populations do not show divergence but distant populations do (Nosil et al. 2009). Outliers identified in only one geographic region could be involved in local adaptation of populations or, more conservatively, may comprise false positives.

Results

Genetic structure and variation

Two clusters were shown to best fit the data in STRUCTURE using the ad hoc statistic ΔK (Evanno et al. 2005), which correspond to the two species, *Q. rubra* and *Q. ellipsoidalis* (Suppl. 6). The *Q. ellipsoidalis* populations in the CNF and NNF showed higher levels of potential introgression as compared to the FC populations. However, most individuals could be assigned a species designation with high certainty (posterior probability >0.90). The species assignment for the FC populations displayed similar numbers of hybrids and introgressive forms for both the runs using 44 (data not shown) and 20 markers (Suppl. 6). This indicates that the higher number of hybrids and introgressive forms in the CNF and NNF populations are not due to the lower number of markers used for species assignment. Consequently, outlier screens were conducted with all samples and after excluding potential hybrids and introgressive forms. Results were consistent and differed mainly in the number and identity of loci potentially subject to balancing selection which were not considered (Suppl. 13-15). An F_{ST} based UPGMA unrooted dendrogram showed populations separated by species (Fig. 7). All nodes had high

bootstrap values whether all individuals or only "pure" species (data not shown) were included in the analysis.

At the subset of 20 markers common to all eight populations, all populations exhibited similar levels of genetic variation (Suppl. 16). The mean heterozygosity (H_e) was 0.715 and 0.711 for *Q. rubra* and *Q. ellipsoidalis* respectively, while mean number of alleles per locus was 15 for both species. Additionally, 19 out of 20 markers showed statistically significant differentiation (F_{ST}) between species with values ranging from 0.007 to 0.668 (Suppl. 17). Mean genetic diversity was slightly lower at the 44 marker set only used in the four FC populations with means of 0.660 and 0.642 for *Q. rubra* and *Q. ellipsoidalis,* respectively. Average number of alleles at the 44 markers was also slightly lower than at the 20 markers, averaging between 9 and 10 for both species in the FC populations. This is likely due to the higher number of EST-SSRs used, which show lower diversity than nSSRs. Thirty-five out of 44 markers showed significant differentiation (F_{ST}) between species with a range from 0.0046 to 0.801 (Suppl. 18).

Outlier screens

LOSITAN consistently identified FIR013 as an outlier between species in almost all possible comparisons (Fig. 8, Suppl. 19 and 20). This locus showed two major alleles and was nearly fixed on alternative alleles in all populations of both species (*Q. ellipsoidalis*, 138 bp; *Q. rubra* 141 bp; Fig. 9). Consistent with this pattern of allelic variation, FIR013 was identified as an outlier between species within and among regions, which strongly implicates isolation by adaptation (Table 8). Locus FIR013 is putatively located in a *CONSTANS*-like (*COL*) gene. Together with *CONSTANS* (*CO*), *COL* genes comprise a family encoding transcription factors involved in the photoperiodic regulation of growth and development. Specifically, this family is broken up into three broad groups based on structural differences. In *Arabidopsis*, *CO* and homologous *COL* (1-5) genes belong to Group I and are characterized by the presence of two conserved zinc finger B-boxes near the amino-terminus and a CCT

domain near the carboxy terminus (Griffiths et al. 2003). *COL* genes can be distinguished from *CO* by the presence of poly-Q regions following the B-box zinc finger domain (Almada et al. 2009).

Alignment of the *Quercus* EST unigene element to *Populus COL-1* and *COL-2* genes shows 81% and 84% similarity, respectively, and indicates that the FIR013 locus encodes a poly-Q repeat and is located between a conserved double B-box zinc finger domain and a CCT motif (Yuceer et al. 2002). Additionally, a gene tree created using protein sequences from closely related species and the *Quercus* unigene indicated strong similarity to the *COL* genes in *Populus* spp. with high bootstrap support (Suppl. 21). Together, the presence of characteristic protein domains and the high similarly to putatively homologous genes supports the identity of this *Quercus* locus as a Group I *COL* gene.

Even though FIR013 shows high divergence between species, the LnRH statistic only detected it between species in the NNF population (Table 8, Suppl. 22). The LnRH statistic describes significant reductions of heterozygosity, but both *Q. rubra* and *Q. ellipsoidalis* exhibit very low heterozygosity at this locus overall, with different alleles being nearly fixed in each species (*Q. rubra*: $H_e = 0.08$; *Q. ellipsoidalis*: $H_e = 0.25$). Additional outliers were identified by either LOSITAN or LnRH, but did not show up consistently across all comparisons, which could indicate a role in local adaptations of specific populations (Table 8, Suppl. 23 and 24). However, given the potentially high rate of false positives associated with outlier methods, we used caution in interpreting the results. Thus, we considered the FIR013 locus which was identified as an outlier by the F_{ST}-based method in multiple replicate species comparisons to be the best candidate under divergent selection.

Discussion

By employing replicated population pairs, we identified a marker located within the coding sequence of a *CONSTANS*-like gene (*COL*) as under divergent selection in

almost all interspecific comparisons within and among regions, but not for within species comparisons. Such a pattern of spatially-replicated divergent selection in contrasting environments is consistent with a role of this *COL* gene in ecological speciation, whereby natural selection drives species divergence and reproductive isolation (Nosil et al. 2009; Schluter 2009). Moreover, *COL* genes have been implicated in the regulation of both flowering time and growth, which presents an avenue for the development of ecologically-mediated reproductive barriers. Putative outlier loci are only rarely confirmed in multiple population pairs (e.g., Nielsen et al. 2009) and have never before been confirmed in forest tree species. Interestingly, SNP variation in the same *COL* gene was significantly associated with the timing of vegetative bud burst in a *Q. petraea* provenance trial (Alberto et al. 2013). Also other studies have shown associations between nucleotide variation in *COL* genes and phenotypic traits like flowering time and height in other plant species (e.g., *Medicago sativa*, Herrmann et al. 2010; *Populus nigra*, Fabbrini et al. 2012; *Populus tremula*, Ma et al. 2010). For example, in the association mapping study of *Medicago sativa*, a *CONSTANS*-like gene was shown to be involved in flowering time and plant height (Herrmann et al. 2010).

Together with *CONSTANS*, *COL* genes form a family of transcription factors involved in the photoperiod pathway of floral transition (Amasino 2005). Group I *COL* genes in *Arabidopsis* are regulated by a circadian clock and, in turn, may help regulate the pace of the circadian oscillator in *Arabidopsis* as it has been shown to accelerate the circadian clock (Griffiths et al. 2003; Ledger et al. 2001). However, altered expression in transgenic plants of either *COL-1* or *COL-2* had little effect on flowering time, suggesting that the functions of *CO*, *COL-1*, and *COL-2* may not completely overlap and may have diverged in function in *Arabidopsis* (Ledger et al. 2001).

Functional diversification of *COL* genes is also evident in angiosperm trees. In *Populus trichocarpa*, *PtCOL-2* was shown through RNAi experiments to play a

central role in the *CO*/*Flowering Time* (*FT*) regulon that controls seasonal growth patterns in trees (Böhlenius et al. 2006). While overexpression of *PtCOL-1* and *PtCOL-2* did not alter normal reproductive onset, bud break or bud set in *Populus*, overexpression of *PtCOL-1* in *Arabidopsis* rescued the late flowering phenotype of the *co-1* mutant, suggesting that it may function similarly to *CO* (Hsu et al. 2012). In addition, overexpression of *PtCOL-1* and *PtCOL-2* in *Populus* affected plant height.

Both *Q. rubra* and *Q. ellipsoidalis* are nearly fixed on different alleles at the *COL* microsatellite marker (Fig. 9), which corresponds to the addition or deletion of a glutamine residue as the microsatellite encodes a poly-Q repeat. While the role of poly-Q repeats in human genetic disorders is well studied, their role in normal protein function is largely unknown. However, it has been posited that they may be trans-activation sequences involved in stabilization of protein interactions or in regulation of gene transcription activation (Yuceer et al. 2002). More interestingly, it has been proposed that selection drives the evolution of low-complexity sequences like poly-Q repeats. This has been demonstrated in the Chinook salmon circadian-regulating *CLOCK* gene, where repeat lengths are positively correlated with latitude and thus potentially involved in adaptation of these fish to different latitudes (Haerty and Golding 2010). Notably, one allele of a poly(Q) repeat in the *Populus tremula COL2B* gene was found to be associated with growth cessation, albeit this effect was not independent of other polymorphisms in the photoperiodic pathway (Ma et al. 2010). While the exact function of *COL* genes and the poly-Q length polymorphism in *Quercus* are currently unknown, they provide excellent candidates for underlying functional polymorphisms in flowering time and growth-related traits and warrant further investigation.

Differences in flowering time are a clear mechanism of pre-zygotic isolation. Selection on the photoperiod pathway might have been essential in divergence and maintenance of species differences. While oak flowers are difficult to observe in field studies, the timing of vegetative bud burst is strongly associated with flowering

time and is used to infer phenological patterns in flowering time (Chesnoiu et al. 2009). Latitudinal and altitudinal gradients in sessile oak (Ducousso et al. 1996) and local environmental conditions have been shown to impact timing of bud burst in two interfertile live oak species (Cavender-Bares and Pahlich 2009). Our previous study indicated low levels of introgression between adult and seedling populations of the two species, which may be a result of pre-zygotic isolation via flowering time (Lind and Gailing 2013). Furthermore, significant differences in bud burst were observed in a common garden of *Q. ellipsoidalis* and *Q. rubra* seedlings. *Q. ellipsoidalis* seedling exhibited significantly later bud burst and leaf fall (a proxy for bud set) than *Q. rubra* seedlings, suggesting an underlying genetic mechanism (Gailing 2013). Also *Q. ellipsoidalis* seedlings were significantly smaller and showed a lower survival rate than *Q. rubra* seedlings from neighboring populations (Gailing 2013). Interspecific differences in growth and flowering time could possibly be linked to the divergence between the two species at the putative *COL* gene since similar genes in *Populus* and other plant species have been shown to be involved in growth and development (Herrmann et al. 2010; Hsu et al. 2012). In natural populations, *Q. ellipsoidalis* trees are smaller in habit than *Q. rubra* trees as well (personal field observation). Given the xeric nature of the environment *Q. ellipsoidalis* resides in, a slower growth strategy, including later flowering, may confer a selective advantage by avoiding late frost damage and conserving energy. In fact, frost pockets and slow forest canopy establishment are characteristic in level xeric areas like the outwash plains *Q. ellipsoidalis* grows on (Motzkin et al. 2002).

Further assessment of flowering time in the natural populations of these two species will be of value in assessing the importance of this trait in maintenance of species identity in interfertile oak species. The genetic assignment analysis suggested recurrent gene flow between species and higher rates of introgression in the CNF and NNF populations than in the FC populations. Gene flow estimates via parentage analysis could confirm recurrent gene flow between the species and its dependence on environmental conditions providing additional support for adaptive species

divergence with gene flow. Alternatively, post-zygotic selection on *COL* could be inferred from reciprocal transplant experiments between parental environments of *Q. rubra*, *Q. ellipsoidalis* and hybrid seedlings with different genotypes for *COL* (138/138 bp, 138/141 bp, 141/141 bp). Since adult populations are nearly fixed on the alternative alleles, 138 bp in *Q. ellipsoidalis* and 141 bp in *Q. rubra*, an increase of species-specific alleles from seedling to adult generation and higher growth and survival of *Q. ellipsoidalis* and *Q. rubra* seedlings with species-specific alleles in parental environments would implicate post-zygotic selection on the *Quercus COL* gene. Finally, expression studies of this gene through qPCR could be done to assess differential expression patterns for both species throughout the photoperiod.

Conclusions

We employed outlier screens with three replicated population pairs and identified a putative *CONSTANS*-like (*COL*) gene potentially involved in adaptive divergence and reproductive isolation of *Q. rubra* and *Q. ellipsoidalis*. We show that oaks provide a good model to identify potential genomic regions involved in ecological speciation. In particular, the gene associated with the EST-SSR FIR013, *COL* could be a "magic gene" involved in ecological speciation because of its potential simultaneous involvement in reproductive isolation and local adaptation (Servedio et al. 2011). Variation in putative "speciation genes" can be associated with traits related to adaptive species differences (e.g., in drought tolerance and bud burst) in QTL and association mapping approaches. Alternatively *COL* may not be subject to divergent selection *per se* but could be linked to an ecologically significant gene (Via 2012). The availability of high density linkage maps or whole genome sequence would allow for the identification of clusters of linked genes, which could help elucidate the relative importance of linkage disequilibrium and single-locus effects. Full-sib families in *Q. rubra* and high-density genetic linkage maps are currently being constructed. Next generation sequencing technologies such as restriction-site associated DNA (RAD) sequencing and integration of these and other markers in

genetic linkage maps will also help to identify genome-wide patterns of adaptive species divergence.

Acknowledgements

We are grateful to Linda Parker, Daniel Hinson, and John Lampereur for their assistance in locating appropriate populations in the Chequamegon-Nicolet National Forest. We would also like to thank Jonathan Riehl for his invaluable help with constructing many of the figures, and two anonymous reviewers for suggestions to further improve this paper. This work was supported by the Research Excellent fund of Michigan Technological University, the Hanes Trust, the United States Department of Agriculture's McIntire Stennis fund, the National Science Foundation's Plant Genome Research program (NSF 1025974), the Huron Mountain Wildlife Foundation, the Northern Institute of Applied Climate Science, the United States Department of Education and the European Commission's Directorate General for Education and Culture States' Atlantis Program and the Michigan Technological University Ecosystem Science and Biotechnology Research Centers.

References

Abrams MD. 1988. Comparative water relations of three successional hardwood species in central Wisconsin. *Tree Physiology* **4**:263-273.

Abrams MD. 1990. Adaptations and responses to drought in *Quercus* species of North America. *Tree Physiology* **7**: 227-238.

Alberto FJ, Derory J, Boury C, Frigerio J-M, Zimmermann NE, Kremer A. 2013. Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* **195**:495-512.

Aldrich PR, Michler CH, Sun WL, Romero-Severson J. 2002. Microsatellite markers for northern red oak (Fagaceae : *Quercus rubra*). *Molecular Ecology Notes* **2**: 472-474.

Almada R, Cabrera N, Casaretto J, Ruiz-Lara S, González Villanueva E. 2009. *VvCO* and *VvCOL1*, two *CONSTANS* homologous genes, are regulated during flower induction and dormancy in grapevine buds. *Plant Cell Reporter* **28**: 1193-1203.

Altschul SF, Madden TL, Schäffer AA, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Amasino RM. 2005. Vernalization and flowering time. *Current Opinions in Biotechnology* **16**:154-158.

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G. 2008. LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**: 323.

Beaumont M, Nichols R. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B Biological Sciences* **263**: 1619–1626.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439**: 719-723.

Böhlenius H, Huang T, Charbonnel-Campaa L, et al. 2006. *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**: 1040-1043.

Burns RM, Honkala BH. 1990. *Silvics of North America Volume 2: Hardwoods*. Washington D.C.: United States Forest Service.

Cavender-Bares J, Pahlich A. 2009. Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). *American Journal of Botany* **96**: 1690-1702.

Chesnoiu E, Sofletea N, Curtu AL, Toader A, Radu R, Enescu M. 2009. Bud burst and flowering phenology in a mixed oak forest from Eastern Romania. *Annals of Forest Research* **52**: 199-206.

Curtu A, Gailing O, Finkeldey R. 2009. Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. *BMC Evolutionary Biology* **9**: 284.

Ducousso A, Guyon J, Kremer A. 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). *Annals of Forest Science* **53**: 775-782.

Durand J, Bodenes C, Chancerel E, et al. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* **11**.

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

Fabbrini F, Gaudet M, Bastien C, et al. 2012. Phenotypic plasticity, QTL mapping and genomic characterization of bud set in black poplar. *BMC Plant Biology* **12**: 47.

Gailing O. 2013. Difference in growth, survival and phenology in *Q. rubra* and *Q. ellipsoidalis* seedlings. *Dendrobiology* **70**: 71-79.

Gailing O, Lind J, Lilleskov E. 2012. Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* E.J. Hill populations in contrasting environments. *Plant Systematics & Evolution* **298**: 1533-1545.

Goicoechea, PG, Petit RJ, Kremer A. 2012. Detecting the footprints of divergent selection in oaks with linked markers. Heredity **109**: 361-371.

Griffiths S, Dunford RP, Coupland G, Laurie DA. 2003. The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis. Plant Physiology* **131**: 1855-1867.

Haerty W, Golding GB. 2010. Low-complexity sequences and single amino acid repeats: not just "junk" peptide sequences. *Genome* **53**: 753-762.

Herrmann D, Barre P, Santoni S, Julier B. 2010. Association of a *CONSTANS*-like gene to flowering and height in autotetraploid alfalfa. *Theoretical & Applied Genetics* **121**: 865-876.

Hokanson SC, Isebrands JG, Jensen RJ, Hancock JF. 1993. Isozyme variation in oaks of the Apostle Islands in Wisconsin: genetic structure and levels of inbreeding in *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). *Amercian Journal of Botany* **80**: 1349-1357.

Hsu C-Y, Adams JP, No K, 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.

Huang X and Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Research* **9**: 868-877.

Langella O. 1999. Populations 1.2.30: a population genetic software CNRS UPR9034.

Ledger S, Strayer C, Ashton F, Kay SA, Putterill J. 2001. Analysis of the function of two circadian-regulated *CONSTANS*-like genes. *The Plant Journal* **26**: 15-22.

Lepais O, Gerber S. 2011. Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. *Evolution* **65**: 156-170.

Lind J, Gailing O. 2013. Genetic structure of *Quercus rubra* L and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetic & Genomes* **9**: 707-722.

Ma X-F, Hall D, St. Onge KR, Jansson S, Invgarsson PK. 2010. Genetic differentiation, clinal variation, and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* **186**: 1033- 1044.

Motzkin G, Ciccarello SC, Foster DR. 2002. Frost pockets on a level sand plain: does variation in microclimate help maintain persistent vegetation patterns? *Journal of the Torrey Botanical Society* **129**: 154-163.

Nielsen E, Hemmer-Hansen J, Poulsen N, et al. 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**: 276.

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**: 375-402.

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

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

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

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecology Letters* **8**: 336-352.

Savolainen V, Anstett M-C, Lexer C, et al. 2006. Sympatric speciation in palms on an oceanic island. *Nature* **441**: 210-213.

Schlötterer C, Dieringer D. 2005. *A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity*. In: Nurminsky D (ed) *Selective Sweep*. Landes Bioscience, Georgetown, TX.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* **323**: 737-741.

Schluter D. 2000. *The ecology of adaptive radiation*. Oxford: Oxford University Press.

Scotti-Saintagne C, Mariette S, Porth I, et al. 2004. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**: 1615-1626.

Servedio MR, Doorn GSV, Kopp M, Frame AM, Nosil P. 2011. Magic traits in speciation: magic but not rare? *Trends in Ecology & Evolution* **26**: 389-397.

Steinkellner H, Lexer C, Turetschek E, Glössl J. 1997. Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. *Molecular Ecology* **6**: 1189-1194.

Storz J (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology***14**: 671-688.

R A, Lind JF, McCleary TS, Romero-Severson J, Gailing O. 2013. Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. *Plant Molecular Biology Reporter* **31**: 231-239.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535-538.

Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 451-460.

Yuceer C, Harkess RL, Land Jr SB, Luthe DS. 2002. Structure and developmental regulation of *CONSTANS*-LIKE genes isolated from *Populus deltoides*. *Plant Science* **163**: 615-625.

Table 7 Sampled *Quercus rubra* and *Q. ellipsoidalis* populations

Abbreviation	Region	Population	Species	Sample size (n)	Soil characteristics[a]	Latitude	Longitude
FC-A	FC/Baraga Plains: MI	Stand A	*Q. rubra*	40	2 (78C: Keweenaw-Kalkaska Complex; 1-12% slopes)	46° 39' 9" N	88° 30' 6." W
FC-B	FC/Baraga Plains: MI	Stand B	*Q. rubra*	40	3 (25B: Munising-Yalmer loamy sand; 1-6% slopes)	46° 40' 27 " N	88° 31' 27 " W
FC-C	FC/Baraga Plains: MI	Stand C	*Q. ellipsoidalis*	40	1 (10B: Grayling sand; 0-6% slopes)	46° 39' 14" N	88° 35' 25" W
FC-E	FC/Baraga Plains: MI	Stand E	*Q. ellipsoidalis*	40	1 (10B: Grayling sand; 0-6% slopes)	46° 39' 55" N	88° 33' 19" W
N-QR	Nicolet National Forest: WI	Nicolet QR	*Q. rubra*	40	2 (KaC: Kennan fine sandy loam; 6-15% slopes)	45° 20' 53" N	88° 23' 17" W
N-QE	Nicolet National Forest: WI	Nicolet QE	*Q. ellipsoidalis*	39	2 (RsB: Rousseau fine sand; 1-6% slopes)	45° 19' 19" N	88° 19' 53" W
C-QR	Chequamegon National Forest: WI	Chequamegon QR	*Q. rubra*	40	3 (480B: Portwing-Herbster complex, 0 to 6 percent slopes	46° 42' 54" N	91° 02' 8" W
C-QE	Chequamegon National Forest: WI	Chequamegon QE	*Q. ellipsoidalis*	40	1 (74B/C: Vilas loamy sand; 0-15% slope)	46° 44' 43" N	91° 04' 20" W

[a] Rating developed from drainage classes with 1 being excessively drained, 2 being well drained, and 3 being moderately well drained; soil type and drainage class identified according to the Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Official Soil Series Descriptions available online at http://websoilsurvey.nrcs.usda.gov/app/HomePage.htm Accessed [11/29/2012]

Table 8 Summary of between species identified outliers under divergent selection

Pairwise Comparison	LOSITAN	LnRH
FC (QR vs. QE)	FIR013 POR016	FIR039 GOT040
Chequamegon National Forest (QR vs. QE)	FIR013	GOT021
Nicolet National Forest QR vs. QE)	FIR013	FIR013
ALL (QR vs. QE)	FIR013	NO

NO = no outliers detected

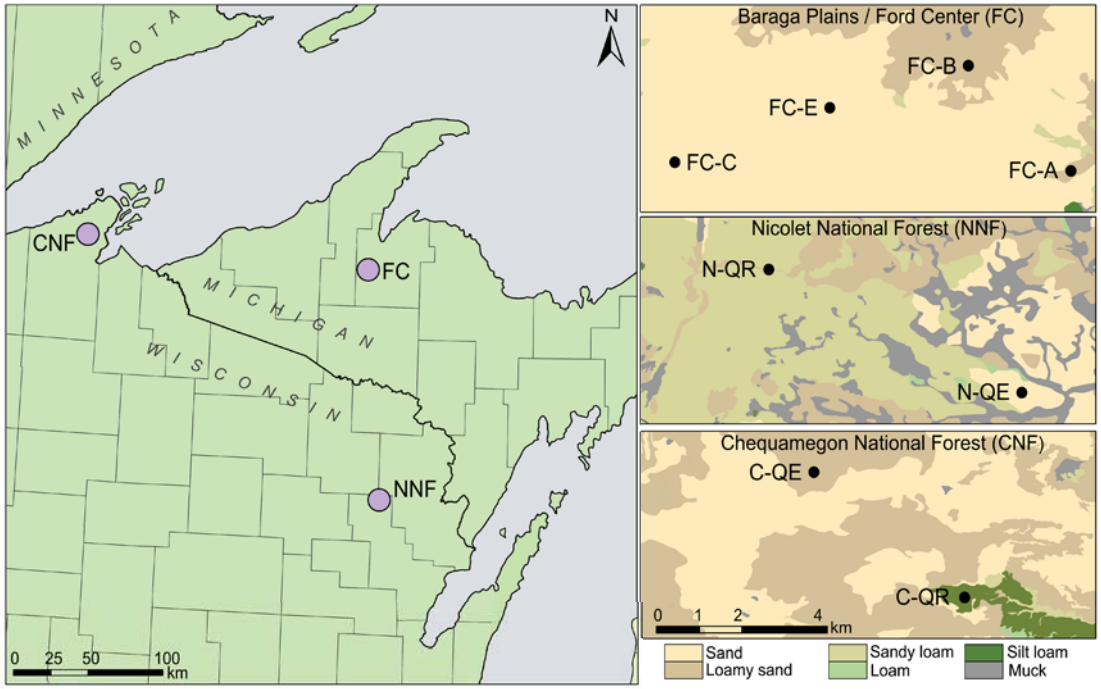


Figure 6 Sampling Locations. Soil data are from Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Available online at **http://soildatamart.nrcs.usda.gov**. Accessed [06/01/2013]. In the case of FC-A, modified slightly to match field soil samples (data not shown)

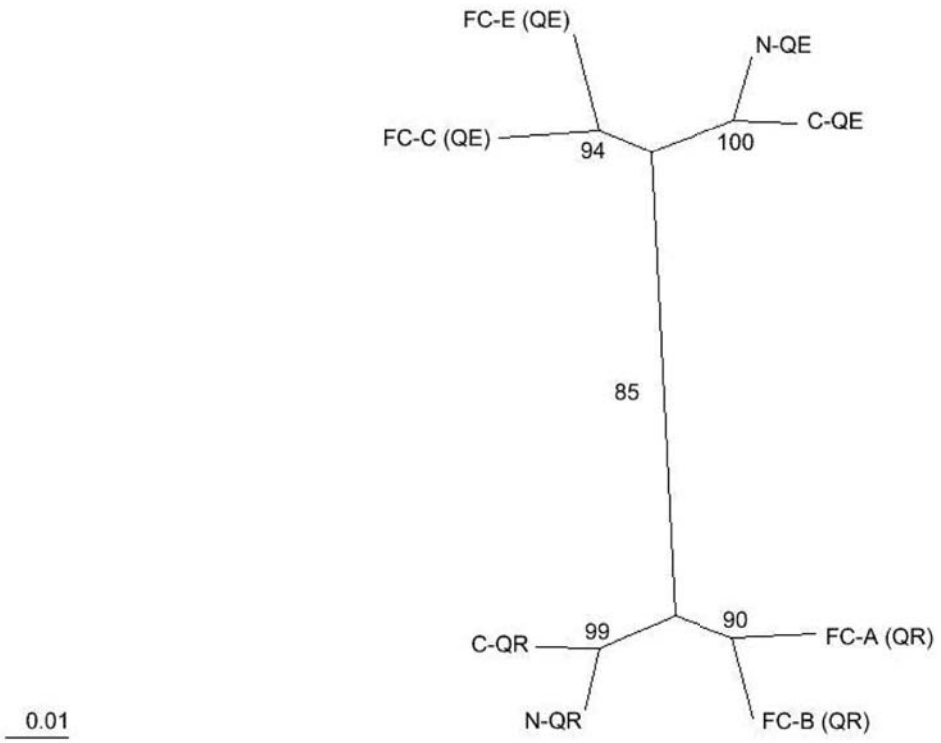


Figure 7 Unrooted tree (UPGMA method) based on genetic distance (F_{ST}) at 20 markers; numbers at nodes are percentages over 1000 bootstrap replicates using Populations 2.0

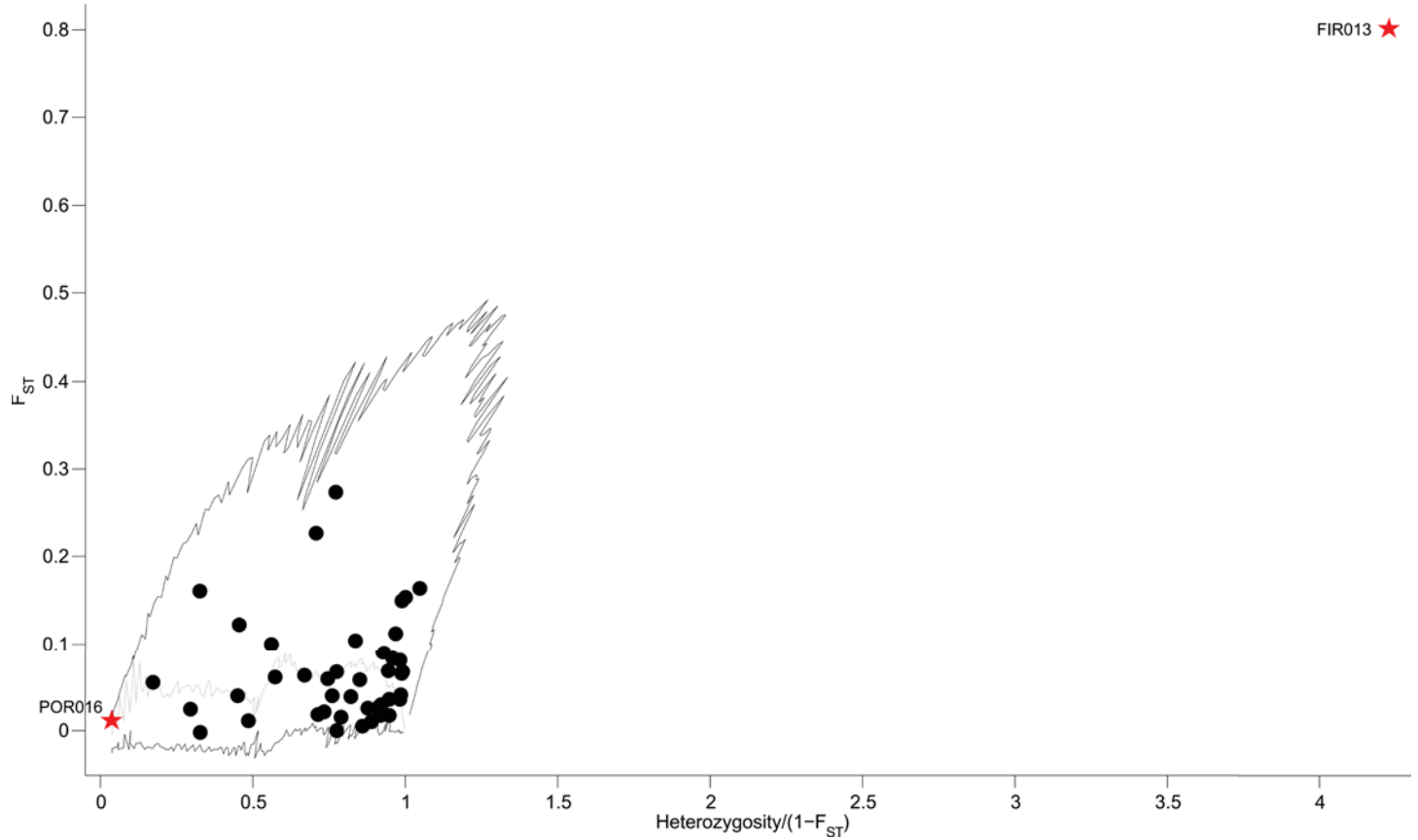


Figure 8 Outlier loci analysis of forty-four markers between species among all four FC/Baraga Plains populations as calculated by LOSITAN (Antao et al. 2008). Outliers are indicated by a filled star shape while all others are filled circles and dark grey lines represent the 5% quantiles with markers outside the lower 5% quantile under balancing selection and the markers outside the upper 5% quantile under divergent selection while the light gray line represents the median (50%)

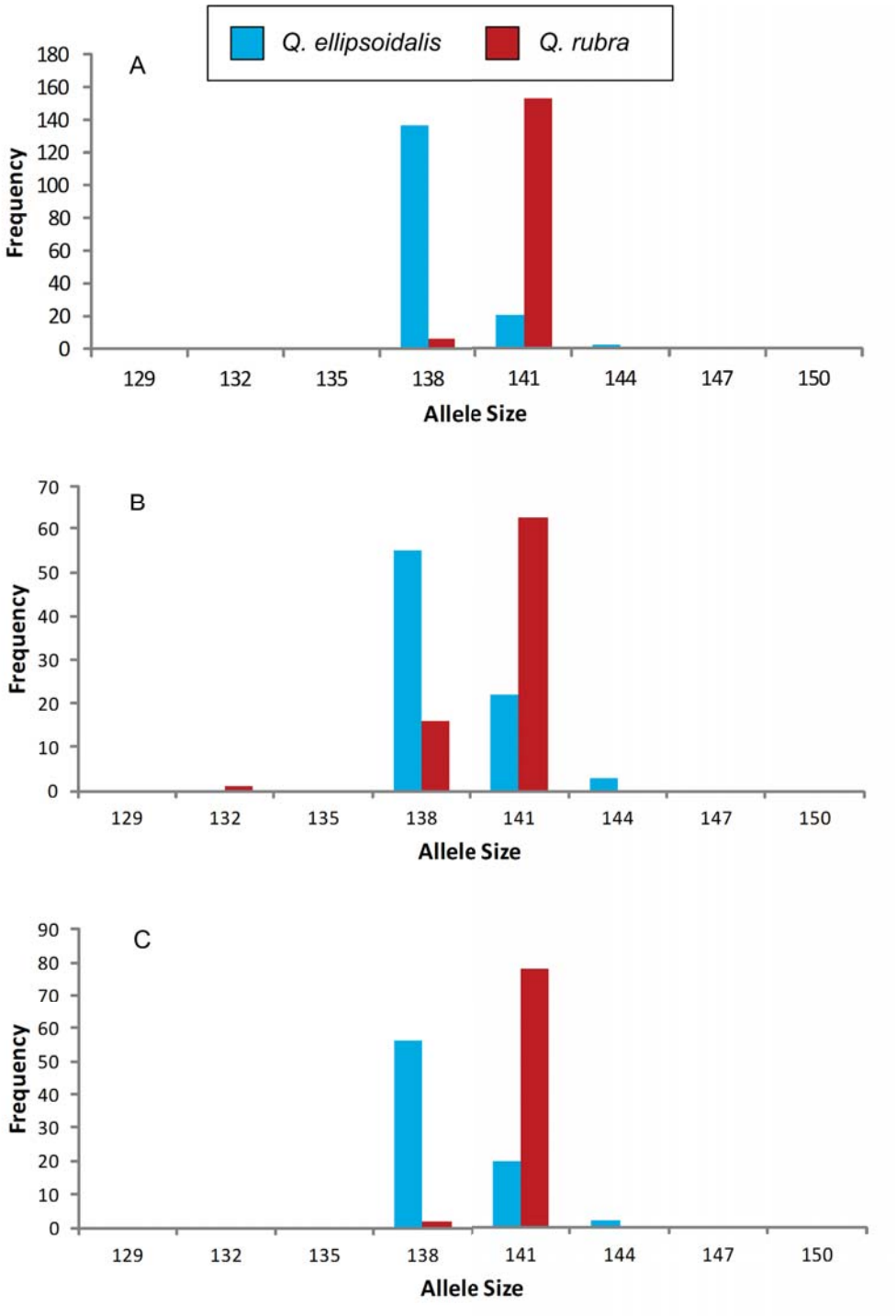
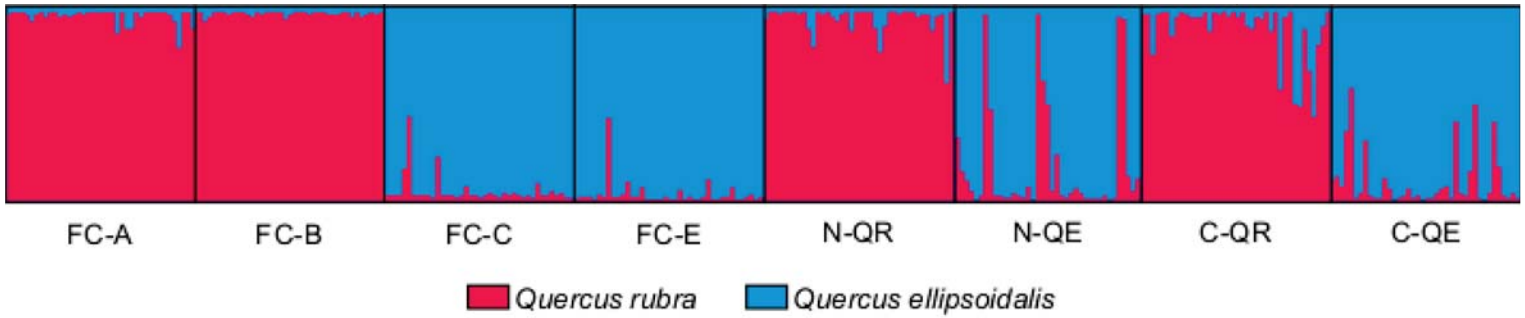


Figure 9 Allele frequency distribution of each species in the a) FC/Baraga Plains populations, b) CNF populations, c) NNF populations for EST-SSR FIR013 (*CONSTANS*-like 1)



Supplement 6 Genetic assignment of individuals and populations according to the Bayesian method implemented in the program STRUCTURE 2.2 (Pritchard, 2000) displayed. One sample of 5 iterated runs considered to best explain the data is shown. Each thin vertical line represents an individual and the proportion of each color is the proportion of ancestry derived from each of the two main genetic groups (K= 2) inferred. Populations are separated by black vertical lines

Supplement 7 Microsatellite marker characteristics

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Functional annotation [b]
1P10[e,h]	$(TG)_{12}$	6FAM-ATTTCTGATGCAGGGTGTCG	TAGGCCAAGGACCAGAGACC	62	237-265	*Q. rubra*	-	-
2P24[e,h]	$(CA)_{14}$	VIC-GCAAGAGATCACACACAAACTAGC	CTTTGGGTTCACCAAACAGC	62	136-172	*Q. rubra*	-	-
3A05[e,h]	$(CT)_{12}$	PET-AACGTGACCTCTCTCACAGC	AGTGCTGGAGTGCTCATGG	62	138-162	*Q. rubra*	-	-
3D15[e,h]	$(CA)_{15}$	NED-GGTGGTGGCAGATACACTGG	GACTCAGACAACCAACTTCAGG	62	208-236	*Q. rubra*	-	-
FIR004[d,h]	$(CT)_{18}$	6FAM-TCTCTCTCAGGGCAGCTTCT	AACCAAACTCAGATCCAGATTCA	59	124-178	*Q. robur*	3	---NA---
FIR013[d,h]	$(CAG)_5$	6FAM-CGGGGAGGTTGATGAGTATT	AACACTGTCACCCCCATAGC	56	133-144	*Q. robur*	2	constans-1

[a] primer pair amplifies two loci

[b] Putative function determined through a BLASTx search following the method in Luro et al. (2008)

[c] SSRs (Aldrich et al. 2002)

[d] EST-SSRs (Durand et al. 2010) (Sullivan et al. 2013)

[e] SSRs (Sullivan et al. 2013)

[f] SSRs (Steinkellner et al. 1997)

[g] on *Q. robur* linkage map in either Durand et al. (2010) or Bodénès et al. (2012)

[h] amplified in CNNF populations

S7 continued

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Putative function[b]
FIR024[d]	$(CCT)_6$	NED-CGCTTCTCCTCATCCTCAAG	CTCAAAAGGCACGATTCTCC	59	214-229	*Q. robur*	5	dna gyrase subunit a
FIR028[d,h]	$(TC)_8$	VIC-GGAAGAGTGTTCGGAAAGCA	CCAGCTCCTCCACAATAGCA	56	201-237	*Q. robur*	1	tropinone reductase homolog at1g07440
FIR030[d]	$(AG)_7$	6FAM-GGACATATTTATCTAGGAGACGAGGT	ATGTCCCATAGCACAGAGCA	57	157-183	*Q. robur*	7	nadh dehydrogenase
FIR031[d]	$(TC)_7$	PET-ACGAGTCCAACGGAAGTTGT	CACAACTTCACAAGGCAAGG	59	135-182	*Q. robur*	11	predicted protein
FIR035[d]	$(AT)_6$	NED-GCTAAGGTTCCGTGTTCCAA	GGCCAGCAACTAAACCAAGA	56	146-152	*Q. robur*	5	chaperone protein dnaj
FIR039[d,h]	$(CT)_7$	PET-GAGCCTCTTTCATCGCTCAC	TCAACACCCCAAAACTCCAT	59	111-132	*Q. robur*	1	histone deacetylase
FIR043[d]	$(TC)_9$	PET-TTCTCCATTTCACACGCTTC	ACGACATCGTTTTGGAGCTT	56	114-146	*Q. robur*	7	metallopeptidase m24 domain-containing protein
FIR048[d,h]	$(CT)_9$	PET-TGCACCAAAATTGGAGGATG	TTGATGCAAGGTGCAGTTTC	56	187-219	*Q. robur*	2	Protein
FIR053[d]	$(GTG)_7$	NED-AGTTTCCCCACATTTGTTGC	TACCATGCACCAAGCAATTC	59	136-150	*Q. robur*	5	glutaredoxin c9

S7 continued

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Putative function[b]
FIR089[d]	$(GA)_6$	6FAM-AGCGACTAACCCAACTTCCA	GCGGATTCGATAGCATTTTT	56	159-181	*Q. robur*	7	abscisic acid receptor pyl4
FIR104[d,h]	$(GGT)_7$	VIC-TTAACTCGGTTTGCGACTCA	AGCACGTGACTCGACCTGTA	59	203-224	*Q. robur*	11	r2r3-myb transcription factor
GOT004[d,h]	$(TG)_{12}$	6FAM-GGGCATATTGATCGCTTAGG	TGAGCATTCATACATTCCATGAT	59	264-294	*Q. robur*	7	---NA---
GOT009[d,h]	$(TC)_7$	6FAM-CACCTCACTAAGCAACCTGTCA	TTTTGGAGGCGGAGATAATG	56	225-249	*Q. robur*	12	predicted protein
GOT011[d]	$(TC)_{11}$	VIC-CCCCACCGTCTACTCTCAAA	GCGTTCACCACGTCCATAAT	56	162-212	*Q. robur*	1	pyridoxal biosynthesis protein
GOT021[d,h]	$(AT)_{13}$	VIC-AGAAAGTTCCAGGGAAAGCA	CTTCGTCCCCAGTTGAATGT	59	95-101	*Q. robur*	3	histidine kinase 4-like
GOT037[d]	$(CT)_{11}$	PET-CCATCCTTTTCATTCTTTCCA	TGTTGTTGTTGCTGTTGTCG	57	239-265	*Q. robur*	5	5 - nucleotidase sure-like
GOT040[d]	$(GA)_{11}$	PET-AAGGCACTCGTCGCTTTCTA	ACCGATTTGAAGCTCGAGAA	59	234-252	*Q. robur*	6	40s ribosomal protein s16
GOT047[d]	$(A)_{10}(CT)_{12}$	NED-AACCCAAACCCAAACCTTTC	TGGTGAATTCGAGGTGTTGA	56	250-268	*Q. robur*	not assigned	zinc finger a20 and an1 domain-containing stress-associated protein 5-like

S7 continued

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Putative function[b]
PIE039[d]	$(CTT)_8$	6FAM-GTAAAACGACG GCCAGTGTCCTC ACCCTCTGCGGT CT	CAGAAAGGGCT GCAAAGC	59	157-178	*Q. robur*	12	Protein
PIE040[d,h]	$(TTC)_8$	NED-GTGAGAGAGAG AGAGACAAAGA AGAAAAA	AAATTCTCCGCC ACATTGAG	59	155-174	*Q. robur*	11	basic leucine zipper transcription factor-like protein
PIE099[d,h]	$(TC)_9$	VIC-GTAAAACGACG GCCAGTGTGGCT ACCGACTACTAC CACTTC	CGGTGGACCCA ATATGTAAC	56	179-209	*Q. robur*	8	at1g05070 t7a14_6
PIE101[d]	$(AT)_{13}$	PET-GCGACAGTCAC AATTAAGCTAC	CACCCAATTTTC AATCTGTG	56	139-173	*Q. robur*	8	histidine kinase 3
PIE125[d]	$(GGAAGC)_3$	PET-AATACAAATCGC AGGAGGTG	CTAACCCATCGT TCATGGAG	57	146-162	*Q. robur*	6	unknown
PIE200[d]	$(CAA)_5$	6FAM-ACAACATGTGCC AAAACTGC	TCGATGATGTGG TTGTTGATG	56	107-119	*Q. robur*	not assigned	zinc finger a20 and an1 domain-containing stress-associated protein 5-like
PIE260[d]	$(AG)_9$	VIC-TTCCTTACTCCT TTCCCACTTC	TGGCTGTTCCAA ATCTTCAA	56	136-171	*Q. robur*	6	sugar transporter erd6-like protein

S7 continued

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Putative function[b]
POR003[d]	$(CT)_8$	6FAM-CTCGCTCTCCTCCCCTAATC	AGCTTTGATCGAGTCCGAAA	59	91-115	*Q. robur*	3	bzip transcription factor bzip109
POR016[d,h]	$(GGT)_6$	NED-AGCAACAGCAGAGCCAAAAT	CAGCGGCTTTGAGGTAATTC	59	110-126	*Q. robur*	6	at1g04190 (uncharacterized TPR-containing protein)
QpZAG15[f,h]	$(AG)_{23}$	HEX-CGATTTGATAATGACACTATGG	CATCGACTCATTGTTAAGCAC	54	103-144	*Q. petraea*	9	
quru-GA-0C11[c,h]	$(GA)_{15}$	VIC-ATACCCAGCTCCCATGACCA	TCCCCAAATTCAGGTAGTGT	59	201-233	*Q. rubra*	-	
quru-GA-0E09[c,h]	$(GA)_{16}$	6FAM-TGCCATCCCTATACACAACCA	CCTCCATCACAAAGTTGCC	59	174-250	*Q. rubra*	-	
quru-GA-1F07[c,h]	$(GA)_{22}$	NED-CCGGTCAAAGAAGTTATCAGA	GGGTGGATTGGGTTTCTACCTA	59	294-352	*Q. rubra*	-	
VIT023[d]	$(ATA)_6$	PET-AATGCGAACGACATGAACAA	CTCTCGTCGGAGACTCAACC	56	115-118	*Q. robur*	5	ap2 erf domain-containing transcription factor
VIT057[d]	$(AACTCG)_3$	VIC-TCAGCAAAATCCCAACTTTGT	ACACTTCGCTGTTCCTCGAT	57	128-153	*Q. robur*	9	ap2 erf domain-containing transcription factor

S7 continued

Locus	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (C°)	Size range (in base pairs)	Species	Linkage group[g]	Putative function[b]
VIT01L1[a,d]	$(CAT)_9$	AATTCAAACCCA GCCAACTG	TCCTCTGGAT GCTCCATCA	56	108-112	*Q. robur*	not assigned	proline-rich protein
VIT081L2[a,d]	-	-	-	-	112-136	-	-	
VIT086[d]	$(CAG)_5$	VIC-AAGAACACCCAT TTCCACCA	TAAAATCCAT TTGCCGGTTC	56	184-207	*Q. robur*	not assigned	chaperone protein dnaj chloroplast
VIT107[d]	$(TA)_{13}$	NED-TGATCACAGATT GGAGCTTAACA	CCCCCACTTA GGAAAGAAGC	59	124-142	*Q. robur*	3	light-harvesting complex i protein lhca2
WAG004[d]	$(TTC)_7$	PET-AAAGCAATTCAA CTGGGACG	ACGACACCGT TTGTTCCTTC	56	250-300	*Q. robur*	2	hypothetical protein VITISV_037831
WAG018[d]	$(GT)_7$	6FAM-GGTTCCGATTTGT TTGAGTCC	TACAAAACCC AAGCTCCCTG	56	132-151	*Q. robur*	6	inactive beta-amylase 9

Supplement 8 HWE Exact Tests (Weir and Cockerham, 1984) by locus in each population. Significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of α= 0.05

Population	F	p-value	Standard Error
1P10			
FC-A	0.0161	0.6187	0.0028
FC-B	-0.0313	0.6771	0.0047
FC-C	0.0499	0.1525	0.0027
FC-E	0.0294	0.2871	0.0032
Mean	0.0160	-	-
2P24			
FC-A*	0.2800	0.0059	0.0002
FC-B*	0.1789	0.0047	0.0003
FC-C*	0.3139	**0.0005**	0.0001
FC-E*	0.2213	0.0108	0.0003
Mean	0.2485	-	-
3A05			
FC-A*	0.3736	0.0011	0.0001
FC-B*	0.2344	0.0332	0.0005
FC-C*	0.3241	0.0046	0.0002
FC-E*	0.5321	**0**	0
Mean	0.3661	-	-
3D15			
FC-A	0.1156	0.5603	0.0027
FC-B*	0.1814	0.0378	0.0010
FC-C	-0.2204	0.5429	0.0009
FC-E	0.0838	0.1436	0.0011
Mean	0.0401	-	-
FIR004			
FC-A*	0.4236	**0**	0
FC-B*	0.3283	**0**	0
FC-C*	0.3243	**0.0001**	0.0001
FC-E*	0.2425	0.0067	0.0007
Mean	0.3297	-	-
FIR013			
FC-A	-0.0174	1	0
FC-B	0.4834	0.0755	0.0001
FC-C	0.2837	0.1632	0.0005
FC-E	-0.0852	1	0
Mean	0.1661	-	-

*potential for null alleles

S8 continued

Population	F	p-value	Standard Error
FIR024			
FC-A*	0.6348	**0**	0
FC-B*	0.3471	0.0117	0.0002
FC-C	0.0930	0.7203	0.0009
FC-E*	0.5058	**0**	0
Mean	0.3952	-	-
FIR028			
FC-A	-0.0437	0.5709	0.0031
FC-B	-0.0301	0.6854	0.0026
FC-C	0.0048	0.9855	0.0002
FC-E	0.1480	0.3043	0.0014
Mean	0.0198	-	-
FIR030			
FC-A*	0.2272	0.0340	0.0017
FC-B*	0.3386	0.0064	0.0004
FC-C	-0.1337	0.8286	0.0011
FC-E	-0.0218	0.5710	0.0014
Mean	0.1026	-	-
FIR031			
FC-A	0.0378	0.1155	0.0025
FC-B*	0.3752	0.0014	0.0002
FC-C*	0.2117	0.0054	0.0004
FC-E*	0.6087	**0**	0
Mean	0.3084	-	-
FIR031			
FC-A	0.0378	0.1155	0.0025
FC-B*	0.3752	0.0014	0.0002
FC-C*	0.2117	0.0054	0.0004
FC-E*	0.6087	**0**	0
Mean	0.3084	-	-
FIR035			
FC-A	0.2481	0.2069	0.0009
FC-B	-0.1642	0.5677	0.0002
FC-C	-0.0263	1	0
FC-E	-0.0541	1	0
Mean	0.0009	-	-
FIR039			
FC-A	0.0792	**0**	0
FC-B*	0.1742	**0.0001**	0
FC-C*	0.2151	0.0174	0.0004
FC-E*	0.5774	**0**	0
Mean	0.2615	-	-

S8 continued

Population	F	p-value	Standard Error
FIR043			
FC-A	0.0091	0.4361	0.0028
FC-B	0.1271	0.5805	0.0030
FC-C*	0.226	0.0165	0.0004
FC-E*	0.3479	**0**	0
Mean	0.1775	-	-
FIR048			
FC-A	-0.0035	0.3552	0.0059
FC-B	0.1450	0.1612	0.0056
FC-C	-0.0701	0.6496	0.0019
FC-E	-0.0241	0.4369	0.0029
Mean	0.0118	-	-
FIR053			
FC-A	-0.0882	0.1597	0.0006
FC-B	0.0186	0.6302	0.0007
FC-C	0.2277	0.1878	0.0002
FC-E	0.0680	0.7496	0.0004
Mean	0.0565	-	-
FIR089			
FC-A	-0.0147	0.4843	0.0016
FC-B	0.0872	0.5245	0.0019
FC-C*	0.2010	0.0894	0.0012
FC-E	0.0500	0.4554	0.0024
Mean	0.0809	-	-
FIR104			
FC-A	0.2383	0.2495	0.0007
FC-B	-0.0894	1	0
FC-C	-0.0328	1	0
FC-E	0.0703	0.9572	0.0004
Mean	0.0466	-	-
GOT004			
FC-A*	0.2166	0.0035	0.0002
FC-B	0.1144	0.0888	0.0010
FC-C	0.1616	0.0242	0.0004
FC-E*	0.7613	**0**	0
Mean	0.3135	-	-
GOT009			
FC-A	0.0061	0.7836	0.0016
FC-B	-0.0566	0.8433	0.0013
FC-C*	0.2014	0.0041	0.0002
FC-E	0.0781	0.2816	0.0015
Mean	0.0573	-	-

S8 continued

Population	F	p-value	Standard Error
GOT009			
FC-A	0.0061	0.7836	0.0016
FC-B	-0.0566	0.8433	0.0013
FC-C*	0.2014	0.0041	0.0002
FC-E	0.0781	0.2816	0.0015
Mean	0.0573	-	-
GOT011			
FC-A*	0.5764	**0**	0
FC-B*	0.6004	**0**	0
FC-C*	0.7358	**0**	0
FC-E*	0.6447	**0**	0
Mean	0.6393	-	-
GOT021			
FC-A	-0.2741	0.1522	0.0002
FC-B	-0.1485	1	0
FC-C	-0.0065	1	0
FC-E	0.3810	0.0147	0.0001
Mean	-0.01203	-	-
GOT037			
FC-A*	0.3283	**0**	0
FC-B*	0.5348	**0**	0
FC-C*	0.5260	**0**	0
FC-E*	0.3166	**0**	0
Mean	0.4264	-	-
GOT040			
FC-A	0.0968	0.7466	0.0013
FC-B	-0.0431	0.2163	0.0013
FC-C	-0.0852	1	0
FC-E	0.2010	0.1338	0.0004
Mean	0.0424	-	-
GOT047			
FC-A*	0.3775	**0**	0
FC-B*	0.4029	**0**	0
FC-C	0.1904	0.0546	0.0010
FC-E*	0.2121	**0.0010**	0.0002
Mean	0.2957	-	-
PIE039			
FC-A	0.0612	0.3528	0.0021
FC-B	-0.1661	0.7787	0.0008
FC-C	-0.0014	0.9204	0.0004
FC-E	0.1896	0.0477	0.0003
Mean	0.0208	-	-

S8 continued

Population	F	p-value	Standard Error
PIE040			
FC-A	0.0913	0.7052	0.0016
FC-B*	0.2530	0.0363	0.0009
FC-C	-0.0135	0.6149	0.0013
FC-E*	0.2028	0.3900	0.0019
Mean	0.1334	-	-
PIE099			
FC-A*	0.4216	**0**	0
FC-B*	0.4513	**0**	0
FC-C	-0.0066	0.6871	0.0025
FC-E*	0.1678	0.0343	0.0006
Mean	0.2585	-	-
PIE101			
FC-A	-0.0030	0.2455	0.0026
FC-B	0.0230	0.1185	0.0013
FC-C	-0.0221	0.1981	0.0018
FC-E	-0.008	0.8622	0.0020
Mean	-0.0025	-	-
PIE125			
FC-A	-0.0468	0.6154	0.0013
FC-B	-0.0448	0.6605	0.0006
FC-C	0.2196	0.0152	0.0003
FC-E	0.0728	0.1471	0.0013
Mean	0.0502	-	-
PIE200			
FC-A	0.2697	**0.0007**	0
FC-B	0.0441	0.0040	0.0001
FC-C	0.2000	**0**	0
FC-E	-0.0581	1	0
Mean	0.1139	-	-
PIE260			
FC-A	0.0537	0.2844	0.0066
FC-B	0.0148	0.4554	0.0062
FC-C	0.0855	0.0027	0.0003
FC-E	-0.0200	0.2248	0.0033
Mean	0.0335	-	-
POR003			
FC-A	0.0370	0.3748	0.0026
FC-B	0.0714	0.1365	0.0028
FC-C	-0.0392	0.7805	0.0019
FC-E	0.1386	**0.0009**	0.0001
Mean	0.0520	-	-

S8 continued

Population	F	p-value	Standard Error
POR016			
FC-A	-0.04	1	0
FC-B	-	-	-
FC-C	-	-	-
FC-E	-	-	-
Mean	-	-	-
QpZAG15			
FC-A	0.1333	0.3569	0.0030
FC-B	0.0970	0.1814	0.0029
FC-C	0.0602	0.1177	0.0030
FC-E	0.0484	0.7223	0.0020
Mean	0.0847	-	-
quru-GA-0C11			
FC-A	0.0440	0.3297	0.0035
FC-B	0.0436	0.2660	0.0032
FC-C	-0.0841	0.6385	0.0026
FC-E	-0.0368	0.2883	0.0042
Mean	-0.0083	-	-
quru-GA-0E09			
FC-A*	0.2000	**0**	0
FC-B*	0.2531	**0.0001**	0.0001
FC-C*	0.2831	**0**	0
FC-E	0.0934	0.02	0.0019
Mean	0.2074	-	-
quru-GA-1F07			
FC-A*	0.2069	0.0076	0.0009
FC-B*	0.1730	0.0746	0.0036
FC-C*	0.1941	0.0432	0.0015
FC-E*	0.4035	**0**	0
Mean	0.2444	-	-
VIT023			
FC-A	0.1152	0.659	0.0002
FC-B	0.0416	1	0
FC-C	-0.0263	1	0
FC-E	-0.2381	0.3124	0.0002
Mean	-0.0269	-	-
VIT057			
FC-A	0.1172	0.1932	0.0005
FC-B	0.0972	0.0458	0.0004
FC-C	-0.0591	0.5275	0.0014
FC-E	0.1376	0.1117	0.0007
Mean	0.0732	-	-

S8 continued

Population	F	p-value	Standard Error
VIT081L1			
FC-A	-0.0130	1	0
FC-B	-0.0400	1	0
FC-C	-0.0400	1	0
FC-E	0.1522	0.3737	0.0003
Mean	0.0148	-	-
VIT081L2			
FC-A	-0.0765	0.8705	0.0010
FC-B	0.1287	0.0251	0.0008
FC-C	0.0102	0.0577	0.0009
FC-E*	0.2257	0.0008	0.0001
Mean	0.0720	-	-
VIT086			
FC-A*	0.3011	**0.0001**	0
FC-B*	0.3316	**0**	0
FC-C*	0.8303	**0**	0
FC-E*	0.7639	**0**	0
Mean	0.5567	-	-
VIT107			
FC-A	0.1798	0.1878	0.0013
FC-B	-0.0106	0.8246	0.0022
FC-C	0.0558	0.4805	0.0022
FC-E	-0.063	0.5301	0.0012
Mean	0.0405	-	-
WAG004			
FC-A	0.0958	0.7754	0.0008
FC-B*	0.2434	0.2247	0.002
FC-C	0.1221	0.0065	0.0003
FC-E	0.1744	0.0112	0.0002
Mean	0.1589	-	-
WAG018			
FC-A	-0.0986	1	0
FC-B	0.2712	0.1402	0.0002
FC-C	-0.1500	0.6004	0.0006
FC-E	0.0905	0.7010	0.0004
Mean	0.0283	-	-

Supplement 9 HWE Exact Tests (Weir and Cockerham, 1984) by locus at each population. Significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of α= 0.05

1P10			
Population	**F**	**p-value**	**Standard Error**
N-QR	0.0119	0.4599	0.0035
N-QE	0.0273	0.6129	0.0041
C-QR	0.0119	0.3384	0.0055
C-QE	-0.0293	0.6748	0.0021
Mean	0.0170	-	-
2P24			
N-QR	-0.046	0.4735	0.0027
N-QE	0.1294	0.0136	0.0007
C-QR	-0.0893	0.0964	0.0019
C-QE	-0.0245	0.8665	0.0019
Mean	-0.0020	-	-
3A05			
N-QR	0.1172	0.1483	0.0017
N-QE*	0.524	**0**	0
C-QR	0.1421	0.2925	0.0024
C-QE*	0.685	**0**	0
Mean	0.2611	-	-
3D15			
N-QR	-0.0969	0.8467	0.0025
N-QE	0.0072	0.0526	0.0008
C-QR	0.0174	0.7978	0.0021
C-QE	0.0355	0.3601	0.0017
Mean	-0.0241	-	-
FIR004			
N-QR*	0.1889	0.0048	0.0007
N-QE*	0.2117	**0**	0
C-QR	0.0651	0.0683	0.0028
C-QE	0.0933	0.1486	0.0053
Mean	0.1552	-	-
FIR013			
N-QR	-0.013	1	0
N-QE*	0.3978	0.0245	0.0002
C-QR*	0.7929	**0**	0
C-QE*	0.6739	**0**	0
Mean	0.3926	-	-

S9 continued

Population	F	p-value	Standard Error
FIR028			
N-QR	0.1078	0.0103	0.0004
N-QE	0.1123	0.0299	0.0017
C-QR	0.0269	0.061	0.0011
C-QE	0.0017	0.3623	0.0027
Mean	0.0823	-	-
FIR039			
N-QR*	0.22	0.0708	0.0005
N-QE*	0.2621	0.0054	0.0002
C-QR	0.0375	**0.0008**	0.0001
C-QE	-0.0364	0.7517	0.0016
Mean	0.1732	-	-
FIR048			
N-QR	0.0865	0.1012	0.0037
N-QE*	0.139	0.0051	0.0006
C-QR	0.0874	0.2248	0.0037
C-QE	0.0426	0.6824	0.0032
Mean	0.1043	-	-
FIR104			
N-QR	-0.0321	0.2051	0.0005
N-QE	0.1403	0.0039	0.0001
C-QR	0.0407	0.6805	0.002
C-QE	-0.043	0.4036	0.0009
Mean	0.0496	-	-
GOT004			
N-QR*	0.2542	0.0383	0.0005
N-QE*	0.4161	**0**	0
C-QR	0.0146	0.3709	0.0026
C-QE*	0.5794	**0**	0
Mean	0.2283	-	-
GOT009			
N-QR	0.116	0.1776	0.0017
N-QE	0.013	0.1907	0.0024
C-QR	-0.0744	0.9208	0.0008
C-QE	0.1246	0.0226	0.0007
Mean	0.0182	-	-
GOT021			
N-QR	-0.0833	1	0
N-QE*	0.7935	**0.0005**	0
C-QR	-0.1304	1	0
C-QE	-	-	-
Mean	0.1933	-	-

S9 continued

Population	F	p-value	Standard Error
PIE041			
N-QR	0.1866	0.0209	0.0002
N-QE	-0.0893	0.1861	0.0017
C-QR	-0.1599	0.5969	0.0018
C-QE	-0.1487	0.7413	0.0016
Mean	-0.0209	-	-
PIE099			
N-QR	0.1067	0.142	0.0022
N-QE	0.0031	0.1665	0.0029
C-QR*	0.448	**0**	0
C-QE	0.0152	0.2415	0.0036
Mean	0.1859	-	-
POR016			
N-QR	-0.0174	1	0
N-QE	-0.027	1	0
C-QR	-0.013	1	0
C-QE	-0.0174	1	0
Mean	-0.0191	-	-
QpZAG15			
N-QR	0.0084	0.4053	0.0048
N-QE	-0.0438	0.6054	0.0052
C-QR	0.0688	0.5977	0.004
C-QE	0.0268	0.0979	0.0029
Mean	0.222	-	-
quru-GA-0C11			
N-QR	0.0626	0.5384	0.0032
N-QE	0.1169	**0.0024**	0.0003
C-QR	0.0134	0.3121	0.0026
C-QE*	0.1789	0.0231	0.0011
Mean	0.0111	-	-
quru-GA-0E09			
N-QR*	0.4511	**0**	0
N-QE	-0.0011	0.3318	0.0092
C-QR*	0.2092	**0**	0
C-QE	-0.0252	0.2673	0.0071
Mean	0.2197	-	-
quru-GA-1F07			
N-QR	0.0874	0.6361	0.0054
N-QE	0.0189	0.1633	0.0046
C-QR	0.0599	0.5849	0.0051
C-QE	-0.0106	0.307	0.0059
Mean	0.0554	-	-

Supplement 10 Summary of between species identified outliers under divergent selection (correction for null alleles)

Pairwise Comparison	LOSITAN	LnRH
FC (QR vs. QE)	FIR013	GOT040
Chequamegon National Forest (QR vs. QE)	NO	GOT021
Nicolet National Forest QR vs. QE)	FIR013	FIR013
ALL (QR vs. QE)	FIR013	NO

NO = no outliers detected

Supplement 11 Detailed outlier results summary for the Ford Center with 44 markers (correction for null alleles)

Method	LOSITAN	LnRH
FC (QR vs. QE)	FIR013	FIR039
FC-A vs. FC-B	NO	PIE200 quru-GA-0C11 VIT081L1
FC-A vs. FC-C	FIR013 GOT021	GOT021 GOT040
FC-A vs. FC-E	FIR013	GOT040
FC-B vs. FC-C	FIR013	VIT081L1
FC-B vs. FC-E	FIR013	NO
FC-C vs. FC-E	NO	PIE260 POR016

NO = no outliers detected

Supplement 12 Detailed outlier results summary for the Ford Center and Chequamegon-Nicolet National Forest with 20 markers (correction for null alleles)

Comparison	LOSITAN	LnRH
ALL (QR vs. QE)	NO	NO
C-QE vs. C-QR	NO	GOT021
C-QE vs. N-QE	NO	GOT021
C-QE vs. N-QR	FIR013	FIR013
C-QR vs. N-QE	GOT021	FIR013
C-QR vs. N-QR	FIR013	FIR013
N-QE vs. N-QR	FIR013	FIR013
FC-A vs. C-QE	FIR013	FIR013 GOT021
FC-A vs. C-QR	FIR013 PIE040	FIR013
FC-A vs. N-QE	FIR013	FIR013 FIR104
FC-A vs. N-QR	PIE040	GOT021
FC-B vs. C-QE	FIR013	FIR013
FC-B vs. C-QR	FIR013 PIE040	FIR013 FIR039
FC-B vs. N-QE	FIR013	FIR013 FIR039
FC-B vs. N-QR	FIR013	NO
FC-C vs. C-QE	NO	FIR013
FC-C vs. C-QR	FIR013 PIE040	GOT021
FC-C vs. N-QE	NO	GOT021
FC-C vs. N-QR	FIR013	FIR013
FC-E vs. C-QE	quru-GA-0C11	FIR013
FC-E vs. C-QR	FIR013	PIE040
FC-E vs. N-QE	PIE040	PIE040
FC-E vs. N-QR	FIR013 POR016	quru-GA-0E09

NO = no outliers detected

Supplement 13 Summary of between species identified outliers under divergent selection (introgressive forms and hybrids excluded)

Pairwise Comparison	LOSITAN	LnRH
FC (QR vs. QE)	FIR013	FIR039 GOT040
Chequamegon National Forest (QR vs. QE)	FIR013	NO
Nicolet National Forest (QR vs. QE)	FIR013	FIR013
ALL (QR vs. QE)	FIR013	FIR013

NO = no outliers detected

Supplement 14 Detailed outlier results summary for the Ford Center with 44 markers (introgressive forms and hybrids excluded)

Method	LOSITAN	LnRH
FC (QR vs. QE)	FIR013	FIR039 GOT040
FC-A vs. FC-B	NO	PIE260 POR016 VIT086
FC-A vs. FC-C	FIR013 GOT021	GOT021 GOT040
FC-A vs. FC-E	FIR013	VIT081L1
FC-B vs. FC-C	FIR013	FIR053 GOT040
FC-B vs. FC-E	FIR013	FIR039
FC-C vs. FC-E	NO	PIE200 VIT081L1

NO = no outliers detected

Supplement 15 Detailed outlier results summary for the Ford Center and Chequamegon-Nicolet National Forest with 20 markers (introgressive forms and hybrids excluded)

Comparison	LOSITAN	LnRH
ALL (QR vs. QE)	FIR013	FIR013
C-QE vs. C-QR	FIR013	NO
C-QE vs. N-QE	NO	quru-GA-0E09
C-QE vs. N-QR	FIR013	FIR013
C-QR vs. N-QE	FIR013	2P24
C-QR vs. N-QR	FIR039	FIR013
N-QE vs. N-QR	FIR013	FIR013
FC-A vs. C-QE	FIR013	FIR039
FC-A vs. C-QR	PIE040	FIR039
FC-A vs. N-QE	FIR013	FIR104 GOT021
FC-A vs. N-QR	NO	GOT021
FC-B vs. C-QE	FIR013	FIR039
FC-B vs. C-QR	PIE040	FIR039
FC-B vs. N-QE	FIR013	FIR039 GOT021
FC-B vs. N-QR	PIE040	FIR039
FC-C vs. C-QE	NO	NO
FC-C vs. C-QR	FIR013	NO
FC-C vs. N-QE	PIE040	NO
FC-C vs. N-QR	FIR013	FIR013
FC-E vs. C-QE	NO	PIE040
FC-E vs. C-QR	FIR013	PIE040
FC-E vs. N-QE	NO	PIE040
FC-E vs. N-QR	FIR013	FIR013

NO = no outliers detected

Supplement 16 Overall genetic variation

44 Microsatellite marker set[a]			
Population	**Gene diversity (H_e)**[b]	**Average Number of Alleles (N_a)**	**Inbreeding Coefficient (F)**
FC-A	0.652	8	0.121
FC-B	0.659	9	0.125
FC-C	0.642	8	0.104
FC-E	0.628	8	0.168
Population mean	0.645	8	0.129
Q. rubra	0.660	10	0.135
Q. ellipsoidalis	0.642	9	0.155
Species mean	0.647	10	0.145
20 Microsatellite marker set[a]			
Population	**Gene diversity (H_e)**	**Average Number of Alleles (N_a)**	**Inbreeding Coefficient (F)**
FC-A	0.700	10	0.111
FC-B	0.695	10	0.126
FC-C	0.684	9	0.087
FC-E	0.660	9	0.193
N-QR	0.698	11	0.073
N-QE	0.732	12	0.145
C-QR	0.713	11	0.066
C-QE	0.700	11	0.094
Population mean	0.698	10	0.112
Q. rubra	0.715	15	0.135
Q. ellipsoidalis	0.711	15	0.168
Species mean	0.713	15	0.152

[a] This refers to the total possible number of markers used to characterize the genetic variation of the population

[b] unbiased expected heterozygosity (Peakall and Smouse 2006)

Supplement 17 Pairwise F_{ST} between species including all eight populations at each of the 20 markers

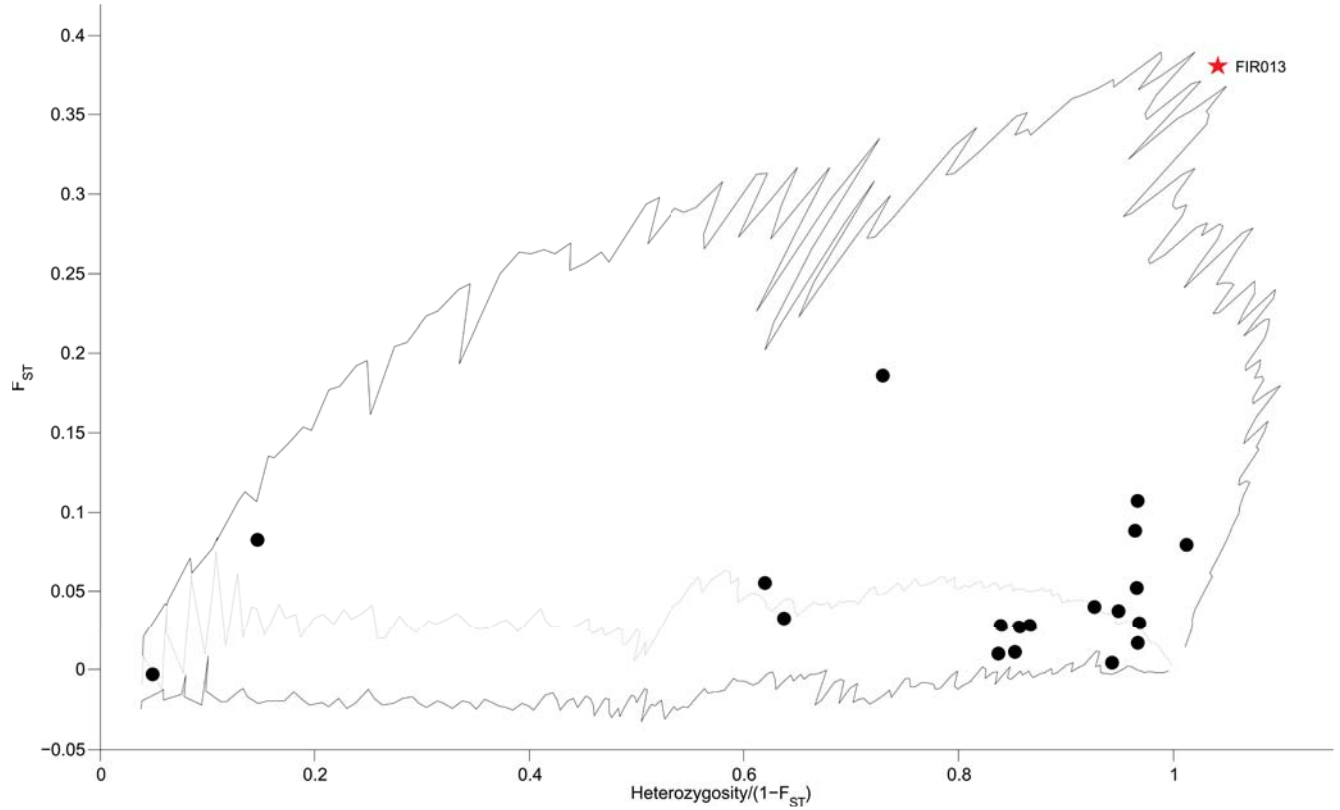
Marker	Pairwise F_{ST} between species[a]
1P10	**0.019**
2P24	**0.067**
3A05	**0.113**
3D15	**0.067**
FIR004	**0.010**
FIR013	**0.668**
FIR028	**0.058**
FIR039	**0.108**
FIR048	**0.049**
FIR104	**0.231**
GOT004	**0.038**
GOT009	**0.013**
GOT021	**0.103**
PIE040	**0.007**
PIE099	**0.038**
POR016	0.000
QpZAG15	**0.040**
quru-GA-0C11	**0.013**
quru-GA-0E09	**0.015**
quru-GA-1F07	**0.008**
ALL	**0.079**

[a] significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of $\alpha = 0.05$

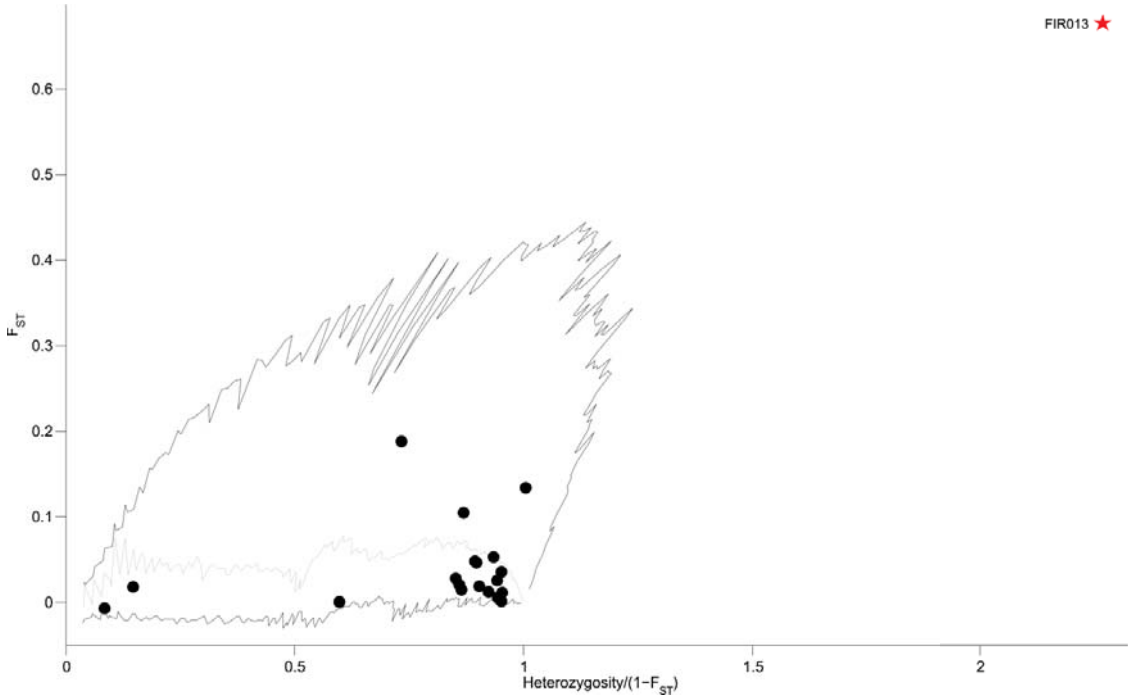
Supplement 18 Pairwise F_{ST} between species including only the four Baraga Plains populations at each of the 44 markers

Marker	Pairwise F_{ST} between species[a]
1P10	**0.0151**
2P24	**0.0809**
3A05	**0.1514**
3D15	**0.1116**
FIR004	**0.0159**
FIR013	**0.8005**
FIR024	**0.0611**
FIR028	**0.0898**
FIR030	**0.0609**
FIR035	0.0242
FIR039	**0.1626**
FIR043	**0.0097**
FIR048	**0.0807**
FIR053	**0.0996**
FIR089	-0.0006
FIR104	**0.2729**
GOT004	**0.0669**
GOT009	**0.039**
GOT011	**0.0237**
G0T021	**0.1615**
GOT037	**0.0623**
GOT040	**0.2267**
GOT047	**0.0566**
PIE039	**0.0592**
PIE040	**0.0179**
PIE099	**0.0277**
PIE101	**0.026**
PIE125	**0.0392**
PIE200	0.039
PIE260	0.0354
POR003	**0.0046**
POR016	0.0117
QpZAG15	**0.0668**
quru-GA-0C11	**0.0354**
quru-GA-0E09	**0.0392**
quru-GA-1F07	**0.0157**
VIT023	-0.0018
VIT057	0.0107
VIT081L1	0.0541
VIT081L2	**0.1032**
VIT086	**0.0638**
VIT107	0.0216
WAG004	**0.149**
WAG018	**0.1216**
ALL	**0.0836**

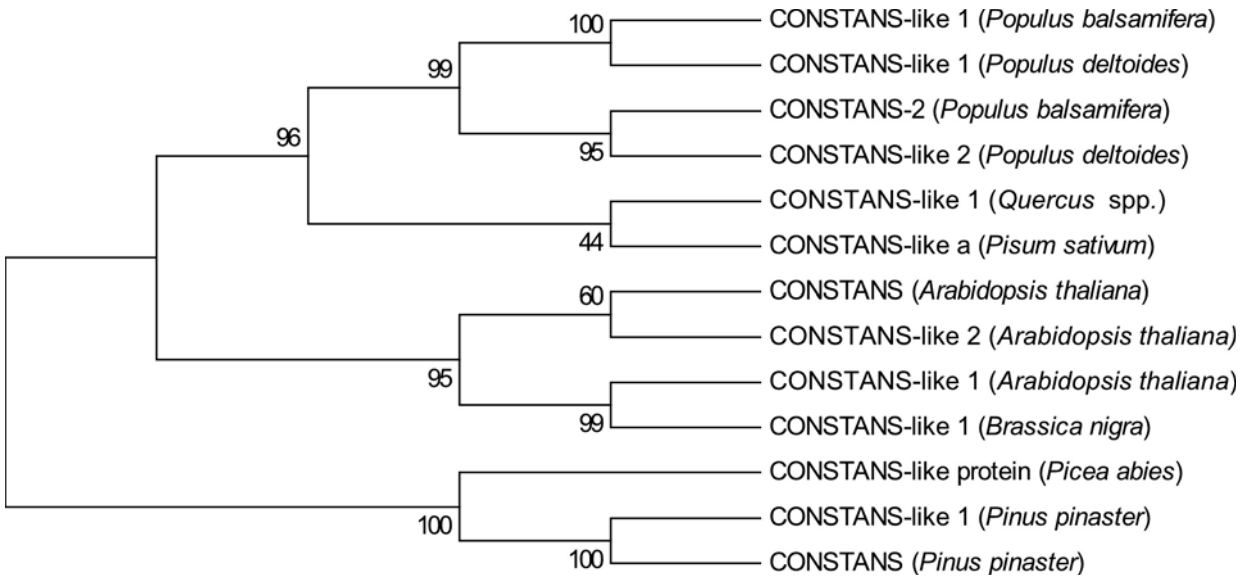
[a] significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of $\alpha = 0.05$



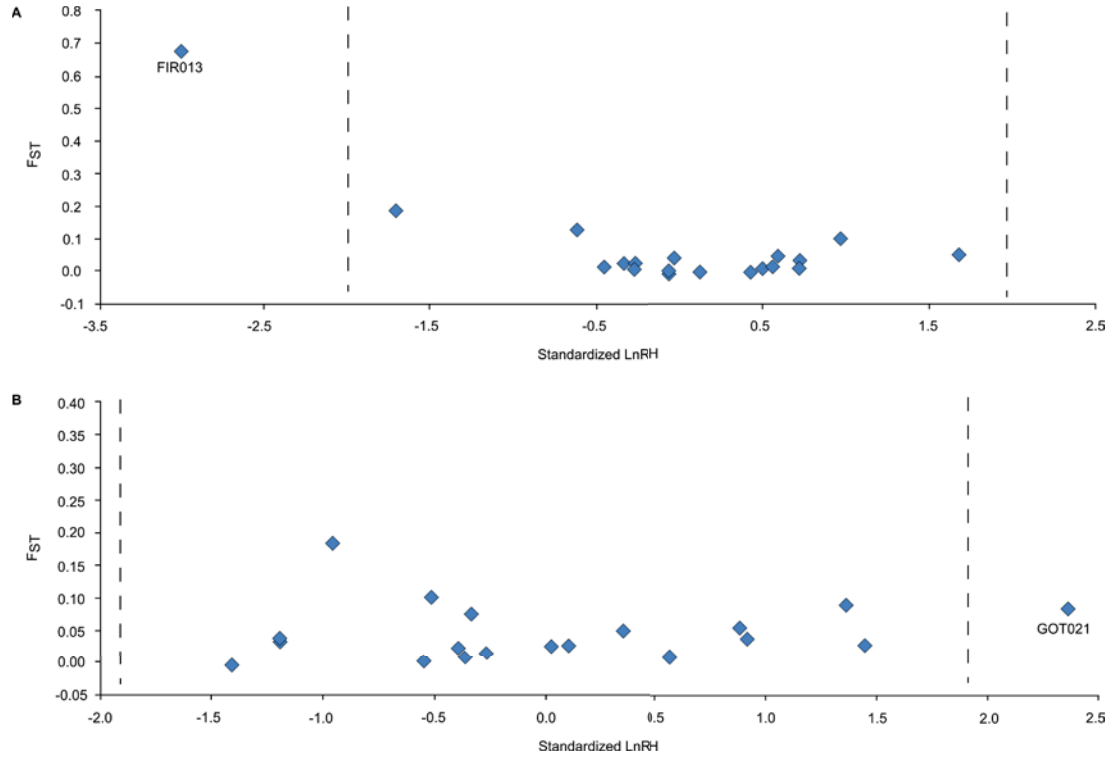
Supplement 19 Outlier loci analysis of forty-four markers between species in the CNF as calculated by LOSITAN (Antao et al. 2008). Outliers are indicated by a filled star shape while all others are filled circles and dark grey lines represent the 5% quantiles with markers outside the lower 5% quantile under balancing selection and the markers outside the upper 5% quantile under divergent selection while the light gray line represents the median (50%)



Supplement 20 Outlier loci analysis of forty-four markers between species in the NNF as calculated by LOSITAN (Antao et al. 2008). Outliers are indicated by a filled star shape while all others are filled circles and dark grey lines represent the 5% quantiles with markers outside the lower 5% quantile under balancing selection and the markers outside the upper 5% quantile under divergent selection while the light gray line represents the median (50%)



Supplement 21 Neighbor-joining phylogeny of aligned protein sequences from a BLASTx search for protein orthologues of the putative *CONSTANS*-like gene



Supplement 22 Standardized LnRH (estimator of variability based on gene diversity) plotted against differentiation values (F_{ST}) between species in the A) NNF populations and B) CNF populations; the 95% confidence interval is designated by a black dashed line

Supplement 23 Detailed outlier results summary for the Ford Center with 44 markers

Method	LOSITAN	LnRH
FC (QR vs. QE)	FIR013 POR016	FIR039 GOT040
FC-A vs. FC-B	NO	PIE260 POR016 VIT086
FC-A vs. FC-C	FIR013	GOT021 GOT040
FC-A vs. FC-E	FIR013	VIT081L1
FC-B vs. FC-C	FIR013	FIR053 GOT040
FC-B vs. FC-E	FIR013	FIR039
FC-C vs. FC-E	FIR104 FIR089 VIT081L1	PIE200 VIT081L1

NO = no outliers detected

Supplement 24 Detailed outlier results summary for the Ford Center and Chequamegon-Nicolet National Forest with 20 markers

Comparison	LOSITAN	LnRH
ALL (QR vs. QE)	FIR013	NO
C-QE vs. C-QR	FIR013	GOT021
C-QE vs. N-QE	NO	GOT021 quru-GA-0E09
C-QE vs. N-QR	FIR013	FIR013
C-QR vs. N-QE	FIR013	POR016
C-QR vs. N-QR	FIR013	FIR013
N-QE vs. N-QR	FIR013	FIR013
FC-A vs. C-QE	FIR013	GOT021
FC-A vs. C-QR	PIE040	FIR013 FIR039
FC-A vs. N-QE	FIR013 FIR104	FIR104 GOT021
FC-A vs. N-QR	PIE040	GOT021
FC-B vs. C-QE	FIR013	FIR039 GOT021
FC-B vs. C-QR	PIE040	FIR013 FIR039
FC-B vs. N-QE	FIR013	FIR039
FC-B vs. N-QR	PIE040	FIR039
FC-C vs. C-QE	NO	NO
FC-C vs. C-QR	NO	NO
FC-C vs. N-QE	PIE040	NO
FC-C vs. N-QR	FIR013	FIR013
FC-E vs. C-QE	quru-GA-0C11	GOT021
FC-E vs. C-QR	FIR013	PIE040
FC-E vs. N-QE	PIE040	PIE040 quru-GA-0E09
FC-E vs. N-QR	FIR013	NO

NO = no outliers detected

License agreement

This is a License Agreement between Jennifer F Lind-Riehl ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.
License number: 3357811157292
License date: Mar 28, 2014
Licensed content publisher: Oxford University Press
Licensed content publication: Annals of Botany
Licensed content author: Jennifer F. Lind-Riehl, Alexis R. Sullivan, Oliver Gailing
Licensed content date first published online: March 10, 2014
Type of use: Thesis/Dissertation
Title of your work: GENETIC VARIATION, LOCAL ADAPTATION AND POPULATION STRUCTURE IN NORTH AMERICAN RED OAK SPECIES, *QUERCUS RUBRA* L. AND *Q. ELLIPSOIDALIS* E. J. HILL

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL

FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent.
Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should

there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions: v1.4

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number: RLNK501263852.

Once you receive your invoice for this order, you may pay your invoice by credit card.

Please follow instructions provided at that time.

Make Payment To:

Copyright Clearance Center

Dept 001

Fine-scale spatial genetic structure of two red oak species, *Quercus rubra* and *Q. ellipsoidalis*[3]

Jennifer Lind-Riehl† and Oliver Gailing†‡

† Michigan Technological University, 1400 Townsend Drive, 49931 Houghton, MI, United States

‡For correspondence: ogailing@mtu.edu or (906) 487-1615

[3] The material in this chapter was previously published material and has been reprinted with kind permission from Springer Science and Business Media from Lind-Riehl JF and Gailing O (2014) Fine-scale spatial genetic structure of two red oak species, *Quercus rubra* and *Q. ellipsoidalis*. Plant Systematics and Evolution 10.1007/s00606-014-1173-y

Abstract

Peripheral populations located at their range edge may be at risk due to geographical isolation, environmental changes, human disturbances or catastrophic events such as wildfires. Fine-scale spatial genetic structure (SGS) investigations provide a way to examine the spatial arrangement of genetic variation within populations. SGS can result from restricted seed and pollen dispersal and might be affected by geographic isolation and environmental changes and disturbances even in outcrossing wind-pollinated species like oaks. Studying the SGS of peripheral populations provides information that can be used to develop improved conservation and management plans at the species' range edge. We assessed the level of genetic variation and SGS in twelve range edge populations in northern Wisconsin and the Upper Peninsula of Michigan (USA): eight *Quercus rubra* and four *Q. ellipsoidalis* populations that were subject to different management regimes and natural disturbances. In contrast to *Q. rubra* populations, the drought tolerant *Q. ellipsoidalis* populations are isolated from the species' main distribution range. These populations are not actively managed but are especially prone to recurring fire events. The four managed and four old growth ("unmanaged") *Q. rubra* populations displayed similar levels of genetic variation. Likewise the *Sp* statistic showed similar SGS levels in managed and unmanaged *Q. rubra* populations (*Sp* = 0.005) comparable to other *Quercus* species (European *Q. robur*: *Sp* = 0.003). *Q. ellipsoidalis* populations showed similar or more pronounced SGS than neighboring *Q. rubra* populations, extending up to 83 m in one population. A significant excess of homozygotes across markers in two of the *Q. ellipsoidalis* populations suggests potential inbreeding. In summary, diverse management activities combined with various natural disturbances are likely both influencing SGS patterns. Outcrossing forest trees like oaks hold large amounts of genetic diversity allowing adaptation to environmental changes over their long life spans. Reductions of these genetic stores, through inbreeding for example, can inhibit a species' ability to adapt to changing environmental conditions.

Introduction

Outcrossing forest tree populations are known to harbor high levels of genetic diversity within populations, with relatively little genetic variation among them due to their long life spans, large neighborhood sizes, and extensive gene dispersal (Petit and Hampe 2006; Hamrick et al. 1979). The spatial arrangement of this genetic diversity is expected to vary widely between different forest tree species given their great variety in reproduction, mating, and dispersal systems (Epperson 1992). Fine-scale spatial genetic structure (SGS) can result from, for example, limited seed or pollen dispersal that leads to the clustering of genetically similar individuals (McCauley 1997; Trapnell and Hamrick 2004). The incidence and strength of SGS is influenced by dispersal efficiency, mating systems, population density and life history (Vekemans and Hardy 2004). In turn, distribution of genetic variation in populations reflects patterns of gene flow through pollen and seed dispersal, and microenvironmental selection (Epperson 2000). However, relatively little is known about the distribution of genetic variation at the population level for peripheral or marginal populations (Jump and Peñuelas 2006). Peripheral populations have been shown to be sources of genetic variation (Jimenez et al. 1999; Lorenzo et al. 2009) and the genetic diversity and distinctness in peripheral populations may be the result of local selective pressures or limited gene flow from other populations (Gibson et al. 2009). They may also be particularly vulnerable to climatic changes resulting in increased competition and fire or drought events (Aitken et al. 2008), and/or anthropogenic changes such as fragmentation (Muir et al. 2004; Gibson et al. 2009), both of which may influence the spatial arrangement of genetic variation. Thus, it is imperative to examine SGS in different species with varied life history traits as well as peripheral populations of these species in order to provide the best information for conservation and management activities of a particular species or population (Epperson 1992).

Studies of fine-scale spatial genetic structure (SGS) in oaks and other wind pollinated forest tree species have yielded generally low SGS with a few exceptions. Thus, many wind pollinated species with gravity or animal dispersed seeds showed low, but significant genetic structure over short distances likely as the result of restricted seed dispersal (Berg and Hamrick 1995; Stefenon et al. 2008; Streiff et al. 1998; Troupin et al. 2006). Spatial aggregation of adult trees and seedlings in *Abies alba* Hill showed reduced SGS, suggesting that wind pollination counterbalanced the effect of restricted seed dispersal (Sagnard et al. 2011). Strong SGS has been observed in studies as the result of clonal reproduction or strong inbreeding in geographic isolation (Berg and Hamrick 1994; Chybicki et al. 2011). For example, strong kinship structure up to 50-100 m was found in highly isolated populations of wind pollinated *Taxus baccata* L., whose seeds are dispersed by gravity or birds (Chybicki et al. 2011).

Studies examining range edge populations are rare, but Pandey and Rajora (2012) compared SGS between two peripheral and two core populations of *Thuja occidentalis* and found relatively higher SGS in peripheral populations. Muir et al. (2004) examined range edge fragmented populations of *Q. petraea* in Ireland using both microsatellite and plastid markers and found similar levels of genetic variation to mainland European populations likely maintained by high levels of outcrossing. They caution that the effects of recent exploitation of the area may not yet be reflected in the current population structure due to limited generations exposed to genetic drift and currently high gene flow levels. In fact, Jump and Peñuelas (2006) have shown that chronic fragmentation of *Fagus sylvatica* populations for over more than 600 years has had significant impacts on genetic diversity and structure.

Studies comparing different oak species have shown differences in extent of SGS that often correspond to differences in life history characteristics such as efficiency of seed dispersal or mating systems (Berg and Hamrick 1994; Cottrell et al. 2003; Streiff et al. 1998). For instance, more pronounced SGS was detected in *Quercus*

petraea than in *Q. robur* which was attributed to *Q. robur*'s potential for longer range seed dispersal and the ability to grow over a wider range of site conditions (Streiff et al. 1998). Additionally, management of stands may influence the spatial arrangement of trees consequently affecting SGS by influencing mating patterns and gene flow through modification of the spatial distribution of trees (Finkeldey and Ziehe 2004). For example, harvesting and thinning projects could impact genetic diversity of populations and their ability to adapt by selecting for certain desired phenotypic traits (Finkeldey and Ziehe 2004). However, little is known about the effects of management on SGS (Cottrell et al. 2003). Cottrell et al. (2003) found a higher significant SGS in populations of *Q. robur* and *Q. petraea* that experienced long-term coppicing and planting as compared to relatively unmanaged populations which they attributed to the influence of artificial regeneration through planting of seeds originating from only a few individuals. Managed (logging and other silvicultural activities) and unmanaged populations of another Fagaceae tree species, *Fagus sylvatica* L., exhibited similar levels of genetic diversity (Rajendra et al. 2014). However, managed *F. sylvatica* stands exhibited reduced SGS in comparison to the unmanaged stands potentially as the result of selective removal of trees breaking down family structures (Rajendra et al., 2014; Paffetti et al. 2012).

Here we focus on the two red oak species, *Q. rubra* and *Q. ellipsoidalis* in northern Wisconsin and the Upper Peninsula of Michigan (USA). While *Q. rubra* is common and has a wide distribution range, *Q. ellipsoidalis* has a much smaller distribution range with only fragmented populations in the study areas. Both species can co-occur, but generally are more likely to be near one another than sympatric as *Q. ellipsoidalis* favors dry sandy soils and *Q. rubra* more mesic conditions (Abrams 1990). Given the dry conditions at these sites, *Q. ellipsoidalis* stands are prone to fire, and stand replacing fires may occur as frequently as every 30 years (Dickmann and Leefers 2003). While both species maintain distinct adaptations to drought, there is potential for hybridization, and genetic assignment analyses were used to identify *Q. rubra* and *Q. ellipsoidalis* samples (Lind and Gailing 2013). *Quercus rubra* is

primarily an outcrossing species with potentially long distance pollen dispersal (Ennos 1994; Moran and Clark 2012). Acorn dispersal by gravity is common and is thought to play a role in limiting gene dispersal since acorns are heavy. Jones et al. (2006) found that *Q. rubra* seedlings showed SGS at small scales (up to 25 m) consistent with limited seed dispersal. However, animals are also important vectors and birds such as blue jays are known to cache acorns and have been implicated in postglacial dispersal of fagaceous trees in eastern North America (Johnson and Thompson Webb 1989). Pollen dispersal has been shown to be farther reaching than seed dispersal for *Q. rubra* as would be expected in an outcrossing and wind pollinated tree species (Moran and Clark 2012). Little is known about gene dispersal in *Q. ellipsoidalis*, but *Q. ellipsoidalis* shows similar life history characteristics to *Q. rubra* and both species are hybridizing; seed dispersal is thought to be carried out by similar animals such as squirrels and blue jays. Additionally, seed production seems to be less frequent in *Q. ellipsoidalis* based on field observations in the Ford Research Forest-Baraga Plains area (FRF-BP) since 2009. Little is known about the SGS of *Q. ellipsoidalis* as it is not considered to be a valuable timber species. Aldrich et al. (2005) described SGS in old growth *Q. rubra* populations in Indiana and found significant spatial structure up to 70 m. The stand had maintained high levels of genetic diversity despite the absence of smaller size classes in the core habitat of the stand. This may be indicative of the early stages of a genetic bottleneck likely created by extensive harvesting during European settlement of the area. However, the decline in oak regeneration has become a concern. Many reasons have been cited for this decline including fire suppression, elevated herbivory, and competition with invasive plant species (Huebner 2003; Lorimer 1993).

Q. rubra is an important component of temperate forests in the Great Lakes region and provides habitat and food for various wildlife (McShea et al. 2007). Economically the species is valuable for many uses including wood production (Aldrich and Cavender-Bares 2011). Considering the importance of *Q. rubra*, both economically and ecologically, further understanding of how management practices

affect genetic diversity of *Q. rubra* and the closely related interfertile *Q. ellipsoidalis* would be beneficial to appropriately address issues such as low regeneration rates. Furthermore, since these populations are at the range edge, it will be important to understand current SGS and genetic variation patterns in order to assess the vulnerability of these populations to climate change and anthropogenic impacts.

Our study aims to characterize fine-scale spatial genetic structure (SGS) in unmanaged and managed peripheral populations of *Q. rubra* and to compare SGS in both *Q. rubra* and *Q. ellipsoidalis* peripheral populations. We want to better understand how differences in the historical management and disturbance regimes affect the level of SGS. We also examine whether there are differences in SGS and levels of inbreeding between the more widely distributed species Q. *rubra* and the more isolated populations of *Q. ellipsoidalis*.

Materials and methods

Sample locations

A total of eight *Q. rubra* and four *Q. ellipsoidalis* populations were sampled from five geographic regions from the northern distribution edge of both species. While both species are at their northern distribution edge, *Q. rubra* remains far more frequent than *Q. ellipsoidalis*. *Quercus ellipsoidalis* populations were geographically isolated and separated from the main distribution range of the species (Lind-Riehl et al. 2014). Eight populations (four *Q. rubra* and four *Q. ellipsoidalis*) were subjected to different degrees of management, while four populations (*Q. rubra* only) were mainly unmanaged pre-European settlement forests (Dickman and Leefers 2003) (Figure 10, Table 9). All *Q. rubra* populations showed signs of natural regeneration. The managed forests consist of two pairs of *Q. rubra* and *Q. ellipsoidalis* populations in the Western Upper Peninsula of Michigan in the Ford Research Forest-Baraga Plains area (FRF-BP) as well as two pairs of *Q. rubra* and *Q. ellipsoidalis* populations from the Chequamegon-Nicolet National Forest in northern Wisconsin

(Fig. 10, Table 9). Like many forests in the temperate regions of the United States both of these areas were heavily logged for pine in the late 1800s and then again for northern hardwoods between the 1920s and 1940s (Dickmann and Leefers 2003; Saetre 1983). Catastrophic wild fires followed the logging of forests in Michigan and Wisconsin. Pine barrens such as in the Baraga Plains and in the Chequamegon-Nicolet National Forest were especially prone to fire and were maintained by fires that could return at intervals of less than 30 years (Dickmann and Leefers 2003; Saetre 1983). Red oaks colonized burnt cutover lands in both Michigan and Wisconsin (Abrams 1992).

At the FRF-BP site, a stand replacing fire occurred sometime around 1910 that burned through the Baraga Plains reaching almost to Alberta, which encompasses the areas occupied by all four sampled stands (James Schmierer, personal communication). Fire scars are still visible on some of the older surviving trees in the FC-A stand (field observation). No major fires have occurred in the area since then. In 1954, the Ford Motor Company donated 1700 acres of land including the current Ford Research Forest. Since then, management for hardwoods, including *Q. rubra*, has been based on a selection system silviculture. In the Baraga Plains jack pine has been managed using even-aged methods and various regeneration techniques including scarification or spot fires, with incidental management of *Q. ellipsoidalis* populations (James Schmierer, personal communication). The *Q. ellipsoidalis* stands, FC-C and FC-E, are both located in pine barrens known as the Baraga Plains with jack pine surrounding them. The *Q. rubra* stand FC-A is in a mixed mesic deciduous forest with major sugar maple and pine components, while FC-B is in a mixed deciduous hardwood forest with maple and hemlock (field observation).

Starting in the late 1920s Wisconsin began acquiring land for the creation of national forests. By 1933, the Chequamegon and Nicolet National Forests (CNF and NNF) were officially named and now consist of more than 1.5 million acres of land across northern Wisconsin. Prior to 1930, fire was common in northern Wisconsin with

about 2500 fires burning half a million acres a year. Since that time fires have decreased, replanting increased and more sustainable management practices have been adopted (Saetre 1983). Jack pine that surrounded the *Q. ellipsoidalis* stand in the NNF has been periodically clear cut (last cut in 1992). The area around the CNF *Q. ellipsoidalis* stand, which is partially on private land, has also been periodically clear cut (last cut in 1994) and is now mostly covered with jack pine. The two *Q. rubra* stands, one in CNF and one in NNF, have been managed through thinning and shelter wood cutting, most recently between 2000 and 2010, allowing for the establishment of even-aged stands. *Q. rubra* is the dominant species in the NNF stand and the CNF stand contains mostly *Q. rubra* with sugar maple and some aspen and spruce (Deborah Veen (USFS) and Alexis Sullivan, personal communication). In general, *Q. rubra* has been directly managed in both the FRF-BP, CNF, and NNF populations, while the *Q. ellipsoidalis* populations in these areas experienced only incidental management.

Additionally, four unmanaged *Q. rubra* populations were collected in the Western Upper Peninsula of Michigan (Fig. 10, Table 9). Compared to those described above, these populations have not been actively managed. Three of them are located in the Huron Mountain Reserve (HMR), which was established as the privately owned Huron Mountain Shooting and Fishing Club in 1889. It contains one of the few remaining extensive tracts of intact hardwood-hemlock-pine forests in the Upper Great Lakes (Dickmann and Leefers 2003). In the 1930s the Club approached Aldo Leopold for advice on how to best manage their lands and his 1938 report focused on preserving natural habitat for native wildlife and as large a sample as possible of uncut timber. In the 1940s, the Club teamed up with the US Forest Service to implement many of Leopold's recommendations which have guided their management strategies to this day (Flaspohler and Meine 2006). Only about 20% of the Club's perimeter lands were selectively logged for white pine in the late 1800s and about 6% of the Club's land consists of oak-pine communities (Davis 1996). None of the sampled populations were located in the selectively logged lands. The

fourth population is located in the Porcupine Mountains Wilderness State Park (PM). The PM was formed in 1945 to protect large stands of old growth northern hardwood forest that are the largest west of the Adirondacks, with 35000 acres of unlogged forest. The forest is primarily composed of sugar maple, American basswood, eastern hemlock and yellow birch and may have been lightly logged for white pine in the late 1800s (Davis 1996). The area is open to the public for recreation as a state park with trail management activity, but timber production is not a focus of the management plan for this forest.

Sample collection

Leaf material was collected and GPS coordinates were recorded for each tree in all sampled populations. Exhaustive sampling was undertaken in the HMR, FRF-BP populations resulting in a large number of comparisons in small distance classes (Suppl. 25). In the CNF, NNF, and PM populations the smaller distance classes were less well represented (Fig. 10, Suppl. 25). While we will not be able to detect SGS at a very small scale due to a lower representation in the small distance classes for the latter populations, we can compare the presence and extent of SGS for populations with similar sampling design. The leaf material was stored at -20°C until DNA extraction. Total genomic DNA (~10-20ng) was extracted using the Qiagen DNeasy96 Plant Kit (Qiagen, Valencia, CA) following the manufacturer's instructions.

Microsatellite genotyping

All populations were characterized at 15 microsatellite markers including eight nuclear simple sequence repeats (nSSRs: 1P10, 2P24, 3A05, 3D15, QpZAG15, quru-GA-0C11, quru-GA-0E09, quru-GA-1F07) and seven expressed sequence tag-SSRs (EST-SSRs: FIR004, FIR048, GOT004, GOT009, GOT021, PIE040, PIE099) described in an earlier study (Lind and Gailing 2013). According to Cavers et al. (2005), 10 microsatellite loci are sufficient to characterize fine-scale SGS. Nuclear

and EST-SSRs developed in *Q. robur* were adapted for use in *Q. rubra* and *Q. ellipsoidalis* and tested for the presence of null alleles (Lind and Gailing 2013; Lind-Riehl et al. 2014, Sullivan et al. 2013). PCR amplification and electrophoretic separation were performed according to Lind and Gailing (2013), but the 10 μL PCR reaction was scaled to a 15 μL reaction.

Genetic variation analyses

GeneAlEx 6.5 (Peakall and Smouse 2006) was used to calculate most genetic diversity parameters including average number of alleles per locus (N_A), number of rare alleles (N_{RARE}, frequency < 5%), effective number of alleles ($N_E = 1/(1-H_S)$), observed heterozygosity (H_O), expected heterozygosity (H_E), and the inbreeding coefficient (F_{IS}) for all populations as well as for both managed and unmanaged populations and for each species. Fisher's exact tests to determine significance of F_{IS} values were calculated in GenePop 4.2 (Raymond and Rousset 1995). FSTAT (Goudet 1995) was used to calculate the measures of genetic diversity allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity within groups (H_E) and genetic differentiation (F_{ST}) for all three groups (*Q. ellipsoidalis* and managed and unmanaged *Q. rubra*). A rarefaction index as suggested by Petit et al. (1998) was used to correct for unequal sample sizes before calculating A_R. Means over all populations were estimated and 5000 permutations were used to test differences between the means in FSTAT (Goudet 1995). This was also done to test for differences of diversity estimates between species and for both managed and unmanaged populations. MICROCHECKER (Van Oosterhout et al. 2004) was used to check for null alleles in each population and markers showed generally low null allele frequencies (see also Lind and Gailing 2013; Lind-Riehl et al. 2014).

Spatial genetic structure analyses

Fine-scale spatial genetic structure (SGS) was analyzed through spatial autocorrelation analysis in SPAGeDi version 1.4 (Hardy and Vekemans 2002).

Spatial autocorrelation analyzes the degree of dependency among observations, genetic variation in this case, in a geographic space. Specifically, allelic information is correlated between pairs of individuals in the same distance class. In order to assess the extent of SGS, the relationship between genetic similarity and geographic distance between individuals in each population was assessed through Moran's I statistics and regression analysis of kinship coefficients (Loiselle et al. 1995) on geographical distances using a jackknife method to estimate standard errors. The results provide insight as to whether genetic structure exists and at what scale (Sokal and Oden 1978). To determine the significance of the coefficients averaged over all loci, they were tested against the null hypothesis of no spatial genetic structure through the creation of a null distribution of randomly permuting individuals among distance classes 10000 times (95% confidence interval). Additionally, the *Sp* statistic (Vekemans and Hardy 2004) was calculated to allow quantitative comparisons of the extent of spatial genetic structure among species and/or populations. This statistic is not affected by the sampling scheme as heavily as the kinship coefficient as it is primarily dependent on the rate of decrease of pairwise kinship coefficients with the logarithm of the distance between individuals. The *Sp* statistic is calculated based on the regression slope of the kinship coefficient, such that $Sp = -b_f/(1 - F_1)$, where F_1 is the mean kinship coefficient over all loci between individuals belonging to the first distance class and b_f is the regression slope of F_1. Statistical significance of F_1 and b_f was determined under a 95% confidence interval of F_{ij} created by 10000 permutations of individuals among distance classes.

The size of the distance classes was determined by SPAGeDi to ensure an equal number of comparisons within each distance class. Initial analyses using various numbers of distance classes were run to determine the optimal number of distance classes to ensure at least 50 comparisons per distance class for all populations and to best capture fine-scale spatial genetic structure. The use of 15 distance classes was determined to be optimal. Runs were performed for each population with only

individuals with a diameter at breast height (DBH) of ≥10cm for *Q. rubra* samples and ≥7cm for *Q. ellipsoidalis* which has a more shrubby growth (Suppl. 26-28).

The area was estimated by taking the area of a polygon created around the sampled area in ArcMap (ESRI 2011) for each population to provide a clearer view of the scale and sampling in each population and to calculate the effective census density according to Stefenon et al. (2008). Effective census density was used to indirectly estimate the gene dispersal rate using SPAGeDi assuming equilibrium of isolation by distance in the fine-scale spatial genetic structure as described by Vekemans and Hardy (2004).

Additionally, to determine the relative diversity of diameter classes, the Shannon Wiener diversity index was determined for each population. The equation is often used to calculate species diversity, but can also be applied to age or diameter diversity within a single species (McPherson and Rowntree, 1989). Populations with individuals evenly distributed among all diameter or age classes will show high values, indicative of high diameter or age class diversity. To calculate the Shannon Wiener diversity index (H), the equation below was used, where p_i is the proportion of the total sample represented by diameter class i:

$$H = -\sum_{i} [p_i \cdot \ln(p_i)]$$

Results

Managed vs. unmanaged Q. rubra *stands*

No significant differences between managed and unmanaged stands were detected for most genetic variation parameters (Table 10). However, managed stands had slightly, but significantly more genetic diversity (H_E) within populations than unmanaged stands (Suppl. 29). Level and extent of SGS was similar in most managed (0-36 m) and unmanaged stands (0-28 m) with unmanaged stand PM-QR

showing weak SGS up to 60 m (Fig. 11l, Table 11 and Suppl. 25). No significant SGS was found in the managed stands CNF-QR and NNF-QR in the first distance class (Fig. 11e, h). Also the unmanaged stand, HMR-MI, with the highest census density showed no significant SGS and low values for the *Sp* statistic (0.002) and Moran's I (0.011) (Table 11). The census densities of the other two unmanaged stands from the same geographic region, HMR-IH and HMR-LP, are lower than in HMR-MI but still much higher than in the managed *Q. rubra* stands (Table 11). Overall there was no significant relationship between Moran's I ($R^2 = 0.0194$) or *Sp* ($R^2 = 0.0053$) and census density. Finally, *Q. rubra Sp* values were similar to other *Quercus* species, such as the European *Q. robur* (Table 12). A linear regression of Moran's I and *Sp* against the Shannon Wiener diversity index (SWI) of diameter classes revealed significant negative associations across all populations ($R^2 = 0.47$, $p< 0.001$; $R^2 = 0.17$, $p< 0.05$). However, our analysis did not include the managed NNF-QR population due to missing diameter data. The inclusion of this even-aged population with non-significant SGS and expected low SWI would have resulted in a non-significant association between SGS and SWI.

Species differences

Genetic variation parameters for both species were similar, but managed *Q. rubra* displayed significantly higher genetic diversity (H_E) within populations than *Q. ellipsoidalis* and unmanaged *Q. rubra* (Table 9 and Suppl. 29). The unmanaged PM-QR population showed the lowest genetic diversity of all populations. Additionally, most of the *Q. ellipsoidalis* populations also displayed higher F_{IS} values across most markers than *Q. rubra* populations, particularly for FC-E and NNF-QE, indicating potential inbreeding (Table 9 and Suppl. 30). The extent of SGS was similar for species pairs of populations in the FRF-BP region (Fig. 11b, g, Table 11: *Q. ellipsoidalis*: up to 26 m and *Q. rubra*: up to 36 m). In the *Q. ellipsoidalis* populations SGS extended to 37 m and 83 m in the CNF and NNF regions, respectively, while no significant SGS was observed in the *Q. rubra* populations in

the first distance class (Fig. 11a, d, e, h). In population NNF-QR significant SGS was found in only one distance class from 90 to 103 m (Fig. 11h). The *Sp* statistic showed the most pronounced SGS for *Q. ellipsoidalis* populations FC-E and NNF-QE (Fig. 11c, d), exceeding the values for neighboring *Q. rubra* (FC-A, FC-B, NNF-QR, Fig. 11f-h) and *Q. ellipsoidalis* populations (FC-C, Fig. 11b) (Table 11). In general, *Q. ellipsoidalis* populations also show higher kinship coefficients than *Q. rubra* populations even though the extent of SGS is similar for both species (Table 11).

Finally, even though gene dispersal estimates did not reach convergence for most populations tested, estimated distances were larger for the *Q. rubra* populations (FC-B: 259 m, CNF-QR: 269 m) than for the *Q. ellipsoidalis* populations (FC-E: 32.6 m, NNF-QE: 158 m). Similar differences in SGS were observed by others for the interfertile sympatric oak species *Q. petraea* and *Q. robur* (Vekemans and Hardy 2004), where the *Sp* value was larger for *Q. petraea* which has a more limited seed dispersal range than *Q. robur* (Table 12).

Discussion

We find that genetic diversity is slightly but significantly higher in managed than in unmanaged *Q. rubra* stands. Although there are not very many studies comparing managed to unmanaged stands, those that do have also generally shown low to no differences in diversity between managed and unmanaged populations of outcrossing temperate forest trees. Thus, investigations of the impact of management on fine-scale spatial genetic structure (SGS) in two European oak species (*Q. robur* and *Q. petraea*) found that only the heavily managed stand had slightly lower genetic diversity than the unmanaged stand (Cottrell et al. 2003). Likewise, several studies in another Fagaceae species, *Fagus sylvatica*, have shown no significant differences in levels of genetic variation as the result of a range of management practices including shelterwood, plantation, and semi-natural regeneration (Buiteveld et al. 2007; Rajendra et al. 2014), but one study noted that some rare alleles were lost in

fragmented populations (Paffetti et al. 2012). *Fagus sylvatica* shares many life history traits with *Quercus* species such as wind pollination, outcrossing nature, and abundant seed and pollen production. The similar levels of genetic variation within both natural and managed populations may be explained by the life history traits of these forest tree species. Hamrick et al. (1979) analyzed relationships between 12 life history traits and various ecological variables and their impact on the levels of genetic variation within populations of 113 different plant taxa. They found that species with the highest genetic diversity tend to be those with large ranges, high fecundities, an outcrossing mode of reproduction, wind pollination, long generation times, and habitats representing later stages of succession. *Quercus rubra*, like many other outcrossing temperate tree species, possesses most of these traits and may be able to maintain high levels of genetic diversity despite natural and anthropogenic interferences. This may also hold true for peripheral populations. Accordingly, similarly high within population genetic diversity was evidenced in peripheral *Q. petraea* populations as in the central distribution range of the species (Muir et al. 2004). Another study that examined the long-term impact of fragmentation on the population genetic structure of 14 *Q. macrocarpa* populations found a lack of large scale population genetic structure with most diversity existing within populations suggesting that wind pollinated trees with large distribution ranges may be resilient to human impacts like fragmentation as well (Craft and Ashley 2007). Most studies examine the impact of fragmentation on genetic diversity of forest tree populations that has occurred within the last 200 years. However, Jump and Peñuelas (2006) found significant effects of long-term (> 600 years) fragmentation on the genetic diversity and population structure of *Fagus sylvatica*. Thus, chronic fragmentation may reduce resilience in the long run and should not be ignored.

While there was no clear difference in SGS between managed and unmanaged *Q. rubra* stands, non-significant SGS was found for the two populations that were subject to even-aged management. Among *Q. rubra* populations, the FRF-BP stands have been managed primarily using a selection system silviculture method (uneven-

aged management), while the CNF and NNF populations were largely managed through shelterwood cutting (even-aged management) where up to 1/3 to 1/2 of the trees are left after harvest (Dickmann and Leefers 2003). Accordingly, the CNF population shows clumps of similar sized trees near each other, while the FC-A and FC-B populations show a more mixed spatial structure of different sized trees (Suppl. 31-33). Additionally, FC-A and FC-B show a higher variation in diameter classes as shown by a higher SWI value than the CNF population (Fig. 12, no data on NNF). The even-aged management in the CNF and NNF populations mimics large scale disturbances such as stand replacing fires, while the uneven-aged management in the FRF-BP populations is similar to natural small scale disturbances. The absence of significant SGS in the CNF and NNF populations (Fig. 11e, h) could be a result of management, but the lower representation in small distance classes in the CNF and NNF populations may also have prevented the detection of fine-scale SGS at smaller distances (see material and methods). Uneven-aged management allows reproduction to occur in overlapping generations and may promote family structures (Finkeldey & Ziehe 2004), which is displayed by the presence of significant SGS up to 36 m in the FRF-BP populations. Similar to the managed stands, the magnitude and extent of SGS in unmanaged stands was highly variable ranging from no SGS in the population with the highest density (HMR-MI, 284 individuals/ha) to extended SGS in the stand with the lowest density (PM-QR). The absence of SGS in HMR-MI might have been caused by overlapping seed shadows. However, across all populations no clear association of SGS was found with either census density or SWI. Since detailed records on management activities and natural disturbances are missing, it is difficult to disentangle the effects of management and natural disturbances on SGS. Long-term silvicultural trials including managed and unmanaged populations might provide a resource to better assess the effect of disturbance regimes on genetic variation and SGS. In the limited number of studies comparing SGS in managed and unmanaged stands, management has been found to both decrease (Paffetti et al. 2012; Rajendra et al. 2014) and

increase SGS (Cottrell et al. 2003). Specifically, Cottrell et al. (2003) found a managed stand to exhibit the highest level of SGS and attributed this to human planting of acorns using seeds from a small number of mother trees. Additionally, there was evidence for pronounced SGS in a clumping of trees within the managed site that represented 50% of the total number of mature trees paired with young trees (DBH < 30 cm) less than 50 m apart.

Studies that have compared SGS between species have generally found that life history characteristics impact the extent and significance of SGS (Berg and Hamrick 1994; Cottrell et al. 2003; Streiff et al. 1998). For example, the more pronounced SGS in *Q. petraea* as compared to *Q. robur* was attributed to *Q. robur*'s potential for longer range seed dispersal and the ability to grow over a wider range of site conditions (Streiff et al. 1998). In our study, differences in level and extent of SGS of *Q. rubra* and *Q. ellipsoidalis* stands suggest that disturbance history (natural and human impact) and species abundance strongly affect SGS. Thus, more pronounced SGS was found in *Q. ellipsoidalis* stands as compared to neighboring *Q. rubra* stands. Oaks show a synchronous reproduction system referred to as masting where at variable intervals (usually every few years) a larger than usual seed crop is produced. If natural regeneration is established in a mast year and that coincides with recent management activity, the new seedling generation could be produced from a limited number of trees and extended family structures might develop. The number of reproducing trees seems to be lower and the intervals between mast years longer in *Q. ellipsoidalis* than in *Q. rubra* populations. Thus, in *Q. ellipsoidalis* stands FC-C and FC-E, fewer reproducing trees were present (multiple year field observations) as compared to neighboring *Q. rubra* stands. While seed production was observed every year since 2009 in *Q. rubra*, only a small group of *Q. ellipsoidalis* trees produced seeds in a single year at the FRF-BP site. Lower numbers of reproducing adults can also lead to inbreeding, which was observed in two of the *Q. ellipsoidalis* populations. This may be characteristic for *Q. ellipsoidalis* which is often found in pine barrens and deals with more harsh conditions such as drought. Fires are also

more frequent in pine barren habitats and while the sampled stands have not had extensive fires in the past 100 years, the *Q. ellipsoidalis* stands have been subject to incidental clear cutting due to their proximity to *P. bansksiana* stands that are managed through this method mimicking the historical fire regime of this type of habitat. It is possible that geographic isolation of the fragmented *Q. ellipsoidalis* populations has contributed to the development of more pronounced family structures and potential inbreeding than in the more widely distributed *Q. rubra* at the species' range edge. These effects of frequent disturbances on SGS might be limited in the main distribution range of *Q. ellipsoidalis*, warranting a comparison of core and peripheral populations to test this hypothesis.

Overall our study suggests that management regimes, natural disturbance, life history traits and geographic isolation can affect SGS and genetic variation of peripheral populations. Isolated *Q. ellipsoidalis* populations showed low seed production, significant SGS and potential evidence for inbreeding. However, it was not possible to disentangle the effects of management and natural disturbances on genetic variation and SGS. Furthermore, life history characteristics of the species in combination with natural disturbance regimes and human mediated management likely leads to variable impacts on a population's SGS. In wind pollinated forest tree species with generally large effective population sizes, the effects of management and disturbance on SGS and inbreeding are likely to be most pronounced in fragmented populations at the species' distribution edge. SGS is also likely to affect the mating patterns in tree populations increasing for example the likelihood of mating between related individuals even in wind pollinated trees (Berg and Hamrick 1994; Chybicki et al. 2011). However, to study the long-term impact of SGS on evolutionary and ecological processes requires long-term monitoring in trees that need 40-50 years to mature.

Acknowledgements

We would like to thank the Huron Mountain Wildlife Foundation, both the Ecosystem Science Center and Biotechnology Research Center in the School of Forest Resources and Environmental Sciences at Michigan Technological University, Michigan Technological University's Finishing Fellowship program, the NSF Plant Genome research program (NSF1025974), the USDA McIntire Stennis fund, the Hanes Trust and the Northern Institute of Applied Climate Science for providing the funding for this work. We would also like to thank Alexis Sullivan for sharing some population data and site information, James Schmierer and Deborah Veen for their willingness to share their knowledge of the management history of the CNF, NNF, and FRF-BP sites, and Jonathan Riehl for his invaluable assistance with the construction of the maps.

References

Abrams MD (1990) Adaptations and responses to drought in *Quercus* species of North America. Tree Physiol 7:227-238

Abrams MD (1992) Fire and the development of oak forests. BioScience 42: 346-353

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-MacLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol App 1: 95-111

Aldrich PR, Cavender-Bares J (2011) Wild crop relatives: genomic and breeding resources. In: Kole C. (Ed), Quercus. Springer Berlin Heidelberg, Berlin, pp 89-129

Aldrich PR, Glaubitz JC, Parker GR, Rhodes OE, Michler CH (2005) Genetic structure inside a declining red oak community in old-growth forest. Journal of Heredity 96: 627-634

Berg EE, Hamrick JL (1994) Spatial and genetic structure of two sandhills oaks: Quercus laevis and Quercus margaretta (Fagaceae). Amer J Bot 81: 7-14

Berg EE, Hamrick JL (1995) Fine-scale genetic structure of a turkey oak forest. Evolution 49: 110-120

Buiteveld J, Vendramin GG, Leonardi S, Kamer K, Geburek T (2007) Genetic diversity and differentiation in European beech (Fagus sylvatica L.) stands varying in management history. For Ecol Manage 247: 98-106

Cavers S, Degen B, Caron H, Lemes MR, Margis R, Salgueiro F, Lowe AJ (2005) Optimal sampling strategy for estimation of spatial genetic structure in tree populations. Heredity 95: 218-289

Chybicki IJ, Oleksa A, Burczyk J (2011) Increased inbreeding and strong kinship structure in Taxus baccata estimated from both AFLP and SSR data. Heredity 107: 589-600

Cottrell JE, Munro RC, Tabbener HE, Milner AD, Forrest GI, Lowe AJ (2003) Comparison of fine-scale genetic structure using nuclear microsatellites within two British oakwoods differing in population history. For Ecol Manage 176: 287-303

Craft KJ, Ashley MV (2007) Landscape genetic structure of bur oak (Quercus macrocarpa) savannas in Illinois. For Ecol and Manage, 239: 13-20

Davis MB (1996) Eastern old-growth forests: prospects for rediscovery and recovery. Island Press, Washington D.C.

Dickmann DI, Leefers LA (2003) The forests of Michigan. University of Michigan Press, Ann Arbor

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72: 250-259

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. In: Adams WT, Strauss S, Copes D, Griffin AR (Eds) Population Genetics of Forest Trees, vol 42. Forestry Sciences. Springer, Netherlands, pp 257-278

Epperson BK (2000) Spatial genetic structure and non-equilibrium demographics within plant populations. Plant Spec Biol 15: 269-279

ESRI (2011) ArcGIS Desktop. 10 edn. Environmental Systems Research Institute. Redlands, CA

Finkeldey R, Ziehe M (2004) Genetic implications of silvicultural regimes. For Ecol Manage 197: 231-244

Flaspohler DJ, Meine C (2006) Planning for wildness: Aldo Leopold's report on Huron Mountain Club. J Forest 104: 32-42

Gibson SY, Van Der Marel RC, Starzomski BM (2009) Climate change and conservation of leading-edge peripheral populations. Conserv Biol 23: 1369-1373

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

Hamrick JL, Linhart YB, Mitton JB (1979) Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. Annu Rev Ecol Syst 10: 173-200

Hardy OJ, Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2: 618-620

Huebner CD (2003) Vulnerability of oak-dominated forests in West Virginia to invasive exotic plants: temporal and spatial patterns of nine exotic species using herbarium records and land classification data. Castanea 68: 1-14

Jimenez P, Agundez D, Alia R, Gil L (1999) Genetic variation in central and marginal populations of Quercus suber L. Silvae Genet 48: 278–284

Johnson WC,Webb, T (1989) The role of blue jays (Cyanocitta cristata L.) in the postglacial dispersal of fagaceous trees in eastern North America. J Biogeogr 16: 561-571

Jones FA, Hamrick JL, Peterson CJ, Squiers ER (2006) Inferring colonization history from analyses of spatial genetic structure within populations of Pinus strobus and Quercus rubra. Mol Ecol 15: 851-861

Jump AS, Peñuelas J (2006) Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. PNAS 103: 8096-8100

Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a CONSTANS-like gene between two red oak species. Ann Bot London 113: 967-975

Lind JF, Gailing O (2013) Genetic structure of Quercus rubra L and Quercus ellipsoidalis E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genet Genomes 9: 707-722

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). Amer J Bot 82: 1420-1425

Lorenzo Z, Burgarella C, Lopez de Heredia U, Lumaret R, Petit RJ, Soto A, Gil L (2009) Relevance of genetics for conservation policies: the case of Minorcan cork oaks. Ann Bot London 104: 1069–1076

Lorimer CG (1993) Causes of oak regeneration problem. United Staes Department of Agriculture, Washington D.C.

McCauley DE (1997) The relative contributions of seed and pollen movement to the local genetic structure of Silene alba. Heredity 88: 257-263

McPherson EG, Rowntree, RA (1989) Using structure measures to compare twenty-two U.S. street tree populations. Landscape 8: 13-23

McShea WJ, Healy WM, Devers P, Fearer T, Koch FH, Stauffer D, Waldon J (2007) Forestry matters: decline of oaks will impact wildlife in hardwood forests. J Wildlife Manage 71: 1717-1728

Moran EV, Clark JS (2012) Between-site differences in the scale of dispersal and gene flow in red oak. PLoS ONE 7: e36492

Muir G, Lowe AJ, Fleming CC, Vogl C (2004) High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of Quercus petraea at the margin of its range in Ireland. Ann Bot London 93: 691-697

Paffetti D, Travaglini D, Buonamici A, Nocentini S, Vendramin GG, Giannini R, Vettori C (2012) The influence of forest management on beech (Fagus sylvatica L.) stand structure and genetic diversity. For Ecol Manage 284: 34-44

Peakall ROD, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288-295

Pandey M and Rajora OP (2012) Higher fine-scale genetic structure in peripheral than in core populations of a long-lived and mixed-mating conifer- eastern white cedar (Thuja occidentalis L.). BMC Evol Biol 12: 48

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst 37: 187-214

Petit RJ, el Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12: 844-855

Rajendra KC, Seifert S, Prinz K, Gailing O, Finkeldey R (2014) Subtle human impacts on patterns of neutral genetic variation in European beech (Fagus sylvatica). For Ecol Manage 319: 138-149

Raymond M, Rousset F (1995) GenePop (Version-1.2) - Population genetics software for exact tests and ecumenicism. Heredity 86: 248-249

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43: 223–225

Saetre S (1983) Chequamegon: The making of a forest. manuscript on file at the Supervisor Office, Chequamegon-Nicolet National Forest, Rhinelander

Sagnard, F, Oddou-Muratorio S, Pichot C, Vendramin GG, Fady B (2011) Effects of seed dispersal, adult tree and seedling density on the spatial genetic structure of regeneration at fine temporal and spatial scales. Tree Genet Genomes 7: 37-48

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology: 1. Methodology. Biol J Linn Soc 10: 199-228

Stefenon VM, Gailing O, Finkeldey R (2008) The role of gene flow in shaping genetic structures of the subtropical conifer species Araucaria angustifolia. Plant Biol 10: 356-364

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within-population genetic structure in Quercus robur L. and Quercus petraea (Matt.) Liebl. assessed with isozymes and microsatellites. Mol Ecol 7: 317-328

Sullivan AR, Lind JF, McCleary TS, Romero-Severson J, Gailing O (2013) Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. Plant Mol Biol Rep 31:231-239

Trapnell DW, Hamrick JL (2004) Partitioning nuclear and chloroplast variation at multiple spatial scales in the neotropical epiphytic orchid, Laelia rubescens. Mol Ecol 13: 2655-2666

Troupin D, Nathan R, Vendramin GG (2006) Analysis of spatial genetic structure in an expanding Pinus halepensis population reveals development of fine-scale genetic clustering over time. Mol Ecol 15: 3617-3630

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4: 535–538

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

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370

Table 9 Sampled population characteristics

	Population	Region	Species	Sample Size (n)	Latitude (S)	Longitude (W)	Altitude (m)	Sampled area (m^2)	Mean DBH (cm)
Managed Stands	CNF-QE	Chequamegon National Forest (CNF)	*Q. ellipsoidalis*	40	46° 44' 43" N	91° 04' 20" W	384	33076	19.1
	FC-C	Ford Research Forest-Baraga Plains (FRF-BP)	*Q. ellipsoidalis*	50	46° 39' 14.454" N	88° 35' 25.616" W	394	41354	19.2
	FC-E	Ford Research Forest-Baraga Plains (FRF-BP)	*Q. ellipsoidalis*	47	46° 39' 55.879" N	88° 33' 19.775" W	398	2575	12.3
	NNF-QE	Nicolet National Forest (NNF)	*Q. ellipsoidalis*	39	45° 19' 19" N	88° 19' 53" W	301	38755	18.7
	CNF-QR	Chequamegon National Forest (CNF)	*Q. rubra*	40	46° 42' 54" N	91° 02' 8" W	323	69969	25.5
	FC-A	Ford Research Forest-Baraga Plains (FRF-BP)	*Q. rubra*	48	46° 39' 9.407" N	88° 30' 6.962" W	297	29743	35.3
	FC-B	Ford Research Forest-Baraga Plains (FRF-BP)	*Q. rubra*	48	46° 40' 27.937" N	88° 31' 27.397" W	423	87754	32.0
	NNF-QR	Nicolet National Forest (NNF)	*Q. rubra*	40	45° 20' 53" N	88° 23' 17" W	354	36596	NA
Unmanaged Stands	HMR-IH	Huron Mountain Reserve (HMR)	*Q. rubra*	76	46° 51' 12.884" N	87° 50' 42.824" W	257	11696	20.5
	HMR-LP	Huron Mountain Reserve (HMR)	*Q. rubra*	52	46° 50' 59.813" N	87° 49' 48.806" W	246	8570	21.4
	HMR-MI	Huron Mountain Reserve (HMR)	*Q. rubra*	60	46° 51' 20.783" N	87° 51' 24.026" W-	307	2359	31.8
	PM-QR	Porcupine Mountains Wilderness State Park (PMWSP)	*Q. rubra*	40	46 44' 23.070"N	89 46' 23.460"W	NA	100931	44.8

NA = information not available

Table 10 Genetic variation at 15 microsatellites loci in 12 managed and unmanaged *Quercus rubra* and *Q. ellipsoidalis* populations (the population abbreviations follow those in Table 9)

Population	N	N_A	N_E	N_{RARE}	A_R[b]	H_O	H_E[a]	F_{IS}
Quercus rubra								
Unmanaged								
HMR-IH	76	13	6	5	12	0.698	0.778	0.080
HMR-LP	52	13	6	5	13	0.666	0.780	0.120
HMR-MI	60	12	6	6	12	0.685	0.776	0.092
PM-QR[c]	36	11	5	6	12	0.635	0.723	0.105
Mean	56	12	6	6	12	0.671	0.757	0.099
Managed								
CNF-QR	40	13	7	6	13	0.748	0.797	0.033
FC-A	48	12	6	6	11	0.669	0.805	0.144
FC-B	48	12	6	6	12	0.652	0.787	0.150
NNF-QR	40	13	7	7	13	0.713	0.798	0.084
Mean	44	12	7	6	13	0.696	0.797	0.103
Quercus ellipsoidalis								
CNF-QE	40	12	6	6	12	0.692	0.777	0.091
FC-C	50	11	6	6	11	0.690	0.767	0.083
FC-E	47	10	5	5	10	0.590	0.749	0.214
NNF-QE	39	14	7	6	14	0.711	0.801	0.139
Mean	44	12	6	6	13	0.671	0.773	0.132

[a] unbiased expected heterozygosity (Peakall and Smouse 2006)
[b] corrected for unequal sample sizes using the rarefaction index suggested by Petit et al. (1998)
[c] only analyzed at 14 markers
N = sample size
N_A = number of alleles averaged over all loci
N_E = number of effective alleles
N_{RARE} = number of different alleles with a frequency of ≥5%
A_R = allelic richness
H_O = observed heterozygosity
H_E = expected heterozygosity
F_{IS} = inbreeding coefficient

Table 11 Estimation of the fine-scale genetic structure[a] at 15 microsatellite markers in managed and unmanaged *Q. rubra* and *Q. ellipsoidalis* populations for sampled trees with a DBH ≥ 10cm for *Q. rubra* and a DBH ≥ 7cm for *Q. ellipsoidalis* (the population abbreviations follow those in Table 9)

Population	CD	F1[b]	b_F	Sp	Moran's I[b]	Extent of SGS	SWI
Quercus rubra							
Unmanaged							
HMR-IH	75	0.026***	-0.009	0.009	0.049***	0-28 m	0.29
HMR-LP	81	0.013*	-0.002	0.002	0.026*	0-13 m	1.17
HMR-MI	284	0.005	-0.002	0.002	0.011	ns	1.92
PM-QR[c]	4	0.016*	-0.006	0.006	0.032*	0-60 m	1.80
Managed							
CNF-QR	6	0.012	-0.006	0.006	0.025	ns	1.23
FC-A	16	0.016**	-0.007	0.007	0.030**	0-28 m	1.81
FC-B	6	0.017**	-0.004	0.004	0.032**	0-36 m	1.92
NNF-QR	11	-0.005	-0.004	0.004	-0.006	ns	-
Quercus ellipsoidalis							
CNF-QE	12	0.022**	-0.006	0.006	0.042**	0-37 m	0.94
FC-C	14	0.020**	-0.005	0.005	0.039**	0-26 m	1.47
FC-E	194	0.023*	-0.014	0.014	0.042*	0-10 m and 20-25 m	0.86
NNF-QE	10	0.027**	-0.017	0.017	0.039**	0-83 m	1.34

[a] 15 distance classes with a minimum of 50 pairs per distance class

[b] level of significance after 10,000 permutations ($p<0.05$*, $p<0.01$**, $p<0.001$***)

[c] only analyzed at 14 markers

CD = census density (individuals per hectare)

F_{IS} = inbreeding coefficient calculated in SPAGeDi

N_A = number of alleles averaged over all loci

F_1 = multilocus kinship coefficient between individuals of the first distance class (Loiselle et al. 1995)

b_F = regression slope of F on natural log distance

Sp = quantification of the SGS

SWI = Shannon Weiner Index (using diameter class)

Table 12 Quantification of the fine-scale genetic structure (Sp values) for selected Fagaceae species

Species	Sp	Reference
Quercus rubra	0.005	-
Quercus ellipsoidalis	0.011	-
Quercus petraea	0.008	Vekemans & Hardy (2004)
Quercus robur	0.003	Vekemans & Hardy (2004)
Fagus sylvatica	0.011	Rajendra et al. (2014)

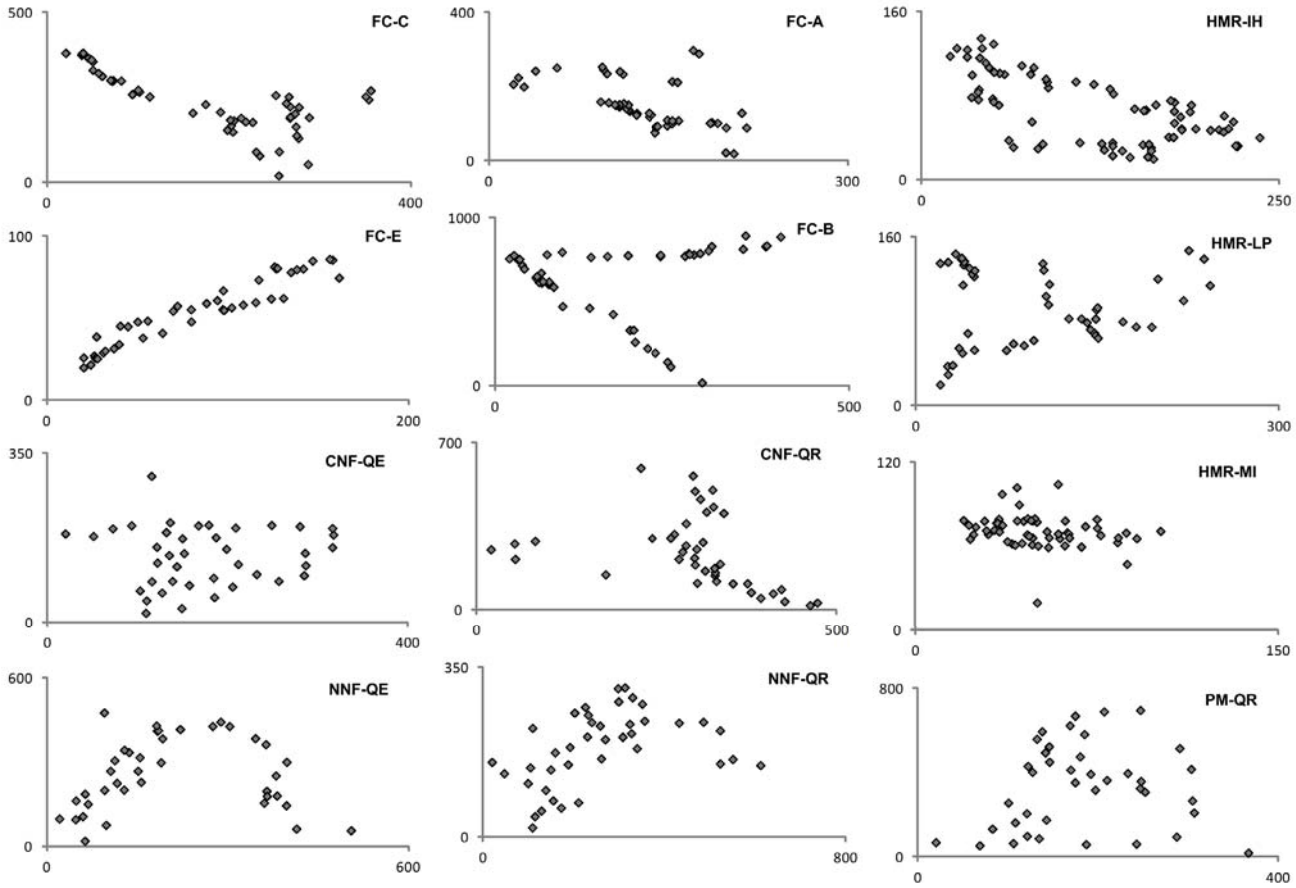


Figure 10 Spatial distribution of sampled populations in northern Wisconsin and the Upper Peninsula of Michigan (USA) (the population abbreviations follow those in Table 9)

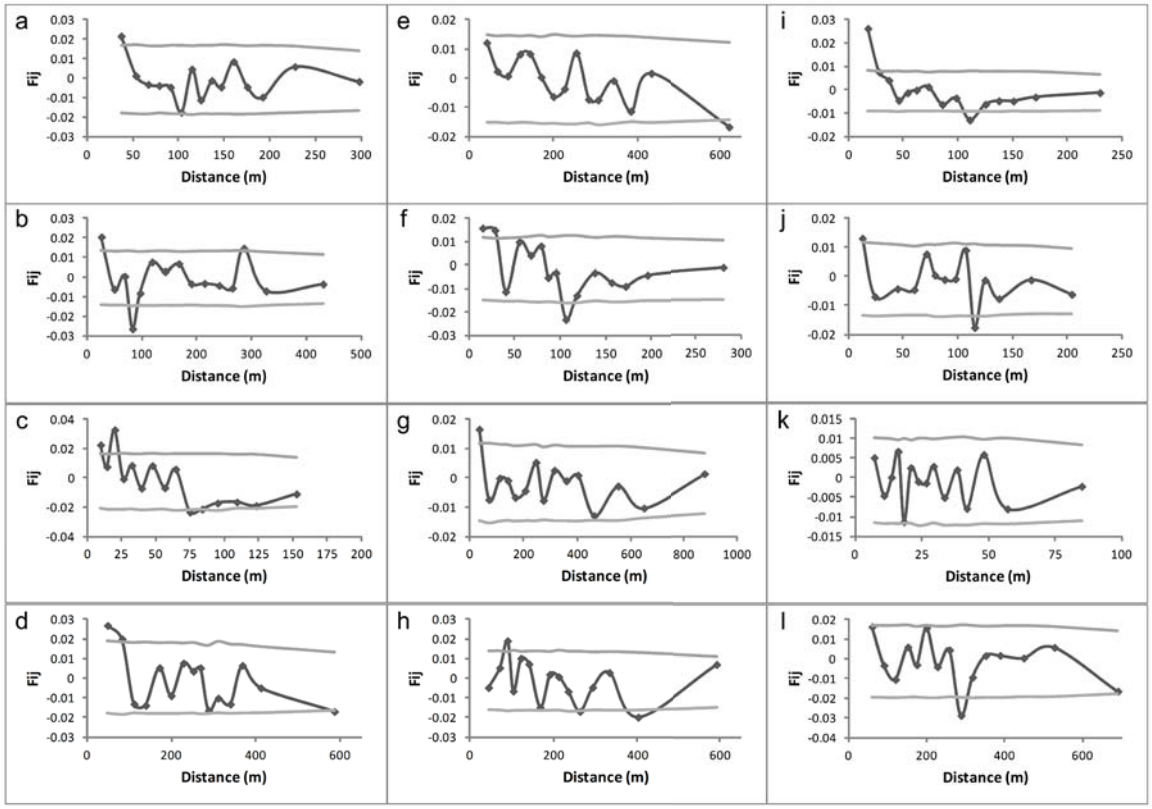


Figure 11 Correlograms showing fine-scale spatial genetic structure (SGS) of studied populations (the population abbreviations follow those in Table 9): a) CNF-QE, b) FC-C, c) FC-E, d) NNF-QE, e) CNF-QR, f) FC-A, g) FC-B, h) NNF-QR, i) HMR-IH, j) HMR-LP, k) HMR-MI, l) PM-QR

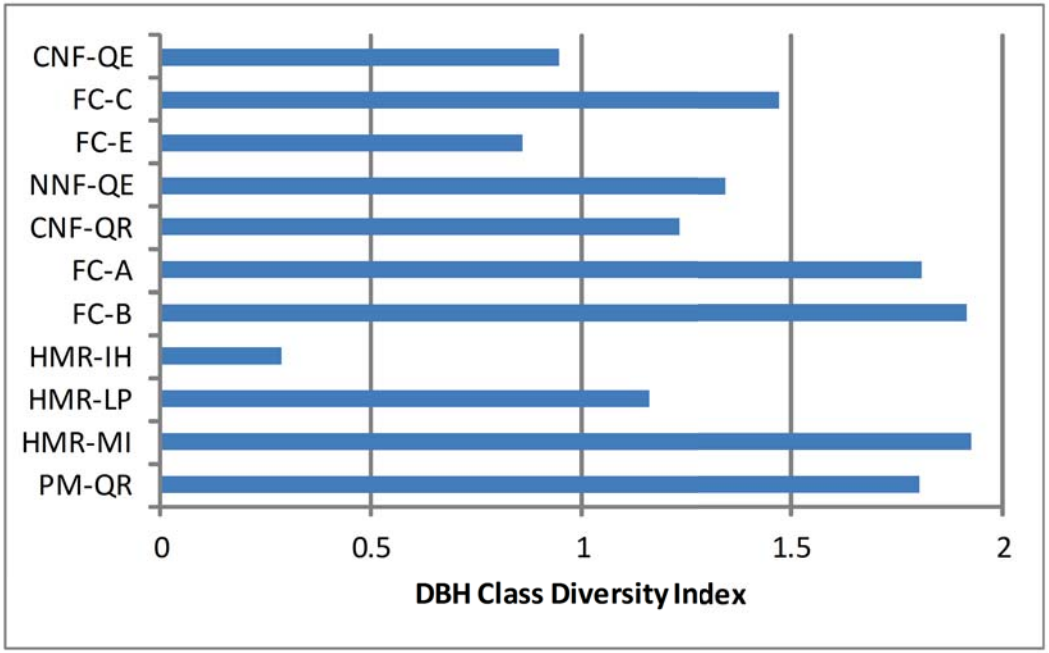
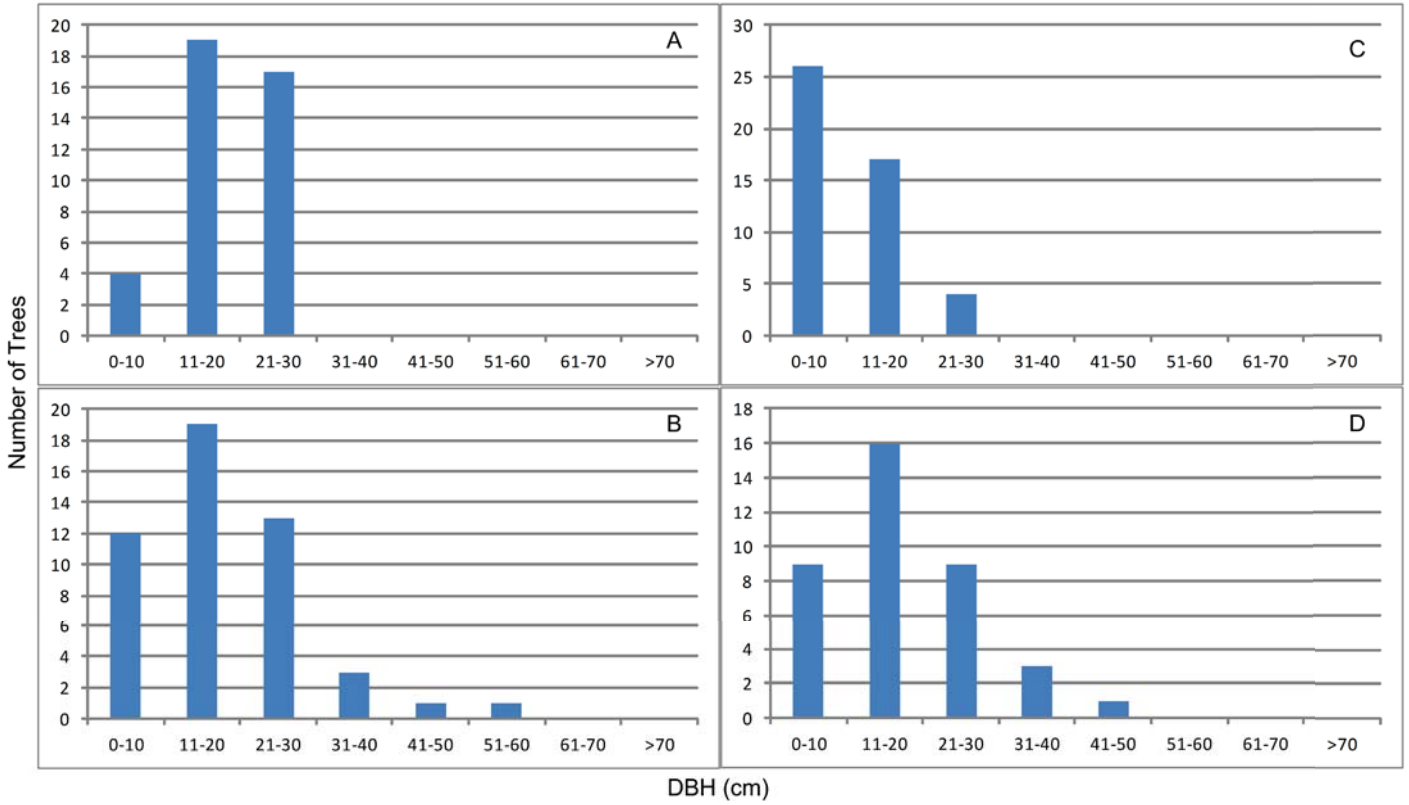


Figure 12 Diameter class diversity for all sampled populations of *Quercus rubra* and *Q. ellipsoidalis* (the population abbreviations follow those in Table 9). DBH – diameter at breast height

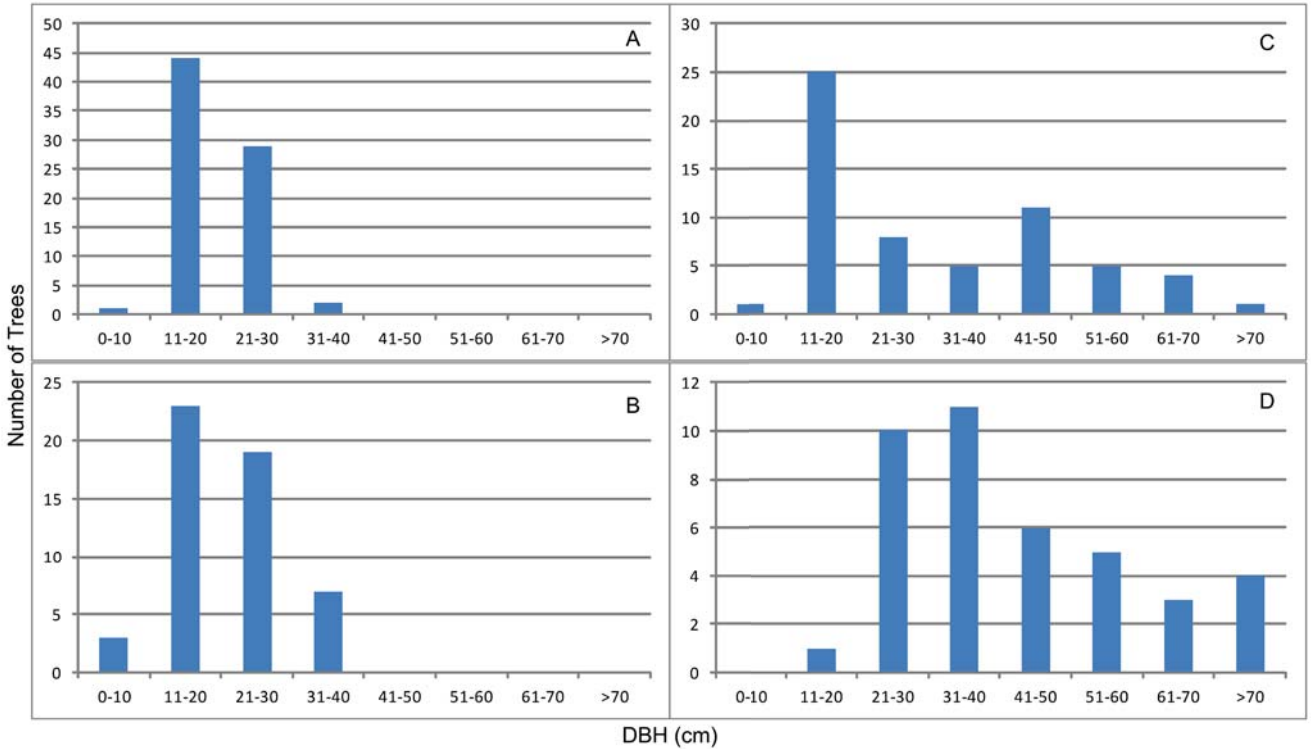
Supplement 25 Size of distance classes and number of pairwise comparisons for each of 15 distance classes for sampled trees with a diameter at breast height (DBH) ≥ 10cm for *Q. rubra* and a DBH ≥ 7cm for *Q. ellipsoidalis* (the population abbreviations follow those in Table 9)

				Distance Class														
	Population	**Sample Size (n)**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Managed Stands	**CNF-QE**	40	Max Distance (m)[a]	***37***	54	67	79	91	103	115	124	136	147	160	175	192	227	298
			# of Pairs	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52
	FC-C	50	Max Distance (m)[a]	***26***	50	69	83	97	119	142	167	190	214	241	265	*286*	327	430
			# of Pairs	81	82	82	81	82	82	81	82	82	81	82	82	81	82	82
	FC-E	47	Max Distance (m)[a]	***10***	14	***20***	26	33	40	48	57	65	75	84	95	109	123	153
			# of Pairs	57	57	58	57	58	57	57	58	57	58	57	57	58	57	58
	NNF-QE	39	Max Distance (m)[a]	***49***	***83***	112	139	172	199	228	252	269	288	311	340	368	412	586
			# of Pairs	49	49	50	49	50	49	49	50	49	50	49	49	50	49	50
	CNF-QR	40	Max Distance (m)[a]	42	67	92	121	145	171	201	227	255	284	307	343	384	433	622
			# of Pairs	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52
	FC-A	39	Max Distance (m)[a]	***15***	***28***	40	55	68	79	87	95	106	118	138	156	172	196	281
			# of Pairs	75	75	75	75	75	76	74	76	75	75	75	75	75	75	76
	FC-B	40	Max Distance (m)[a]	***36***	72	112	143	169	206	247	274	315	359	402	463	551	647	878
			# of Pairs	75	75	75	75	76	75	75	75	75	76	75	75	75	75	76
	NNF-QR	40	Max Distance (m)[a]	45	72	***90***	104	122	141	168	191	212	235	263	293	333	401	592
			# of Pairs	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52
Unmanaged Stands	**HMR-IH**	76	Max Distance (m)[a]	***18***	***28***	37	46	54	62	73	86	99	111	125	138	151	171	230
			# of Pairs	190	190	190	190	190	190	190	190	190	190	190	190	190	190	190
	HMR-LP	52	Max Distance (m)[a]	***13***	24	45	60	71	79	88	98	107	115	125	138	167	204	1361
			# of Pairs	88	88	89	88	88	89	88	88	90	87	88	89	88	88	89
	HMR-MI	61	Max Distance (m)[a]	7	11	14	16	18	21	24	27	29	33	38	42	48	57	85
			# of Pairs	117	119	117	118	118	117	118	118	118	118	118	117	118	118	118
	PM-QR	41	Max Distance (m)[a]	***60***	90	120	152	175	200	227	259	288	318	351	388	449	527	690
			# of Pairs	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52

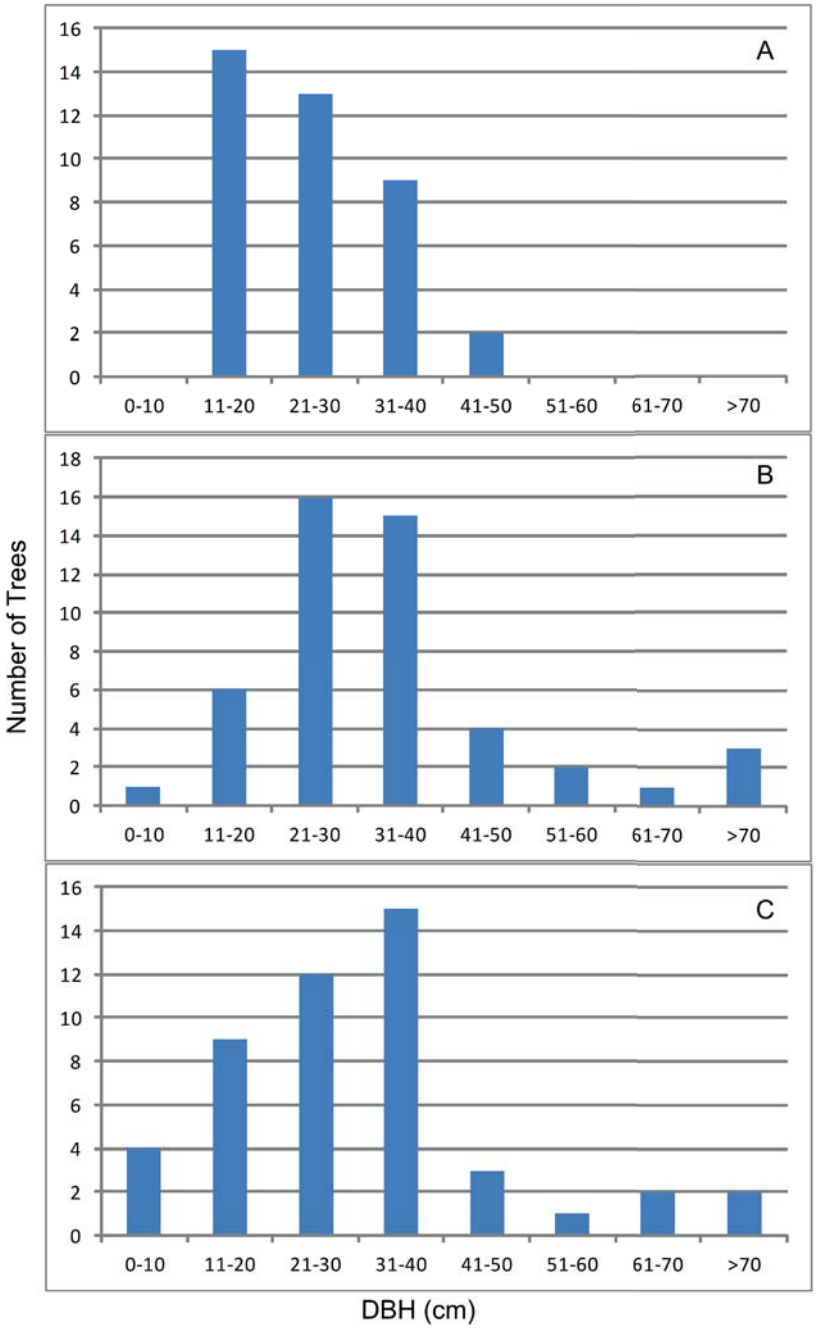
[a] significance after 10,000 permutations indicated by bolded italicized numbers



Supplement 26 Frequency distribution of the diameter at breast height (DBH) for the sampled populations of *Q. ellipsoidalis*: (A) CNF-QE, (B) FC-C, (C) FC-E, and (D) NNF-QE (the population abbreviations follow those in Table 9)



Supplement 27 Frequency distribution of the diameter at breast height (DBH) for the sampled populations of unmanaged *Q. rubra*: (A) HMR-IH, (B) HMR-LP, (C) HMR-MI, and (D) PM-QR (the population abbreviations follow those in Table 9)



Supplement 28 Frequency distribution of the diameter at breast height (DBH) for the sampled populations of managed *Q. rubra*: (A) CNF-QR, (B) FC-A, and (C) FC-B (the population abbreviations follow those in Table 9)

Supplement 29 Influence of management activities and species on mean genetic variation

	Number of populations	A_R [a]	H_O	H_E	F_{IS}	F_{ST}
Q. rubra	4	12	0.693	0.798	0.132	0.022
Q. ellipsoidalis	4	12	0.668	0.773	0.135	0.033
p-value[c]	-	0.631	0.686	**0.027**	0.769	0.803
***Q. rubra* (managed)**[b]	4	12	0.684	0.792	0.137	0.023
***Q. rubra*(unmanaged)**[b]	4	11	0.674	0.772	0.127	0.042
Q. ellipsoidalis[b]	4	11	0.655	0.767	0.146	0.033
p-value[c]	-	0.906	0.655	**0.031**	0.824	0.635

[a] corrected for unequal sample sizes using the rarefaction index suggested by Petit et al. (1998)
[b] only analyzed at 14 markers
[c] significance after 5000 permutations indicated by bolded numbers
A_R = Allelic richness
H_O = observed heterozygosity
H_E = expected heterozygosity
F_{IS} = inbreeding coefficient
F_{ST} = genetic differentiation

Supplement 30 HWE Exact Tests (Weir and Cockerham, 1984) by locus in each population. Significant values in boldface type are based on a sequential Bonferroni correction (Rice 1989) applied to the original significance level of α= 0.05

1P10			
Population	**F**	**p-value**	**Standard Error**
CNF-QE	-0.0293	0.6733	0.0021
FC-C	0.057	0.3137	0.0037
FC-E	0.0655	0.1718	0.0021
NNF-QE	0.0273	0.6154	0.0042
CNF-QR	0.0119	0.345	0.0054
FC-A	0.0383	0.6568	0.0024
FC-B	-0.0521	0.6463	0.0044
NNF-QR	0.0119	0.4601	0.0038
HMR-IH	0.1068	0.2362	0.0056
HMR-LP	0.0721	0.5368	0.0055
HMR-MI	0.1726	0.0316	0.0016
PM-QR	0.0479	0.1828	0.0025
2P24			
CNF-QE	-0.0245	0.8689	0.0018
FC-C	0.2667	**0.0018**	0.0002
FC-E	0.2052	0.1784	0.0016
NNF-QE	0.1294	0.0148	0.0007
CNF-QR	-0.0893	0.0948	0.002
FC-A	0.2532	**0.0019**	0.0001
FC-B	0.1684	0.0082	0.0004
NNF-QR	-0.046	0.4739	0.0027
HMR-IH	0.1262	0.0591	0.001
HMR-LP	0.1782	0.0367	0.0009
HMR-MI	0.0492	0.5029	0.0024
PM-QR	0.0099	0.8607	0.0015
3A05			
CNF-QE	0.685	**0**	0
FC-C	0.3132	0.0108	0.0003
FC-E	0.6485	**0**	0
NNF-QE	0.524	**0**	0
CNF-QR	0.1421	0.2901	0.0024
FC-A	0.3223	**0.002**	0.0002
FC-B	0.2427	0.0315	0.0005
NNF-QR	0.1172	0.1457	0.0017
HMR-IH	0.1442	0.2085	0.0025
HMR-LP	0.2287	0.0081	0.0003
HMR-MI	-0.0156	0.7506	0.0018
PM-QR	0.2037	**0.0015**	0.0001

*PM-QR only analyzed at 14 markers

S30 continued

3D15			
Population	**F**	**p-value**	**Standard Error**
CNF-QE	0.0355	0.3612	0.0018
FC-C	-0.1315	0.2782	0.0013
FC-E	0.1678	0.1132	0.0019
NNF-QE	0.0072	0.0534	0.0008
CNF-QR	0.0174	0.794	0.0021
FC-A	0.1623	0.1106	0.0017
FC-B	0.1671	0.0241	0.001
NNF-QR	-0.0969	0.848	0.0026
HMR-IH	0.0649	0.4009	0.0043
HMR-LP	0.2214	**0.0006**	0.0001
HMR-MI	0.1316	0.5041	0.0027
PM-QR	0.1148	0.0122	0.0005
FIR004			
CNF-QE	0.0933	0.1802	0.0061
FC-C	0.2234	**0**	0
FC-E	0.3413	**0**	0
NNF-QE	0.2117	**0**	0
CNF-QR	0.0651	0.0667	0.0029
FC-A	0.4115	**0**	0
FC-B	0.399	**0**	0
NNF-QR	0.1889	0.0053	0.0007
HMR-IH	0.2381	**0**	0
HMR-LP	0.156	0.02	0.0014
HMR-MI	0.2975	**0**	0
PM-QR	0.1543	0.0142	0.0009
FIR048			
CNF-QE	0.0933	0.1802	0.0061
FC-C	0.2234	**0**	0
FC-E	0.3413	**0**	0
NNF-QE	0.2117	**0**	0
CNF-QR	0.0651	0.0667	0.0029
FC-A	0.4115	**0**	0
FC-B	0.399	**0**	0
NNF-QR	0.1889	0.0053	0.0007
HMR-IH	0.2381	**0**	0
HMR-LP	0.156	0.02	0.0014
HMR-MI	0.2975	**0**	0
PM-QR	0.1543	0.0142	0.0009

S30 continued

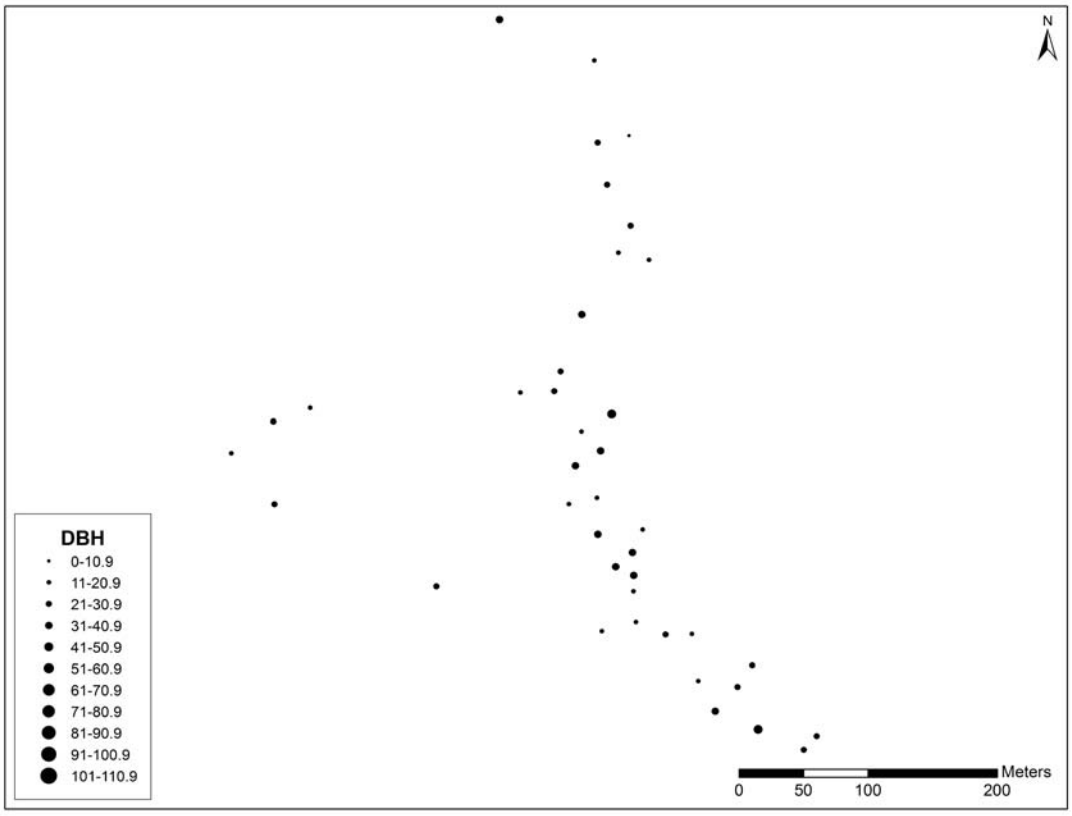
Population	F	p-value	Standard Error
GOT004			
CNF-QE	0.5794	**0**	0
FC-C	0.2888	**0**	0
FC-E	0.7999	**0**	0
NNF-QE	0.4161	**0**	0
CNF-QR	0.0146	0.3676	0.0027
FC-A	0.2181	**0.0027**	0.0002
FC-B	0.2225	0.0072	0.0002
NNF-QR	0.2542	0.0381	0.0005
HMR-IH	-0.0438	0.6678	0.0033
HMR-LP	0.0781	0.0125	0.0007
HMR-MI	0.1163	0.0332	0.0012
PM-QR	0.3233	**0**	0
GOT009			
CNF-QE	0.1246	0.0245	0.0007
FC-C	0.0891	0.5319	0.0023
FC-E	-0.0023	0.8331	0.0011
NNF-QE	0.013	0.1894	0.0024
CNF-QR	-0.0744	0.9228	0.0008
FC-A	-0.0334	0.638	0.002
FC-B	-0.0476	0.4586	0.0025
NNF-QR	0.116	0.1776	0.0018
HMR-IH	-0.0663	0.9394	0.0013
HMR-LP	0.0402	0.1456	0.0018
HMR-MI	0.0575	0.5825	0.003
PM-QR	0.0127	0.7071	0.0037
GOT021			
CNF-QE	-	-	-
FC-C	-0.0051	1	0
FC-E	0.3096	0.0657	0.0003
NNF-QE	0.7935	**0.0005**	0
CNF-QR	-0.1304	1	0
FC-A	-0.2577	0.1193	0.0006
FC-B	-0.0383	1	0
NNF-QR	-0.0833	1	0
HMR-IH	-0.1441	0.2804	0.0003
HMR-LP	-0.0851	1	0
HMR-MI	-0.2165	0.1852	0.0002
PM-QR	-0.1088	1	0

S30 continued

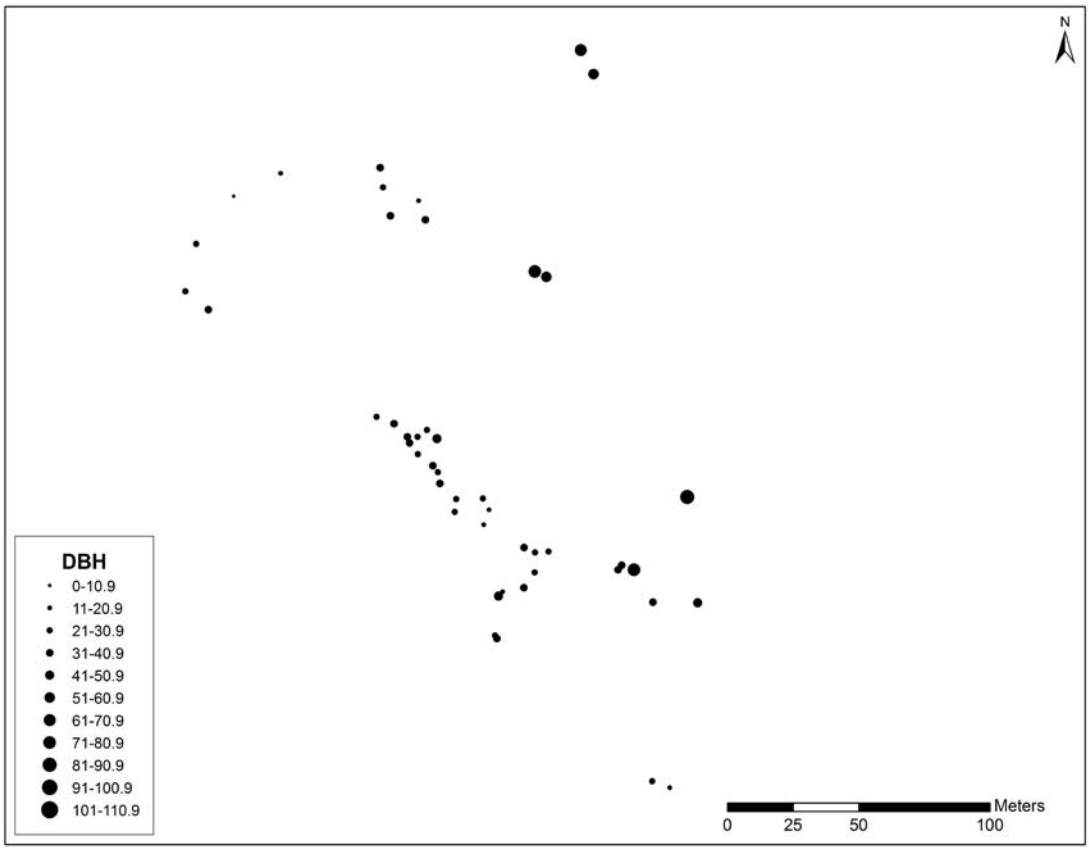
Population	F	p-value	Standard Error
PIE040			
CNF-QE	-0.1487	0.7425	0.0016
FC-C	-0.0566	0.2576	0.0012
FC-E	0.1362	0.7608	0.0014
NNF-QE	-0.0893	0.1874	0.0016
CNF-QR	-0.1599	0.5963	0.0018
FC-A	0.137	0.1685	0.0028
FC-B	0.1693	0.1671	0.002
NNF-QR	0.1866	0.0213	0.0002
HMR-IH	-0.016	0.901	0.0009
HMR-LP	0.0808	0.1259	0.0018
HMR-MI	0.0709	0.0961	0.0017
PM-QR	0.4072	**0.0012**	0.0001
PIE099			
CNF-QE	0.0152	0.2394	0.0036
FC-C	0	0.2692	0.0025
FC-E	0.1106	0.1189	0.0011
NNF-QE	0.0031	0.1619	0.0028
CNF-QR	0.448	**0**	0
FC-A	0.4178	**0**	0
FC-B	0.3842	**0**	0
NNF-QR	0.1067	0.1477	0.0022
HMR-IH	0.3428	**0**	0
HMR-LP	0.3786	**0**	0
HMR-MI	0.3211	**0**	0
PM-QR	0.2868	0.0178	0.001
QpZAG15			
CNF-QE	0.0268	0.1	0.003
FC-C	0.0005	0.0583	0.0025
FC-E	0.0229	0.4789	0.0032
NNF-QE	-0.0438	0.6131	0.0051
CNF-QR	0.0688	0.6001	0.0039
FC-A	0.1341	0.3855	0.0025
FC-B	0.077	0.5208	0.0032
NNF-QR	0.0084	0.4084	0.0049
HMR-IH	-0.0173	0.8127	0.003
HMR-LP	0.1066	0.5929	0.0034
HMR-MI	0.0162	0.8167	0.0031
PM-QR*	-	-	-

S30 continued

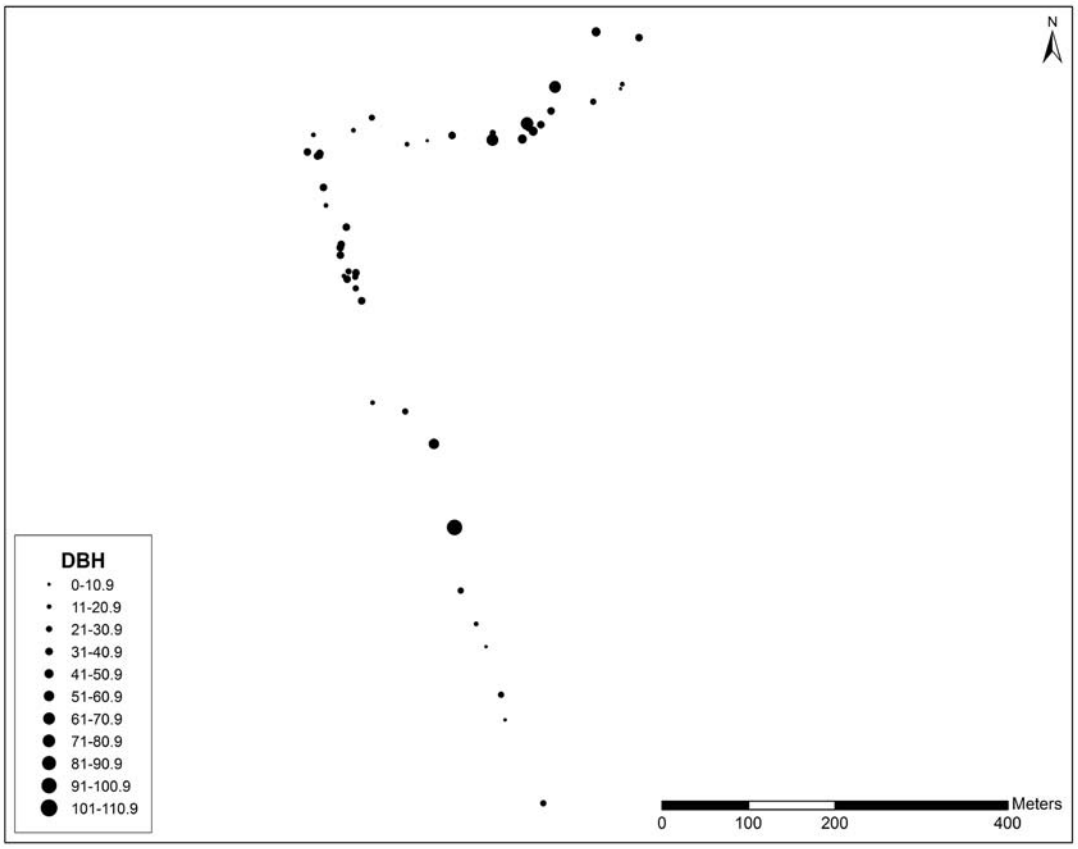
quru-GA-0C11			
Population	**F**	**p-value**	**Standard Error**
CNF-QE	0.1789	0.0209	0.0011
FC-C	-0.0861	0.3828	0.0036
FC-E	0.1147	**0.0033**	0.0002
NNF-QE	0.1169	**0.0031**	0.0005
CNF-QR	0.0134	0.3034	0.0026
FC-A	0.0542	0.0685	0.0021
FC-B	0.0329	0.2865	0.0034
NNF-QR	0.0626	0.5419	0.0033
HMR-IH	0.0392	0.2662	0.0025
HMR-LP	0.0011	0.8259	0.0029
HMR-MI	0.1558	0.0054	0.0003
PM-QR	-0.0514	0.9569	0.0008
quru-GA-0E09			
CNF-QE	-0.0252	0.2771	0.0076
FC-C	0.2512	**0**	0
FC-E	0.0983	0.3233	0.0065
NNF-QE	-0.0011	0.3164	0.009
CNF-QR	0.2092	**0**	0
FC-A	0.2231	**0**	0
FC-B	0.3341	**0**	0
NNF-QR	0.4511	**0**	0
HMR-IH	0.2641	**0**	0
HMR-LP	0.4	**0**	0
HMR-MI	0.2315	**0**	0
PM-QR	0.2746	**0.0012**	0.0003
quru-GA-1F07			
CNF-QE	-0.0106	0.3181	0.0057
FC-C	0.1954	0.0056	0.0006
FC-E	0.3189	**0**	0
NNF-QE	0.0189	0.1522	0.0044
CNF-QR	0.0599	0.5818	0.0052
FC-A	0.2088	**0.0005**	0.0002
FC-B	0.2225	**0.0019**	0.0005
NNF-QR	0.0874	0.6477	0.0054
HMR-IH	0.2103	**0**	0
HMR-LP	0.1567	0.0446	0.003
HMR-MI	0.121	0.0383	0.0023
PM-QR	0.0818	0.5103	0.0073



Supplement 31 Spatial and size distribution of managed *Q. rubra* population CNF-QR (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals

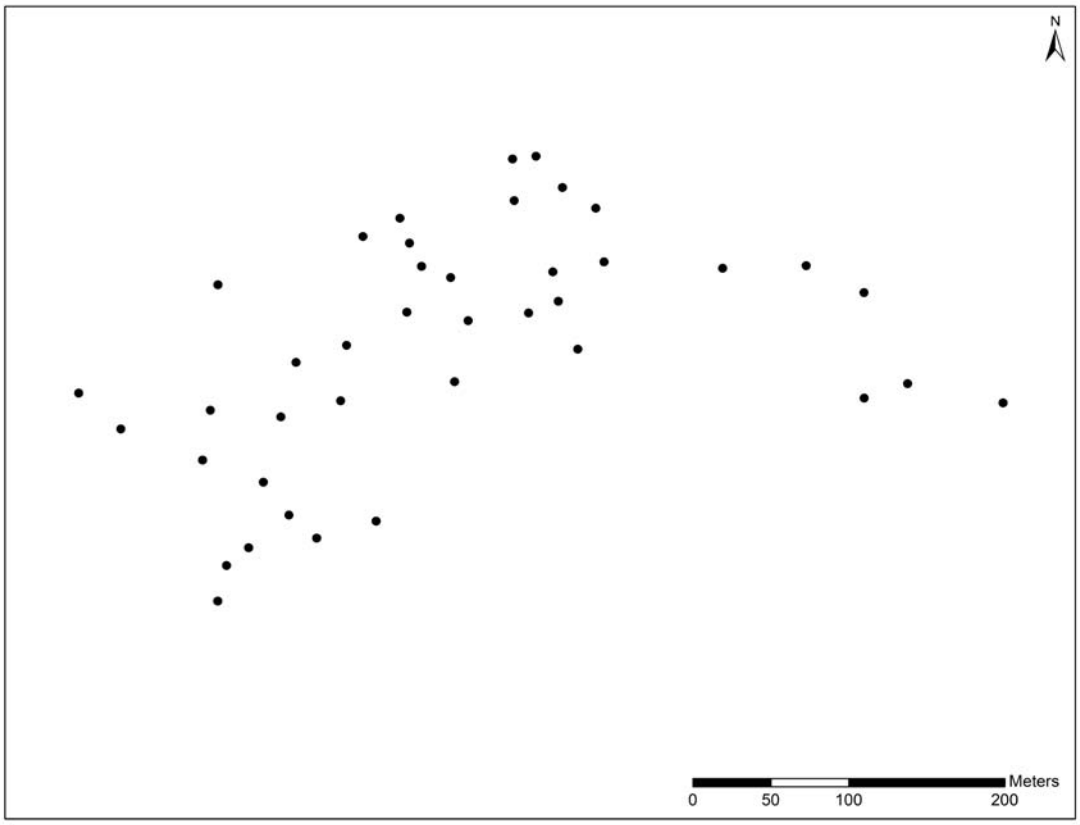


Supplement 32 Spatial and size distribution of managed *Q. rubra* population FC-A (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals

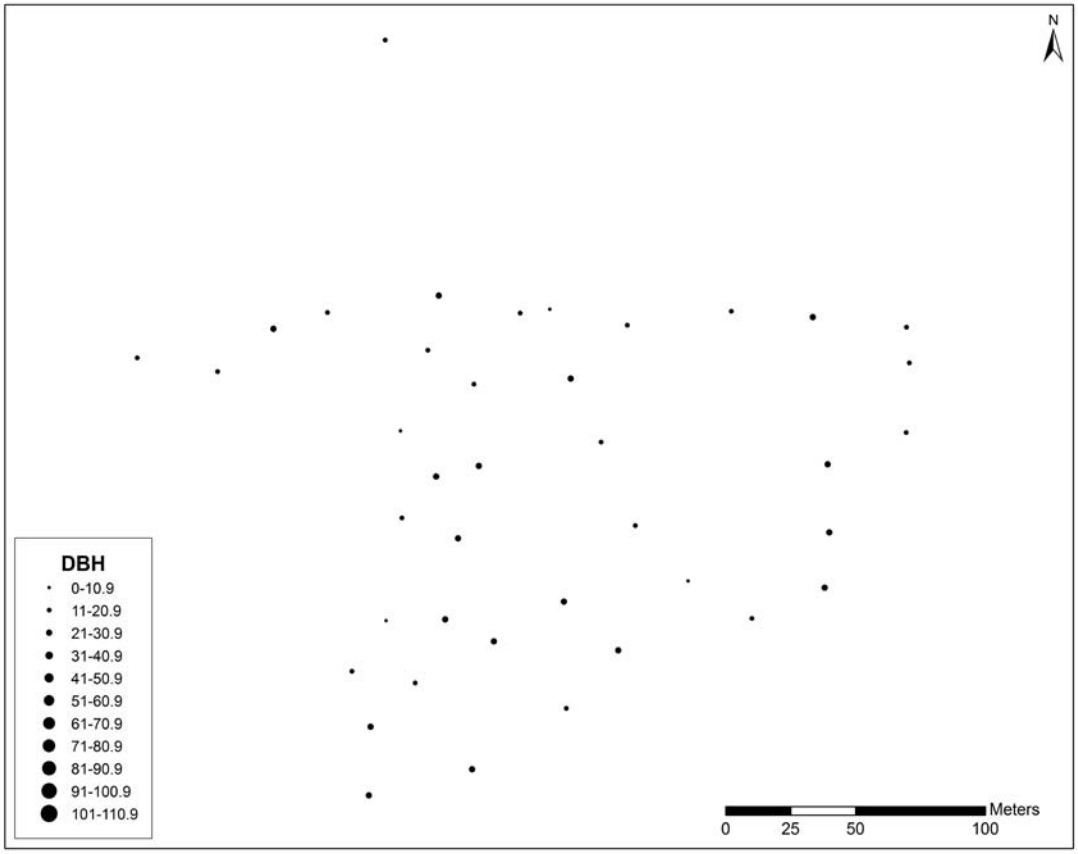


Supplement 33 Spatial and size distribution of managed *Q. rubra* population FC-B (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals

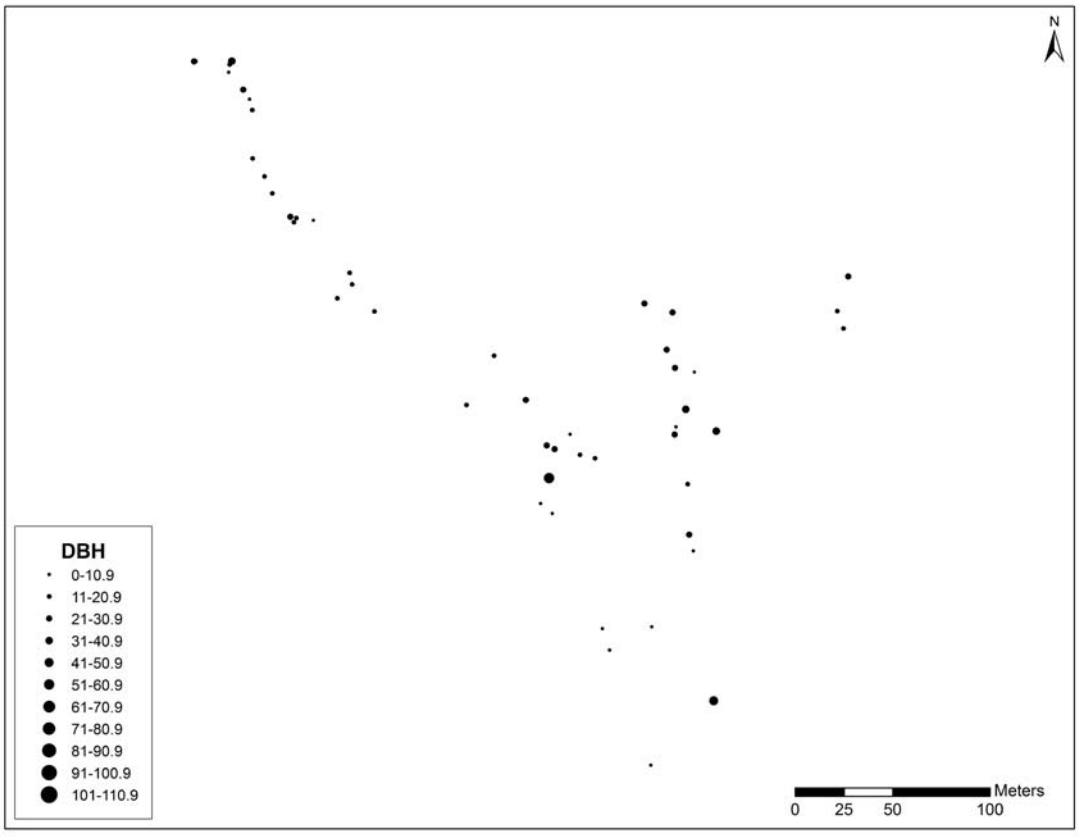
Special Note: In addition to Supplements 31-33 discussed in the text, Supplements 34-42 are also included to provide access to a detailed look at the spatial and size class distribution of all individuals in each population.



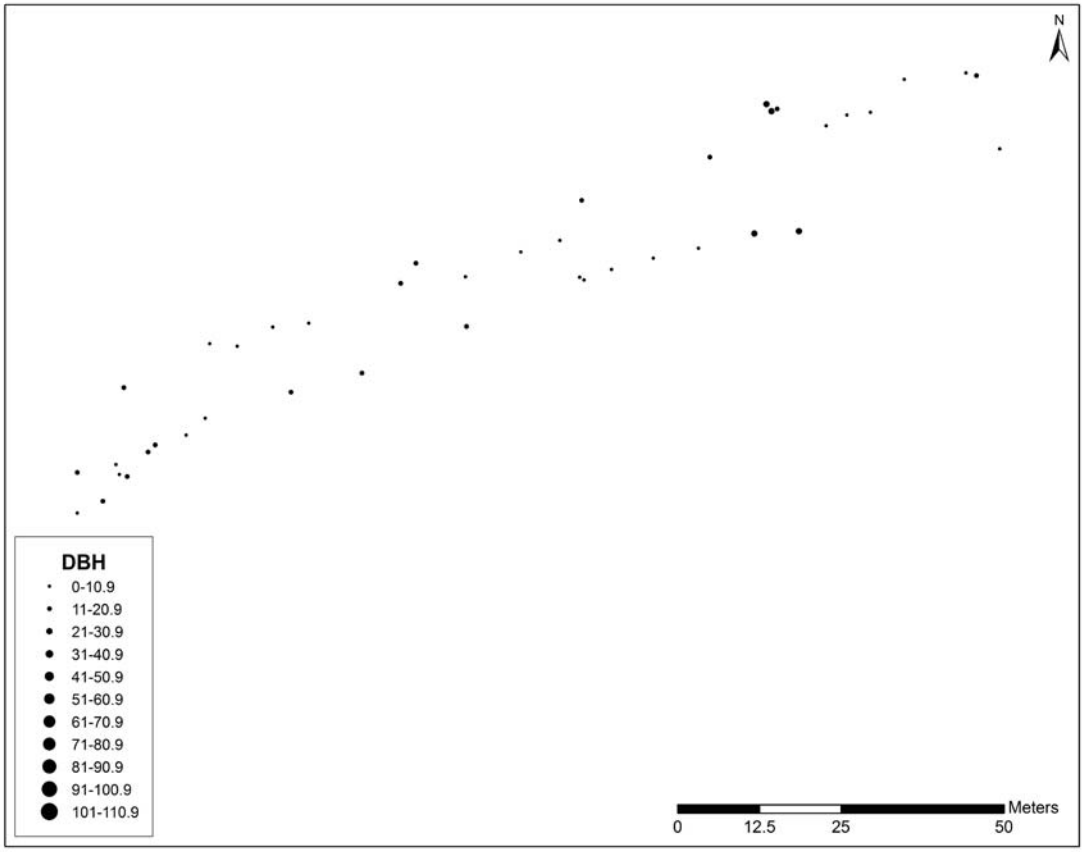
Supplement 34 Spatial distribution of managed *Q. rubra* population NNF-QR (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals (size data unavailable)



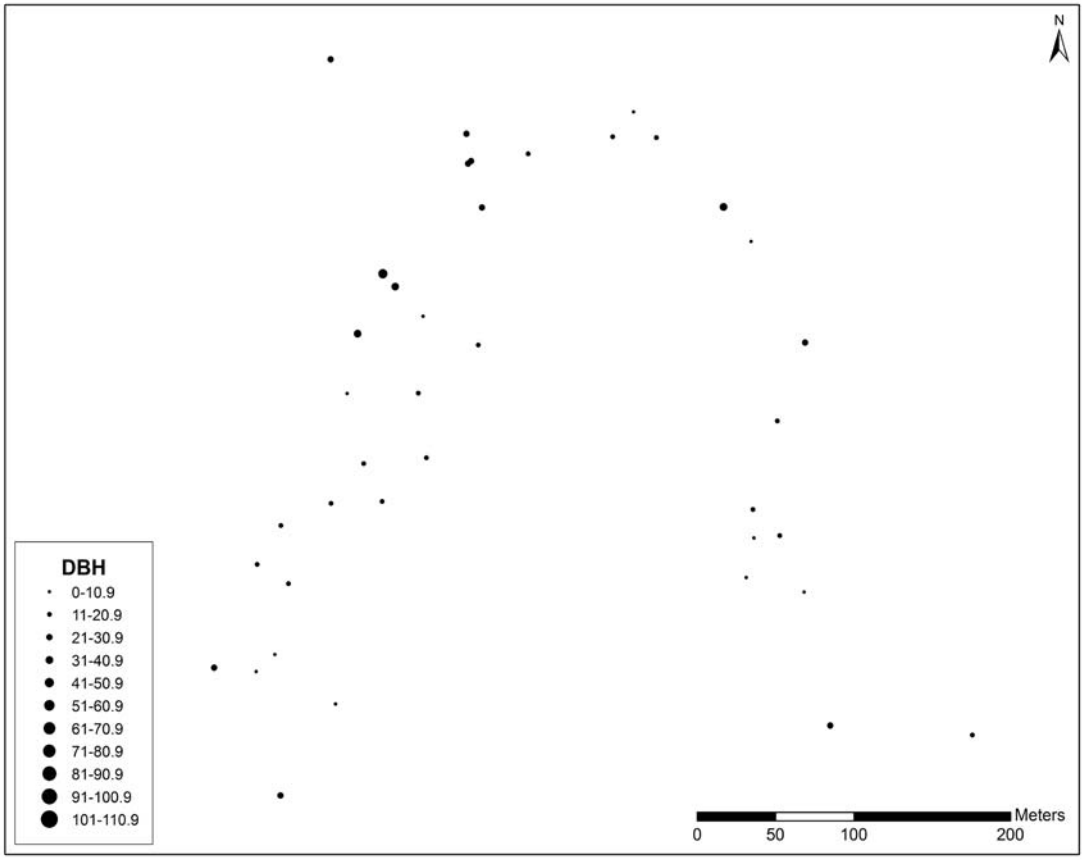
Supplement 35 Spatial and size distribution of *Q. ellipsoidalis* population CNF-QE (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



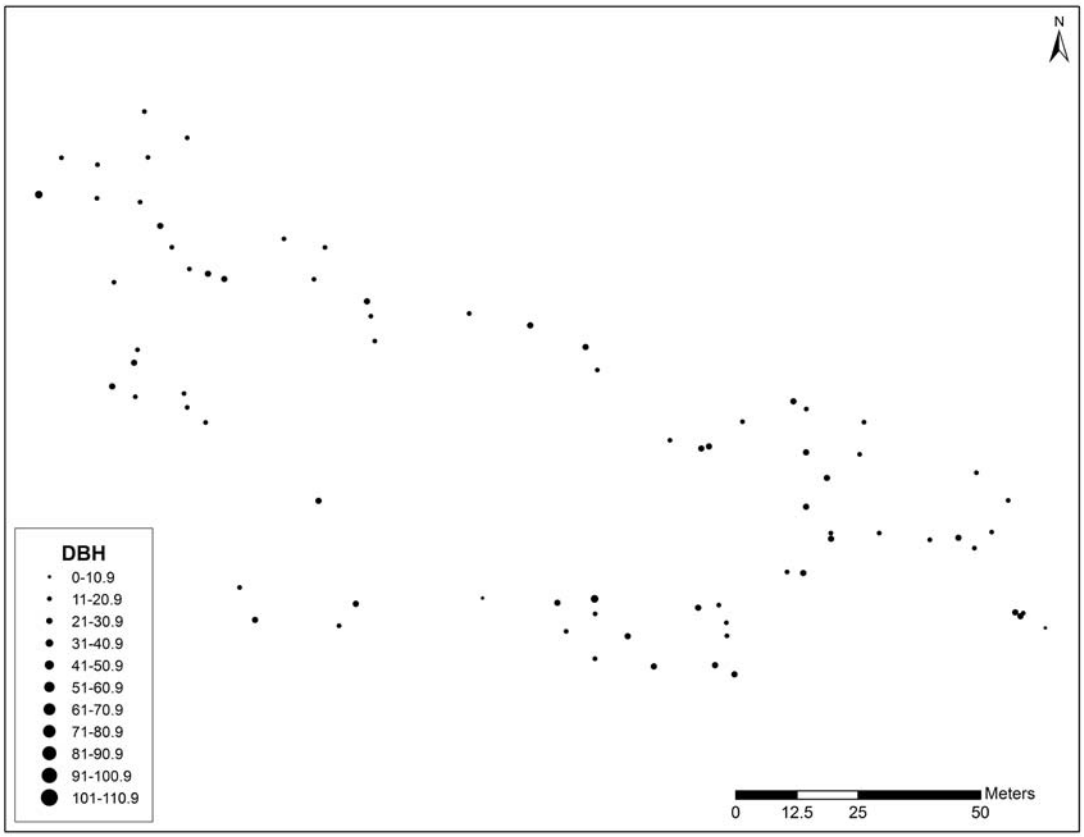
Supplement 36 Spatial and size distribution of *Q. ellipsoidalis* population FC-C (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



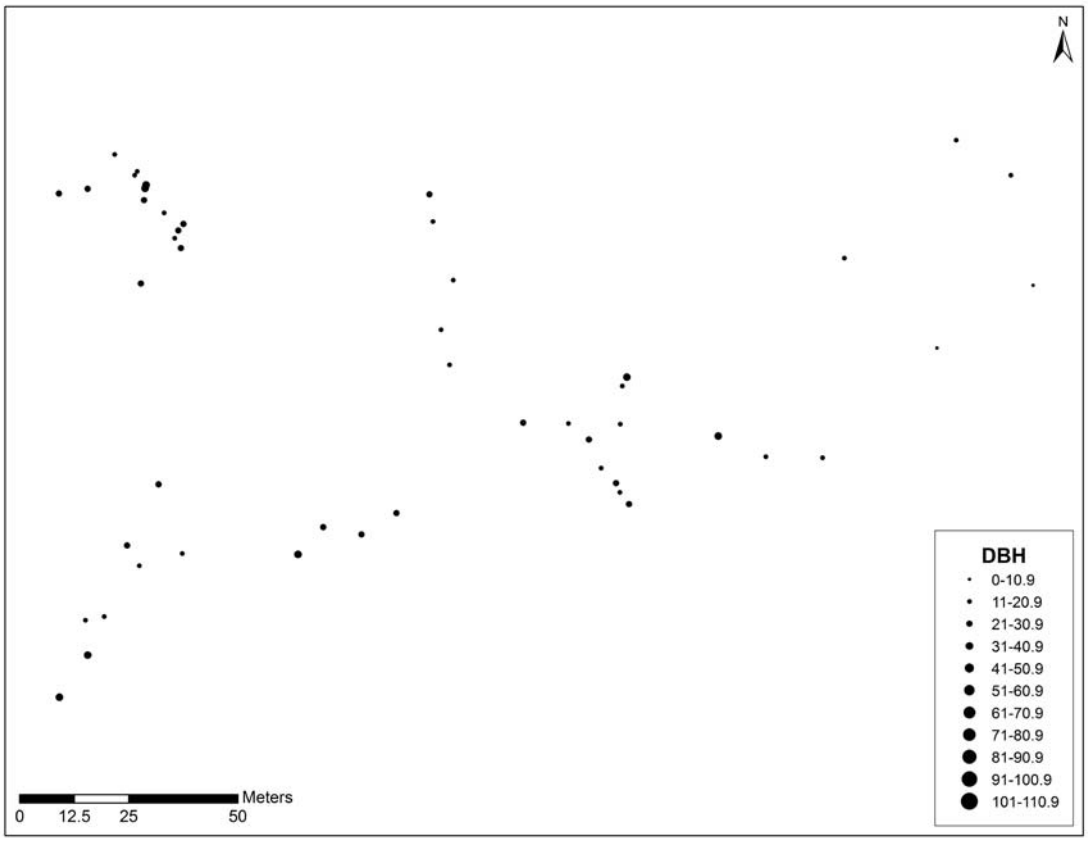
Supplement 37 Spatial and size distribution of *Q. ellipsoidalis* population FC-E (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



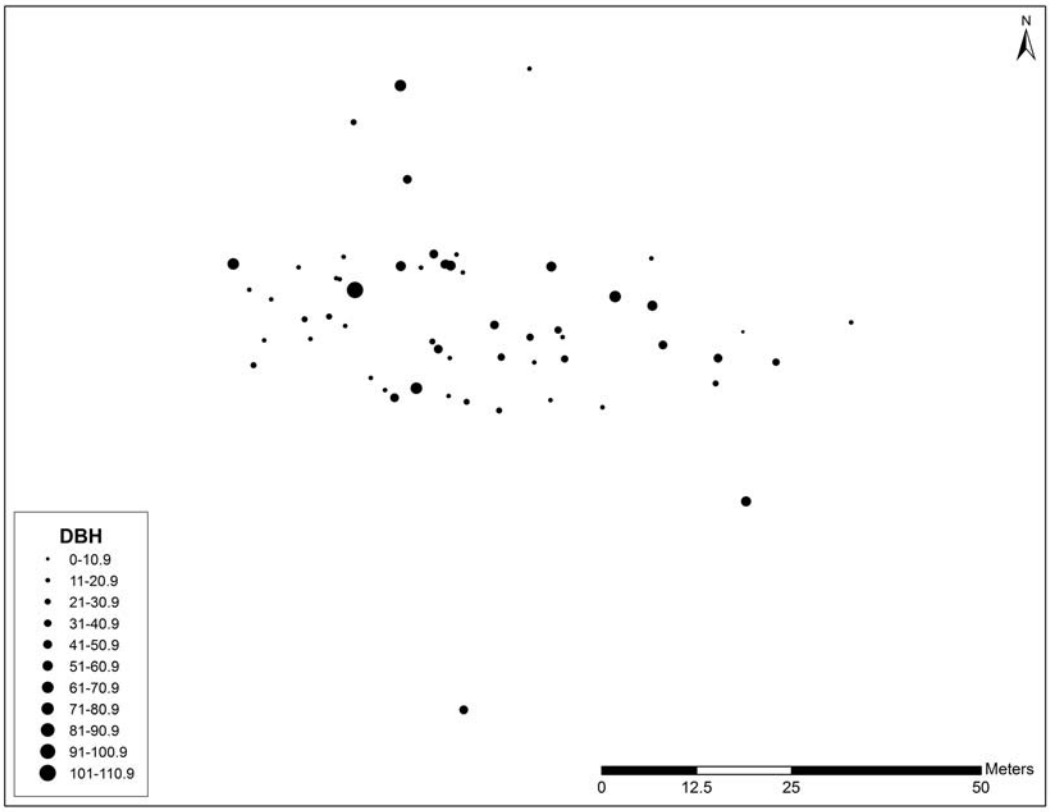
Supplement 38 Spatial and size distribution of *Q. ellipsoidalis* population NNF-QE (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



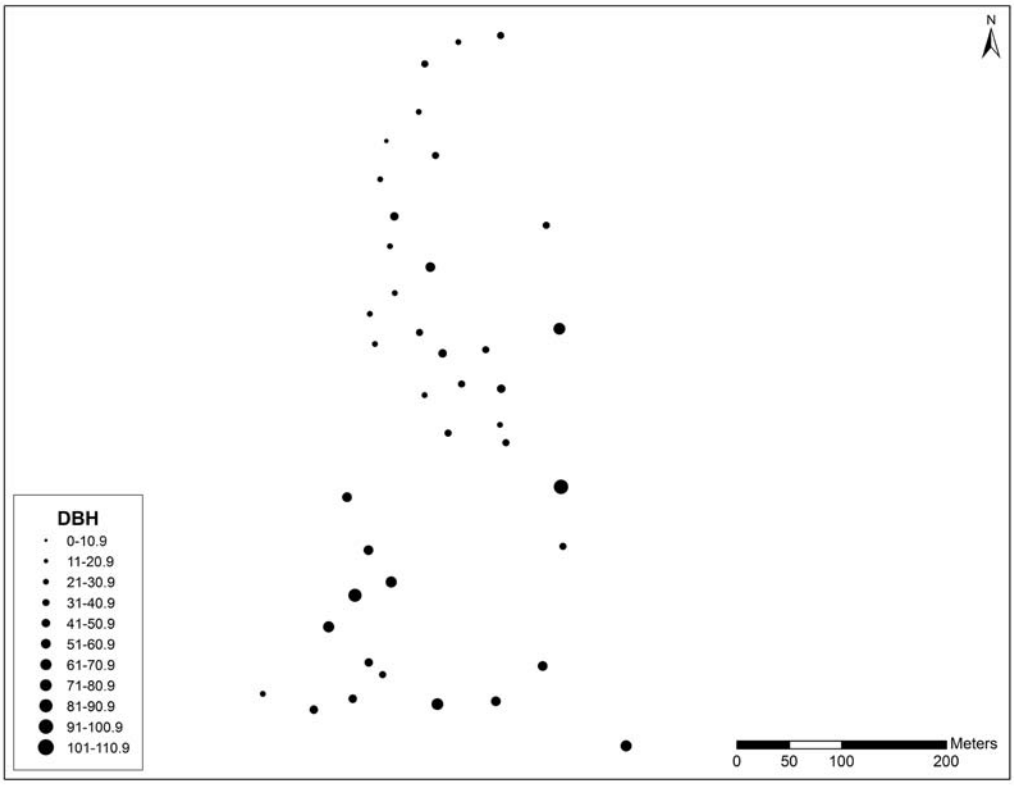
Supplement 39 Spatial and size distribution of unmanaged *Q. rubra* population HMR-IH (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



Supplement 40 Spatial and size distribution of unmanaged *Q. rubra* population HMR-LP (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



Supplement 41 Spatial and size distribution of unmanaged *Q. rubra* population HMR-MI (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals



Supplement 42 Spatial and size distribution of unmanaged *Q. rubra* population PM-QR (the population abbreviations follow those in Table 9). The size of the symbols is related to diameter at breast height (DBH) of studied individuals

License Agreement

This is a License Agreement between Jennifer F Lind-Riehl ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 3521980990321
License date Dec 04, 2014
Licensed content publisher Springer
Licensed content publication Plant Systematics and Evolution
Licensed content title Fine-scale spatial genetic structure of two red oak species, Quercus rubra and Quercus ellipsoidalis
Licensed content author Jennifer Lind-Riehl
Licensed content date Jan 1, 2014
Type of Use Thesis/Dissertation
Portion Full text
Number of copies 1
Author of this Springer article Yes and you are the sole author of the new work
Order reference number None
Title of your thesis /dissertation GENETIC VARIATION, LOCAL ADAPTATION AND POPULATION STRUCTURE IN NORTH AMERICAN RED OAK SPECIES, QUERCUS RUBRA L. AND Q. ELLIPSOIDALIS E. J. HILL
Expected completion date Jan 2015
Estimated size (pages) 220
Total 0.00 USD

Terms and Conditions

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reprint in your thesis material on which Springer Science Rightslink and Business Media control the copyright, permission is granted,

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).OR:
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.