PHENOTYPIC AND MOLECULAR INSIGHT INTO GENETIC DIFFERENTIATION, INTROGRESSION AND SELECTION IN *QUERCUS RUBRA* AT A FINE SPATIAL SCALE

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA

BY

Maria Jose Gomez Quijano

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Briana L. Gross , Julie R. Etterson

July 2021

Acknowledgements

I would like to thank my advisors, Briana Gross and Julie Etterson, for the immense and unconditional support during experimental design, data collection, data analysis and development of my thesis. I would also like to acknowledge the support received by my committee members, Jessica Savage and Amanda Grusz, for their valuable suggestions. I also want to recognize Frank Maragi who provided invaluable assistance with construction of supplies and Matt Jahnke for the advice on oak seedling growth and significant assistance in the greenhouse. I would like to recognize the assistance from the Gross Lab members, Sophie Leaf for assistance with leaf collection and DNA extraction, Michael Laidlaw, Samantha Ykema for the valuable assistance with acorn collections, greenhouse measurements and data entry and all the undergraduate, graduate and lab members who assisted with this project.

I would like to acknowledge my funding agencies: the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Department of Commerce provided to the Minnesota Department of Natural Resources (DNR) for Minnesota's Lake Superior Coastal Program (Project Number 18-306-10H; NOAA Award NA18NOS4190081).

Lastly, I would like to thank the University of Minnesota Duluth Biology Department for providing funding through Teaching and Research Assistantships. And the University of Minnesota Duluth as a whole institution for the unconditional support I received as an international student. I would also like to acknowledge all my family and friends who have assisted me and encouraged me throughout this process.

Table of Contents

List of Tables

List of Figures

Abstract

The massive scale and cold temperature of Lake Superior creates unique microclimates in coastal terrestrial environments resulting in cooler summers, an extended fall season, warmer winters, and a reduced risk of spring frost. This gives rise to a steep climate gradient from coastal to inland regions that could lead to genetic differentiation among populations. To test this hypothesis, we studied Northern red oak (*Quercus rubra* L.) to examine phenotypic and molecular differentiation among populations ranging from 1–160 km from the lake shore. In a common garden experiment, we found 30% of germination and juvenile traits differed significantly from expectation. We also used restriction site associated DNA sequencing (RAD-seq) to examine population structure and genomic signatures of selection in these populations. Our results suggest that, in contrast to quantitative traits, *Q. rubra* populations are not differentiated at neutral genetic markers according to their distance from Lake Superior. However, unexpectedly, we also found evidence of increasing levels of introgression from the closely related species *Quercus ellipsoidalis* E.J. Hill into *Q. rubr*a as species overlap and population distance from the lake increased. Our scan for selection and environmental association analysis identified one outlier locus in common, and this locus is associated with the precipitation of the wettest month. Overall, despite the lack of molecular population structure, the common garden experiment revealed that *Q. rubra* populations differ for key phenotypic traits. This, in combination with the genomic scans for selection, suggests the influence of natural selection driven by climate heterogeneity with increasing distance from the lake. Moreover, this is the first study that has jointly leveraged quantitative and molecular genetics to dissect signatures of selection in *Q. rubra* across a fine geographical scale.

Introduction

Within a single species, genetic differentiation can occur among populations as a consequence of stochastic or adaptive evolutionary processes (Joshi et al., 2001). These differences can often be identified at both the molecular and phenotypic level (De Kort et al., 2014; de Villemereuil et al., 2016; Dempewolf et al., 2015; Ramírez-Valiente et al., 2018; Sork et al., 2013). Ideally, combining molecular and quantitative techniques can provide high inferential power and evidence of the potential for plant species to adapt to their local environment (Berlin et al., 2014; Räsänen & Hendry, 2008). The combination of genomic and quantitative techniques can unravel the confounding effect of hidden population structure, and the combined effects of genetic drift, gene flow, natural selection, and phenotypic plasticity. Molecular markers can be used to identify population structure, to allow the identification of gene flow and selection, while the confounding effect of phenotypic plasticity can be controlled using common garden experiments (de Villemereuil et al., 2016; Sork et al., 2013). Moreover, a comprehensive understanding of genetic variation and adaptive response to a changing climate can be increased when the integration of molecular and quantitative approaches includes environmental data that expose associations between the environment and the observed genetic variation (De Kort et al., 2014; Lepais & Bacles, 2014). Therefore, genomic techniques and spatial genetic patterns, paired with common garden experiments that investigate variation based on quantitative traits, can a be powerful tools to understand the effects of evolutionary forces and improve species management under a changing environment (Gugger et al., 2021; Hoffmann et al., 2015; Murray et al., 2019; Sork et al., 2013).

Phenotypic differentiation: At the quantitative trait level, population differentiation can be detected in common garden experiments where differences could be due to either restricted gene flow, hybridization, and/or natural selection (Kirk & Freeland, 2011). Reciprocal transplant experiments are the most powerful technique for identifying instances of local adaptation to the environment and the traits that contribute to that adaptation. However, phenotypic variation that parallels climate gradients also bolsters

inferences about adaptation (Lascoux et al., 2016; Savolainen et al., 2013). In natural populations, environmental heterogeneity can lead to divergent selection that favors the emergence of specialized organisms and local adaptation (Papaïx et al., 2013).

Raising multiple populations in a common environment and measuring quantitative traits is a valuable approach to detangle the pervasive effects of evolutionary and environmental phenomena. If populations are significantly different when environmental effects are equalized, the phenotypic differences are inferred to be genetically based (de Villemereuil et al., 2016). This classic experimental design has been used to discover population differentiation in a broad range of organisms from plants (Browne et al., 2019; Colautti et al., 2010; Etterson et al., 2016) to vertebrates (Bassar et al., 2010; Conover & Baumann, 2009; Harvey et al., 2016; Sorci et al., 1996; While et al., 2015). These studies have identified phenotypic differences that reflect locally adapted populations, or populations with the potential for local adaptation (Ramírez-Valiente et al., 2018, 2019; Sork et al., 1993). However, even when plants are grown in a controlled environment, transgenerational plasticity, where the offspring's phenotypic response is based on the parental growth environment, can complicate the powerful common garden experimental approach (Donelson et al., 2018). Although transgenerational plasticity can be adaptive (Galloway & Etterson, 2007), it is typically considered to be "noise" in genetic experiments that can give a false impression of genetic differences among populations when none exist (Donelson et al., 2018). Transgenerational plasticity can be accounted for through a refresher generation or a multiple year growing experiment. When neither of these approaches is possible, traits that are sensitive to the maternal environment, such as seed size, can be used as a covariate to control for environmental carryover effects (de Villemereuil et al., 2016; Yang et al., 2015).

Molecular differentiation: At the molecular level, allele frequencies differ across geographical space because of variation in gene flow patterns and selective pressures (Murray et al., 2019). Restricted gene flow and genetic drift between populations can lead to differentiation and the formation of genetic clusters (Greenbaum et al., 2016). Moreover, gene flow between species can introduce new variants into populations,

affecting both genetic variation and population structure. Hybridization can impact the process of adaptation to a local environment and the effectiveness of selection, by providing sources of new genetic material that can aid with adaptation and species radiation (Hipp et al., 2020; Mitchell et al., 2019; Moran et al., 2012; Rieseberg and Willis, 2007). Simultaneously, natural selection can lead to genomic regions or individual loci exhibiting reduced levels of diversity, an altered allele frequency spectrum, or significant associations with environmental factors (Excoffier et al., 2009; Frichot et al., 2013; Gugger et al., 2021; Savolainen et al., 2013).

The power to detect population differences and genomic regions under selection has exploded in recent decades, even for non-model systems. High-throughput next generation sequencing technologies such as RAD-seq (restriction associated DNA sequencing) allow for genotyping and identification of SNPs for any species and do not require prior genomic information for the populations that are being studied (Andrews et al., 2016; Cariou et al., 2013). RAD loci can occur in both coding or non-coding regions of the genome, and loci are often conserved between organisms of closely related species (Andrews et al., 2016). The generation of RAD loci allows for the calculation of genomic diversity indexes, the detection of hybridization or introgression between species, and the identification of the signature of selection (Savolainen et al., 2013).

The combination of molecular and phenotypic data: In natural populations, variation in quantitative traits across an environmental gradient are often associated with ecological factors that reflect selective pressures on individual phenotypes (Anderson et al., 2014), and molecular markers can be used to pinpoint loci across the genome that underlie these adaptive changes. For example, forest tree species with large population sizes frequently occupy highly heterogeneous environments, resulting in high genetic variation given the large range of pollen dispersal that can spread adaptive traits across the occupied landscape. Despite this, strong population differentiation for key adaptive traits is frequently documented, and selection for local phenotypes appears to be strong (Ramirez-Valiente et al., 2018 and 2019; Sork et al., 2016; Sork et al., 2013). However, caution must be used when interpreting some measures of selection, because strong neutral molecular population genetic structure that has arisen through neutral processes, like

genetic drift, interspecific gene flow, and mutation, can mimic patterns expected under non-neutral processes that are assessed through either F_{ST} or environmental association analyses (Excoffier et al., 2009; Rellstab et al., 2016). Therefore, when looking for molecular signals of local adaptation associated with environmental factors, the effect of population structure as well as the evaluation of phenotypic variation should be considered to avoid a high rate of false positives (Rellstab et al., 2015; de Villemereuil et al., 2016; Frichot et al., 2013).

Differentiation in Quercus:

There is substantial evidence from the literature that species of oak (*Quercus* L.) can adapt to different environments and potentially to rapid climate change (Du et al., 2020; Gugger et al., 2016; Rellstab et al., 2016; Sork, 2016; Sork et al., 1993). Recently, genomic studies at broad scales (>200 km) have revealed patterns of local adaptation of European white oaks (Rellstab et al., 2016), North American valley oaks (Gugger et al., 2021; Sork, Squire, et al., 2016) and evergreen oaks in the Qinghai Tibetan Plateau (Du et al., 2020). However, patterns of gene flow and selection that can lead to population structure and adaptation at a very fine spatial scale (<200 km) is relatively unexplored, even though some oak species occupy steep environmental gradients.

Quercus rubra L. has an expansive range in North America, and one small part of this range occurs across a sharp environmental gradient along the north shore of Lake Superior. Lake Superior has a significant effect on climate patterns and allows for an extended growing season by producing warmer falls and winters, reduced frost in the spring, and cooler summers by the shore (Moen, 2018). Along the north shore of Lake Superior, four different hardiness zones (3a–4b) can be found within a small geographical space (<50 km). Regional microclimates produced along the Lake Superior shoreline could contribute to the formation of locally adapted populations, given the diversity of microclimates that occur at different distances from the lake. Similar conditions and examples have been observed in various habitats close to large bodies of water, where plant community variation and biodiversity differentiation are found at different distances

from the shore, due to the unique environmental conditions (Chen & Schemske, 2015; Nevo, 1995; Vanwalleghem & Meentemeyer, 2009; Vallés et al., 2011).

Like many other oak species, *Q. rubra* is prone to interspecific hybridization (Moran et al., 2012; Sander, 1990). Northern red oak has been shown to hybridize with at least seven other species in the subgenus *Quercus* section *Lobatae* (Moran et al., 2012; Sander, 1990). One of the species known to hybridize with *Q. rubra* is the closely related species *Quercus ellipsoidalis* E.J. Hill. These two species' ranges overlap in the Great Lakes region (including Wisconsin, Michigan, and Minnesota). In regions east of Lake Superior, like the upper peninsula of Michigan and the Apostle Islands, several studies using morphometric traits and genetic markers have shown varying rates of hybridization between the two species (Gailing et al., 2012; Gailing & Zhang, 2018; Jensen et al., 1993; Lind & Gailing, 2013). These past studies have been limited to a low number of molecular markers and mostly focused on few candidate genes. However, in the past decade RAD-seq has revolutionized the study of selection, and interspecific hybridization, even in the absence of fully assembled genomes (Cavender-Bares et al., 2015; Cavender-Bares & Bazzaz, 2000; Eaton et al., 2015; Hipp et al., 2020), making oaks a new model system for understanding interspecific gene flow and adaptation. The state of Minnesota, located west of Lake Superior, is within the natural distribution of both *Q. rubra* and *Q. ellipsoidalis*, which could lead to interspecific gene flow in areas where their ranges overlap. However, introgression between these two species in this region has not yet been studied.

To examine the potential for local adaptation across a steep environmental gradient in Northern red oak populations west of Lake Superior, we characterized population structure and genetic divergence among *Q. rubra* populations located at varying distances from the lake, across a small geographical scale (160 km). We also identified signatures of selection based on genetic distance, as well as associations between genomic regions and environmental factors, using multiple Bayesian, distance-based approaches, and statistical models. Using this powerful combination of population genomics and quantitative genetic approaches, we interrogate the following questions:

1. *Is there population structure based on both molecular markers and phenotypic traits?* Here, populations would be strongly differentiated because of reproductive isolation based on divergent life-history traits. These genetic differences would be the product of natural selection on traits that have led to local adaptation (Jordan et al., 2005; Räsänen & Hendry, 2008).
2. *Is there population structure based on phenotypic traits but not on molecular markers?* Here, populations would have high gene flow, usually because of high dispersal range that is sufficient to overcome genetic drift. However, natural selection would be stronger than gene flow resulting in adaptation to a local environment despite the high gene flow (Petit & Hampe, 2006; Zanella et al., 2011).
3. *Is there population structure based on molecular markers but not on phenotypic traits?* Here, populations would have low or no gene flow, but are phenotypically similar due to convergent phenotypic evolution or divergent evolution with phenotypic maintenance. These evolutionary processes would lead to what is known as cryptic population structure, where distinct genetic lineages are phenotypically identical (Baker et al., 1995; Ribeiro et al., 2011).

In this study, we account for the challenges of transgenerational plasticity, hybridization between *Q. rubra* and *Q. ellipsoidalis*, and population structure. Here, we provide a comprehensive analysis that elucidates the genetic variation underlying potential local adaptation within a fine geographical scale with highly heterogeneous environments in the western edge of the Great Lakes region in the state of Minnesota.

Materials and Methods

Study system

Quercus (Fagaceae) is a widespread genus that has radiated across the northern hemisphere since its estimated origin 56 million years ago (Ma) (Kremer & Hipp, 2020). Their range expands from the equator to boreal regions at a latitude of 60ºN in Europe and 4000 m above sea level in China (Kremer & Hipp, 2020). Northern red oak (Quercus rubra) is a deciduous tree that is native to North America and is widely distributed through the eastern United States and southeastern Canada (Dey et al., 2007). The species inhabits a diverse range of hardiness zones (3–8) (USDA, 2019) and grows in acidic, loamy, moist, clay, and well-drained soils (Tirmenstein, 1991). The hardwood lumber produced from harvesting the species has high economic value, and the fruit (acorns) is the main source of food for many birds and mammals (Sander, 1990). Acorn production is a highly energetically costly process for oak trees—large acorn crops occur every 3–5 years and a single tree can produce up to 10,000 acorns in a year (Sander, 1990).

Population sampling and plant material

We collected acorns and leaf tissue from three regions that differed with respect to distance from Lake Superior: “Coastal” (0–16 km), “Inland” (17–80 km), and “Interior” (81–160 km) (Fig. 1, Supplementary table 1). We used Minnesota Department of Natural Resources Relevé data (MN-DNR, 2013) to locate 30 populations of *Q. rubra*. However, like many oak species, *Q. rubra* trees produce acorns every three to five years (Sander, 1990); therefore, while leaf tissue collection was possible at all sites, acorn collection was not.

Greenhouse common garden experiment: We collected acorns from ten populations (five coastal, four inland, one interior). Within each population, 10–15 trees were sampled, and 20–30 acorns were collected per tree (Fig. 1a, Supplementary Table 1). Because of low acorn availability in interior populations, we pooled the inland and interior populations and focused on identifying differences between “Coastal” (0–16 km) and “Noncoastal”(17–160 km) regions. To determine viability, we subjected the acorns to a float test (Gribko & Jones, 1995) and then stratified them for three months in a cold room at 4°C until planting.

Molecular marker study: We collected leaf tissue from 30 populations: ten each from coastal, inland, and interior regions. These 30 populations included the ten populations included in our greenhouse study, although collections were made at different times of the year and the individuals sampled from each population in these two phases of our study were not identical. Within each population, we sampled 15–18 individuals. To broaden the scope of our collection, we collected leaf tissue from 73 seedlings of *Q. rubra* that had been grown in the summer 2019 greenhouse common garden experiment. We obtained *Q. rubra* populations from the MN-DNR Badoura Nursery, originally sourced from two sites in southern Minnesota and four sites in northern Iowa. We also collected tissue from three other *Quercus* species (*Quercus imbricaria* Michx., *Quercus macrocarpa* Michx., *Quercus palustris* Münchh.) from the Minnesota Landscape Arboretum (Chaska, MN) and from the Morton Arboretum (Lisle, IL). Finally, to detect any potential hybridization between *Q. rubra* and *Q. ellipsoidalis*, we collected two *Q. ellipsoidalis* populations from locations near Minnesota (Fig. 1). We stored all the collected tissue at -80°C until genomic DNA extractions were performed.

Greenhouse common garden experimental design and measurements

During the summer of 2019 (May – August), we conducted a common garden experiment at the University of Minnesota Duluth greenhouse (46°49’00.7”N 92°05’12.1”W). We arranged a total 1438 acorns from 102 maternal lines from ten populations in a randomized block design with four blocks. We distributed collections from populations and maternal lines as evenly as possible across the blocks. Prior to planting, we recorded seed mass and whether an acorn had an emergent radicle. We planted the acorns in low water retention soil (Promix BRK), maintained the seedlings at a temperature between 11°C and 20°C (+ 5°C), with no supplemental light, and watered them every other day for 12 weeks.

We recorded seedling emergence as a measure of germination every other day until 80% of the seedlings had emerged and then weekly until the end of the experiment. We measured stem height, stem diameter, leaf number and survival weekly for the first five

weeks of the experiment and then every other week until the end of the experiment as seedling growth slowed. To calculate specific leaf area (SLA), we collected the uppermost fully expanded leaf at week ten from each available seedling. To obtain leaf area we took a scaled picture of each leaf and analyzed the images with the Easy Leaf Area (Easlon & Bloom, 2014) software. We then dried the leaves at 60°C for one week, weighed them, and calculated SLA by dividing leaf area by dry leaf mass (cm^2/g). Oak seedlings had a rapid growth period during the first four weeks of the experiment, followed by a period of relative stasis; therefore, we estimated the total increase in growth measured by stem height and diameter during the early season (Weeks 1–4) and the late season (Weeks 5–11). We calculated growth rate for each growing period by subtracting the last measurement from the first week measurement and dividing it by the length of the growing period.

Greenhouse common garden data analysis

We leveraged two approaches to analyze our phenotypic data, using: 1) distance categories from Lake Superior (coastal and noncoastal regions), and 2) climate data from the seedlings' home environments as predictors. Because of insufficient sample size (n = 2) , two coastal populations were removed from all analyses. In our regional analyses, we used ANCOVA for continuous response variables (seed mass, germination phenology, growth, and SLA). We used seed mass and the presence of an acorn root as covariates for the analysis of phenology, growth, and SLA. Block was a random factor, and region and population nested within region were fixed factors. To achieve normality of the residuals, we log transformed the SLA data. We used logistic regression for the binomial response variables (presence of acorn root, germination, and survival). All statistical models were constructed in JMP Pro 16 software (SAS Institute Inc, 1989-2021) and all graphs were constructed using the least square means and standard errors in R version 4.1.0 (R Core Team, 2017).

For climate analyses, we obtained 19 site-specific bioclimatic variables (http://www.worldclim.org) and determined which variables to include using a backwards stepwise model reduction approach with AIC, BIC, and factor collinearity as

criteria. The factors that were retained were: (1) mean annual temperature, (2) temperature seasonality, (3) maximum temperature of the warmest month, (4) minimum temperature of the coldest month, (5) annual precipitation, and (6) precipitation of the wettest month. The model structure for the regional and the climate analysis were the same, aside from the substitution of regional categories for climate data. To determine the environmental variable that accounted for most variation in the data, we performed a non-metric multidimensional scaling (NMDS) model with 100 permutations in the 'vegan' package in R (Oksanen et al., 2020). After ordination of the phenotypic traits, we used the *envfit* function in the 'vegan' package to estimate the strength and significance of the relationships between the environmental vectors and the phenotypic traits.

DNA isolation and genotyping

We used a modified CTAB extraction (Sork Lab: Protocols, 2018) to isolate genomic DNA for all the populations using 0.1 g of frozen tissue per sample. The modified CTAB extraction requires a prewash step to break down sugars and secondary metabolites that are found in high quantities in the *Quercus* species (Sork Lab: Protocols, 2018). We quantified the extracted DNA using a Qubit 3.0 fluorometer and excluded any samples with DNA concentrations lower than ~10 ng/µl. Genotyping was performed by restriction associated DNA sequencing (RAD-seq) and completed at the University of Minnesota Genomic Center (UMGC). Genomic DNA of 526 *Q. rubra*, 45 *Q. ellipsoidalis* and 18 other *Quercus* species samples were used for library preparation and Illumina sequencing. RAD libraries were prepared using *BamHI* + *NsiI* restriction enzymes. All generated libraries were combined into a single pool and sequenced on a NextSeq High-output 1x150-bp flow cell. After sequencing, samples were demultiplexed and adaptors were removed using standard UMGC data processing pipelines.

SNP identification and validation

We performed variant calling on the demultiplexed trimmed reads using the standard ipyRAD pipeline (Eaton & Overcast, 2020). To construct the assembly, we used the *Quercus lobata* Neé reference genome (Sork, Fitz-Gibbon, et al., 2016) with a clustering threshold of 0.85 and a minimum of eight samples per locus; all other parameters

followed the standard workflow as suggested in Eaton and Overcast (2020). We constructed the ipyRAD assembly for the following combinations of samples: (1) field collected *Q. rubra* (coastal, inland, and interior regions), (2) field collected *Q. rubra* and *Q. ellipsoidalis*, and (3) field collected *Q. rubra*, greenhouse *Q. rubra* seedlings, nursery *Q. rubra*, and field collected *Q. ellipsoidalis*.

We performed further SNP filtering using VCFtools (Danecek et al., 2011). To remove low-confidence genotypes, we removed sites with a Phred quality score of ≤20 (--minQ), a sequencing depth of ≤5 reads per genotype (--minDP), a minimum mean depth of five reads per site (--min-meanDP), and a minor allele frequency (MAF) of ≤0.03 (--maf). An MAF of ≤0.03 allows a variant to be called when it is present as a heterozygote in an average of 12 samples, which is equivalent to most of the population sizes in the final dataset; this was designed to capture variants that were potentially unique to single populations. To reduce the rate of missing data, we retained individuals with ≤20% missing data (--missing-indv) and retained sites with ≥80% of data present (--max-missing). Finally, we removed indels and retained only SNPs (--remove-indels). We performed filtering on the three previously described datasets, and the number of retained SNPs ranged between 8,210 and 6,906 (Supplementary Table 2). We used PGDspider (Lischer & Excoffier, 2012) and plink v.1.9 (Purcell et al., 2007) to convert VCF files to the required formats for subsequent analyses.

Molecular population structure and interspecific gene flow

We calculated allele frequencies for each population in VCFtools (Danecek et al., 2011). To characterize genetic differentiation among populations, we calculated pairwise genetic differentiation (F_{ST}) within and among populations in ARLEQUIN v3.5.2.1 (Excoffier & Lischer, 2010). We used STRUCTURE v2.3.4 (Pritchard et al., 2000) to infer subdivisions within and admixture among samples. We performed STRUCTURE analyses using the three previously described datasets with ten iterations of each level *K* (2–6) and a burn-in period of 10,000 followed by 20,000 MCMC reps. To visualize likelihood values and determine the number of clusters that best fit the data, we implemented the delta K method (Evanno et al., 2005) in STRUCTURE HARVESTER

(Earl & vonHoldt, 2012). We used Pophelper (Francis, 2017) to graphically visualize population clustering. To quantify admixture between *Q. rubra* and *Q. ellipsoidalis*, we investigated patterns of population structure using the field collected *Q. rubra* and *Q. ellipsoidalis* samples at $K = 2$. We summarized the independent STRUCTURE runs using CLUMMP (Jakobsson & Rosenberg, 2007) and calculated the average Q-values for each population and region. To assess significance of the *Q. ellipsoidalis* admixture proportions within each *Q. rubra* region, we performed ANOVA on the $K = 2$ Q-values using region and population nested within regions as predictors using the JMP Pro 16 software (SAS Institute Inc, 1989-2021). We also analyzed genetic groupings using Identity-By-Descent measures and constructed principal component analyses (PCA) with the R package 'SNPrelate' (Zheng et al., 2012).

Genomic signatures of selection

To assess signatures of selection, we constructed a dataset of unlinked SNPs. We selected one SNP from each RAD locus using the '--thin' option with a distance threshold of 280 bp in VCFtools (Danecek et al., 2011). This process retained a total of 2,560 sites with one SNP for each locus and reduced the physical linkage between SNPS that could potentially bias the selection results. Because linked SNPs could still be present in the dataset, we calculated pairwise linkage disequilibrium of the 2,560 SNPs using plink v1.9 (Purcell et al., 2007) implementing windows of 50 markers and ten marker sliding windows. We then removed loci with an r^2 above 0.05 resulting in 2,162 retained SNPs.

Distance based outlier analysis: We identified outlier loci by implementing a per-locus pairwise F_{ST} approach, using two methods. The first method was FDIST, as implemented in ARLEQUIN v3.5.2.1 (Excoffier & Lischer, 2010), allowing a maximum of 20% missing data and conducting 10,000 simulations with standard parameters. Loci identified in FDIST were considered significant outliers using a 0.05 false discovery rate (FDR) correction with the R package 'qqman' (Turner, n.d.). The second method was BayeScan 2.1 (Foll & Gaggiotti, 2008) with a burn-in of 50,000, a thinning interval of 10, a sample size of 5,000, and 20 pilot runs with 5,000 iterations each for a total of 100,000 iterations. We identified outliers from BayeScan using the R script provided by the program authors

and a 0.05 FDR correction threshold (Benjamini & Hochberg, 1995). We searched for functional annotation of outlier loci that showed potential signs of selection using the NCBI BLAST tool and genome viewer using the *Q. lobata* genome as a reference. Lastly, we calculated allele frequencies for all outlier loci in each region using VCFtools (Danecek et al., 2011).

Environmental Association Analysis: To identify loci associated with environmental variables in the three *Q. rubra* regions, we performed an environmental association analysis (EAA). We obtained climate data for each population from the 19 bioclimatic variables in the Worldclim dataset (http://www.worldclim.org). We ensured that the six variables used in the original greenhouse analysis were still significant for our EAA, by including them in the calculation for the correlation matrix and variance inflation factor (VIF) analysis. We removed variables with an $R^2 > 0.7$ and a VIF > 10. Of the original six environmental variables, temperature seasonality was the only factor that was not retained due to high collinearity. The bioclimatic variables retained for the EAA analysis were: (1) mean annual temperature, (2) mean diurnal range, (3) isothermality, (4) maximum temperature of the warmest month, (5) minimum temperature of the coldest month, (6) annual precipitation, (7) precipitation of the wettest month, and (8) precipitation seasonality.

We performed the EAA using Bayenv2 (Günther & Coop, 2013) and a Latent Factor Mixed Model (LFMM) analysis as implemented in the R package 'LEA' (Frichot & François, 2015). Bayenv2 uses a Bayesian approach to identify correlations between loci and environmental variables by comparing the fit of a null model that includes only neutral genetic structure to the fit of a model that includes each environmental factor (Günther & Coop, 2013). We generated SNP covariance matrices with 10,000 iterations using the 2,162 unlinked SNPs. We then averaged the output of ten covariance matrices to control for matrix variation. Bayes factor (BF), absolute Spearman's rank (ρ), and Pearson correlation coefficient (r) for each locus of four independent Bayenv2 runs were calculated. To identify significant putative adaptive loci driven by environmental

variables, we retained SNPs that fell in the top 1% of the BF, ρ and *r* across all four independent Bayenv2 iterations.

The LFMM approach (Frichot et al., 2013) identifies loci associated with environmental factors while including the effect of population structure as a random factor (latent factor). For this approach, the user needs to determine the number (*K*) of latent factors by performing multiple runs with different levels *K* or extrapolating the best value of K obtained via STRUCTURE (Pritchard et al., 2000) analyses. We performed five independent runs for $K = 2, 4, 6, 8, 10$ for each of the eight previously selected environmental factors with a burn-in period of 5,000 and a total of 10,000 MCMC. The association results for each value of *K* were very similar; thus, we chose $K = 2$ to avoid overparameterizing the model. We averaged the z-scores for each independent run and calculated an adjusted p-value for the correlations between each locus and environmental variable. To reduce the number of false positives, a 0.05 FDR correction with the R package 'qqman' (Turner, n.d.) was performed.

Results

Population differentiation with respect to distance from Lake Superior

Seed mass did not differ significantly between regions; however, a difference between populations within regions was statistically significant (Table 1). More specifically, coastal populations showed higher average seed mass variation, ranging between 3.41 g and 5.01 g, whereas noncoastal populations had less variation, ranging between 4.08 g and 4.51 g. The presence or absence of an emergent radicle was significantly different between regions and populations within regions (Fig. 2a, Table 1). Coastal populations had high variation in emergent radicle presence ranging between 47% and 91%. Similarly, noncoastal populations had high variation in emergent radicle presence ranging between 25% and 72%. Higher seed mass was associated with higher rates of emergent radicle presence.

Germination percentage and germination phenology did not differ significantly between regions. However, populations within each region showed significant differences for both traits (Table 2). Coastal populations had lower germination percentage variation ranging between 66% and 79%, whereas noncoastal populations had higher variation ranging between 55% and 89%. Regarding germination phenology, coastal populations germinated within four days of each other, while noncoastal populations germinated within one day of each other. Germination rates were positively associated with presence of emergent radicles and seed mass. In contrast, germination phenology was positively associated with presence of emergent radicles but negatively associated with seed mass. Oak seedlings had a rapid growth period during the first four weeks of the experiment, followed by a period of relative stasis. Stem height analysis at week four showed significant differences between coastal and noncoastal regions (Fig. 2b, Table 1). Stem height also showed significant variation between populations within each region. Coastal populations had a 7% difference in stem height. Meanwhile, noncoastal populations had a 17% difference. Higher stem height was positively associated with higher seed mass.

Stem diameter analysis at week four showed no significant differences between coastal and noncoastal regions, however, populations within each region were significantly different (Table 1). Stem diameter for coastal populations ranged between 1.8 cm and 2.18 cm. Lower significant variation in stem diameter for noncoastal populations was identified, with stems ranging between 1.99 cm and 2.20 cm in diameter. Higher stem diameter was positively associated with the presence of radicles and higher seed mass.

The number of leaves at week four was non-significant between regions, but variation between populations within each region was significant (Table 1). Significant variation between populations within region was driven by coastal populations, and no significant differences between noncoastal populations was observed. The number of leaves between coastal populations showed a 21.5% difference. Higher leaf number was positively associated with seed mass and negatively associated with the presence of radicles. Specific leaf area (SLA) was non-significant between regions, but significant between populations within each region (Table 1). Specific leaf area for coastal populations showed a 2.7% difference and for noncoastal populations the analysis showed a 3.0% difference. Higher SLA was associated with lower seed mass and with a higher rate of radicle presence.

The change in stem height during the period of rapid growth, quantified as early height growth rate, showed significant differences between regions and between populations within each region (Fig. 2c, Table 1). Coastal populations' early height growth rate ranged from 1.47 cm/week to 1.99 cm/week; meanwhile, noncoastal populations' early growth in height ranged from 1.78 cm/week to 2.32 cm/week. Early height growth rate was positively associated with presence of radicles. In contrast, the change in stem height during the period of slow growth, quantified as late height growth rate was nonsignificant between regions and between populations within regions (Table 1). Late height growth rate was positively associated with the presence of radicles. The change in stem diameter during each period of rapid and slow growth, quantified as early and late stem diameter growth rate, did not show significant differences between regions or between populations within each region (Table 1). Both traits were positively associated with seed mass.

Therefore, variation in plant growth was mostly explained by changes in stem height during the first four weeks of growth.

Lastly, the percentage of seedlings that survived to the end of the common garden experiment was also significantly different between regions (Fig. 2d, Table 1). Survival was significant between populations within each region. Coastal populations had a 28% difference in survival. Whereas noncoastal populations showed higher variation in survival with a 37% difference. Consistent with earlier trends, higher survival rates were associated with higher rates of radicle formation.

Population differentiation with respect to climate variables

The analysis of phenotypic traits using population-specific environmental factors as predictors and the nonmetric multidimensional scaling analysis (NMDS) showed significant differences influenced by the environment. For each phenotypic trait there was at least one environmental variable that explained a significant amount of variation. Acorns that had an emergent radicle before planting germinated an average of seven days earlier than acorns with no radicles. Seedlings that had an emergent radicle had 14.8% slower stem height growth during the early growth period and 1% lower SLA than seedlings with no emergent radicle (Table 2). Acorns that did not have an emergent radicle before planting, comparatively, had 7% lower seed mass, were 50% less likely to germinate, showed a 7.5% reduced stem diameter, and were 50.5% less likely to survive the season (Table 2).

Higher seed mass had positive and significant associations with the presence of radicles, percent germination, stem height, stem diameter, leaf number, late stem height growth rate, and early stem diameter growth rate (Table 2). In contrast, acorns with higher mass were negatively associated with germination phenology, SLA, and late stem diameter growth rate. Seed mass was the factor sharing the most associations with the phenotypic traits measured; only two traits (early stem height growth rate and survival) did not show significant associations. However, early stem height growth rate was marginally significant (p-value = 0.08). Planting block as a random factor was significantly different and showed positive associations for all traits except for SLA (Table 2).

Higher mean annual temperature was associated with higher percent germination, seedlings with larger stem height and diameter, a higher stem height growth during the rapid growing period in the early growing season, and higher survival rates (Table 2). However, seedlings from populations with higher mean annual temperature had lower seed mass, reduced presence of radicles before planting, and fewer leaves with lower SLA (Table 2). Temperature seasonality was positively associated with higher germination and survival rates. Populations with higher temperature seasonality showed lower rates of emergent radicle and leaves with lower SLA. No other phenotypic traits were significantly associated with temperature seasonality.

Higher maximum temperature of the warmest month was associated with acorns that had high seed mass and high percentages of radicle presence. Moreover, higher maximum temperature of the warmest month was associated with seedlings that had a higher leaf count and leaves with higher SLA (Table 2). Maximum temperature of the warmest month also showed negative associations with germination rates, stem height, stem diameter, slower growth during the early growth period, and survival rates (Table 2). Lower minimum temperature of the coldest month was associated with acorns that had higher seed mass, seedlings with more leaves and leaves with higher SLA. Lower minimum temperatures of the coldest month were associated with low germination rates, seedlings with low stem height and diameter, and low survival rates. Lower minimum temperatures of the coldest month were also associated with slower height growth rates during the rapid growth period and slower stem diameter growth rate during the slow growth period (Table 2).

Increased annual precipitation showed associations with acorns that had higher mass and seedlings with fewer leaves. In contrast, increased annual precipitation showed associations with lower rates of acorns with radicle presence (Table 2). No other phenotypic traits were significantly associated with annual precipitation. Lastly, increased precipitation during the wettest month was associated with higher germination rates and survival rates. Increased precipitation of the wettest month was associated with lower seed mass acorns and seedlings with fewer leaves (Table 2).

Non-metric multidimensional scaling (NMDS) analysis showed correlations between environmental factors and *Q. rubra* seedling phenotypic traits (Fig. 3). Correlation scores showed that mean annual temperatures (r^2 = 0.12) and maximum temperatures of the warmest month (r^2 = 0.1) had the greatest influence on *Q. rubra* seedlings' phenotypic variation. The remaining four environmental factors had weaker associations (ranging from r^2 = 0.089–0.097); however, all the environmental variables were statistically significant and positively associated with germination dates, stem height and diameter at week 4, leaf number and leaf SLA growth rates in early and late seasons, and survival rates.

Molecular population structure and interspecific gene flow

After filtering, we retained 434 individuals and 2,495 loci and identified a total of 7,683 SNPs for our dataset that included all field-collected, greenhouse, and nursery *Q. rubra* samples and field-collected *Q. ellipsoidalis*. Interspecific F_{ST} was 0.15 and pairwise F_{ST} among all *Q. rubra* and *Q. ellipsoidalis* populations ranged between 0.01 and 0.25. Lower pairwise F_{ST} values were between *Q. rubra* populations and higher values between *Q. rubra* and *Q. ellipsoidalis* populations. However, four populations (SP, BL, II, GW) that had been identified as *Q. rubra* in the field showed elevated pairwise F_{ST} values ranging between 0.08 and 0.25 with other *Q. rubra* populations. They also showed low interspecific pairwise F_{ST} values ranging between 0.009 and 0.04 with *Q. ellipsoidalis* populations (Fig 4a). Principal component analysis also showed that the four *Q. rubra* populations with high F_{ST} levels clustered more closely with *Q. ellipsoidalis* populations than with other *Q. rubra* populations (Supplementary Figure 1). The four high F_{ST} populations (SP, BL, II, GW) were therefore considered as misidentified and were removed from subsequent analyses. After removing the four misidentified high F_{ST} populations, the global F_{ST} for *Q. rubra* was 0.03 and the pairwise F_{ST} for the remaining *Q. rubra* populations ranged between 0.0052 and 0.066.

Admixture between *Q. rubra* and *Q. ellipsoidalis* individuals and populations were calculated based on STRUCTURE assignments at level of K = 2. The average proportion of admixture from *Q. ellipsoidalis* into *Q. rubra* populations ranged between 4.8% and 10.6% (Fig. 4b). Higher rates of admixture were detected as distance from Lake Superior

increased, and *Q. ellipsoidalis* populations showed no evidence of *Q. rubra* identity proportions, indicating unidirectional gene flow from *Q. ellipsoidalis* intro *Q. rubra.* Admixture proportions were significantly different between coastal and noncoastal regions (F = 20.7, P <0.0001) and between populations within regions (F = 3.3, P <0.0001) (Fig. 4c). Five *Q. rubra* individuals (two inland and three interior) showed a *Q. ellipsoidalis* admixture proportion above 45%. When the analysis was performed without these five individuals, patterns of admixture proportion were consistent and retained significance (F = 45.5, P <0.0001 and F = 4.65, P <0.0001).

Population structure analysis was performed using only coastal, inland, and interior *Q. rubra* populations and excluding misidentified populations. The optimum value of K for the STRUCTURE analysis was K = 3, but there was no evidence of differentiation between regions (Supplementary Figure 2). Similarly, a STRUCTURE analysis performed with K = 2 showed no evidence of population differentiation based on region (Supplementary Figure 2). The PCA explained a small amount of the variance between regions (1.8% across the first two PCAs) and clustered all regions into one group, though some samples from inland and interior regions diverged from the main cluster along PC1 and PC2 (Fig. 4d). Nonetheless, there was no evidence of strong differentiation between *Q. rubra* regions or populations, which is consistent with the low pairwise F_{ST} values.

Genomic signatures of selection

We identified 2,245 sites with one SNP per locus in linkage equilibrium across 413 field collected *Q. rubra* individuals. We identified a low level of linkage between loci; of the initial 2,599 loci, only 13.6% exhibited high levels of linkage ($r^2 > 0.5$) and were removed.

Distance-based outlier analysis

Outlier analysis using BayeScan and FDIST identified ten and seven loci, respectively, with elevated levels of divergence at a 5% significance level (Fig. 5). Five of these outlier loci overlapped between analyses and exhibited F_{ST} levels ranging between 0.143 and 0.667. Two of the five outliers were in Chromosome 1 (position 29790677 and

29936530), one in Chromosome 4 (position 38317234), one in Chromosome 5 (position 53461008), and one in Chromosome 9 (position 45446014) (all positions are based on the *Q. lobata* reference genome). We calculated allele frequencies for the five outlier loci in each *Q. rubra* region and identified patterns consistent with strong directional selection in Chromosome 1 (Supplementary Table 4). Locus 29790677 in Chromosome 1 was fixed for a single allele in the interior region but polymorphic for the coastal and inland regions. Locus 29936530, also located in Chromosome 1, was fixed for a single allele in the coastal region but polymorphic for the inland and interior region. In contrast, all other loci under selection were polymorphic within each region (Supplementary Table 4).

Environmental Association Analysis (EAA)

Using Bayenv2, we detected six loci with a significant association to one or more environmental variables when the four independent runs were combined (Table 3). No loci were associated with the minimum temperature of the coldest month or annual precipitation. We identified one site in Chromosome 1 (29936530) that was associated with the precipitation of the wettest month. This locus was also significant in both outlier analysis and had the highest Fst level (0.66) in that analysis. Two loci in Chromosome 10 (14049740 and 37795457) were associated with mean annual temperature. One locus in Chromosome 8 (24442158) was associated with isothermality. Two loci in Chromosome 2 (62044137 and 63576952) were associated with precipitation seasonality and mean diurnal range, respectively. Of the six associated loci, only site 37795457 located in Chromosome 10 exhibited association with more than one variable (mean annual temperature and maximum temperature of the warmest month).

The LFMM analysis revealed a total of 521 unique loci associated with eight environmental variables at 5% significance (Fig. 6). While this number is significantly higher than the number of loci found using the other programs, this result is not surprising as LFMM accounts for any signs of population structure and similar studies have found elevate numbers of LFMM-associated loci (Abebe et al., 2015; Gugger et al., 2021; Rellstab et al., 2017). Of the 521 loci identified by the LFMM analysis, two loci were also identified by Bayenv2 and eight were identified by BayeScan and FDIST (Fig.

6). Temperature-related factors had an overall lower number of associated loci, but higher numbers of overlapping loci between variables, when compared to precipitation factors. Precipitation seasonality was associated with the highest number of loci (231), while isothermality had the lowest number of associated loci (106) (Table 3). Locus 29936530, located in Chromosome 1, which had been identified as an outlier, was also found to be associated with the precipitation of the wettest month, consistent with the association between this locus and the variable obtained by the Bayenv2 program.

Discussion

The potential for local adaptation in long-lived oak species is supported by mounting evidence within the literature that combines molecular genetics and phenotypic data (Sork et al., 1993; Ramirez-Valiente et al., 2018 and 2019; Sork, 2016; Gugger et al., 2016; Rellstab et al., 2016). For example, a study on *Quercus oleoides* Schltdl. & Cham. found no molecular population structure and low F_{ST} values, but significant phenotypic differentiation in functional traits related to drought tolerance and environmental variability (Ramírez-Valiente et al., 2018, 2019). Other studies performed on *Q. lobata* have found high F_{ST} values at some candidate loci related to flowering and temperature stress genes, indicating loci under potential diversifying selection between populations, and *Q. lobata*'s ability to respond to changing climate conditions (Sork et al., 2016). Here, we combined molecular and phenotypic data to investigate patterns of gene flow and natural selection on coastal and noncoastal *Q. rubra* populations at a fine spatial scale (0–160 km). We framed our study to identify patterns of gene flow and natural selection on the genomic and quantitative level, while accounting for transgenerational plasticity, hybridization, and population structure. Our study provides phenotypic evidence of genetic differences and integrates a molecular approach to dissect the genomic underpinnings of the observed variation of *Q. rubra* populations west of the Great Lakes. Although our study was performed across a small spatial scale, our results are consistent with low population differentiation, but strong evidence of selection correlated with the environment.

Do populations show genetic differentiation based on phenotypic traits?

Within the common garden greenhouse experiment, coastal seedlings had higher radicle presence before planting, slower stem growth during the early season resulting in shorter stems at week four, and lower survival rates (Table 1, Fig. 2). We also identified high variation between populations within each region, showing that each population had a unique response to germination and growth under controlled environmental conditions. Maintenance of genetic variation among individuals and populations is crucial for adaptive evolution to occur (Anderson et al., 2012). Due to the high population variation

within the region, we showed that some of the detected phenotypic differences in *Q. rubra* seedlings are not driven by regional trends, but rather by site-specific environmental factors.

Environmental heterogeneity drives local adaptation (Papaïx et al., 2013); but adaptive evolution has been measured as being slower than the rate at which the climate is changing (Shaw & Etterson, 2012). Thus, evolution in some organisms is unable to keep up with quickly changing climatic conditions (Shaw & Etterson, 2012), resulting in strong directional selection that favors extreme genotypes (Anderson et al, 2012). To delve more deeply into the potential factors that explain observed patterns of genetic variation, we conducted analyses in which we used regions as a proxy for climate. When investigating the effects of temperature and precipitation on seedling germination, growth, and survival, we identified that all the measured traits, except germination phenology and stem diameter at week four, were influenced by at least one environmental variable. Germination phenology and stem diameter at week four were only influenced by presence of an emergent radicle and seed mass (Table 2).

No environmental variable had consistent positive or negative effects for all phenotypic traits. No phenotypic trait had an association in the same direction across all environmental variables. Although we did not observe strong correlations ($r^2>0.5$) for any trait, mean annual temperature and maximum temperature of the warmest month appear to be the variables with the strongest correlations with phenotypic variation (Fig. 3). Studies of *Q. lobata* have shown an adaptation lag to temperature (Browne et al., 2019). For example, acorns planted in climates warmer than where they originated had a 5.6% slower growth over a 3-year period (Browne et al., 2019). Our analyses cannot distinguish whether there is an adaptation lag or not. Yet, we demonstrated that high mean annual temperature is associated with seedlings producing fewer leaves with low SLA, and high maximum temperature of the warmest month is associated with seedlings producing short, narrow stems with slower growth during the early season (Table 2). More experiments are necessary to determine the direction and effects of natural selection on these traits; however, we can infer that local adaptation driven by natural selection

might act differently on each population, given the unique environmental conditions exhibited by the coast of Lake Superior.

The moderating effect of Lake Superior produces warmer spring temperatures by the shore (March–May) and cooler temperatures in inland areas (Moen, 2018). We observed high germination success under controlled environmental conditions, with over 70% germination success for both coastal and noncoastal regions. Even though seedling germination, as measured by emergence, was non-significant, coastal acorns had a higher percent of radicle presence (Fig. 2a). The presence of an emergent radicle is evidence for early germination, given its positive association with germination phenology in the region and environment-based analyses (Table 2). Our analysis showed that seedlings with an emergent radicle germinated 7 days earlier than those without. Early germination, quantified as the presence of an emergent radicle, might give *Q. rubra* coastal seedlings a germination advantage in coastal environments where spring temperatures are mild, and risk of freezing is low. Phenology of life history events is a key component of climate adaptation. Although radicle emergence has been studied as a measure of germination success (Romero-Rodríguez et al., 2018; Wardle et al., 1991), to our knowledge, no evidence has previously shown how the presence of an emergent radicle before planting affects juvenile growth. However, this does not reflect an accelerated growth in the coastal region under greenhouse conditions; instead coastal seedlings growth is slower when compared to noncoastal seedlings.

We showed numerous genetic differences between coastal and inland regions, and associations between seedling growth and home environment (Tables 1, 2). However, there is a possible influence from the maternal growing environment that might persist, even when the seed is grown in different conditions (Wulff et al., 1994). This transgenerational plasticity can affect the expressed phenotype of an individual grown in a common garden (Galloway & Etterson, 2007) and can last up to the late seedling stage, which in *Q. rubra* individuals can take up to five years. This same slow growth makes performing a refresher generation as a control for transgenerational plasticity practically impossible. Instead, to control for seed provisioning from the maternal environment, we

used seed mass as a covariate in our analyses. As a covariate, seed mass was a factor that significantly influenced germination and all the measured growth traits (Table 2). However, we note that seed mass did not influence the survival rates of seedlings, showing that *Q. rubra* seedlings from coastal populations were not poorly provisioned, despite their slower growth.

Do populations show genetic differentiation based on molecular markers?

We did not detect strong genetic differentiation among *Q. rubra* populations based on F_{ST} values or clustering analyses (Fig. 4a,d). The low levels of population differentiation are consistent with high rates of gene flow, likely due to the long generation time and the extended pollen dispersal range (recorded up to 100 km) that characterize oaks (Schueler & Schlünzen, 2006). Both factors would ensure the dispersal and maintenance of allelic diversity across populations. Indeed, our global F_{ST} value (F_{ST} = 0.03) is similar to those found in *Quercus* and other tree species such as *Q. oleoides* (F_{ST} = 0.09) (Ramirez-Valiente et al., 2018), *Quercus robur* L. (F_{ST} = 0.07) (Vakkari et al., 2006), *Eucalyptus albens* Benth. (F_{ST} = 0.018) (Murray et al., 2019), and *Pinus taeda* L. (F_{ST} = 0.04) (Bassar et al., 2010).

In contrast to the low F_{ST} levels calculated between *Q. rubra* populations, we identified four populations that had higher genomic identity with *Q. ellipsoidalis* than with *Q. rubra*. Although our primary interest was to characterize population structure within Northern red oak west of Lake Superior, we also included two Northern pin oak (*Q. ellipsoidalis*) populations as reference samples to confirm species identity. This was necessary, given the morphological similarity of *Q. rubra* and *Q. ellipsoidalis* leaves in the wild, the overlapping range between the two species, and evidence of *Q. rubra* hybridizing with *Q. ellipsoidalis* in contact zones in eastern areas of the Great Lakes region (Lind & Gailing, 2013; Jensen et al., 1993; Gailing et al., 2018). The inclusion of these reference populations allowed us to identify four putative *Q. rubra* populations in our original collections (SP, BL, GW, II) with high *Q. ellipsoidalis* genetic identity (Fig. 4a). We note that our collections were based on the occurrence of *Q. rubra* in MN-DNR surveys, and the survey data did not indicate the presence of *Q. ellipsoidalis* in any of the

four misidentified populations. For the SP and II populations, there was evidence of bur oak (*Q. macrocarpa*) presence, but the vast morphological differences between *Q. rubra* and *Q. macrocarpa* eliminate the possibility that we collected *Q. macrocarpa* individuals. One possible explanation for the collection of *Q. ellipsoidalis* at these sites is that trees were misidentified in the original surveys, but another is that our samples were from a stand of *Q. ellipsoidalis* very close to, but not included in, the original survey site. The latter possibility suggests that the two species might occur in close proximity, making these locations in Minnesota potential oak hybrid zones.

Because of the occurrence of interspecific gene flow between *Q. rubra* and *Q. ellipsoidalis* in other regions (Lind & Gailing, 2013; Jensen et al., 1993; Gailing et al., 2018), the identification of the four misidentified populations, and the possibility of hybrid zones, we used admixture proportions from STRUCTURE to evaluate the extent of gene flow between these two species. The STRUCTURE analysis was consistent with unidirectional gene flow from *Q. ellipsoidalis* into *Q. rubra*, but no *Q. rubra* identity was found in the *Q. ellipsoidalis* populations (Fig. 4b). We identified an increasing *Q. ellipsoidalis* proportion in the *Q. rubra* populations with greater distance from the shore of Lake Superior, which corresponds with a growing overlap between the species (Fig. 4c). Interior populations had the highest average *Q. ellipsoidalis* proportion (10%), and three interior individuals had an admixture proportion above 45%, indicating the presence of early generation hybrids. The inland populations had slightly lower admixture proportions (9.3%) than interior populations, and two inland individuals had more that 45% admixture with *Q. ellipsoidalis*. However, coastal populations had no individuals with admixture proportions above 45% and the lowest *Q. ellipsoidalis* percent identity (4.8%). Although gene flow between these two species is not unexpected, this is the first time that introgression from *Q. ellipsoidalis* into *Q. rubra* has been documented west of the Great Lakes. Other studies have demonstrated the adaptive potential of introgression from *Q. rubra* into *Q. ellipsoidalis* in parapatric populations, despite symmetric gene flow between the two species, using a CONSTANS-like gene (Khodwekar & Gailing, 2017; Lind-Riehl & Gailing, 2016). Therefore, investigating the directionality of gene

flow, and whether introgression is adaptive in populations in regions west of Lake Superior, is an interesting area for future study.

Evidence for selection despite gene flow

Despite the low population differentiation that is consistent with high rates of gene flow, we found genomic evidence of natural selection. These findings align with evidence from the phenotypic analysis, suggesting a scenario of population differentiation due to strong natural selection acting against the homogenizing effects of gene flow. In this study, we identified five overlapping loci that showed signatures of selection in the distance-based outlier approaches (Fig. 5), and two overlapping loci that showed a significant association with environmental variables in the two EAA approaches. Our results are consistent with similar studies performed on *Quercus* species on broader geographical scales (>200 km). These studies identified the signature of selection for genes related to drought adaptation and various metabolic processes (Gugger et al., 2021; Lind-Riehl et al., 2014; Pettenkofer et al., 2020). Previous studies that have used distance-based approaches to identify signatures of selection in oak species have also found loci under selection, ranging from a few (Pettenkofer et al., 2020; Lind-Riehl et al., 2014) to >500 loci (Gugger et al., 2021). Only two studies have used environmental association analysis in *Quercus;* these studies, performed in *Q. lobata*, identified 34 loci using SNPs and 43 sites using single methylation variants (SMVs). (Gugger et al., 2016 and 2021). Environmental association analyses using SNP data in other species have found between 54 and 1011 loci associated with the environment (e.g., Christmas et al., 2016; Pais et al., 2017; Rellstab et al., 2016). However, all of these studies are conducted on broader geographical scales than our study, ranging from 200–800 km. In comparison to previous studies, this study leverages geographical scale (160 km), an expanded sampling of genomic loci for tests of selection, as well as a diverse suite of association analyses to provide robust evidence of selection driven by the environment.

The environmental association analyses, based on phenotypic traits and molecular markers, revealed how a highly heterogenous environment can influence selection. However, our results show a contrasting pattern of environmental factors associated with

phenotypic and molecular variation. According to our environmental-based analyses of phenotypic differentiation, the factors with the highest correlations are mean annual temperature and maximum temperature of the warmest months. Meanwhile, the LFMM of environmental association with molecular markers revealed that precipitation of the wettest month and precipitation seasonality had the highest number of associations. To our knowledge, the contrast between environmental factors associated with phenotypic vs. molecular variation presented here have not been demonstrated previously. We hypothesize that these differences may be due to the different life stages that were included in each analysis or the greenhouse conditions. In our study, the phenotypic analyses were performed using juvenile seedlings, while the EAA used data from adult trees. Juvenile growth correlated to temperature, and this might aid in establishment and rapid growth within the first few years of seedling development (Hatfield & Prueger, 2015). In contrast, adult growth is correlated with precipitation that can alter soil water availability, influencing plant survival and reproduction (Larios & Venable, 2018). Clearly, further analyses that integrate multiple approaches at different developmental stages are needed to illuminate what factors are most important for growth and survival across life stages of *Q. rubra* trees.

Finally, we identified only one locus (Chromosome 1, position 29936530) that overlapped between the outlier and the EAA analyses. Locus 29936530 was associated with precipitation of the wettest month and had a significantly higher F_{ST} ($F_{ST} = 0.6$) when compared to the global F_{ST} for *Q. rubra* populations ($F_{ST} = 0.03$) (Fig. 6). While there is not a clear pattern of fixation for most of the other outlier loci, it is important to note that locus 29936530 was fixed in the coastal population but remained polymorphic in the inland and interior populations. This pattern is consistent with a scenario of an adaptive allele being driven to fixation in the coastal populations compared to the inland and interior populations, although further research is necessary to verify this possibility. Moreover, this locus was also alone in having a functional annotation based on the *Q. lobata* reference genome. This locus is found within the *MT-ND1* gene, and it encodes the NADH-ubiquinone oxidoreductase chain 1 protein (Ostaszewska-Bugajska & Juszczuk, 2016). This protein catalyzes electron transfer from NADH during cellular

respiration and uses ubiquinone as an electron acceptor (Ostaszewska-Bugajska & Juszczuk, 2016). Studies in *Arabidopsis thaliana* have shown that long-term sulphur deficiency is associated with changes in oxidative phosphorylation and the alteration of the NADH-ubiquinone oxidoreductase chain 1 protein (Ostaszewska-Bugajska & Juszcuk, 2016). There is also evidence that high precipitation can lead to sulphur deficiency in crops (Kaiser and Vetsch, 2020). However, further studies are necessary to determine whether sulphur deficiency is related to precipitation for *Q. rubra* in coastal environments.

Conclusion

In this study, we evaluated the differentiation of *Q. rubra* populations across a fine spatial scale, spanning an environmental gradient west of Lake Superior, using both molecular markers and quantitative traits. We were also able to identify substantial evidence of interspecific hybridization between *Q. rubra* and *Q. ellipsoidalis*, for the first time in this region, and we found evidence of strong selection associated with environmental variables within *Q. rubra* populations. Our analyses revealed gene flow and selection patterns consistent with a scenario of population differentiation due to strong natural selection despite high gene flow. While the populations are connected by high levels of gene flow, they differ for key traits regarding germination success, growth rates, and survival. This elucidates how natural selection can outcompete the homogenizing effects of gene flow and act on key traits, allowing for populations to adapt in their local environment. We present substantial evidence of environmental-dependent selection based on quantitative traits, exposing multiple genomic regions that are broadly associated with temperature and precipitation. We provide evidence of the signatures of selection and the potential for *Q. rubra* to adapt to local microclimates, even on a small spatial scale, which can further aid seed-sourcing guidelines and continued efforts related to forest conservation and management in the Great Lakes.

Tables and Figures

Table 1. Analysis of variance test statistics from the region-based model, used to analyze genetic differences based on quantitative traits of *Q. rubra* populations from three regions (coastal, inland, and interior)

	Radicle Presence (*df* = 1)		Seed Mass (*df* = 1430)		Block (*df* = 3)		Region (*df* = 1)		Population (region) (*df* = 6)	
	***F* / χ^2**	***p-value***	***F* / χ^2**	***p-value***	***F* / χ^2**	***p-value***	***F* / χ^2**	***p-value***	***F* / χ^2**	***p-value***
Seed Mass	20.56	**<0.01**	-	-	-	-	1.87	0.17	15.87	**<0.001**
Radicle Presence	-	-	23.34	**<0.001**	-	-	46.2	**<0.001**	168.4	**<0.001**
Percent Germination	292.66	**<0.001**	12.12	**0.0007**	43.87	**<0.001**	1.46	0.23	81.19	**<0.001**
Germination Phenology	113.8	**<0.001**	6.47	**0.01**	19.52	**<0.001**	0.09	0.76	2.67	**0.01**
Stem Height (W4)	2.91	0.09	48.22	**<0.001**	12.47	**<0.001**	17.2	**<0.001**	5.49	**<0.001**
Stem Diameter (W4)	21.74	**<0.001**	215.31	**<0.001**	127.4	**<0.001**	3.1	0.08	5.08	**<0.001**
Leaf Number (W4)	0.23	0.63	14.91	**0.001**	4.88	**0.002**	1.52	0.22	3.62	**0.001**
Specific Leaf Area (SLA) (log)	6.58	**0.01**	4.48	**0.03**	0.09	0.76	0.59	0.44	5.33	**<0.001**
Early Stem Height Growth Rate	4.86	**0.02**	2.92	0.08	8.15	**<0.001**	7.79	**0.005**	5.05	**<0.001**
Late Stem Height Growth Rate	0.73	0.39	5.16	**0.02**	26.79	**<0.001**	0.17	0.67	0.54	0.77
Early Stem Diameter Growth Rate	3.18	0.07	41.12	**<0.001**	94.78	**<0.001**	0.49	0.48	0.4	0.87
Late Stem Diameter Growth Rate	2.92	0.08	5.98	**0.014**	40.85	**<0.001**	0.11	0.73	1.12	0.34
Season Survival	221.17	**<0.001**	1.41	0.23	40.41	**<0.001**	4.75	**0.03**	73.22	**<0.001**

Table 2. Analysis of variance test statistics from the environmental-based model, used to analyze genetic differences based on quantitative traits of *Q. rubra* populations using temperature and precipitation factors as predictors.

	Radicle Presence (*df* = 1)			Seed Mass (*df* = 1430)			Block (*df* = 3)		
	β	***F* / χ^2**	***p-value***	***β***	***F* / χ^2**	***p-value***	***β***	***F* / χ^2**	***p-value***
Seed Mass	-0.16	18.56	**<0.0001**	-	-	-	-	-	-
Radicle Presence	-	-	-	0.29	21.03	**<0.0001**	-	118.06	**<0.0001**
Percent Germination	-1.57	292.3	**<0.0001**	0.26	12.96	**0.0003**	-	43.61	**<0.0001**
Germination Phenology	3.55	113.4	**<0.0001**	-0.5	6.89	**0.008**	-	19.37	**<0.0001**
Stem Height (W4)	-0.25	2.83	0.09	0.6	49.95	**<0.0001**	-	12.4	**<0.0001**
Stem Diameter (W4)	-0.08	19.1	**<0.0001**	0.16	221.9	**<0.0001**	-	123.83	**<0.0001**
Leaf Number (W4)	-0.03	0.28	0.59	0.15	14.81	**0.0001**	-	4.87	**0.002**
SLA (log)	0.02	5.76	**0.01**	-0.01	5.13	**0.02**	-	0.16	0.68
Early Stem Height Growth Rate	0.13	4.87	**0.02**	0.04	2.92	0.08	-	8.17	**<0.0001**
Late Stem Height Growth Rate	-0.02	0.79	0.37	0.04	7.54	**0.006**	-	23.37	**<0.0001**
Early Stem Diameter Growth Rate	0.008	3.47	0.06	0.01	41.98	**<0.0001**	-	94.49	**<0.0001**
Late Stem Diameter Growth Rate	0.005	2.73	0.09	-0.004	6.35	**0.01**	-	40.74	**<0.0001**
Survival	-1.22	219.4	**<0.0001**	0.08	1.97	0.16	-	39.59	**<0.0001**

Table 2 continued. Analysis of variance test statistics from the environmental-based model, used to analyze genetic differences based on quantitative traits of *Q. rubra* populations using temperature and precipitation factors as predictors.

	Mean Annual Temperature (*df* = 1)			Temperature Seasonality (*df* = 1)			Max Temperature Warmest Month (*df* = 1)		
	β	***F / χ²***	***p-value***	***β***	***F / χ²***	***p-value***	***β***	***F / χ²***	***p-value***
Seed Mass	-2.6	23.55	**<0.0001**	-0.009	2.13	0.14	2.73	22.91	**<0.0001**
Radicle Presence	-4.03	14.82	**0.0001**	-0.05	17.51	**<0.0001**	2.23	3.93	**0.04**
Percent Germination	10.02	52.57	**<0.0001**	0.08	29.97	**<0.0001**	-11	56.99	**<0.0001**
Germination Phenology	1.03	0.05	0.88	0.003	0.005	0.94	0.5	0.01	0.9
Stem Height (W4)	7.53	15.35	**<0.0001**	0.05	5.55	**0.01**	-6.72	10.81	**0.001**
Stem Diameter (W4)	0.55	4.84	**0.02**	0.006	5.45	**0.01**	-0.6	5.0	**0.02**
Leaf Number (W4)	-2.21	5.8	**0.01**	0.008	0.55	0.45	1.89	3.76	**0.05**
SLA (log)	-0.54	13.82	**0.0002**	-0.006	15.2	**0.0001**	0.58	14.47	**0.0002**
Early Stem Height Growth Rate	2.04	11.36	**0.0008**	0.01	4.99	**0.02**	-1.77	7.34	**0.006**
Late Stem Height Growth Rate	0.3	0.66	0.41	$7x10^{-4}$	0.03	0.86	-0.3	0.59	0.44
Early Stem Diameter Growth Rate	0.02	0.11	0.73	$8x10^{-5}$	0.01	0.9	-0.01	0.01	0.91
Late Stem Diameter Growth Rate	0.06	2.64	0.1	$4x10^{-4}$	0.53	0.46	-0.07	2.37	0.12
Survival	7.7	39.66	**<0.0001**	0.07	27.07	**<0.0001**	-8.54	43.85	**<0.0001**

Table 2 continued. Analysis of variance test statistics from the environmental-based model, used to analyze genetic differences based on quantitative traits of *Q. rubra* populations using temperature and precipitation factors as predictors.

	Min Temperature Coldest Month (*df* = 1)			Annual Precipitation (*df* = 1)			Precipitation of Wettest Month (*df* = 1)		
	β	***F / χ²***	***p-value***	***β***	***F / χ²***	***p-value***	***β***	***F / χ²***	***p-value***
Seed Mass	0.88	9.65	**0.002**	0.02	9.4	**0.002**	-0.37	27.17	**<0.0001**
Radicle Presence	0.36	0.43	0.51	-0.11	35.06	**<0.0001**	0.28	3.67	0.05
Percent Germination	-5.3	53.46	**<0.0001**	-9.6×10^{-6}	1×10^{-6}	0.99	0.99	31.44	**<0.0001**
Germination Phenology	1.85	0.7	0.4	-0.03	0.21	0.64	-0.09	0.03	0.86
Stem Height (W4)	-4.6	21.8	**<0.0001**	-0.008	0.07	0.78	0.53	4.29	**0.03**
Stem Diameter (W4)	-0.39	9.18	**0.002**	0.002	0.33	0.56	0.03	0.86	0.35
Leaf Number (W4)	0.92	3.86	**0.05**	0.03	4.7	**0.03**	-0.34	7.84	**0.005**
SLA (log)	0.29	15.23	**0.0001**	-0.001	0.19	0.66	-0.03	3.35	0.06
Early Stem Height Growth Rate	-0.86	6.97	**0.008**	0.01	1.04	0.31	0.1	1.59	0.2
Late Stem Height Growth Rate	-0.14	0.53	0.46	0.002	0.09	0.76	0.02	0.23	0.62
Early Stem Diameter Growth Rate	-0.01	0.19	0.66	0.0007	0.51	0.47	-0	0.15	0.69
Late Stem Diameter Growth Rate	-0.04	4.54	**0.03**	0.0004	0.24	0.61	0.005	1.05	0.3
Survival	-4.21	42.17	**<0.0001**	0.01	0.52	0.47	0.75	22.97	**<0.0001**

Table 3. Number of loci associated with environmental variables for the EAA using two different approaches. For each approach the number of loci *n* is the total number of associated sites in the analysis.

	Bayenv2 (*n= 6*)	**LFMM** (*n = 521*)
Mean annual temperature	2	118
Mean diurnal range	1	112
Isothermality	1	106
Maximum temperature of the warmest month	1	109
Minimum temperature of the coldest month	0	127
Annual precipitation	0	107
Precipitation of the wettest month	1	227
Precipitation seasonality	1	231

***The sum of the loci associated exceeds the number of *n* for each approach because of associations with more than one variable per locus.**

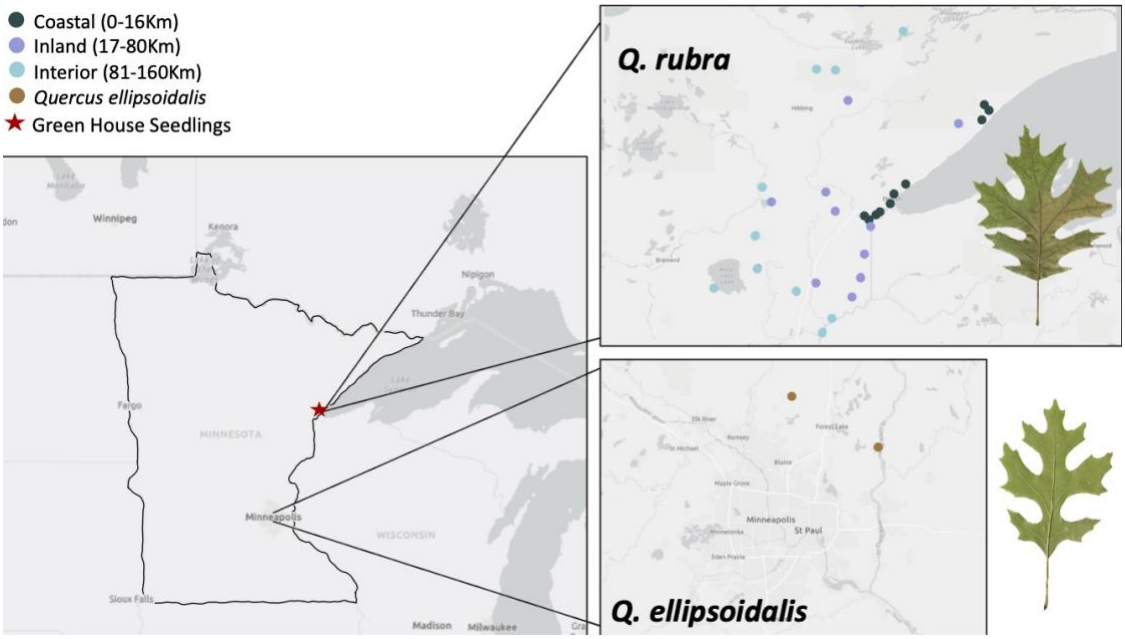


Figure 1. Map of sampled locations; this map includes all the populations that were sampled for either the common garden, the molecular analysis or both (For specifics on each population refer to Supplementary table 1).Top panel includes *Q. rubra* populations from coastal (0–16 km), inland (17–80 km), and interior (81–160 km) from the coast of Lake Superior. Bottom panel contains the two *Q. ellipsoidalis* populations included in the analysis. The star represents greenhouse seedling samples included in the molecular analysis. Examples of morphological differences between species are included to the right of each panel.

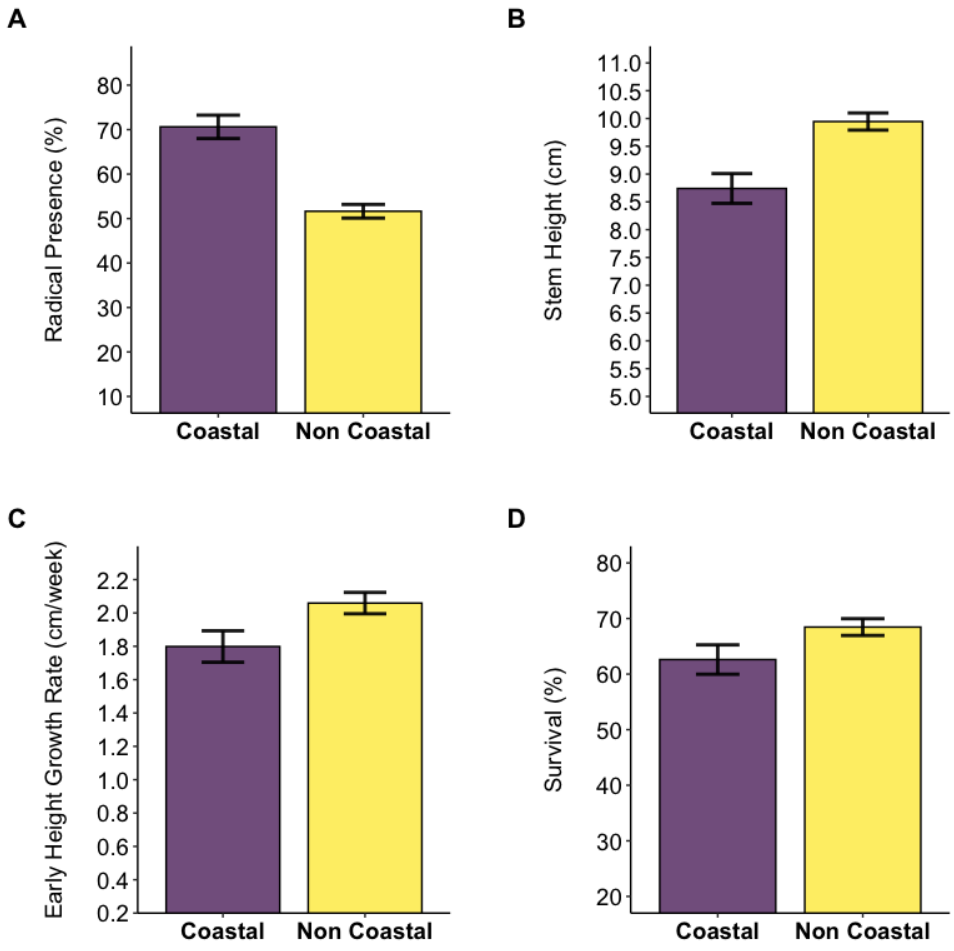


Figure 2. Bar graphs of the least square of significant traits based on region in the common garden experiment. Each bar in represents either a coastal or a noncoastal population, error bars are based on standard error for each trait. The traits that were significantly different based on region the common garden experiment that were presence of radicle (a) Stem height (cm) at week 4 (b) growth rate during the early season (weeks 1–4) (c) and survival rates (d).

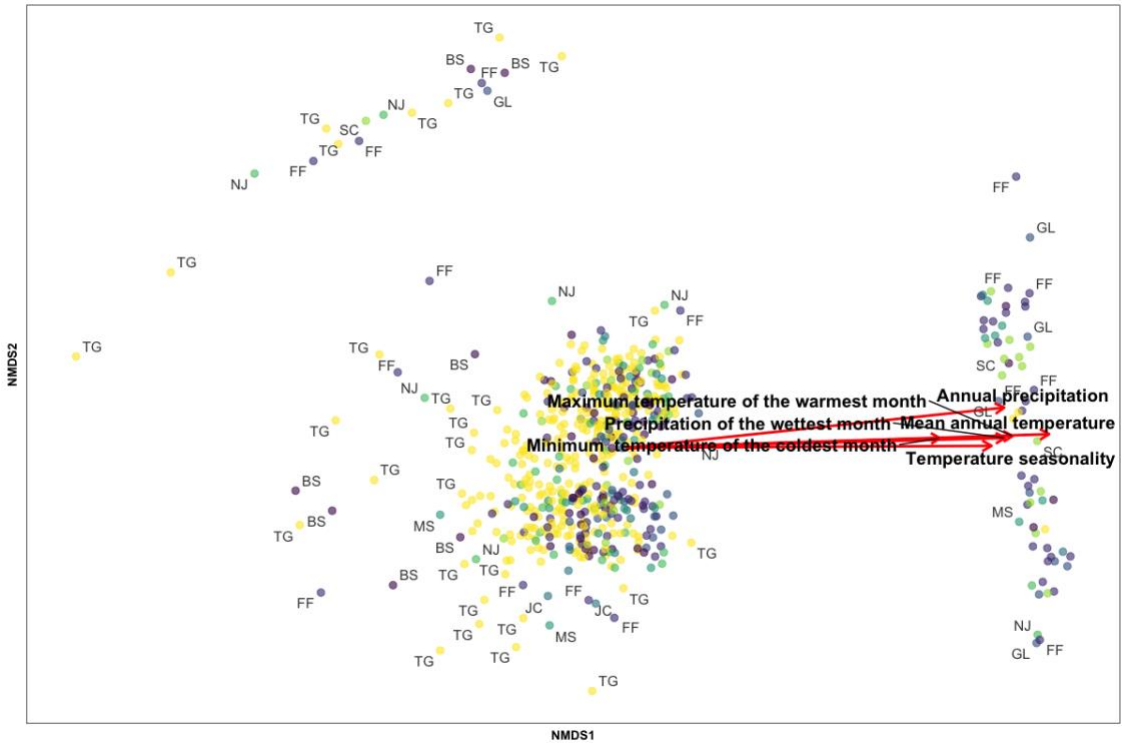


Figure 3. Non-metric multidimensional scaling (NMDS) in a *k = 2* space using "Euclidean" distances, of phenotypic data for all individuals with population specific environmental variables fitted to the ordination. Each point represents an individual from each population (represented by color), few individuals per population are labeled to reduce label overlap. The length of the arrow is proportional to the correlation between the environmental variable and the ordination. The correlation for each variable is: mean annual temperature ($r^2 = 0.12$, $P > 0.001$), temperature seasonality ($r^2 = 0.095$, $P > 0.001$), maximum temperature of the warmest month ($r^2 = 0.1$, $P > 0.001$), minimum temperature of the coldest month ($r^2 = 0.089$, $P > 0.001$), annual precipitation ($r^2 = 0.094$, $P > 0.001$), and precipitation of the wettest month ($r^2 = 0.097$, $P > 0.001$).

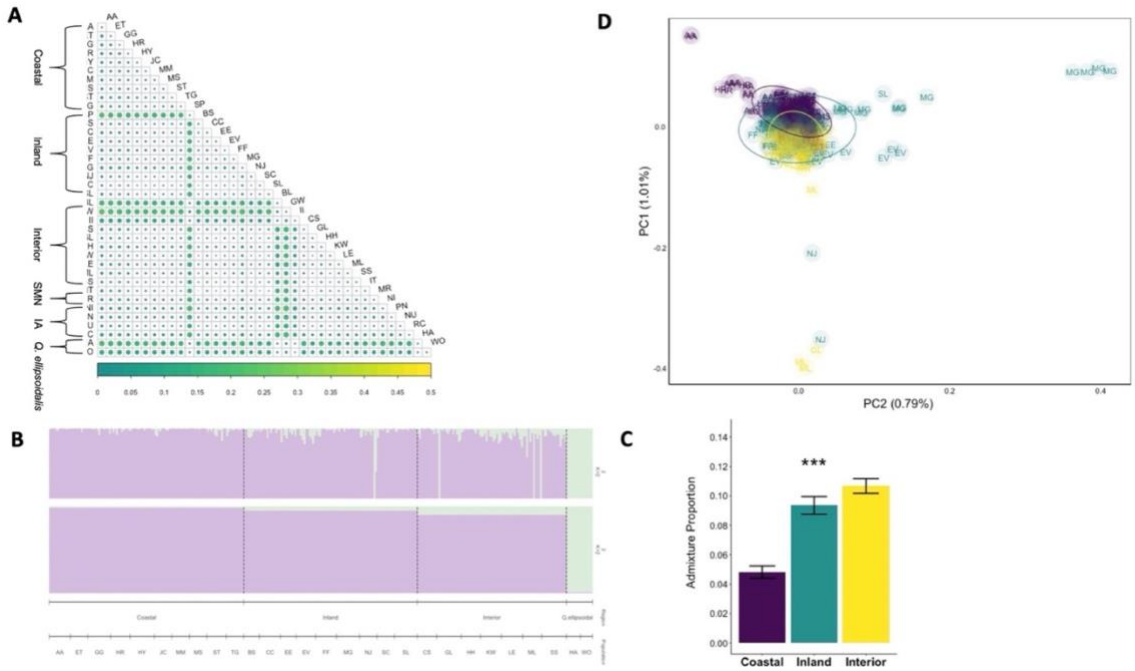


Figure 4. Population structure and evidence of interspecific hybridization based on neutral molecular markers. (A) Pairwise F_{ST} matrix for *Q. rubra* and *Q. ellipsoidalis* populations; circle size represents how large is the pairwise F_{ST} value for each set of populations. Because the calculated values are small, the matrix is scaled from 0 to 0.5 rather than the upper limit of F_{ST} of 1. (B) STRUCTURE plot of *Q. rubra* and *Q. ellipsoidalis* at $K = 2$ to identify admixture proportions between the two species. the top panel has the proportions for each individual, while the bottom panel has the average proportions across each region. (C) admixture proportions between coastal inland and interior populations. *** indicates a significant difference between coastal and the noncoastal regions. (D) Principal component analysis of coastal inland and interior populations, each region is represented by a color and each population is represented by a two-letter code (Supplementary Table 1). Ellipses represent 95% CI. Ellipse overall represents no difference in the clustering for each population, which is consistent with no population structure.

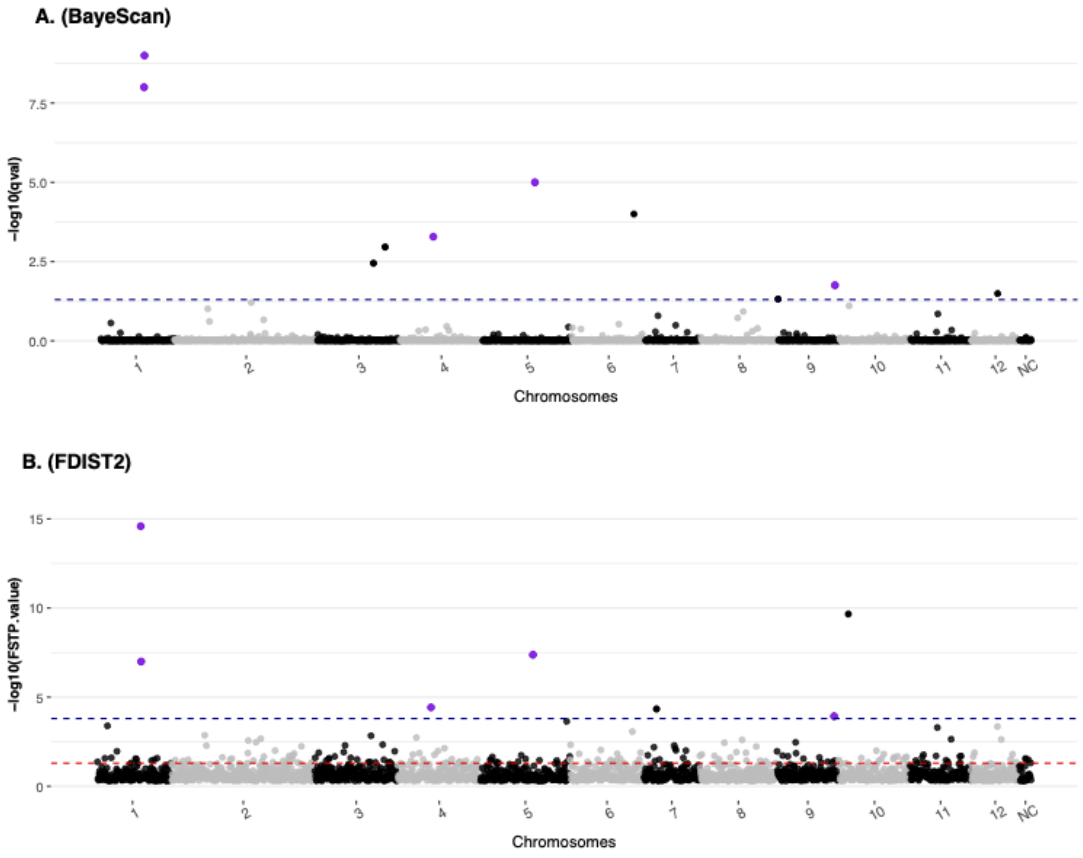


Figure 5. Manhattan plot of outlier loci using two distance-based approaches: BayeScan (A) and FDIST (B). Each point in the graph represents a the –log of the F_{ST} value for each locus within its respective chromosome based on the *Q. lobata* reference genome . NC are non-characterized chromosomal regions. Points above the dotted line are loci that have been identified as outliers after a 0.05% FDR correction (represented by the blue line threshold). The red dotted line in the FDIST approach represent a p = 0.05

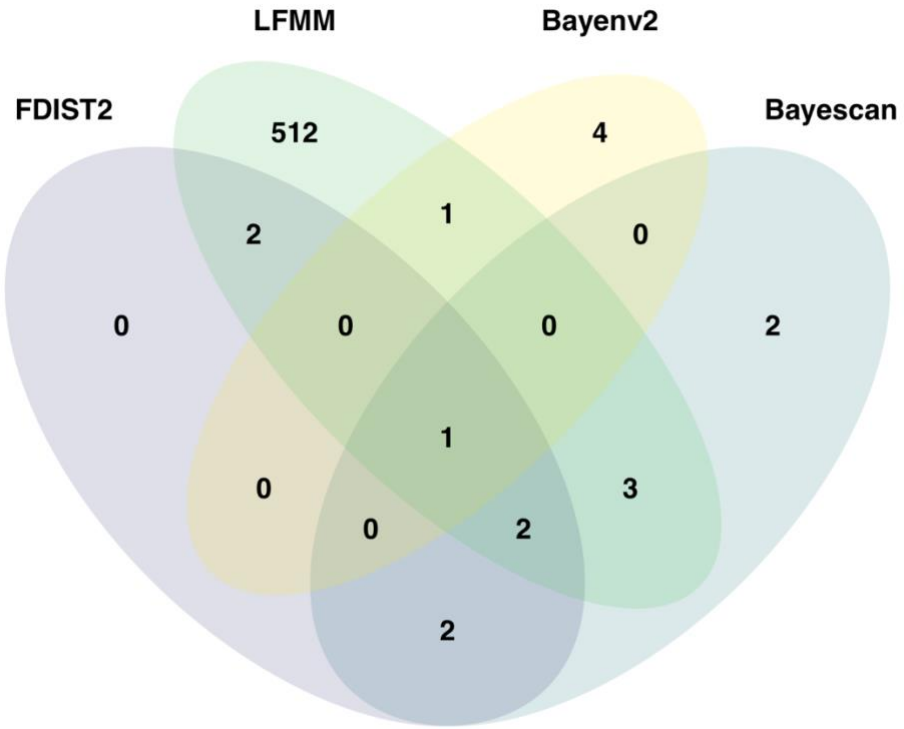


Figure 6. Venn diagram of loci identified as outliers by distance-based methods (FDIST and BayeScan) and loci associated with environmental variables (Bayenv2 and LFM) overlapping of ellipses represent the number of overlapping loci between software's and approaches. Only one locus was found to be under selection and associated with an environmental variable between the four different approaches.

Literature Cited

Abebe, T. D., Naz, A. A., & Léon, J. (2015). Landscape genomics reveal signatures of local adaptation in barley (Hordeum vulgare L.). *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00813

Anderson, J. T., Inouye, D. W., McKinney, A. M., Colautti, R. I., & Mitchell-Olds, T. (2012). Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1743), 3843–3852. https://doi.org/10.1098/rspb.2012.1051

Anderson, J. T., Wagner, M. R., Rushworth, C. A., Prasad, K. V. S. K., & Mitchell-Olds, T. (2014). The evolution of quantitative traits in complex environments. *Heredity*, *112*(1), 4–12. https://doi.org/10.1038/hdy.2013.33

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Baker, A. J., Daugherty, C. H., Colbourne, R., & McLennan, J. L. (1995). Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. *Proceedings of the National Academy of Sciences*, *92*(18), 8254–8258. https://doi.org/10.1073/pnas.92.18.8254

Bassar, R. D., Marshall, M. C., López-Sepulcre, A., Zandonà, E., Auer, S. K., Travis, J., Pringle, C. M., Flecker, A. S., Thomas, S. A., Fraser, D. F., & Reznick, D. N. (2010). Local adaptation in Trinidadian guppies alters ecosystem processes. *Proceedings of the National Academy of Sciences*, *107*(8), 3616–3621. https://doi.org/10.1073/pnas.0908023107

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berlin, S., Trybush, S. O., Fogelqvist, J., Gyllenstrand, N., Hallingbäck, H. R., Åhman, I., Nordh, N.-E., Shield, I., Powers, S. J., Weih, M., Lagercrantz, U., Rönnberg-Wästljung, A.-C., Karp, A., & Hanley, S. J. (2014). Genetic diversity, population structure and phenotypic variation in European Salix viminalis L. (Salicaceae). *Tree Genetics & Genomes*, *10*(6), 1595–1610. https://doi.org/10.1007/s11295-014-0782-5

Browne, L., Wright, J. W., Fitz-Gibbon, S., Gugger, P. F., & Sork, V. L. (2019). Adaptational lag to temperature in valley oak (*Quercus lobata*) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences*, *116*(50), 25179–25185. https://doi.org/10.1073/pnas.1908771116

Cariou, M., Duret, L., & Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, *3*(4), 846–852. https://doi.org/10.1002/ece3.512

Cavender-Bares, J., & Bazzaz, F. A. (2000). Changes in drought response strategies with ontogeny in Quercus rubra: Implications for scaling from seedlings to mature trees. *Oecologia*, *124*(1), 8–18. https://doi.org/10.1007/PL00008865

Cavender-Bares, J., González-Rodríguez, A., Eaton, D. A. R., Hipp, A. A. L., Beulke, A., & Manos, P. S. (2015). Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): A genomic and population genetics approach. *Molecular Ecology*, *24*(14), 3668–3687. https://doi.org/10.1111/mec.13269

Chen, G. F., & Schemske, D. W. (2015). Ecological differentiation and local adaptation in two sister species of Neotropical *Costus* (Costaceae). *Ecology*, *96*(2), 440–449. https://doi.org/10.1890/14-0428.1

Christmas, M. J., Biffin, E., Breed, M. F., & Lowe, A. J. (2016). Finding needles in a genomic haystack: Targeted capture identifies clear signatures of selection in a nonmodel plant species. *Molecular Ecology*, *25*(17), 4216–4233. https://doi.org/10.1111/mec.13750

Colautti, R. I., White, N. A., & Barrett, S. C. H. (2010). Variation of Self-Incompatibility within Invasive Populations of Purple Loosestrife (*Lythrum salicaria* L.) from Eastern North America. *International Journal of Plant Sciences*, *171*(2), 158–166. https://doi.org/10.1086/649023

Conover, D. O., & Baumann, H. (2009). PERSPECTIVE: The role of experiments in understanding fishery-induced evolution: Experiments to understand fishery-induced evolution. *Evolutionary Applications*, *2*(3), 276–290. https://doi.org/10.1111/j.1752-4571.2009.00079.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Kort, H., Vandepitte, K., Bruun, H. H., Closset-Kopp, D., Honnay, O., & Mergeay, J. (2014). Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology*, *23*(19), 4709–4721. https://doi.org/10.1111/mec.12813

de Villemereuil, P., Gaggiotti, O. E., Mouterde, M., & Till-Bottraud, I. (2016). Common garden experiments in the genomic era: New perspectives and opportunities. *Heredity*, *116*(3), 249–254. https://doi.org/10.1038/hdy.2015.93

Dempewolf, H., Tesfaye, M., Teshome, A., Bjorkman, A. D., Andrew, R. L., Scascitelli, M., Black, S., Bekele, E., Engels, J. M. M., Cronk, Q. C. B., & Rieseberg, L. H. (2015). Patterns of domestication in the Ethiopian oil-seed crop noug (*Guizotia abyssinica*). *Evolutionary Applications*, *8*(5), 464–475. https://doi.org/10.1111/eva.12256

Dey, D. C., Miller, G. W., & Kabrick, J. M. (2007). *Sustaining Northern Red Oak Forests: Managing Oak from Regeneration to Canopy Dominance in Mature Stands*. 15.

Donelson, J. M., Salinas, S., Munday, P. L., & Shama, L. N. S. (2018). Transgenerational plasticity and climate change experiments: Where do we go from here? *Global Change Biology*, *24*(1), 13–34. https://doi.org/10.1111/gcb.13903

Du, F. K., Wang, T., Wang, Y., Ueno, S., & Lafontaine, G. (2020). Contrasted patterns of local adaptation to climate change across the range of an evergreen oak, *Quercus aquifolioides*. *Evolutionary Applications*, *13*(9), 2377–2391. https://doi.org/10.1111/eva.13030

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Easlon, H. M., & Bloom, A. J. (2014). Easy Leaf Area: Automated digital image analysis for rapid and accurate measurement of leaf area. *Applications in Plant Sciences*, *2*(7), 1400033. https://doi.org/10.3732/apps.1400033

Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression: introgression in the american live oaks. *Evolution*, *69*(10), 2587–2601. https://doi.org/10.1111/evo.12758

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Etterson, J. R., Toczydlowski, R. H., Winkler, K. J., Kirschbaum, J. A., & McAulay, T. S. (2016). *Solidago altissima* differs with respect to ploidy frequency and clinal variation across the prairie-forest biome border in Minnesota. *American Journal of Botany*, *103*(1), 22–32. https://doi.org/10.3732/ajb.1500146

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285–298. https://doi.org/10.1038/hdy.2009.74

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Francis, R. M. (2017). POPHELPER: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

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 and Evolution*, *298*(8), 1533–1545. https://doi.org/10.1007/s00606-012-0656-y

Gailing, O., & Zhang, R. (2018). Experimental evidence for selection against hybrids between two interfertile red oak species. *Silvae Genetica*, *67*(1), 106–110. https://doi.org/10.2478/sg-2018-0015

Galloway, L. F., & Etterson, J. R. (2007). Transgenerational Plasticity Is Adaptive in the Wild. *Science*, *318*(5853), 1134–1136. https://doi.org/10.1126/science.1148766

Greenbaum, G., Templeton, A. R., & Bar-David, S. (2016). Inference and Analysis of Population Structure Using Genetic Data and Network Theory. *Genetics*, *202*(4), 1299–1312. https://doi.org/10.1534/genetics.115.182626

Gribko, L. S., & Jones, W. E. (1995). *Test of the Float Method of Assessing Northern Red Oak Acorn Condition*. 5.

Gugger, P. F., Fitz-Gibbon, S., PellEgrini, M., & Sork, V. L. (2016). Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Molecular Ecology*, *25*(8), 1665–1680. https://doi.org/10.1111/mec.13563

Gugger, P. F., Fitz-Gibbon, S. T., Albarrán-Lara, A., Wright, J. W., & Sork, V. L. (2021). Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. *Molecular Ecology*, *30*(2), 406–423. https://doi.org/10.1111/mec.15731

Günther, T., & Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, *195*(1), 205–220. https://doi.org/10.1534/genetics.113.152462

Harvey, A. C., Glover, K. A., Taylor, M. I., Creer, S., & Carvalho, G. R. (2016). A common garden design reveals population-specific variability in potential impacts of hybridization between populations of farmed and wild Atlantic salmon *, Salmo salar* L. *Evolutionary Applications*, *9*(3), 435–449. https://doi.org/10.1111/eva.12346

Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, *10*, 4–10. https://doi.org/10.1016/j.wace.2015.08.001

Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A. A., Deng, M., Denk, T., Fitz-Gibbon, S., Gailing, O., González-Elizondo, M. S., González-Rodríguez, A., Grimm, G. W., Jiang, X., Kremer, A., Lesur, I., McVay, J. D., Plomion, C., … Valencia-Avalos, S. (2020). Genomic landscape of the global oak phylogeny. *New Phytologist*, *226*(4), 1198–1212. https://doi.org/10.1111/nph.16162

Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., Jordan, R., Oakeshott, J., Weeks, A., Joseph, L., Lockhart, P., Borevitz, J., & Sgrò, C. (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, *2*(1), 1. https://doi.org/10.1186/s40665-014-0009-x

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos,

M. Henry H. Stevens,Eduard Szoecs and Helene Wagner (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Jensen, R., Hokanson, S., Isebrands, J., & Hancock, J. (1993). Morphometric Variation in Oaks of the Apostle Islands in Wisconsin: Evidence of Hybridization Between ' *Quercus* rubra and *Q. ellipsoidalis* (Fagaceae). *American Journal of Botany,80*(11), 1358-1366. doi:10.2307/2445721

JMP®, Version *16*. SAS Institute Inc., Cary, NC, 1989–2021.

Jordan, M. A., Snell, H. L., Snell, H. M., & Jordan, W. C. (2005). Phenotypic divergence despite high levels of gene flow in Galápagos lava lizards (Microlophus albemarlensis): gene flow in galápagos lava lizards. *Molecular Ecology*, *14*(3), 859–867. https://doi.org/10.1111/j.1365-294X.2005.02452.x

Joshi, J., Schmid, B., Caldeira, M. C., Dimitrakopoulos, P. G., Good, J., Harris, R., Hector, A., Huss-Danell, K., Jumpponen, A., Minns, A., Mulder, C. P. H., Pereira, J. S., Prinz, A., Scherer-Lorenzen, M., Siamantziouras, A.-S. D., Terry, A. C., Troumbis, A. Y., & Lawton, J. H. (2001). Local adaptation enhances performance of common plant species. *Ecology Letters*, *4*(6), 536–544. https://doi.org/10.1046/j.1461-0248.2001.00262.x

Kaiser, D. & Vetsch, J. Sulphur for Minnesota Soils (2020). Available at: https://extension.umn.edu/micro-and-secondary-macronutrients/sulfur-minnesota-soils.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Khodwekar, S., & Gailing, O. (2017). Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations. *American Journal of Botany*, *104*(7), 1088–1098. https://doi.org/10.3732/ajb.1700060

Kirk, H., & Freeland, J. R. (2011). Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. *International Journal of Molecular Sciences*, *12*(6), 3966–3988. https://doi.org/10.3390/ijms12063966

Kremer, A., & Hipp, A. L. (2020). Oaks: An evolutionary success story. *New Phytologist*, *226*(4), 987–1011. https://doi.org/10.1111/nph.16274

Larios, E., & Venable, D. L. (2018). Selection for seed size: The unexpected effects of water availability and density. *Functional Ecology*, *32*(9), 2216–2224. https://doi.org/10.1111/1365-2435.13138

Lascoux, M., Glémin, S., & Savolainen, O. (2016). Local Adaptation in Plants. In John Wiley & Sons Ltd (Ed.), *ELS* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0025270

Lepais, O., & Bacles, C. F. E. (2014). Two are better than one: Combining landscape genomics and common gardens for detecting local adaptation in forest trees. *Molecular Ecology*, *23*(19), 4671–4673. https://doi.org/10.1111/mec.12906

Lind, J. F., & 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 & Genomes*, *9*(3), 707–722. https://doi.org/10.1007/s11295-012-0586-4

Lind-Riehl, J. F., Sullivan, A. R., & Gailing, O. (2014). Evidence for selection on a CONSTANS-like gene between two red oak species. *Annals of Botany*, *113*(6), 967–975. https://doi.org/10.1093/aob/mcu019

Lind-Riehl, J., & Gailing, O. (2016). Adaptive Variation and Introgression of a CONSTANS-Like Gene in North American Red Oaks. *Forests*, *8*(1), 3. https://doi.org/10.3390/f8010003

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Minnesota Department of Natural Resources (2013). The Releve method 2nd edition. Available at: https://files.dnr.state.mn.us/eco/mcbs/releve/releve_booklet.pdf

Mitchell, N., Owens, G. L., Hovick, S. M., Rieseberg, L. H., & Whitney, K. D. (2019). Hybridization speeds adaptive evolution in an eight-year field experiment. *Scientific Reports*, *9*(1), 6746. https://doi.org/10.1038/s41598-019-43119-4

Moen, S. Minnesaota Sea Grant - Lake Superior's Natural process. (2018)Available at: http://www.seagrant.umn.edu/superior/processes.

Moran, E. V., Willis, J., & Clark, J. S. (2012). Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae* , Fagaceae). *American Journal of Botany*, *99*(1), 92–100. https://doi.org/10.3732/ajb.1100023

Murray, K. D., Janes, J. K., Jones, A., Bothwell, H. M., Andrew, R. L., & Borevitz, J. O. (2019). Landscape drivers of genomic diversity and divergence in woodland Eucalyptus. *Molecular Ecology*, *28*(24), 5232–5247. https://doi.org/10.1111/mec.15287

Nevo, E. (1995). Asian, African and European biota meet at 'evolution canyon' Israel: local tests of global biodiversity and genetic diversity patterns. Proceedings: Biological Sciences,262(1364), 149-155.

Ostaszewska-Bugajska, M., & Juszczuk, I. M. (2016). Changes in the OXPHOS system in leaf and root mitochondria of Arabidopsis thaliana subjected to long-term sulphur deficiency. *Acta Physiologiae Plantarum*, *38*(6), 141. https://doi.org/10.1007/s11738-016-2155-1

Pais, A. L., Whetten, R. W., & Xiang, Q. (Jenny). (2017). Ecological genomics of local adaptation in *Cornus florida* L. by genotyping by sequencing. *Ecology and Evolution*, *7*(1), 441–465. https://doi.org/10.1002/ece3.2623

Papaïx, J., David, O., Lannou, C., & Monod, H. (2013). Dynamics of Adaptation in Spatially Heterogeneous Metapopulations. *PLoS ONE*, *8*(2), e54697. https://doi.org/10.1371/journal.pone.0054697

Petit, R. J., & Hampe, A. (2006). Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 187–214. https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Pettenkofer, T., Finkeldey, R., Müller, M., Krutovsky, K. V., Vornam, B., Leinemann, L., & Gailing, O. (2020). Genetic variation of introduced red oak (*Quercus* rubra) stands in Germany compared to North American populations. *European Journal of Forest Research*, *139*(2), 321–331. https://doi.org/10.1007/s10342-019-01256-5

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramírez-Valiente, J. A., Deacon, N. J., Etterson, J., Center, A., Sparks, J. P., Sparks, K. L., Longwell, T., Pilz, G., & Cavender-Bares, J. (2018). Natural selection and neutral evolutionary processes contribute to genetic divergence in leaf traits across a precipitation gradient in the tropical oak *Quercus oleoides*. *Molecular Ecology*, *27*(9), 2176–2192. https://doi.org/10.1111/mec.14566

Ramírez-Valiente, J. A., Etterson, J. R., Deacon, N. J., & Cavender-Bares, J. (2019). Evolutionary potential varies across populations and traits in the neotropical oak *Quercus oleoides*. *Tree Physiology*, *39*(3), 427–439. https://doi.org/10.1093/treephys/tpy108

Räsänen, K., & Hendry, A. P. (2008). Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification: Adaptive divergence and gene flow. *Ecology Letters*, *11*(6), 624–636. https://doi.org/10.1111/j.1461-0248.2008.01176.x

Rellstab, C., Fischer, M. C., Zoller, S., Graf, R., Tedder, A., Shimizu, K. K., Widmer, A., Holderegger, R., & Gugerli, F. (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. *Heredity*, *118*(2), 193–201. https://doi.org/10.1038/hdy.2016.82

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., Bodénès, C., Sperisen, C., Kremer, A., & Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology*, *25*(23), 5907–5924. https://doi.org/10.1111/mec.13889

Ribeiro, Â. M., Lloyd, P., & Bowie, R. C. K. (2011). A tight balance between natural selection and gene flow in a southern african arid-zone endemic bird: spatial heterogeneous environments and gene flow. *Evolution*, *65*(12), 3499–3514. https://doi.org/10.1111/j.1558-5646.2011.01397.x

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science (New York, N.Y.)*, *317*(5840), 910–914. https://doi.org/10.1126/science.1137729

Romero-Rodríguez, M. C., Archidona-Yuste, A., Abril, N., Gil-Serrano, A. M., Meijón, M., & Jorrín-Novo, J. V. (2018). Germination and Early Seedling Development in *Quercus ilex* Recalcitrant and Non-dormant Seeds: Targeted Transcriptional,

Hormonal, and Sugar Analysis. *Frontiers in Plant Science*, *9*, 1508. https://doi.org/10.3389/fpls.2018.01508

Sander, I.L. 1990. Northern red oak. In: Burns, R.M.; Honkala, B.H., eds. Silvics of North America. Volume 2, Hardwoods. Agriculture Handbook 654.Washington, DC: U.S. Department of Agriculture. Forest Service: 727-733.

Sanders, T., Pitman, R., & Broadmeadow, M. (2014). Species-specific climate response of oaks (*Quercus* spp.) under identical environmental conditions. *IForest - Biogeosciences and Forestry*, *7*(2), 61–69. https://doi.org/10.3832/ifor0911-007

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. https://doi.org/10.1038/nrg3522

Schueler, S., & Schlünzen, K. H. (2006). Modeling of oak pollen dispersal on the landscape level with a mesoscale atmospheric model. *Environmental Modeling & Assessment*, *11*(3), 179–194. https://doi.org/10.1007/s10666-006-9044-8

Shaw, R. G., & Etterson, J. R. (2012). Rapid climate change and the rate of adaptation: Insight from experimental quantitative genetics. *New Phytologist*, *195*(4), 752–765. https://doi.org/10.1111/j.1469-8137.2012.04230.x

Sorci, G., Clobert, J., & Belichon, S. (1996). Phenotypic Plasticity of Growth and Survival in the Common Lizard Lacerta vivipara. *The Journal of Animal Ecology*, *65*(6), 781. https://doi.org/10.2307/5676

Sork, V., Huang, S., & Wiener, E. (1993). Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. *Annales Des Sciences Forestières*, *50*(Supplement), 261s–270s. https://doi.org/10.1051/forest:19930726

Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., & Neale, D. B. (2013). Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, *9*(4), 901–911. https://doi.org/10.1007/s11295-013-0596-x

Sork, V. L., Fitz-Gibbon, S. T., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R., Stevens, K., Langley, C. H., Pellegrini, M., & Salzberg, S. L. (2016). First Draft Assembly and Annotation of the Genome of a California Endemic Oak *Quercus lobata* Née (Fagaceae). *G3 Genes|Genomes|Genetics*, *6*(11), 3485–3495. https://doi.org/10.1534/g3.116.030411

Sork, V. L., Squire, K., Gugger, P. F., Steele, S. E., Levy, E. D., & Eckert, A. J. (2016). Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany*, *103*(1), 33–46. https://doi.org/10.3732/ajb.1500162

Sork, V. L., Stowe, K. A., & Hochwender, C. (1993). Evidence for Local Adaptation in Closely Adjacent Subpopulations of Northern Red Oak (*Quercus rubra L.*) Expressed as Resistance to Leaf Herbivores. *The American Naturalist*, *142*(6), 928–936. https://doi.org/10.1086/285581

Tirmenstein, D. A. 1991. Quercus rubra. In: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Available: https://www.fs.fed.us/database/feis/plants/tree/querub/all.html [2021, July 25].

Turner, S. D. (n.d.). *qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots*. 2.

USDA. USDA Hardiness Zones. (2019) Available at: https://planthardiness.ars.usda.gov/PHZMWeb/.

Vakkari, P., Blom, A., Rusanen, M., Raisio, J., & Toivonen, H. (2006). Genetic variability of fragmented stands of pedunculate oak (*Quercus robur*) in Finland. *Genetica*, *127*(1–3), 231–241. https://doi.org/10.1007/s10709-005-4014-7

Vanwalleghem, T., & Meentemeyer, R. (2009). Predicting forest microclimate in heterogeneous landscapes. Ecosystems, 12(7), 1158-1172.

Vallés, S., Fernández, J., Dellafiore, C., & Cambrollé, J. (2011). Effects on soil, microclimate and vegetation of the native-invasive Retama monosperma (L.) in coastal dunes. Plant Ecology, 212(2), 169-179.

Wardle, D. A., Ahmed, M., & Nicholson, K. S. (1991). Allelopathic influence of nodding thistle (*Carduus nutans* L.) seeds on germination and radicle growth of pasture plants. *New Zealand Journal of Agricultural Research*, *34*(2), 185–191. https://doi.org/10.1080/00288233.1991.10423358

While, G. M., Michaelides, S., Heathcote, R. J. P., MacGregor, H. E. A., Zajac, N., Beninde, J., Carazo, P., Pérez i de Lanuza, G., Sacchi, R., Zuffi, M. A. L., Horváthová, T., Fresnillo, B., Schulte, U., Veith, M., Hochkirch, A., & Uller, T. (2015). Sexual selection drives asymmetric introgression in wall lizards. *Ecology Letters*, *18*(12), 1366–1375. https://doi.org/10.1111/ele.12531

Wulff, R. D., Caceres, A., & Schmitt, J. (1994). Seed and Seedling Responses to Maternal and Offspring Environments in Plantago lanceolata. *Functional Ecology*, *8*(6), 763. https://doi.org/10.2307/2390236

Yang, F., Yang, X., Baskin, J. M., Baskin, C. C., Cao, D., & Huang, Z. (2015). Transgenerational plasticity provides ecological diversity for a seed heteromorphic species in response to environmental heterogeneity. *Perspectives in Plant Ecology, Evolution and Systematics*, *17*(3), 201–208. https://doi.org/10.1016/j.ppees.2015.03.003

Zanella, C. M., Bruxel, M., Paggi, G. M., Goetze, M., Buttow, M. V., Cidade, F. W., & Bered, F. (2011). Genetic structure and phenotypic variation in wild populations of the medicinal tetraploid species *Bromelia antiacantha* (Bromeliaceae). *American Journal of Botany*, *98*(9), 1511–1519. https://doi.org/10.3732/ajb.1000458

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Appendices

Supplementary Table 1. Northern red oak (*Quercus rubra*) and Northern pin oak (*Quercus ellipsoidalis*) sites and tissues collected for both the greenhouse common garden experiment and molecular marker analyses. Each population is represented by a two-letter code, the region it belongs to, latitude and longitude coordinates, distance from the lake and the type of tissue collected at each site. Specific spatial location can be found in Fig. 1

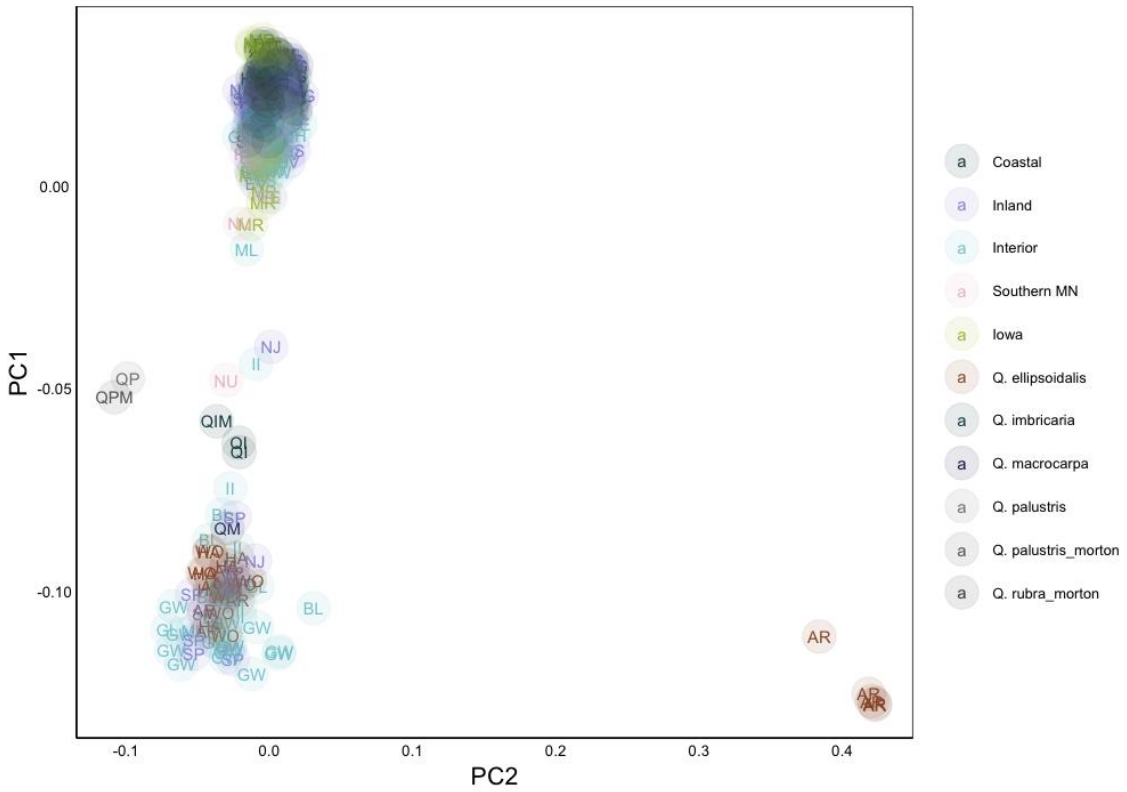
Species name	Population Name	Region	Latitude	Longitude	Distance from the lake (km)	Tissue Collected
Q. rubra	ET	Coastal	46.776132	-92.12431	2.05	Leaf /Acorns
Q. rubra	MM	Coastal	46.8843774	-91.975458	2.65	Leaf
Q. rubra	HY	Coastal	46.8303967	-92.08472	3.07	Leaf
Q. rubra	AA	Coastal	47.428242	-91.20655	4.18	Leaf /Acorns
Q. rubra	TG	Coastal	47.3368634	-91.270446	4.28	Leaf /Acorns
Q. rubra	GG	Coastal	47.44995	-91.2339	9.81	Leaf
Q. rubra	MS	Coastal	46.6952234	-92.231578	14.04	Leaf /Acorns
Q. rubra	ST	Coastal	46.673576	-92.273644	16.57	Leaf
Q. rubra	JC	Coastal	46.642718	-92.322142	16.89	Leaf /Acorns
Q. rubra	HR	Coastal	46.668554	-92.351586	17.38	Leaf
Q. rubra	EE	Inland	47.31909	-91.48048	18.8	Leaf
Q. rubra	MG	Inland	46.617996	-92.316025	23	Leaf
Q. rubra	CC	Inland	46.702444	-92.638807	41.8	Leaf
Q. rubra	NJ	Inland	46.4077174	-92.371857	43.4	Leaf /Acorns
Q. rubra	SL	Inland	46.8388372	-92.719922	49.5	Leaf
Q. rubra	FF	Inland	46.247724	-92.3913	58.9	Leaf /Acorns
Q. rubra	SC	Inland	46.1041104	-92.478812	76.7	Leaf /Acorns
Q. rubra	BS	Inland	46.2071725	-92.819646	82	Leaf /Acorns
Q. rubra	SP	Inland	46.7785583	-93.228834	82	Leaf
Q. rubra	EV	Inland	47.479658	-92.517111	82.7	Leaf
Q. rubra	BL	Interior	47.96372	-92.000742	91.4	Leaf
Q. rubra	HH	Interior	46.8718874	-93.325495	94.9	Leaf
Q. rubra	GL	Interior	46.1243969	-92.997152	98.17	Leaf /Acorns
Q. rubra	KW	Interior	46.541621	-93.373846	101	Leaf
Q. rubra	LE	Interior	47.6920825	-92.629149	104.7	Leaf
Q. rubra	CS	Interior	45.853561	-92.748816	110.2	Leaf
Q. rubra	SS	Interior	46.31162	-93.352731	110.56	Leaf
Q. rubra	GW	Interior	47.694641	-92.81142	112	Leaf
Q. rubra	ML	Interior	46.16718	-93.778303	147.57	Leaf
Q. rubra	II	Interior	45.951059	-92.662983	156.75	Leaf
Q. ellipsoidalis	HA	Pin oak	45.38362	-93.16834	173.8	Leaf
Q. ellipsoidalis	WO	Pin oak	45.21747	-92.77972	178.6	Leaf

Supplementary Table 2. Number of individuals and SNPs retained from each dataset. Datasets include the following populations: (1) field collected *Q. rubra* (coastal, inland, and interior regions), (2) field collected *Q. rubra* and *Q. ellipsoidalis*. Each dataset includes the total number of SNPs after initial filtering, the number of SNPs after removing sites in linkage disequilibrium, the total number of SNPs after removing multiple SNPs per locus , and the number of total SNPs after removing sites in linkage disequilibrium from one SNP per locus dataset.

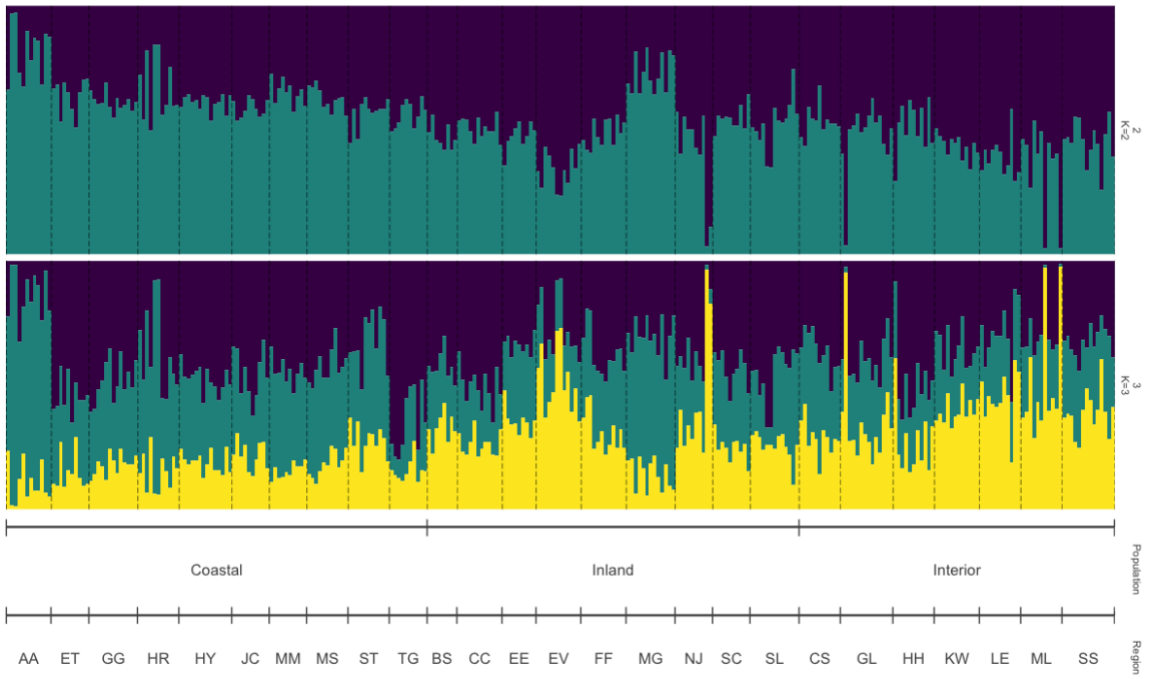
Dataset	**Number of Individuals**	**Total SNPs**	**SNPs in Linkage Equilibrium**	**One SNP per locus**	**One SNP per Locus in Linkage Equilibrium**
		Number of SNPs	**Number of SNPs**	**Number of SNPs**	**Number of SNPs**
1	363	8,210	3,770	2,568	2,162
2	413	7,022	3,347	2,324	1,946

Supplementary Table 3. Allele Frequencies for overlapping outlier loci using BayeScan and FDIST methods. The position and chromosome location are based on the publicly available *Q. lobata* reference genome. F_{ST} value reported corresponds to the value obtained by the BayeScan analysis the FDIST F_{ST} value is not reported given the high similarity in F_{ST} for both approaches.

				Allele Frequency		
Chromosome	Position	F_{ST}	Allele	Coastal	Inland	Interior
1	29790677	0.579	T/G	0.94	0.9	1
1	29936530	0.667	G/T	1	0.79	0.65
4	38317234	0.149	T/C	0.59	0.6	0.75
5	53461008	0.175	C/T	0.53	0.61	0.63
9	45446014	0.143	C/T	0.8	0.68	0.76



Supplementary Figure 1. Principal component analysis of coastal, inland, interior, field and greenhouse collected *Q. rubra* populations, Badoura nursery populations (Southern MN and Iowa), *Q. ellipsoidalis* populations, and other closely related *Quercus* species (Q. *imbricaria, Q. macrocarpa and Q. palustris*) collected from the Minnesota Landscape Arboretum and the Morton Arboretum. For each principal component analysis populations are coded by their region color and by the population two letter codes (Supplementary Table 1) as individual points.



Supplementary Figure 2. STRUCTURE plot analysis for genetic assignment of *Q. rubra* populations for two values of *K*. Genetic assignment using *K* = 2 and *K* = 3 does not provide evidence of population clusters and does not show evidence of strong population structure. The proportion of each cluster is shown in different colors for all the individuals in the analysis.